Lancaster University

School of Computing and Communications
Lancaster University

Recommending Access Control Decisions To Social Media Users

Gaurav Misra
B.Tech.
M.Sc.

Submitted for the degree of
Doctor of Philosophy
September 2017

This work is dedicated to my parents

Abstract

Social media has become an integral part of the Internet and has revolutionized interpersonal communication. The lines of separation between content creators and content consumers have blurred as normal users have platforms such as social media sites, blogs and microblogs at their disposal on which they can create and consume content as well as have the opportunity to interact with other users. This change has also led to several well documented privacy problems for the users. The privacy problems faced by social media users can be categorized into institutional privacy (related to the social network provider) and social privacy (related to the interpersonal communication between social media users) problems.

The work presented in this thesis focuses on the social privacy issues that affect users on social media due to their interactions with members in their network who may represent various facets of their lives (such as work, family, school, etc.). In such a scenario, it is imperative for them to be able to appropriately control access to their information such that it reaches the appropriate audience. For example, a person may not want to share the same piece of information with their boss at work and their family members. These boundaries are defined by the nature of relationships people share with each other and are enforced by controlling access during communication. In real life, people are accustomed to do this but it becomes a greater challenge while interacting online. The primary contribution of the work presented in this thesis is to design an access control recommendation mechanism for social media users which would ease the burden on the user while sharing information with their contacts on the social network.

The recommendation mechanism presented in this thesis, ***REACT (RE**commending*

***A**ccess **C**ontrol decisions **T**o social media users)*, leverages information defining interpersonal relationships between social media users in conjunction with information about the content in order to appropriately represent the context of information disclosure. Prior research has pointed towards ways in which to employ information residing in the social network to represent social relationships between individuals. REACT relies on extensive empirical evaluation of such information in order to identify the most suitable types of information which can be used to predict access control decisions made by social media users. In particular, the work in this thesis advances the state of art in the following ways: (i) An empirical study to identify the most appropriate network based community detection algorithm to represent the type of interpersonal relationships in the resulting access control recommendation mechanism. This empirical study examines a goodness of fit of the communities produced by 8 popular network based community detection algorithms with the access control decisions made by social media users. (ii) Systematic feature engineering to derive the most appropriate profile attribute to represent the strength or closeness between social media users. The relationship strength is an essential indicator of access control preferences and the endeavor is to identify the minimal subset of attributes which can accurately represent this in the resulting access control recommendation mechanism. (iii) The suitable representation of interpersonal relationships in conjunction with information about the content that result in the design of an access control recommendation mechanism, REACT, which considers the overall context of information disclosure and is shown to produce *highly accurate recommendations.*

Declaration

The work presented in my thesis is my original research work. No part of this thesis has been submitted elsewhere for any other degree or qualification. All work is my own and all other contributors have been credited wherever applicable. The entire research was carried out under the supervision of Dr. Jose M. Such at Security Lancaster, School of Computing and Communications at Lancaster University.

27th September, 2017

Gaurav Misra

Acknowledgements

First and foremost, I would like to thank my supervisor, Jose M. Such, for his endless guidance and support. He was always available and willing to help me solve all the problems I faced during my PhD. I feel extremely fortunate to have had a supervisor like him.

I would like to thank my examiners, Arosha Bandara and Maria Angela Ferrario, for providing invaluable feedback which has helped improve my thesis immensely.

I want to express my gratitude to Awais Rashid and Paul Rayson who have provided feedback about my work at very important junctures of my PhD journey. I am also grateful to our current and former department staff members Aimee, Christine, Charlotte, Claire Anne, Debbie, Gillian, Karen and Vicky for their support throughout my time at Lancaster. I would also like to thank my previous supervisors, Marianne Junger and Pieter Hartel, at University of Twente, who encouraged me to pursue a career in research.

I thank Hamed for his assistance in developing the Facebook application for the user study which became an integral part of my PhD. This research would have been significantly more challenging without his help and hard work.

I have been lucky to have had some wonderful friends who have made my life a lot better at Lancaster. I would like to thank my friends Arooj, Ethem and Mahmoud for always being there for me and helping me remain motivated throughout this long journey. I also want to thank my office colleagues Asad, Ben, Joe, Marvin and Rob with whom I have had endless discussions. I will also remember the great time I had with Benny, Kevin and Vaselios who were wonderful flatmates and made this journey a lot more enjoyable.

I want to thank my cousins Pritam and Payal for inspiring me to pursue a career in research and guiding me through major career decisions. I would also like to thank my cousins Tanay and Akanksha for their wishes and support.

Lastly, and most importantly, a big thanks to my parents for their endless prayers. Their love and support means everything to me and has brought me this far in my life.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

In the past decade, the Internet has undergone a radical transformation. It has migrated from being a primarily informative medium and has become much more interactive. This has been brought about by the deluge of social media sites which contribute to a large amount of content on the Internet. Most of the content on these sites is created by general public and the lines between creators and consumers have blurred. This massive shift has been stimulated by the advent of platforms such as blogging sites where users can create content and interact with other users. This has been taken to a completely new level by the introduction of social media sites. These sites allow users to create accounts and then connect and interact with other "users" of the site. Users can share textual content or multimedia content in the form of photos, videos, etc. The popularity of social media sites has grown massively and an ever increasing number of people are joining these sites. Facebook, by far the most popular social media site, is estimated to have 1.18 billion daily active users (users who log in to Facebook daily) as per latest estimates[1]. When so many users are potentially producing content, the amount of content itself is enormous. For example, there are 300 million photos uploaded to Facebook every day and the total data posted per day is estimated to be 500 Terabytes[2].

Having a profile and interacting with friends on social media is a wonderful new way of communicating for people but it has not come without its share of complications.

[1] https://zephoria.com/top-15-valuable-facebook-statistics/

[2] http://gizmodo.com/5937143

The vast amount of personal data being shared brings several privacy challenges for the users. These privacy issues depend on how users use the medium and what expectations they have from it [MO11]. Some people may use social media as a marketing tool and would like to connect and share information with as many people as possible while others may want to only connect to their real life contacts and stay in touch with them [MO11]. Nevertheless, most social media users share a lot of personal data on these platforms [GA05] and experience privacy problems which result in unpleasant outcomes [EOK11]. The effects of such privacy breaches depend on a lot of factors such as who the attacker is and what information was revealed and can range from mild embarrassment (employer accessing embarrassing details or photos about a night-out with friends) to truly dire consequences (sexual predators or criminals using social media to track, monitor and identify a user as a potential victim) [JEB12].

Privacy problems on social media are of different types and can be broadly defined in terms of *Social Privacy* and *Institutional Privacy*. The institutional privacy threats to users' privacy are often the result of the business practices of the social media infrastructure providers which, although legal, may cause a privacy violation for the users. A particular example is when social media providers track users on other websites, often in accordance with their own privacy policies and hence legally, using social widgets which the users can log into with their social media profiles [MM12]. This creates a possible situation where the social media provider can link the users' browsing activity with the data on their social media profiles, which are of course accessible to the social media provider. Monitoring users' online activity, extraneous to the social media site, can be lucrative for the social media provider as it can then provide services such as targeted advertising and tailored content which can benefit the business [AEE$^+$14]. In addition to the possible business motive of invading users' privacy in this way, the often ambiguous privacy related documentation fails to adequately equip users with information about safeguarding their privacy and leaves room for exploitation by either the providers or third-parties [S$^+$10, Ant14].

While institutional privacy problems are important and have been studied extensively in existing literature, users also face many social privacy issues. The social privacy problems faced by users are a result of their failure to appropriately control access to the information they share on the social media site with their social network, often consisting of a vast number of people who represent multiple life-facets (such as work, family, friends, etc.) [MO11]. Such failures can leave a user exposed to a "context collapse" which means that information meant to be viewed by someone they intended and expected was viewed by "unintended recipients" with an associated risk of being viewed out of its context [HLL11]. Thus, a failure to preserve the "contextual integrity" of the disclosed information results in a privacy breach for the users [Nis04]. Users often mitigate such concerns by employing *coping mechanisms* such as "self-censorship" (not sharing something due to the fear of a privacy breach) and "un-friending contacts" [WLW12]. Such mechanisms are counter-productive for the user and diminish the utility of having a profile on social media altogether.

The disclosure to unintended or unwanted recipients often arises from the users' failure to make appropriate access control decisions when sharing the content. One of the reasons that users make erroneous access control decisions has been found to be the complexity of the access control mechanisms [WKR14, JEB12]. The work presented in this thesis focuses on addressing such "unintended disclosure" arising from the users' inability to make appropriate access control decisions by providing an access control recommendation mechanism which will assist them in making appropriate access control decisions while sharing information on social media sites.

The users' information may be disclosed to unintended recipients even if they make appropriate access control decisions. Examples of such instances are "co-privacy" breaches (or "multi-party" privacy problems), where the content (e.g. a photo in which multiple users appear) is disclosed by another user who may be part of that content [SC16], and "dissemination" by originally intended recipients to someone who is considered unintended by the original sharer of the content [Sol06]. These access control problems are beyond the scope of the work presented in this thesis and are discussed in Section 2.1.3 in addition to other social privacy problems in social media.

1.1 Motivation and Thesis Aims

Controlling access to one's information on social media sites is an essential task for users in order to preserve their privacy. However, it has been found that privacy controls afforded to social media users have been found to be burdensome and they have been found to struggle in appropriately configuring access controls while disclosing their information [WKR14]. Therefore, it is imperative to improve existing access control mechanisms and assist users in making appropriate access control decisions and preserve the "contextual integrity" of their content to safeguard their privacy [LHL$^+$09]. A major drawback of existing access control mechanisms in social media sites is the lack of support for interpersonal relationships [MS16]. The interpersonal relationships between users are often an integral component of the access control decisions making process followed by users and failure to adequately accommodate this aspect diminishes the utility of any such mechanism [FSEGF15, LLLT11]. This has led to the development of Relationship based Access Control (ReBAC) mechanisms which rely on information residing in the social network, either in the profiles created by users or in their communication (messages, posts, etc.) with other users, to define the nature and strength of interpersonal relationships [FSEGF15]. There have been proposed approaches to assist users by using relationship based attributes to predict access control decisions made by them [AFW12, ML12, FL10]. In this way, the user is not burdened with the entire responsibility of making access control decisions and can rely on the recommendations made by the mechanism [AFW12]. A major drawback of employing such ReBAC systems, however, is that they typically employ large sets of attributes to define the relationships, which can help users better identify the social context of their disclosures, but can diminish usability due to challenges such as computational overhead of gathering and processing the required information. Previous ReBAC approaches have failed to provide a systematic evaluation of the relevant information required to represent social relationships between social media users which leaves a gap in terms of identifying the most suitable representation of relationships in an access control recommendation mechanism.

In addition to the "who", determined by social relationships, the "what", in terms of information about the content, also needs to be considered to make informed access control decisions. Therefore, content also forms a major part of the overall context of the information disclosure and needs to be accounted for in any access control mechanism [SSLW11, KLM$^+$12]. Previous efforts directed towards recommending access control decisions to users have failed to combine a ReBAC approach with information about the content in order to have more information about the context of the information disclosure.

The work presented in this thesis aims to enhance existing access control mechanisms in social media by leveraging interpersonal relationships between users in conjunction with the information about the content being shared to recommend appropriate access control decisions to users. The work systematically considers various sources of information relevant to modeling interpersonal relationships on social media and identify the set of attributes which contribute the most in enabling prediction of access control decisions. These attributes, representing the social context, are then augmented with information about the content being shared to create an informed approach of recommending access control decisions to social media users which would preserve the contextual integrity of the disclosed information and safeguard their privacy as a result. It is demonstrated in this thesis that the proposed recommendation mechanism provides highly accurate recommendations while using less information as compared to previous similar approaches. This is achieved as a result of empirical evaluation of relevant attributes to arrive at the optimal set of attributes which enable creation of an access control mechanism based on prediction of access control decisions made by users.

1.2 Contributions

The major contributions of the work presented in this thesis are:

1. Systematization of existing access control mechanisms in mainstream social media infrastructures and related literature.

2. Design of an access control recommendation mechanism, REACT (REcommending Access Control decisions To social media users), leveraging interpersonal relationships and information about the content.

3. An empirical study of community detection algorithms to evaluate a goodness of fit with the access control decisions made by users and identify the most suitable algorithm.

4. A systematic study of information present in social media profiles to identify the attributes which, when representing relationship strength, can provide accurate access control recommendations.

5. Implementation and empirical evaluation of REACT using the most ideal subset of attributes identified and empirically evaluate its performance

1.2.1 Systematization of Literature and Social Media Infrastructures

In order to provide a usable and effective solution to the access control issues faced by social media users, the work provides a systematization of previous approaches in literature which discuss social media privacy issues in general and access control recommendation mechanisms in particular. Additionally, an analysis of social media infrastructures which looks at the support, or lack thereof, they provide users to maintain their interpersonal relationships. The analysis found that popular social media sites were moving towards better support for interpersonal relationships but gaps were identified in the lack of usability of the access control mechanisms. The identification of these gaps, in addition to the ones identified in literature, facilitated the design of the access control recommendation mechanism presented in this thesis.

1.2.2 REACT Design

The primary contribution of this thesis is to provide a design for an access control recommendation mechanism, REACT, which adequately accounts for interpersonal relationships between users as well as information about the content being shared in order to represent the overall context of the disclosure. In REACT, relationships are represented in terms of their type by the means of communities created by network based community detection algorithms. The strength of relationships or closeness between individuals is represented by examining a similarity between profile attributes between individuals. Finally, the information about the content is represented by annotations about categories and any social tags made by the user.

1.2.3 Empirical Study of Community Detection Algorithms for Access Control

There is existing research which proposes leveraging network based community detection for access control recommendations. However, there is an absence of a thorough comparison of community detection algorithms and their suitability in enhancing access control mechanisms. The work presented in this thesis provides, to the best of my knowledge, the first empirical evaluation of 8 well known community detection algorithms where a goodness of fit is examined with access control decisions made by users during a user study. This comparison leads to an analysis of the suitability of community detection algorithms for access control in social media.

1.2.4 Identifying Most Suitable Attributes for Profile Similarity

Users' activities on social media sites result in a large amount of information being available to be potentially leveraged for calculating similarity between their profiles. Information including various profile attributes containing personal information such as *location, workplace, education, age* or information arising from interpersonal communication such as *messages, comments, likes, posts* can potentially be leveraged to estimate the closeness (or similarity) between individuals. Given such a broad range of options, it is imperative to identify an optimal set of attributes which can be utilized in any access control mechanism. The endeavor is to accurately portray the strength of relationship or "closeness" between individuals with the minimal amount of information such that the computational burden of retrieving the required information to be processed and the intrusiveness for the user is minimized. The findings presented in this thesis enable the identification of such a minimal set of profile attributes which can be used to predict access control decisions made by users and be used to represent strength of interpersonal relationships in REACT.

1.2.5 Evaluation of the Performance of REACT

REACT was implemented using the most appropriate community detection algorithm, as found by the evaluation shown in this thesis, the appropriate subset of profile attributes identified for representing relationship strength and information about the content provided by users in the form of annotations or "tags". The performance of REACT was empirically evaluated by conducting a user study in order to obtain ground truth access control decisions as well as the relevant information, friend network to enable creation of communities and profile information to represent relationship type, from the users' profiles. The performance was evaluated by using established evaluation metrics used in literature to evaluate recommendation mechanisms. The results of the evaluation presented in this thesis show that REACT produces highly accurate access control recommendations for all users.

1.3 Thesis Outline

The rest of this thesis is organized as follows:

Chapter 2 begins with a description of the various stakeholders in a typical social media ecosystem before describing the major privacy issues faced by social media users. It highlights the particular privacy issues which the work in this thesis aims to address before providing an analysis of social media infrastructures which highlights the absence of support they provide for maintaining interpersonal relationships between users. Finally, related work in the area of access control recommendation are discussed and the challenges addressed by the work in this thesis are summarized.

Chapter 3 describes the design of REACT and discusses the different components it uses to provide access control recommendations to the user, namely, community membership for representing relationship type, profile attributes to represent relationship strength and information about the content.

Chapter 4 describes the user study which led to the collection of all relevant data used for the analyses in this thesis. In particular, the process of the experiment and the demographics of participants are described.

Chapter 5 presents a comparison of 8 well known community detection algorithms where a goodness of fit between the communities created by the algorithms and the access control decisions made by users is examined.

Chapter 6 presents findings which enable identification of the most suitable profile attributes for representing relationship strength in REACT by evaluating a baseline of 30 profile attributes and comparing their ability to predict access control decisions.

Chapter 7 shows an implementation of REACT and provides detailed results of the empirical evaluation of its performance.

1.4 Publications

The work presented in this thesis has led to several papers in prestigious refereed conferences and journals:

- Misra, Gaurav, and Jose M. Such. "Social computing privacy and online relationships," *In Proceedings of the Symposium on Social Aspects of Computing and Cognition, Artificial Intelligence and Simulated Behavior (AISB) Convention*, April, 2015.

- Misra, Gaurav, and Jose M. Such. "How socially aware are social media privacy controls?," *IEEE Computer*, 49(3), pages 96-99, 2016.

- Misra, Gaurav, Jose M. Such, and Hamed Balogun. "Non-sharing communities? An empirical study of community detection for access control decisions," *In Proceedings of IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 49-56, 2016.

- Misra, Gaurav, Jose M. Such, and Hamed Balogun. "IMPROVE: Identifying Minimal PROfile VEctors for similarity based access control," *In Proceedings* *of The 15th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (IEEE TrustCom)*, 2016.

- Misra, Gaurav, and Jose M. Such. "REACT: REcommending Access Control decisions To social media users," *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, In press, 2017.
- Misra, Gaurav, and Jose M. Such. "PACMAN: Personal Agent for Access Control in Social Media," *IEEE Internet Computing*, In Press, 2017.

Chapter 2

Background and Related Work

This chapter provides a background to the work presented in this thesis by introducing the social media ecosystem and the various privacy problems faced by users. This is done through a systematization of existing literature on social media privacy as well as providing an analysis of state of art social media privacy controls available to users. After this, the chapter looks at related work in the area of access control recommendation in social media to provide an overview of previous efforts in this area and also highlight how the work presented in this thesis advances the state of art.

2.1 Background

This section outlines the social media ecosystem and provides an overview of the privacy problems faced by social media users. It starts with Section 2.1.1 where the various stakeholders of a typical social media ecosystem are described before providing a systematization of literature in Section 2.1.2 which helps in enumerating the previous work done in the area of social media privacy. This helps in identifying the gaps in previous work and positioning the work presented in this thesis in the broader spectrum of social media privacy.

In addition to systematizing existing relevant literature, this section also looks at the access control and contact grouping mechanisms available in the popular social media sites to enhance the understanding of state of the art access control mechanisms available to social media users. Section 2.2 provides an evaluation of 30 popular social media sites in terms of the support provided to users for defining and maintaining social relationships by the means of the access control and contact grouping mechanisms afforded to them by the social media infrastructure.

2.1.1 Social Media Ecosystem and Stakeholders

Before detailing the various privacy problems faced by social media users, it is important to understand the ecosystem of a typical social media site. Most popular social media sites have three types of primary stakeholders:

- **Users** create profiles on the social media sites by providing their information such as personal details (name, age, location, etc.), photos, multimedia content, etc. They are also able to share information such as text, photos, etc., in the form of “posts” or “updates” with people who they connect with on these sites. Most of the content on social media sites is created by users which makes them a key element of the social media ecosystem.

- **Providers** run the social media infrastructure, store users’ information and manage its distribution to other users and third parties. Notably, providers are in charge of enforcing their own privacy policies while also communicating these policies to the users in the form of documentation as well as privacy controls. Most popular social media sites have a centralized structure where the provider is in charge of the entire infrastructure. There are some decentralized social networks such as *Diaspora*[1] and *Friendica*[2] but are not as extensively used when compared to popular centralized social networks such as *Facebook* and *Google+*.

- **Third parties** are neither users nor providers. They add to the basic functionality of social media sites by providing services to the users. For example, many popular social media sites allow third parties to provide applications such as social games which can be used by the users to add to their overall social experience.

[1] `https://diasporafoundation.org/`

[2] `http://friendica.com/`

In terms of the overall ecosystem of social media sites, users need to interact with both the social media *Providers* as well as the *Third-parties* via the user interface of the social media site. They use these interfaces to update their profile, connect with people by adding them as "friends" (or "connections", "followers", etc.) and interact with these friends by sharing content with them and making access control decisions. The access control behavior of the users depends to a large extent on the mechanisms offered to them by the social media site. This is discussed in more detail in Section 2.2.

2.1.2 Privacy Problems in Social Media

The social media activity of users often entails disclosure of personal information and hence brings with it risks and threats to their privacy. Such social media privacy problems can be, and indeed have been, categorized in many different ways. Such categorizations and classifications help in an overall understanding of the overall spectrum of privacy problems in social media. For this chapter, the social media privacy problems are classified in two categories:

- ***Social Privacy*** problems arise due to the interaction between social media users on the social media sites. The most important aspect of social media for users is the ability to interact with vast networks of friends who represent different life facets (such as friends, family, co-workers, etc.). In such a situation, failure to appropriately control access to their content can lead to information being revealed to unintended audiences which may lead to a breach of their privacy. Such situations may arise either due to the users' failure to configure access controls appropriately or even in a situation where they made the appropriate access control decisions but actions by the members of their audience result in a breach of privacy.

- ***Institutional Privacy*** problems arise due to the social media infrastructure provided to users. These can be due to the business interests of the social media providers or simply failure to address conceptual gaps in the infrastructure or policy which leaves the privacy of the users at a risk. For this classification, institutional privacy will encompass threats arising directly due to the social media infrastructure as well as the third party applications used by social media users. A distinction is not made between these two for this classification as they often overlap, for example, privacy policies created by social media providers govern how user data can be used by third parties. Moreover, the users often fail to distinguish between third parties and the social media infrastructure and have demonstrated disapproval of third-party access to their data [MC10] while also freely using such applications on their social media profiles[3].

A brief of overview of the various privacy problems and possible methods to mitigate them, as suggested in literature, are provided in Table 2.1. The various social media privacy problems are classified into "Institutional" and "Social" privacy problems in the table. The rest of this section provides a brief description of each of the privacy problems shown in the table. The social privacy problems are discussed in Section 2.1.3 followed by a discussion on institutional privacy problems in Section 2.1.4.

2.1.3 Social Privacy Problems

Social privacy problems occur generally due to the interaction between users on social media sites. Each of the social privacy problems shown in Table 2.1 are discussed below.

Unintended Disclosure *Sensitive information is revealed by social media users to unintended audiences.*

Description: Social media users often misunderstand the nature and extent of their audience on social media and end up *oversharing* content with unintended recipients inside their network [JEB12, MO11, LGKM11, CGNP12, HJ10]. As the users often connect with a large number of people who represent life-facets (work, family, etc.), such unintended disclosure leaves them at risk of their information being processed out of context [HLL11, MO11]. Potential consequences of such "context collapse" can include embarrassing content (e.g., comments, pictures) being publicly visible, or personal data being disclosed to professional contacts [WNK^{+}11, EOK11].

[3] `http://mashable.com/2014/03/19/facebook-games-stats`

Table 2.1: Social media privacy problems and mitigation at a glance

Type	Privacy Problem	Mitigation
Social	***Unintended Disclosure***	Access Control Recommendations [SSLW11, FL10]; Audience Visualization [MLA12, LBW08] and Visual Cues [WLA$^+$14]
	Dissemination	Sticky Policies [PM11]
	Multi-party Privacy	Negotiation techniques: Bidding [SSP09]; Voting [CF11, TGN10]; Game-theory [SR16, HAZY14]
	Infiltration	Relationship-based Access Controls (ReBAC) [HJ10, FSEGF14]
	Inference Attacks	Obfuscation methods [HKT13]
Institutional	*Policy Deficiency*	Machine-readable privacy policies [Cra03]; Representation of users' preferences [SHC$^+$09]
	Tracking	Decentralization [YLL$^+$09]; Anonymous browsing [RSG98]
	Crawling	Limiting queries [WSB$^+$10]; Behavioral analysis [MVC$^+$12]
	De-anonymization	Privacy-preserving data publishing [BS08]

Solutions: Unintended disclosure happens due to the lack of usable access control mechanisms available to social media users. Development of access control mechanisms which reduce the burden on the user by recommending appropriate access control decisions can mitigate this problem [SSLW11, FL10]. Better visualization of the potential audience with whom the content is being shared can also enhance the comprehension of the access control decisions for the user and hence prevent unintended disclosure [MLA12, LBW08].

Dissemination *: A user shares data with a legitimate recipient who in turn "re-shares" it to unintended audiences [Sol06].*

Description: Even when a social media user appropriately configures an access control policy, and only shares with an intended audience, a member of that audience may further share the content with an audience which is unintended for the original user and can therefore compromise their privacy [MTKC]. Dissemination can happen via more than one hop in the social graph [LGKM11, JEB12].

Solutions: There are no clear defences for such privacy problems in social media but "sticky policies", which enable the user to control the flow of information they share on the network [PM11], is a possible mitigation. However, such mechanisms have not been implemented in social media thus far.

Multi-party Privacy *: Information shared on social media may be related to several individuals beyond the owner. Disclosure to certain individuals or groups, while intended by the owner, may result in a breach of privacy for other individuals with conflicting access control preferences [TGN10].*

Description: A user might share a photo where other members of the social network are depicted who may have different access control preferences. This can result in multi-party privacy conflict [IPA$^+$15]. These conflicts, if not resolved to the satisfaction of all parties, can result in a breach of privacy [TGN10].

Solutions: There have been proposed approaches for resolving multi-party privacy conflicts which occur in social media. These include bidding mechanisms, similar to an auction system, where each party bids for their preferred access control policy and the policy with most bids is adopted [SSP09], voting mechanisms, where each user can vote to select a policy which is acceptable to a majority [CF11] or veto voting [TGN10] and game-theoretic approaches [SR16, HAZY14].

Infiltration *: An attacker, posing as a fake user, becomes part of a user's network in order to access data that is only available to restricted audiences.*

Description: Social media users often share a lot of personal information with their connections on social media sites. Users are often provided with access controls mechanisms to deny access to people outside their friend networks [JEB12]. Therefore, a potential attacker may try different ways to access this personal information by strategically befriending the target user's known friends [Fon11b, PCNR10]. Other methods include joining open regional networks of the target user [KW08] or befriening highly connected users (those who have lots of friends), who are more likely to accept requests from a stranger [PCNR10, BAD09], that help identifying targets to launch such infiltration attacks [BAAS09].

Solutions: Infiltration attacks may be prevented by users if they can avoid befriending strangers and are provided with fine-grained access controls such that they can exclude individuals they are unsure about (even if they have befriended them). This can be enhanced by improving the visualization of social relationships in terms of their strength or "closeness" [FSEGF14] and implementing Relationship based Access Control (ReBAC) mechanisms [HJ10].

Inference Attacks *: An attacker - either another user, the provider, or a third-party - infers a user's private information based on the information disclosed publicly by that user.*

Description: Inference attacks can stem from institutional as well as social privacy problems depending on who the attacker is (a friend or the social media provider) and how they infer the undisclosed information. The attacker could infer characteristics of a user based on who they know or connect with on the social media site [HCL06, LHKT09, TGN10, ZG09, GL16, MVGD10, DTRS12, LWMH13, HSGH13]. A user's interests or

activity such as “check-ins” or attendance of events could also potentially lead to inference of private information such as age, gender, religion or political affiliation [WBIT12, AÁK12]. Meta-data created by mobile interfaces and applications can be used to infer location by spatial inference on proximity services [PAP+15].
Solutions: Inference attacks can be prevented or at least inhibited by using obfuscation methods such as hiding certain personal attributes or generalizing details (when disclosing preferences or location) [HKT13]. These techniques would, however, only safeguard against large scale inference attacks and more sophisticated and directed inference attacks may not be mitigated by obfuscation.

2.1.4 Institutional Privacy Problems

The privacy of social media users may be put at risk due to the practices and policies of the social media providers and these problems, classified as institutional privacy problems, are discussed here.

Policy Deficiency *: The policies set out by the social media provider, which govern how user data can be collected, fail to safeguard users' privacy.*
Description: Privacy policies provided to social media users may be considered deficient and insufficient as users may simply avoid reading what they consider as legalese and even when they do read them, may fail to understand the contents [SSM11, FFB15]. Moreover, privacy policies are often incomplete as they acknowledge nominal mechanisms behind data collection by a provider or a third party, but do not sufficiently elaborate the privacy implications necessary for users to make informed decisions [S+10]. Another problem is that privacy policies may not be correctly implemented at the infrastructure level as this can be a challenging engineering task [AHB04] due to the evolving nature of policies which are prone to modifications over time.
Solutions: Possible enhancements of privacy policies include creating machine-readable privacy policies aimed to find a match with users' privacy settings [Cra03] and better representation of the users' privacy preferences using a “privacy persona” [SHC+09].

Tracking *: Collecting information about users to identify repeated visits and track user activity over multiple web pages.*

Description: Traditional tracking techniques can be augmented or enhanced by including information about users' social media activity [EN16]. Social media sites offer their users an opportunity to link their browsing activity (extraneous to the social media site) with their social media profiles. For example, social widgets on websites allow the user to "log in" using their social media credentials and receive a personalized browsing experience but they also provide the social media provider an opportunity to track their browsing activity [MM12]. Personally identifiable information of the user may also be leaked by the provider to third party applications [KW09] and this can be used and correlated with external sites for effective tracking of the users' browsing activity.

Solutions: Tracking by social media provider can only effectively be thwarted by having decentralized social networks such as *Diaspora*[4] and *Friendica*[5] [YLL+09]. Tracking of activity external to the social media site may be prevented by using anonymous browsing techniques [RSG98]. This has an obvious trade-off as avoiding social widgets and other third-party tools will also diminish the personalized browsing experience.

Crawling *: User information is harvested on a large scale [BAD09], using either data published by the social media provider to third parties, or data visible to users of the social media site.*

Description: Crawling is often forbidden by the social media providers' terms and conditions [Fac16]. However, it has been found that it is possible for attackers to harvest data such as email addresses of social media users just by knowing the names they use [PKA+10] without having a social media profile themselves [LHKT09].

Solutions: Social media providers can prevent crawling attacks by limiting querying for public profiles as normal users would not need to access a lot of profiles in a given time period [WSB+10]. Another approach is to use behavioral analysis to identify crawlers which often have a non-regular browsing pattern [MVC+12].

[4] https://diasporafoundation.org/

[5] http://friendica.com/

De-anonymization *: An attacker uniquely identifies a user from an anonymized (or pseudonymized) dataset, using one or several sources.*

Description: In the context of social media, de-anonymization can happen by learning sensitive or undisclosed information about users from anonymized or aggregated graph data, often released by social media providers to third-parties [ANH16]. De-anonymization can be achieved by analyzing the topology of the social graph [BDK07, NS09, PLZW14, BFK14], observing communication between users [DTS08] or combining information across different datasets such as their social media account and their online browsing history [WHKK10].

Solutions: The social media providers can adopt privacy-preserving data publishing techniques while releasing anonymized datasets of users. However, it has been shown that the richness of data shared on social media often defeats these techniques and fails to prevent de-anonymization [BS08].

2.1.5 Focus of this Thesis

It is clear from the above discussion that there are numerous privacy problems faced by social media users. The work presented in this thesis focuses specifically on mitigating the problem of *Unintended Disclosure* in social media by easing the burden of making access control decisions for social media users. As discussed earlier, this is a *Social Privacy* problem which arises due to information being disclosed to unintended members of a user's friend network which leads to a "context collapse" and the information can be processed outside its intended or imagined context by the user [HLL11, MO11]. These problems arise mainly due to the inability of users to appropriately control access to their information, often due to the complexity and lack of usability of access control mechanisms provided to them by existing social media infrastructures [WKR14]. This is more evident from the analysis of the existing access control mechanisms in social media infrastructures presented next, in Section 2.2. After that, Section 2.3 presents related work in the area of access control recommendations which helps in identifying how the work presented in this thesis improves upon previously proposed approaches and advances the state of the art

2.2 Access Control in Current Social Media Infrastructures

In addition to understanding the various privacy threats and mitigation detailed in previous work mentioned in existing literature, it is also essential to examine the status-quo in terms of the access controls afforded to users by the current social media infrastructures in order to provide a holistic analysis. This is especially important as the nature of such mechanisms often shape the users' ability to safeguard their privacy.

This section provides a systematic evaluation of the top 30 social media sites [Ale14] (as of January, 2015) and classifies them into categories offering similar support to users by modeling social contexts through contact grouping mechanisms. Dating sites, online shopping sites and sites that were too specific to particular populations (for example, Classmates for US graduates and Naijapals for Nigerians) were excluded from this evaluation.

2.2.1 Categories

The evaluation of the 30 most popular social media sites in terms of the social aspect of the contact grouping mechanisms offered to the users resulted in five levels or categories. These categories are shown in Figure 2.1. The figure depicts that each level of contact grouping extends the previous level and builds on it. Therefore, a category at a higher level, for instance "predefined grouping", would incorporate the mechanisms of a lower level such as "binary grouping" and add further granularity for the user. In the rest of this section, the five contact grouping mechanisms in each category are described in detail and the 30 social media sites are categorized into these categories.

The individual social media sites were evaluated by looking at their privacy policies and settings offered to the users, especially in terms of managing contacts and supporting relationships. In most cases, user accounts were created to examine the actual available options for users, because this information was not directly evident from privacy policies. The 30 social media sites were thus classified into the five categories as shown in Figure

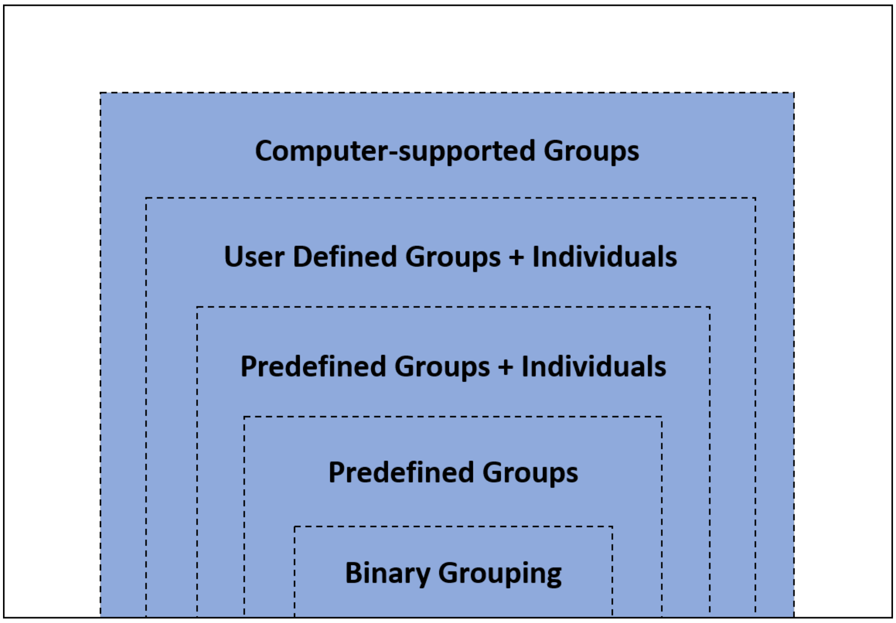

Computer-supported Groups
User Defined Groups + Individuals
Predefined Groups + Individuals
Predefined Groups
Binary Grouping

Figure 2.1: The different levels of contact grouping in social media sites

2.2.

Binary Grouping These social media sites have a binary distinction between a user's "friends" and "everyone else". This results in a scenario where everyone a user connects with is classified as a "friend" (other terms such as "connections" may be used in specific sites) and thus offering no granularity. The user does not have the opportunity to distinguish between two friends in terms of the relationship shared while implementing access control policies. If the user wants to share something on the network, she can either share it with all her friends or make it publicly available to everyone else. Such a situation clearly does not reflect real-life social relationships which are often more complex and varied. According to the evaluation, 14 out of the top 30 social media sites offered this kind of coarse grained friend networks to their users (see Figure 2.2).

Predefined Groups Some sites, such as *LinkedIn*, allow for predefined groups, which help users organize their contacts. Users on these sites can create groups to represent social relationships (for example, "colleagues" or "family") and use these groups while making access control decisions (for example, by sharing content to a specific group while not sharing it with others). While this is certainly better and more granular than the previously discussed binary classification, it still compels the user to treat all members within a group in a similar way. For example, if a user shares something to a "family" group, all members, without exceptions, will have access to that content. Figure 2.2 shows that there were 4 social media sites with such a mechanism.

Predefined Groups + Individuals Some social media sites such as *Tagged* and *Hi5* were found to provide their users with the opportunity of treating individuals separately from predefined groups. This provides more granularity to the user and is more akin to real life social relationships where even relationships of the same type such as a family member may not necessarily mean similar treatment when it comes to information disclosure. Four out of the 30 social media sites which were evaluated exhibited such grouping mechanisms (Figure 2.2).

User Defined Groups + Individuals The contact grouping mechanisms discussed so far have been predefined groups created by the social media site. Thus, the users of those social media sites would have empty groups such as "family", "colleagues", etc., and would populate them with the people they connect with on the network. Such grouping mechanisms offer no opportunity to the users to create their own groups in addition to or in place of the predefined groups provided by the social media site. An improvement is seen in sites such as *Google+* and *Twitter* where users have the ability to define their own groups ("circles", "lists", etc.). This enables them to create groups to mirror relationship types which they consider important while interacting on the social network. Figure 2.2 shows that 7 social media sites had such a grouping mechanism.

Computer Supported Grouping Figure 2.2 shows *Facebook* separately as it provides an enhanced grouping mechanism which is absent in any of the other social media sites which were evaluated. *Facebook* profiles contain a lot of personal information such as "location", "workplace", etc., and it uses this information to automatically create and populate "lists"[6] which can be leveraged by its users to make fine-grained access control

[6] `https://en-gb.facebook.com/help/135312293276793/`

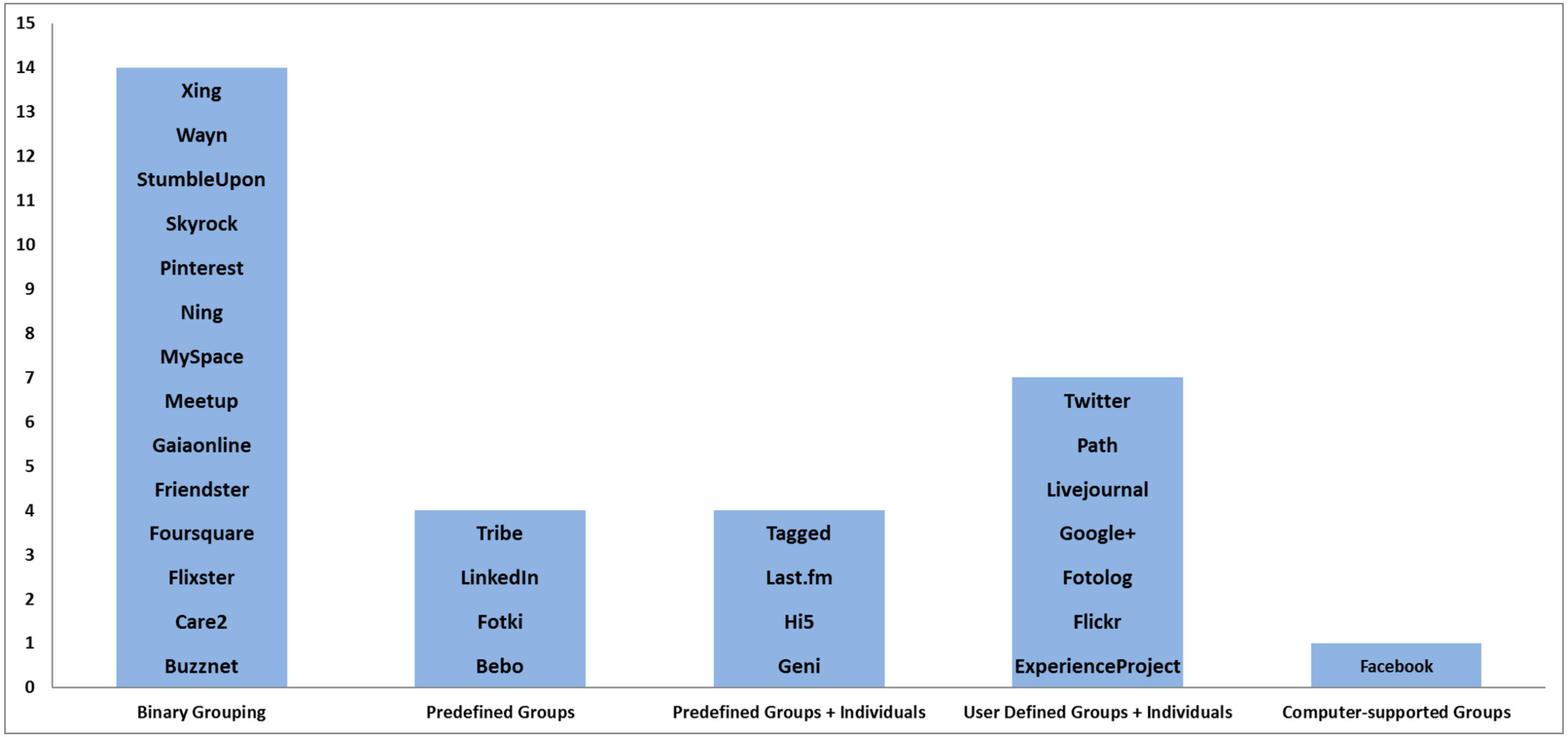


Figure 2.2: The top 30 social media sites classified according to the social aspect of their privacy controls and ordered alphabetically for each level

policies. It also provides the traditional pre-defined groups such as “friends” and “family” like the other social media sites. Thus, *Facebook* users have the opportunity to create their own groups, rely on the “lists” created from their contacts’ profile information as well as rely on the predefined groups to define the type of relationship they share with their contacts. This was found to be the most sophisticated approach out of all the social media sites that were evaluated in this analysis. Another interesting point to note was that *Facebook* provides the opportunity to users to identify friends as “close friends” or “acquaintances” which signifies an acknowledgement of the “strength” of the relationships shared between individuals on the site. Thus, it can be said that *Facebook* seems to be ahead of the other social media sites in terms of enabling its users to handle and manage social relationships.

2.2.2 Discussion

The evaluation of the top 30 social media sites presented in this section shows the different types of contact grouping mechanisms provided to the users by each of the social media sites. The access control mechanisms of these social media sites can leverage these contact grouping mechanisms, enabling the users to employ these groups to create access control policies. It is important to note that Twitter is an exception, as its “lists” are not used for access control but for subscription of content by the users. Twitter users create lists to “follow” (or subscribe to) the content they consider relevant or interesting. Moreover, Twitter is primarily a “broadcast” social medium and access control is not considered to be at the forefront of the users’ minds [MO11]. Therefore, an important distinction needs to be made between social media sites which employ contact grouping for access control mechanisms such as Facebook and those which do not, such as Twitter.

2.2.3 Summary

The evaluation of the top 30 social media sites shows that some popular social media sites are indeed moving in the right direction as far as supporting social relationships is concerned. *Facebook* is ahead of the others as it seemingly combines the type and strength of the interpersonal relationships in its grouping mechanism which can be utilized by its users when making access control decisions. However, it has been observed that users do not usually employ these mechanisms when making access control decisions [WKR14]. A possible reason for this could be the fact that the responsibility of maintaining the appropriateness of these communities lies solely on the users which puts a *cognitive burden* on them. Thus, the next step for social media sites would be to incorporate access control recommendations mechanisms which assist the users in making access control decisions while factoring in the relationship based factors such as type and strength. Such recommendations are generally based on learning from the access control decisions made by users. The primary contribution of the work presented in this thesis is an access control recommendation mechanism, REACT, which leverages interpersonal relationships in conjunction with the content being shared to recommend access control decisions to social media users. Such mechanisms have maximum utility in social media sites such as Facebook or Google+ where the contact grouping mechanisms are used to define interpersonal relationships and these relationships may be leveraged by REACT while providing recommendations. More details about the design of REACT and the types of information it relies on are described in Chapter 3. Section 2.3 discusses other previously proposed access control recommendation mechanisms which have been described in existing literature.

2.3 Related Work in Access Control Recommendation

This chapter has discussed a variety of privacy problems faced by social media users. One particular problem, “Unintended Disclosure” can be mitigated by providing access control recommendations to social media users. This would ease the burden of having to appropriately configure access controls to their information which they have often found to be struggling with [JEB12]. It was also seen in Section 2.2 that none of the popular social media sites provide such access control recommendations to their users. This section discusses and summarizes the research efforts which propose access control recommendation mechanisms.

There are previously proposed works in literature which try to address the problem of accurately predicting access control decisions made by social media users. Indeed, prediction of access control decisions and policies has been a well researched topic in several domains and there have been efforts to provide access control mechanisms in mobile and ubiquitous applications using factors such as location [ACD$^+$06, RKY06], temporal effects [JBLG05] and interconnection of devices [SKW15]. These factors, while being useful in other systems, have little scope of being utilized for access control in social media sites, where this information may not be always available. In such settings, the social context of information disclosure is considered essential to enable the formulation of access control policies in a way which preserves the "contextual integrity" of the information [MS16].

Previous work has leveraged the "social identity theory" [Hog16] to identify the privacy norms which enable definition of the social context of disclosure to assist the user in configuring appropriate access control policies [CLB$^+$16]. The social context can also be derived from information which facilitate the definition of social relationships on these media. These interpersonal relationships can be defined in terms of their "type" (friend, colleague, family, etc.) which is often represented using communities [FL10] and strength or "closeness" which is represented by similarity of profile attributes [AFW12, ML12]. The information to represent these relationships is therefore available in the social network itself and can be leveraged in any potential access control recommendation mechanism. In addition to social relationships, the content which is being disclosed is an integral part of the context of the disclosure and also plays an important role in the formulation of a desired access control policy. Therefore, the information about the content being shared (text, photos, etc.) can be used to predict and recommend access control decisions [SSLW11, KLM$^+$12]. The rest of this section identifies and summarizes some of the important efforts in literature which focus on access control prediction in social media, particularly leveraging social relationships and information about the content. Table 2.2, at the end of the section, shows the type of information that was considered some of the approaches discussed in this section.

2.3.1 Relationship Type

Social media users often interact with vast networks of friends with whom they share various types of social relationships outside the social network. An effective way of representing these interpersonal relationships between social media users is by partitioning their friend networks into communities. These communities can be created by using the network connections between the friends of a user by employing community detection algorithms [Dan09, FL10]. It has been proposed by several previous works that such communities can be leveraged to ease the burden on the user while making access control decisions while sharing on social media sites [CS14, JO10, Dan09, FL10]. It has also been found that users prefer to select their audience from predefined communities as compared to their entire friend lists [JO10] as they visualize their audience in the form of partitions or sub-structures [KBHC12]. This further enhances the argument of partitioning their friend networks into communities and leveraging these communities to enhance access control mechanisms. A particular way of reducing user effort in creating access control policies is to ask them to make decisions with respect to one or some members in a particular "community" (created by the algorithm), and then implement that decision for the other members in that community [FL10, CS14]. In this way, the user does not have to make decisions with respect to all their friends.

The problem with much of the existing work in this area is that the underlying assumption that members of the same community will be treated similarly has not been adequately examined empirically. This assumption depends heavily on the goodness of fit between a user's conception of their audience and the communities created by the algorithm. Moreover, most of these works implement only one algorithm for their experiments and a comparison and an evaluation of community detection algorithms in an access control context is absent. Such a comparison is essential before making any conclusions about the quality of fit that the communities created by the algorithms may have with access control decisions made by users. The analysis presented in Chapter 5 of this thesis shows a detailed empirical evaluation conducted to identify the best among 8 popular network based community detection algorithms in terms of a goodness of fit with access control decisions made by social media users. This analysis enables the identification of the most appropriate community detection algorithm to be used to represent relationship type in the access control recommendation mechanism presented in this thesis.

Alternative to using network based community detection, profile information may also be used to create communities in the friend networks of social media users [AFW12]. Indeed, as mentioned earlier in this chapter when describing the evaluation of social media infrastructure (Section 2.2), there is a shift towards better representation of relationships in mainstream social media sites such as Facebook and Google+ who have made an effort to enable users to represent their interpersonal relationships into communities by creating Lists [7] and Circles [KBHC12] respectively. For the recommendation mechanism presented in this thesis, only network based community detection algorithms were evaluated to represent relationship type as profile information is considered to represent strength of relationships (discussed later in this section). Thus, in this way, the access control recommendation mechanism considers both the network structure (in the form of network based community detection) as well as the appropriate profile information to represent the type and strength of relationships between individuals.

2.3.2 Relationship Strength

Users often provide a lot of information about themselves in the profiles they create on some social media sites such as Facebook. Of course, the extent and nature of the information provided depends on the particular social media site. This profile information can be used to inform and enhance access control mechanisms. One particular method of using profile information is by calculating similarity between users' profiles which can be used as a proxy for the strength of their relationship or the "closeness" between them [ML12, SKLD14, LLWG11, FSEGF14, ACF12]. This closeness can then be used to inform access control decisions making [FSEGF14, SKLD14].

The success of access control recommendation mechanisms relying on profile infor-

[7] https://en-gb.facebook.com/help/135312293276793/

mation depends on several factors. The effectiveness of profile based mechanisms can suffer due to missing information which can happen due to the users not providing information for certain attributes such as “religion”, “relationship status”, “political affiliation”, “age”, etc., as they may consider them to be sensitive [AFW12]. This would compromise the accuracy of the calculation of “closeness” and the subsequent access control recommendation as a result. When providing dynamic assistance, the time required to fetch the required information from the profiles of a user’s friends also needs to be kept to a minimum. Additionally, profile attributes which require personal information (such as personal communication, identifying information such as location, relationship status, etc.) may be considered sensitive by the user and can have privacy implications as a result. Hence, employing them in an attempt to safeguard privacy is obviously not ideal. Moreover, to create a solution which can be used across all platforms, the validity of the profile attributes across platforms needs to be considered as well. Considering platform specific attributes will not result in a holistic solution. Thus, achieving desirable results with profile attribute based mechanisms depends heavily on the choice of profile attributes. Chapter 6 presents detailed results of an empirical evaluation of 30 potential profile attributes which could be used to represent relationship strength in an access control recommendation mechanism. The findings of the evaluation enable the identification of the minimal subset of attributes which can represent relationship strength while overcoming the discussed challenges associated with profile information based mechanisms.

2.3.3 Content

In addition to the “who”, determined by social relationships, the “what”, in terms of information about the content, also needs to be considered to make informed access control decisions [SSLW11, KLM^{+}12]. Social media users share text and multimedia on social media sites and can provide a lot of useful metadata in the form of tags [LGZ08]. They have been found to annotate a lot of the information they post on social media sites to express various emotions, provide topical cues, target content to other users, etc. [MNBD06]. Such manually provided tags are also found to be very valuable for recommending relevant content to social media users [GZR$^+$10]. More pertinently for the work presented in this thesis, it has also been shown that such manually provided tags can be used to create access control policies in a way which is considered minimally intrusive for users [YKGS09, KLM$^+$12].

An alternative method to manual tags is to employ tools to automatically analyze the content and classify or categorize it to leverage this classification to inform access control decisions [SSLW11, SLSW15]. Such automatic classification can be beneficial as it removes the burden of annotating content from the user. However, it also means that the mechanism needs to analyze the content which can create a computational overhead. The computational complexity will of course depend on the nature and the amount of the content being posted. For example, the accuracy of image classification may be different to the analysis of text updates and mechanisms need to be in place which can handle multiple types of content being posted. Moreover, such analysis has to be performed in real-time in order to provide dynamic access control recommendations to the user. Another possible stumbling block associated with automated analysis of content is that such approaches require access to the content being posted. Mechanisms relying on the user to provide tags do not require to access and process the content itself and hence can be absolved of any privacy implications emanating from accessing it. Processing and analyzing the content being shared by social media users in order to recommend access control policies may lead to other privacy problems, especially from an "institutional privacy" perspective if the mechanism is part of the social media infrastructure itself.

2.3.4 Summary and Challenges Addressed by this Thesis

This section has discussed the previous works which focus on learning and recommending access control decisions to social media users by leveraging social relationships and the information about the content being shared. Table 2.2 summarizes these works according to the types of information they leveraged. It can be clearly seen that there is an absence of a holistic approach that combines all these types of information together to accurately

Table 2.2: The type of attributes which were considered by previous research

	Relationship Type	**Relationship Strength**	**Content**
Amershi et al. [AFW12]	✖	✔	✖
Cheek et al. [CS14]	✔	✖	✖
Danezis [Dan09]	✔	✖	✖
Fang et al. [FL10]	✔	✔	✖
Jones et al. [JO10]	✔	✖	✖
Li et al. [LLWG11]	✖	✔	✖
McAuley et al. [ML12]	✖	✔	✖
Misra et al. [MS16]	✖	✔	✖
Squicciarini et al. [SKLD14]	✖	✔	✔
Squicciarini et al. [SLSW15]	✖	✖	✔
Squicciarini et al. [SSLW11]	✖	✖	✔
Yildiz et al. [YK12]	✖	✖	✔

represent the overall context of information disclosure.

The primary contribution of this thesis is an access control recommendation mechanism, REACT, which is aimed towards filling this important gap by providing accurate access control recommendations which can be adapted to formulate a desired access control policy which can preserve the "contextual integrity" of the information being shared by the user. REACT uses information which helps in defining the type and strength of interpersonal relationships between users in conjunction with the information about the content to provide "allow"/"deny" type access control recommendations to social media users. However, as shown later in this thesis, while using all three types of information, it actually requires fewer attributes, and hence less information, when compared to previous approaches. Moreover, it is also shown that the attributes required by REACT are also minimally intrusive and always present in social media infrastructures.

Chapter 3

REACT

This chapter describes the overall design and the different components of REACT, an access control recommendation mechanism which leverages the type and strength of interpersonal relationships and information about the content to recommend access control decisions to the users.

Learning from access control decisions made by social media users and providing recommendations is an established method of trying to assist them by easing the burden of having to make appropriate access control decisions. For any such mechanism to be useful to the users, it is important to consider the social context of the information disclosure in order to provide meaningful access control recommendations which preserve the "contextual integrity" of the information [MS16]. The social context of the information disclosure can be derived from the interpersonal relationships between the sharer and the audience of the content. These interpersonal relationships can be defined in terms of their type (friend, family, etc.) as well as the strength or closeness (close friend, acquaintance, etc.). These factors are often considered important by the social media users when making access control decisions about information they share on the social media site [Fon11a]. In addition to the "who" in terms of the potential audience of the content and the users' relationship with them, the "what", the content itself, has an impact on the access control policy. For example, a person may not feel comfortable in sharing intimate details about embarrassing events with a person in their friend network

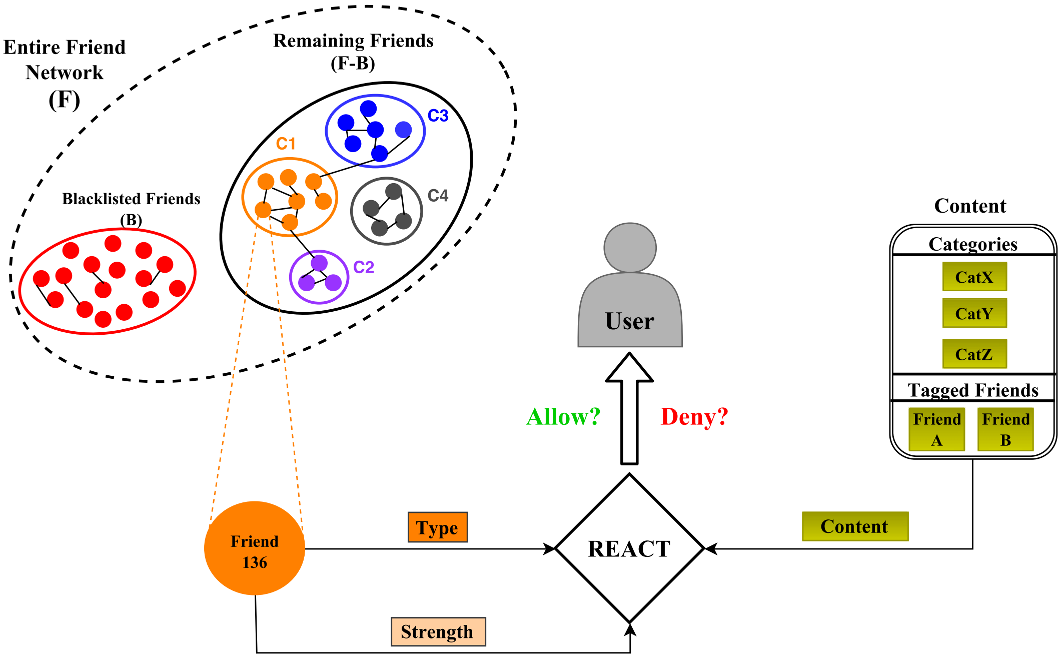


Figure 3.1: Various components of REACT

but might share mundane details with them. Thus, the nature of the content has an impact on the user's desired access control policy in terms of selecting an appropriate audience [SLSW15]. Therefore, in order to completely represent the overall context of the information disclosure, information about the content being shared needs to be considered in conjunction with information enabling definition of interpersonal relationships with respect to type and strength.

Figure 3.1 depicts REACT and the components it uses to make access control recommendations to the users. It has three components to account for the three types of information it uses to produce these "allow"/"deny" type access control recommendations to the user. The *Relationship Type* is represented by community membership of each friend which denotes which community, created by the community detection algorithm, they belong to [FL10]. The *Relationship Strength* is represented by similarity of profile attributes which denotes "closeness" between individuals [AFW12, ML12]. REACT considers the representation of the *Content* in the form of categories or "tags". As shown in the figure, REACT also leverages "blacklisted friends" in order to minimize the burden on the users as it only produces access control recommendations for the non-blacklisted friends. The rest of this chapter discusses the different components of REACT individually.

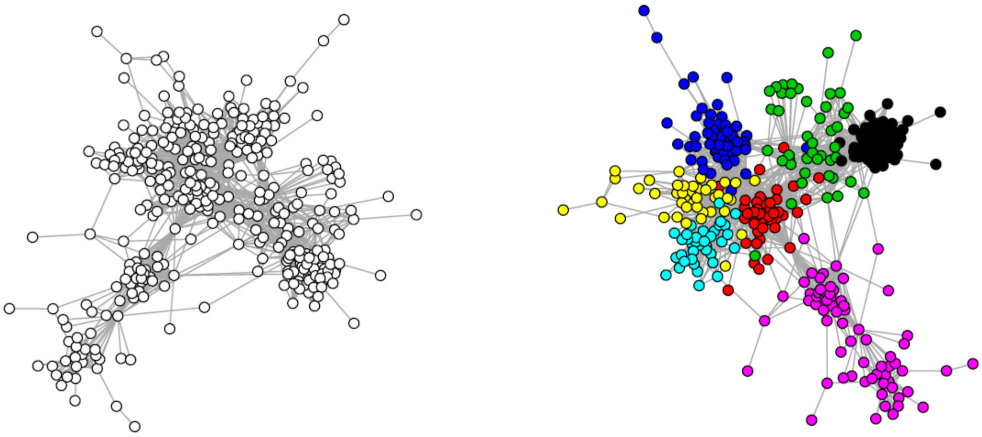
3.1 Components

This section introduces these different components of REACT and discusses why they are important to the design of REACT. The specific details about implementing each of the components is described in detail in Chapter 7 of this thesis.

3.1.1 Relationship Type

Social media users have vast friend networks of people who they connect with and may or may not know in real life [JEB12]. This means that they share different types of relationships (such as friends, colleagues, family, acquaintances, etc.) with members in their network. It is important to acknowledge this variation in the relationships between users and members in their friend network as it does have an influence on the access control behavior. For instance, users may be comfortable in sharing work related information with their office colleagues but not personal events which they may feel comfortable only sharing with their close friends or family.

REACT uses community membership to represent relationship type between a user and each of their friends. To do so, it requires the social network of the user, often represented as a graph of “nodes” representing each of the user’s friends and “edges” which represent the connections between them. REACT takes a network-based approach to partition the social network graph into substructures called “cliques” or communities [PKVS12]. Figure 3.2 shows an example of a friend network before and after the use of community detection algorithms. The communities that are produced from the friend network will not be directly used by the users while making access control decisions but only leveraged by REACT as an attribute (community membership of each friend of the user) in its mechanism to recommend access control decisions to them (refer Figure 3.1). Most community detection algorithms partition the network such that the connections between



(a) Friend Network without communities **(b)** Friend Network after community detection

Figure 3.2: Friend network shown before and after community detection using *Clique Percolation Method* [DPV05]

members within a community are more than the connections between members of different communities [PKVS12].

Network-based community detection is an extensively researched problem and has led to the development of many algorithms [PKVS12]. In such a scenario, it is essential to identify the community detection algorithm which is best suited to be used to represent relationship type in an access control recommendation mechanism such as REACT. Chapter 5 presents a detailed empirical evaluation of 8 well known community detection algorithms by examining a goodness of fit of the communities produced by the algorithms with the access control decisions made by users. More details about the community detection process as well as different types of network based community detection algorithms are also discussed there.

3.1.2 Relationship Strength

Social media profiles often contain a variety of information such as personal details, number of friends, photos, etc. The similarity in profile information of individuals can be used to calculate "closeness" or strength of the relationship between them [GK09]. Interpersonal communication such as messages, comments or posts can also be assimilated for this purpose. This relationship strength can be leveraged to inform users about the access control decisions they make while disclosing the information on the social media sites [FSEGF14] due to the known interplay between relationship strength and access control decisions [WKC$^+$11]. However, there are several challenges when using profile information in any access control enhancing mechanism:

- Many users leave profile attributes blank or set privacy controls so they cannot be retrieved

- There is a huge amount of potential attributes to be used and the associated time to fetch and analyze them needs to be minimal

- Some of the attributes could be seen as too privacy intrusive (e.g. entire conversations exchanged between users) to be used precisely to improve privacy and access control

- Some of the attributes are dependent on the specific social media platform.

Looking at these challenges, it seems imperative to identify the most suitable subset of profile attributes which can be used to represent relationship strength or closeness between individuals. To the best of my knowledge, such a systematic evaluation is missing in literature. Chapter 6 of this thesis presents a detailed empirical study of profile attributes which can be used to represent relationship strength and identifies a minimal subset of attributes which are used in REACT for this purpose.

3.1.3 Content

The design of REACT is agnostic to the type of content being shared, and it considers all “tags” or information about the content available, whether manually provided by users or automatically inferred by tools. REACT can easily be adapted to the type of content or indeed the method used to provide information about the content (manual or automatic) depending on the particular implementation of the design. Modern social media sites often afford the users the opportunity to provide such information through tags and research has shown that users often enhance the content by providing additional topical information through them [MNBD06]. Moreover, it has also been found that such manually provided "tags" are a minimally disruptive experience for the users [YKGS09]. These tags can be about the topic of the content (e.g., Flickr has a list of potential topic categories for the photos users upload) or further information about the social context (e.g., friends being tagged in photos or mentions being made to them in text posts). Topical tags, which help in categorizing the content, as well as social tags (friends being tagged) can help define the content and further contribute to the context of the disclosure and hence enhance access control recommendation.

An alternative method to manually tagging content is to use tools to automatically infer information about the content [KST$^+$08, SSLW11]. This can depend on the type of information being shared and the techniques would vary (for e.g. natural language processing for text or image processing techniques for photos). Both approaches, manual tagging and automatic analysis of content, have their advantages and disadvantages. While automatic analysis of content absolves the user of the responsibility of providing the information about the content, hence removing the burden off them, it may not accurately represent the users' idea of the content. Moreover, it would also necessitate the mechanism to access the content itself which may have privacy implications for the user. Facebook, for example, uses face-recognition techniques to automatically suggest social tags to users in the photos they upload on the site [ZSM11]. While such algorithms may work for social tags, depending on their accuracy, it has been found that users often interpret categories differently and automatic categorizing of photos into categories or topics may not be aligned with their notion of the content [LHY$^+$09].

3.1.4 Blacklisted Friends

Social media users often have vast friend networks consisting of many friends. However, it has been empirically found that social media users often intend to share their content with a limited subset of their entire friend network [KBHC12, MO11]. REACT leverages this intuition by maintaining a "blacklist" of friends for each user. These blacklisted friends can be assigned a default "deny" access control recommendation. Note that the specific number of blacklisted friends depends on various factors for individual users such as their access control behavior and their network size. Different social media users use the medium differently and their access control behavior and patterns are driven by their privacy preferences and their intentions [MO11]. For example, some users may want to use social media as a "broadcast medium" and connect to and interact with as many people as possible. For these users, the number of friends in the blacklist would be minimal. Alternatively, users may choose to interact with only a small subset of their friends and the blacklist for these users may contain a majority of their friends. Thus, the blacklist construction in REACT is a personalized and dynamic process in which a blacklist is learned based on previous access controls decisions and, hence, it entirely depends on the preferences and access control behavior of individual users.

3.2 Example

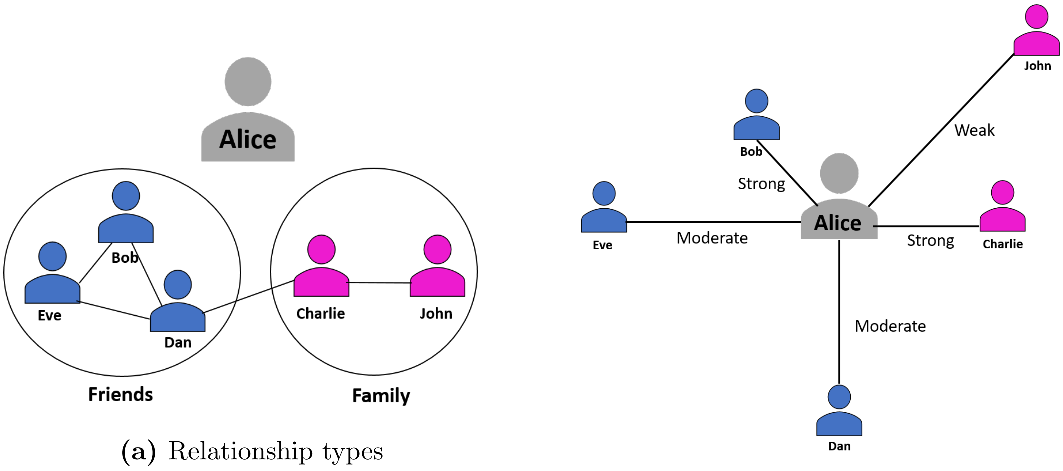
In this section, a minimal example scenario is described where the information representing the type and strength of relationships is used in conjunction with the category of the content by REACT in order to provide access control recommendations to the user.

Figure 3.3 shows a user *Alice* who has 5 friends, namely, *Bob*, *Charlie*, *Dan*, *Eve* and *John*. In Figure 3.3a, it can be seen that REACT produces two types of relationships between *Alice* and her friends after the community detection process. *Charlie* and *John* belong to the "Family" community while *Bob*, *Dan* and *Eve* are "Friends".

Figure 3.3b shows the strength of relationships between *Alice* and all of her friends which would be computed by REACT after processing the relevant profile attributes. *Bob* and *Charlie* are seen to be sharing a "strong" relationship with *Alice* while her relationship with *John* is seen to be "weak".

3.2.1 Access Control Scenario

Given the attributes related to relationship type and strength just described, it is important to understand how REACT would recommend access control decisions to *Alice* in a real-world access control scenario. Consider a situation where *Alice* wants to



(a) Relationship types

(b) Relationship strengths

Figure 3.3: The type and strength of relationships Alice has with her friends

share a photo about a party she had with friends. This photo may contain some privacy implication as she may not want to reveal to everyone in her friend network that she got drunk at the party. While sharing this photo through REACT, she categorizes it as "Party" and "Personal". REACT has learned from the previous access control decisions made by *Alice* that she only shares photos showing her in parties with people who are "close" to her but not with any members of her family.

It is clear that *John*, *Dan* and *Eve* will not be recommended an "allow" decision as none of them have a "strong" relationship with *Alice* (refer Figure 3.3b). *Charlie*, while having a "strong" relationship with *Alice*, is a family member (Figure 3.3a) and hence will be recommended a "deny" decision as well. This leaves only *Bob* who shares a "strong" relationship with *Alice* while not being a family member and will therefore be recommended an "allow" decision as shown in Figure 3.4. It is important to note that none of the 5 friends shown in this example were blacklisted based on previous decisions made by *Alice* and hence REACT had to provide access control recommendations for each one of them. This example illustrates how REACT considers the relationship type and strength between *Alice* and each of the members of her friend network in conjunction with the category of the content, as provided by *Alice*, to recommend access control decisions to her.

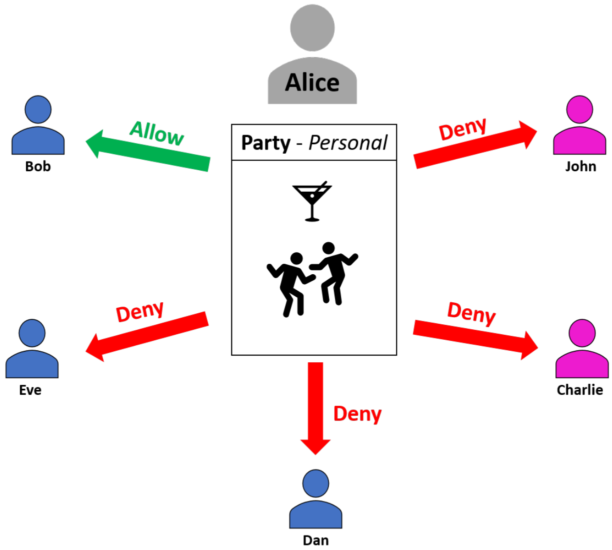


Figure 3.4: Access control recommendations made by REACT for each of Alice's friends

3.3 Chapter Summary

This chapter presented the overall design and the different components of REACT which leverages interpersonal relationships between the users and members of their friend network as well as information about the content being shared to recommend "allow"/"deny" type access control decisions. It also demonstrated a minimal example which described a simplified representation of how REACT can assist users by making access control recommendations by processing the relevant information about the interpersonal relationships and the content being shared. Chapter 4 presents the user study which was conducted to obtain the ground truth dataset of access control decisions made by social media users as well as the relevant information for the different components of REACT (relationship defining information as well as information about the content) in order to evaluate the performance of REACT. Chapter 5 presents the empirical study conducted to identify the most appropriate community detection algorithm to represent relationship type in REACT while Chapter 6 presents the empirical study conducted to identify the most suitable subset of profile attributes used by REACT to represent relationship strength. The results of the evaluation and the description of the particular implementation of REACT are presented in Chapter 7.

Chapter 4

User Study

The primary contribution of the work done in this thesis is an access control recommendation mechanism, REACT, the design of which was presented in the previous chapter. In order to understand whether REACT would be useful to social media users, it needed to be evaluated in an actual access control scenario. Obtaining ground truth access control decisions to evaluate access control mechanisms is a challenging task in itself and is further complicated when information from social media profiles is required [AFW12], as in the case of REACT in order to represent relationship strength. Conducting user studies are an established method of evaluating access control mechanisms [FL10, AFW12, SSLW11]. However, there have been several previous studies in which users were not required to make access control decisions but the evaluations relied on either qualitative feedback of using the mechanism [AFW12] or actions made by the user which were not in an access control scenario (e.g., creating groups to evaluate community detection algorithm without any given context) [FSEGF14]. It was considered important to evaluate REACT in an actual access control scenario where users would make access control decisions while disclosing information. There was an absence of a public dataset of such ground truth access control decisions and the information required to create the attributes necessary for the decision making mechanism of REACT and hence a user study was necessary to provide a thorough evaluation of the various components of REACT.

4.1 Experiment

The user study was conducted in one of the laboratories in the School of Computing and Communications at Lancaster University. The participants were required to use a Facebook application specifically designed and developed for this experiment. This application, and all the data collected during the course of the study, was stored on a secure server. The experiment was piloted by asking 4 colleagues to participate with the aim of ironing out any issues with the design of the application. This was an important phase as it led to minor alterations to the design of the database, which was storing the user data to be collected during the user study, making it easier to eventually analyze the information collected during the actual study. The pilot phase also reaffirmed that the application could handle simultaneous participation of multiple users without adversely affecting the integrity of the collected data. All data collected during the pilot phase was discarded before the commencement of the experiment with the actual participants.

The experiment was conducted after an ethical review by the Research Ethics Committee of Lancaster University. The review process required the submission of a description of the experiment, an information sheet which would be provided to the participants as well as consent forms which would be signed by participants prior to commencing their participation in the user study. The ethical review was an interactive process and the Ethics Committee asked for minor clarifications about the motivations of the user study which were duly communicated to them and which helped to improve the information sheet handed to participants (shown in Appendix A). The participants were informed a-priori that the experiment was being conducted in a simulated environment and none of their activity during the study would get uploaded to Facebook or affect their profiles in any way. They were, however, encouraged to make their access control decisions as if that was the case, i.e., if the photos were to end up being shared on Facebook. Each user provided informed consent before commencing their participation in the user study (consent form shown in Appendix B). The application created for the user study used Facebook Query Language (FQL) and the Facebook Graph API to interact with the Facebook databases. The application was kept alive for a duration of 5 days (in April

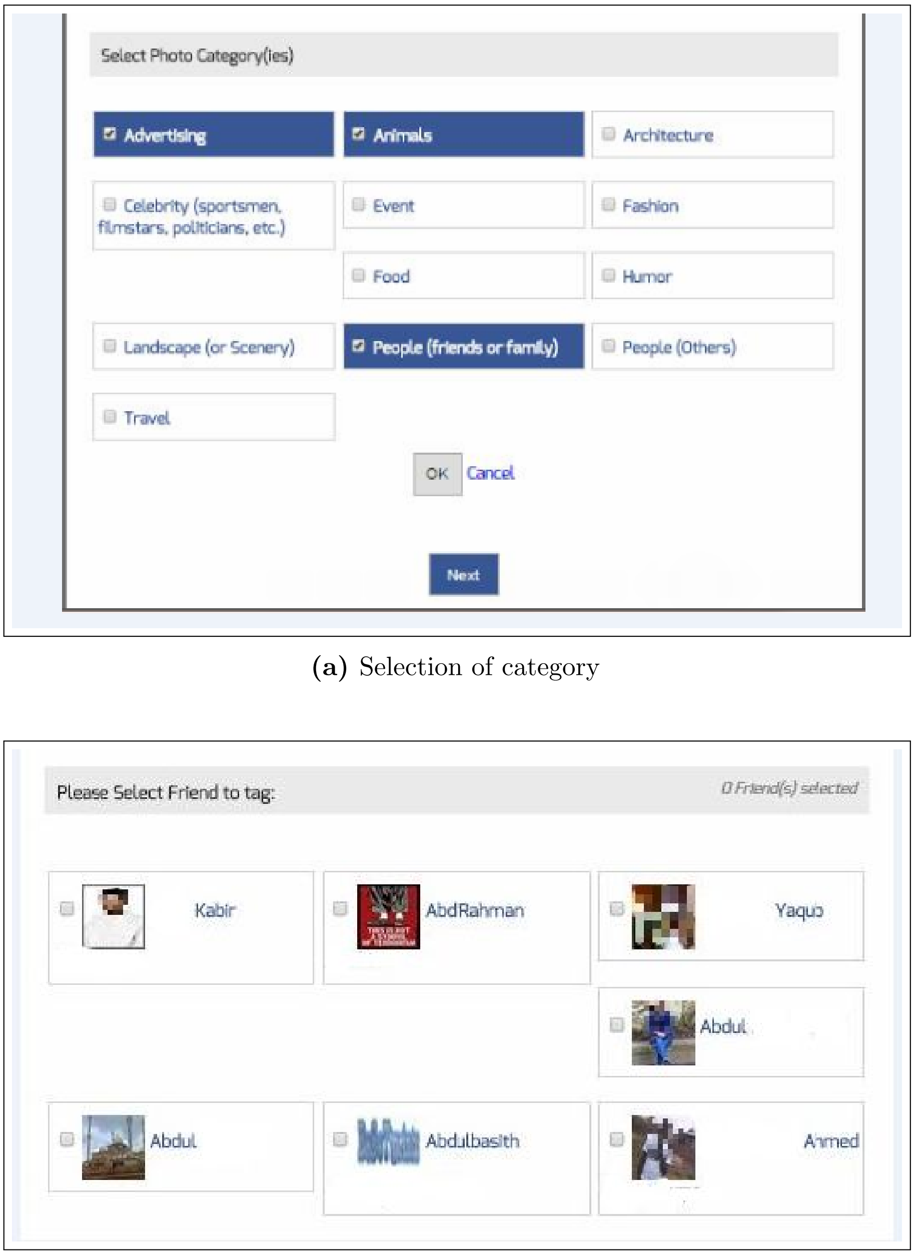
2015) during which the user study was completed.

In addition to accessing the users' profile data as well as the profile data of their friends, the application also downloaded five randomly chosen photos from each user's Facebook profiles. These photos were downloaded in order to be shown to the user during the study for them to make access control decisions. In addition, the participants were asked to select and bring 5 other photos which they had not yet uploaded on Facebook. This was done to avoid a scenario where a user makes access control decisions for all photos during the study for which they had already received comments and likes before as that may have influenced their decisions. Therefore, we wanted them to make some access control decisions about content they had never previously shared on Facebook. The participants were also advised to bring photos which they considered to be personal (either included them or a family member) or considered sensitive so that they had a privacy implication. The different stages of the user study were:

1. The participants logged into the application using their Facebook credentials. They were then alerted about the data that would be accessed and asked for explicit permissions before moving on.

2. The participants were shown 10 photos (5 from Facebook and 5 they brought as detailed earlier) sequentially on the screen, each on an individual page. They were asked to select categories for the photos from a predefined list of 15 popular photo categories (taken from Flickr, shown in Figure 4.1a). They were also given the opportunity to tag any friends in the particular photo (Figure 4.1b).

3. Once they had selected categories and tagged their friends, the users were asked to select each friend who they wanted to disclose the photo to (Figure 4.2). The friend list was shown alphabetically to the participants to imitate the organization Facebook uses to show friend lists to its users. They were explicitly told that any friend who was not selected would be denied access to the photo. All the selections



Select Photo Category(ies)

☑ Advertising
☑ Animals
□ Architecture
□ Celebrity (sportsmen, filmstars, politicians, etc.)
□ Event
□ Fashion
□ Food
□ Humor
□ Landscape (or Scenery)
☑ People (friends or family)
□ People (Others)
□ Travel

(a) Selection of category

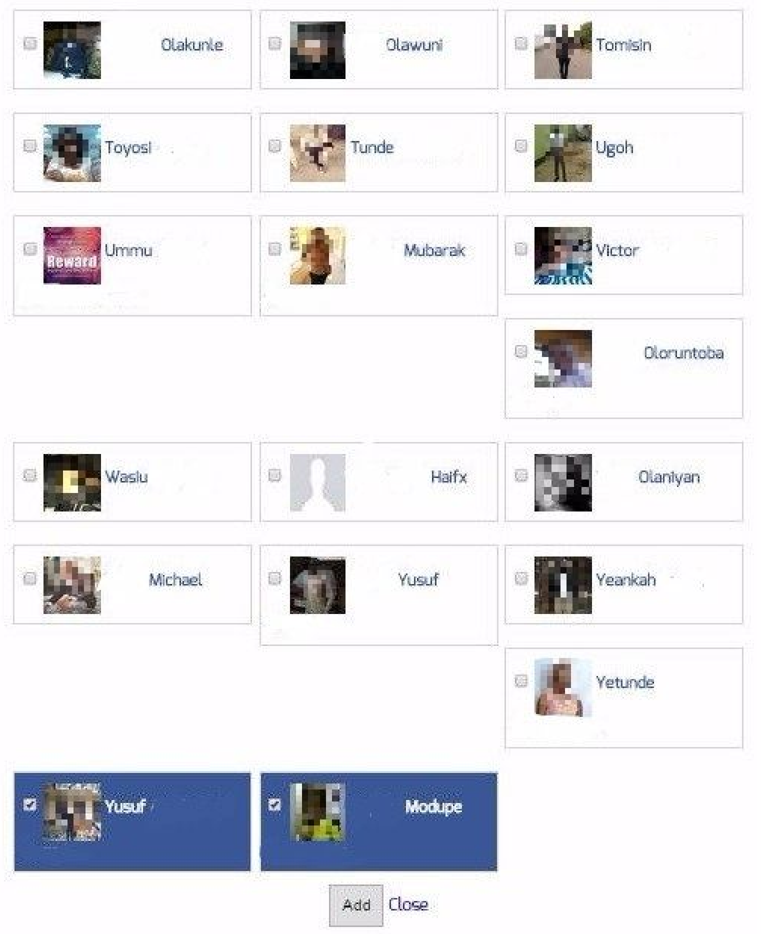
Please Select Friend to tag:

0 Friend(s) selected

□ Kabir
□ AbdRahman
□ Yaqub
□ Abdul
□ Abdul
□ Abdulbasith
□ Ahmed

(b) Selecting friends to tag (faces and parts of names hidden to safeguard privacy of people appearing in the photos)

Figure 4.1: Selecting categories and tagging friends



☐ Olakunle ☐ Olawuni ☐ Tomisin

☐ Toyosi ☐ Tunde ☐ Ugoh

☐ Ummu ☐ Mubarak ☐ Victor

☐ Oloruntoba

☐ Wasiu ☐ Haifx ☐ Olaniyan

☐ Michael ☐ Yusuf ☐ Yeankah

☐ Yetunde

☑ Yusuf ☑ Modupe

Figure 4.2: Selecting an audience for the photo (faces and parts of names hidden to safeguard privacy of people appearing in the photos)

made by the users for each of the 10 photos were stored in the database.

While the users were completing the above mentioned process, the application fetched their friend network and the profile information for each of their friends. The friend network was used as input to the *Community Detection* process which was performed for each user after the completion of the study. The profile information of each of their friends was later converted to generate the relevant *Profile Attributes* for our analysis presented in Chapter 6. The *Photo Categories* were selected by the users during the process as mentioned. In addition to the categories, the *Tagged Friends* were also recorded to complete the information about the content.

4.2 Participants

The eventual sample considered for the analyses presented in this thesis consisted of **26 participants**. The participants were recruited primarily from among the staff and students of Lancaster University. Additionally, there were some participants who were external to the university and were invited through personal communication channels such as email, social networks, etc. Details about the privacy implications and the overall objectives of the experiment were communicated to the registered participants and it was conducted only after the explicit consent of the participants. All participants were compensated with £10 for their involvement in the study.

Typical pre- and post-experiment checks were performed after completion of the user study in order to maximize data quality. In particular, before the experiment participants were screened and everyone who had a Facebook account having 100 friends and had uploaded at least 10 photos before the study was invited to participate in the user study. After an initial registration phase where the participants expressed their interest in being involved in the study, 31 participants were selected to take part.

After completion of the user study, 4 participants were identified who had granted access to groups of alphabetically sorted friends for each photo. These groups were different for each photo and it seems that these participants just randomly scrolled to

	Age	**No. of Friends**
Average	29	265
Std. Dev.	6	121
Median	28	265
Max.	55	519
Min.	21	153

Table 4.1: Demographic details of participants

a section of their friend list and granted access to consecutively listed friends. There was another participant who granted access to 1 friend for each photo but it was a different friend each time. As the participants were explained that they were to make access control decisions and that the friends who they were not selecting would be denied access to the information, it is implausible that a user would grant access to only one friend and it would be a different friend each time. It was concluded that the access control decisions made by these 5 participants were not according to the directions of the user study. The data collected from the participation of the remaining 26 users has been considered to produce the results reported in this thesis. The 26 users which were considered for the analyses consisted of 15 males (57.7%) and 11 females (42.3%). The details of the participants' age and size of friend network are shown in Table 4.1. We find that the average age of the participants was 29 years (s.d = 6) and the average size of network was 265 friends (s.d = 121).

4.3 Dataset

The user study was conducted with the dual aim of obtaining the relevant information to implement the different components of REACT (described in Chapter 3) as well as gathering the ground truth access control decisions to evaluate the access control recommendations made by it. This section provides a description of the data that was collected during the user study and how the different components were represented.

As discussed earlier in Section 4.1, each participant had to make access control decisions with respect to each of their friends for 10 photos, one after the other. There were 26 participants who made access control decisions for 6884 friends in total. The total

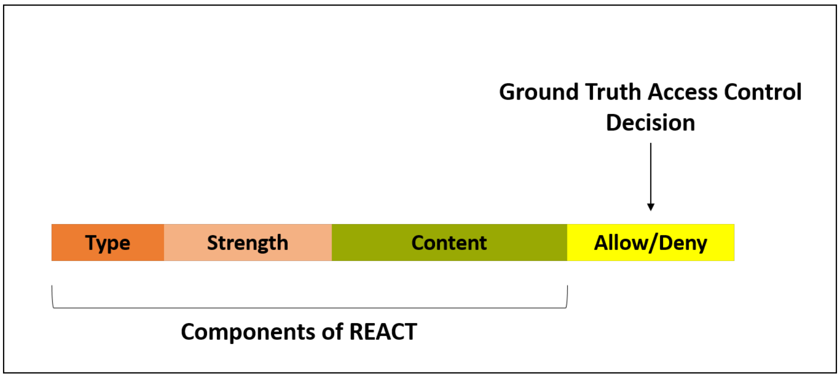


Figure 4.3: Representation of a particular instance in the dataset

number of access control decisions made by all participants during the user study, and therefore the size of the ground truth dataset, was **68,840 access control decisions**. Each instance in the dataset represented information required to create attributes representing the different components of REACT (which were shown in Figure 3.1 in Chapter 3) as well as the ground truth access control decision made by the user during the user study. An illustration of how each instance was represented in the dataset is shown in Figure 4.3.

Relationship Type is represented using community membership in REACT. To create communities for each user, their friend networks were used as input to the community detection process. In this way, each and every friend of a particular user was assigned a community membership which constitutes the relationship type information in each instance as shown in Figure 4.3. More details about the choice of the particular community detection algorithm are provided in Chapter 5 where an empirical study is described. That study compares 8 popular community detection algorithms and chooses the best performing algorithm which is then used to implement relationship type in REACT. The implementation of the community detection algorithm and the representation of community membership is discussed in more detail in Section 7.1 of Chapter 7 alongside the implementation of other components of REACT.

The *Relationship Strength* information shown in Figure 4.3 contains the profile attributes chosen to represent the strength of the relationship between a user and a particular friend. To identify the most suitable set of profile attributes for this purpose, all

No.	Category	Number of Photos	Order
1	People (Friends)	69	3
2	Personal	60	14
3	Travel	54	6
4	Entertainment	51	13
5	Event	35	9
6	Humor	22	10
7	Landscape	21	1
8	Architecture	20	4
9	People (Others)	14	2
10	Food	13	11
11	Animals	11	5
12	Fashion	6	7
13	Technology	6	15
14	Advertising	4	12
15	Celebrity	3	8

Table 4.2: Number of photos from each category as selected by the participants

available profile data of each of the participants' friends was downloaded during the user study and analyzed. Due to the API restrictions imposed by Facebook (discussed in more detail in Chapter 6), the number of friends for whom all profile attributes were available was less than the total 6884 friends. The analysis of profile attributes is described in detail in Chapter 6.

For the *Content* information in REACT, the participants were mandated to select at least one category from a list of 15 predefined categories during the experiment. They were given the option of selecting more than one category for a photo if applicable. Table 4.2 shows the number of photos in each category as selected by the participants during the study. The "order" in the last column shows the order in which they appeared on the list shown to the participants. It is quite clear that the order had no effect on the participants' selections as "Personal" and "Entertainment" appeared low on the list but were selected for many photos by the users as shown in the table. Overall, it can be seen that "People (Friends)", "Personal", "Travel" and "Entertainment" were the most selected categories whereas "Advertising" and "Celebrity" had the lowest number of photos. An important observation here is that the total number of photos, adding up the numbers in column 3, is 389 which is much more than the 260 photos shown to the participants. This is due to the fact that the participants were able to select multiple categories for the photos as was discussed earlier in Section 4.1.

The participants were also given the opportunity of “tagging” any of their friends in the photos during the study. The collected data shows that out of the 260 photos shown to the participants, there were **122 (46.9%)** photos where the participant tagged at least one friend. There were 4 participants who did not tag any friends in any of their 10 photos. On the other hand, 4 out of the 26 participants tagged at least one friend in each of their 10 photos.

Therefore, the content information shown in Figure 4.3 includes the attributes representing the categories of the particular photo as selected by the user during the experiment as well as any social tag made by them.

In addition to information representing each of the components of REACT, the instances in the dataset also included the ground truth access control decisions made by the user during the study (refer to Figure 4.3). These decisions were used to evaluate the performance of the classifier which was trained using the attributes representing each component of REACT. This evaluation and the results are described in detail in Chapter 7 of this thesis.

4.4 Chapter Summary

This chapter presented the details of the user study which was conducted in order to enable evaluation of REACT in an actual access control scenario. The experiment was designed to obtain ground truth access control decisions made by users, their friend networks, to create attributes to represent relationship type, profile information of the users and their friends, to create attributes to represent relationship strength, and annotations representing the photo categories and social tags, to represent information about the content in the implementation of REACT. During the experiment, 26 participants made access control decisions corresponding to each member in their friend network for 10 photos, one after the other, which resulted in a ground truth dataset consisting of 68,840 access control decisions.

Chapter 5

Relationship Type

Chapter 3 showed the design of REACT where it was described that the relationship type component uses community membership. In order for REACT to provide accurate access control recommendations, it is essential to identify the most suitable community detection algorithm which can be used to represent relationship type in an access control scenario. This chapter presents the findings of an empirical evaluation of 8 popular network based community detection algorithms with the aim of finding the most suitable algorithm which has the best fit with access control decisions made by users.

There have been a number of works that proposed using community detection algorithms to facilitate the definition of access control policies [CS14, JO10, Dan09, FL10]. The problem with much of the existing work in leveraging communities for access control mechanisms, however, is that the underlying assumption that members of the same community will be treated similarly has not been adequately examined empirically. This assumption depends heavily on the goodness of fit between a user's conception of their audience and the communities created by the algorithm. Moreover, most of these works implement only one algorithm for their experiments and a comparison and an evaluation of community detection algorithms in an access control context is absent. Such a comparison is essential before making any conclusions about the quality of fit between communities created by the algorithms and the access control decisions made by users. Fogués, et al. [FSEGF14] did test three community detection algorithms, but the algorithms were compared in terms of whether the communities created would be accepted as such by users, not in terms of their direct goodness of fit with access control decisions which is crucial to help automate recommendations as much as possible. The empirical evaluation presented in this chapter examines 8 popular network based community detection algorithms in order to identify the algorithm which has the best fit with access control decisions made by social media users.

5.1 Algorithms Considered

Network based community detection algorithms require the friend network (friends as nodes and connections as edges) as input and produce communities as an output. Such algorithms have been categorized by Papadopoulos, et al. [PKVS12] into the following types depending on their methodology:

1. *Cohesive Subgraph Discovery* - This method involves dividing the network into subgraphs. Generally, the structure of these subgraphs is known a-priori. Some of these structures are cliques, n-cliques, k-cores, LS sets and lambda sets [PKVS12]. Algorithms for enumerating such structures, such as the Bron-Kerbosch algorithm [BK73] and the efficient k-core decomposition algorithm of Batagelj and Zaversnik [BZ03] belong in this category. In addition, methods such as the Clique Percolation Method [PDFV05] and the SCAN algorithm [XYF+07] which lead to the discovery of subgraph structures with well-specified properties fall in the same category.

2. *Vertex Clustering* - A typical means of casting a graph vertex clustering problem to one that can be solved by conventional data clustering methods (such as k-means and hierarchical agglomerative clustering) is by embedding graph vertices in a vector space, where pairwise distances between vertices can be calculated. Another popular method is to use the spectrum of the graph for mapping graph vertices to points in a low-dimensional space, where the cluster structure is more profound [DM04]. Other vertex similarity measures such as the structural equivalence

[BBA75] and the neighborhood overlap have been used to compute similarities between graph vertices [Was94]. A different method, called Walktrap [PL05] makes use of a random-walk based similarity between vertices and between communities and uses modularity in a hierarchical agglomerative clustering scheme to derive an optimal vertex clustering structure.

3. *Community Quality Optimization* - There are a very large number of methods that are founded on the basis of optimizing some graph-based measure of community quality. Subgraph density and cut-based measures, such as normalized cut [SM00] and conductance [KVV04] are some early examples of this approach. The seminal "greedy optimization technique" of Newman [NG04] and speeded up versions of it, such as max-heap based agglomeration [CNM04] and iterative heuristic schemes [BGLL08] are all related to measuring modularity in networks and using it as a metric for dividing the network into groups or communities. A relatively sophisticated method based on "extremal" optimization by Arena, et al. [DA05] is worth a mention in this category as well.

4. *Divisive Algorithms* - This method aims to find relatively disconnected components of the network and create communities such that edges between these communities are reduced. The algorithm by Girvan and Newman [GN02] is an example of this approach and it progressively removes the edges of a network based on some edge betweeness measure until communities emerge as disconnected components of the graph. Several measures of edge betweeness have been devised, for instance, edge, random-walk, and current-flow betweeness [NG04], as well as information centrality [FLM04] and the edge clustering coefficient [RCC$^+$04]. Finally, min-cut/max-flow methods [FLG00, IKN05] adopt a different divisive perspective: they try to identify graph cuts (i.e. sets of edges that separate the graph in pieces) that have a minimum size.

Table 5.1: List of network based community detection algorithms evaluated

Algorithm	**Type**	**Complexity**
FastGreedy (FG)	Community Quality Optim.	$O(n \log^2 n)$
Walktrap (WT)	Vertex Clustering	$O(n^2 \log n)$
Infomap (IM)	Model Based	$O(n \log n)$
Girvan-Newman (GVN)	Divisive	$O(n^3)$
Label Propagation (LP)	Model Based	$O(n)$
Leading Eigenvector (LEV)	Community Quality Optim.	$O(n^2 \log n)$
Multi-level Community (MC)	Community Quality Optim.	$O(n)$
Clique-Percolation (CP)	Cohesive Subgraph Discovery	$O(exp(n))$

5. *Model based Algorithms* - A broad category of methods that either consider a dynamic process taking place on the network, which reveals its communities, or consider an underlying model of statistical nature that can generate the division of the network into communities. Examples of dynamic processes are label propagation [RAK07, LHLC09, Gre10], synchronization of Kuramoto oscillators [ADGPV06], diffusion flow, better known as Markov Cluster Algorithm, and the popular spin model by Reichardt and Bornholdt [RB06]. Other model-based approaches rely on the principle that a good clustering is determined by a low encoding cost, thus they perform community detection by finding the cluster structure that results in the lowest possible cluster encoding cost [Cha04, RB08].

Table 5.1 lists the 8 community detection algorithms which are evaluated in this chapter. The complexity of the different algorithms as shown in the table are obtained from [PKVS12] and [For10].

5.2 Goodness of Fit Metrics

Three metrics were employed in order to examine a goodness of fit between communities produced by the algorithms and the access control decisions made by users. The metrics are defined such that they acknowledge the willingness of users to share *selectively* in communities [KBHC12] but also account for the *effort* required from the user to modify the communities into an audience for their content [FSEGF14].

5.2.1 Number of Communities per Audience

Intuitively, an algorithm is considered more useful if the user needs to select from a small number of communities to build the entire audience for a photo. Thus, the algorithm with a low average number of communities to complete an audience is considered a better fit for audience selection.

Definition 1. Given the set of friends U, the set of communities $\mathbb{C}$, the particular photo p, and the audience for the photo A_p, the number of communities per audience is:

$$G_p = |\{C \mid C \in \mathbb{C} \land \exists u \in A_p, u \in C\}|$$

For example, if a user selects 50 members in an audience for a particular photo and 18 are from community A, 22 from community B and the other 10 from community C according to the communities created by the Fastgreedy algorithm, the value for this metric will be 3 (as 3 communities are represented in the audience) for Fastgreedy for that particular photo.

5.2.2 Ratio of audience in largest community

If the communities produced by an algorithm can be readily used to create an audience, then the burden on the user is minimized. Thus, we calculate the percentage of audience members belonging to the largest community represented in the audience.

Definition 2. Given the set of friends U, the set of communities $\mathbb{C}$, the particular photo p, and the audience for the photo A_p, the ratio of audience in largest community is:

$$R_p = \frac{\max_{C \in \mathbb{C}} |\{u \mid u \in A_p \land u \in C\}|}{|A_p|}$$

The ratio is represented as a percentage of the total number of friends in the audience. For example, if a user selects an audience of 50 friends and 18 of them are from community A, we would calculate the ratio as 36% (18/50).

5.2.3 Penalty for Exclusion

When the user is employing the communities produced by the algorithms for audience selection, it is possible (and probable) that he would need to exclude some of the friends from some of the communities to create an audience of his choice. After all, community membership does not guarantee that all friends in the same community would always be treated in a similar way by the user. Such an exclusion from an audience will require effort from the user. Thus, the algorithms need to be evaluated with a metric that measures the number of friends in a given community who were not included in the audience for a particular photo.

Definition 3. Given the set of friends U, the set of communities $\mathbb{C}$, the particular photo p, and the audience for the photo A_p, the penalty for exclusion is:

$$E_p = \sum_{C \in \mathbb{C}} |\{u \mid u \notin A_p \wedge u \in C \wedge \exists u_2 \in A_p, u \neq u_2 \wedge u_2 \in C\}|$$

Consider the example used to describe Definition 1, where the user selects 18 audience members from community A, 22 from community B and 10 from community C to select a total audience of 50 friends. Suppose community A has 30 total members, B has 25 and C has 10. This would mean that the user would have to manually exclude 12 (30-18) members from community A and 3 (25-22) from community B to achieve the desired audience. Thus, the total penalty for exclusion for selecting such an audience would be 15 (12+3). Note that there is no penalty corresponding to community C as the user selected all 10 of its members in the audience. The penalty is calculated using the equation in Definition 3 as a summation of all those friends who would have to be excluded by the user after selecting their entire community as an audience. Intuitively, a lower penalty value means that the communities provided by the algorithm has a better fit with the access control decisions made by the users.

5.3 Results

The 8 algorithms which are evaluated in this chapter were implemented using iGraph [CN06] and SNAP libraries [LS14]. The communities were created using each of the algorithms for each user from their friend networks which were downloaded during the user study (described in Chapter 4) as a list of "nodes" representing each friend and "edges" representing links between them. The goodness of fit between the communities created by each of the algorithms and the ground truth dataset of 68,840 access control decisions obtained during the user study was evaluated using the metrics described earlier.

5.3.1 Overall Results

For the results discussed in this chapter, the value of each metric is averaged across 10 photos for each individual user and then aggregated for all users for each algorithm.

Communities per Audience

From the results in Fig 5.1, it is clear that Clique Percolation Method (CP) performs slightly better than the other algorithms. The results indicate that a user needs to traverse through less than 4 communities (3.42) on average to complete their audience selection if using the communities produced by CP. Multilevel (3.86), Leading Eigenvector (4.08), Label Propagation (4.15) and Fastgreedy (4.35) produce communities such that a user may need to access less than 5 communities to complete the audience. Infomap produces the least impressive performance with respect to this metric (6.49).

In order to understand whether the dataset followed a normal distribution or not, a Kolmogorov-Smirnov test (KS Test) [Lil67] was conducted. It was found that the dataset was significantly deviant from normal distribution ($p < .001$). This meant that an ANOVA test was not possible in order to examine any statistically significant difference in the performance of the algorithms with respect to the metrics. Therefore, a Kruskal-Wallis H Test [KW52], which checks for significant difference between treatments (the 8 community detection algorithms in our case) and does not require the dataset to be normally distributed, was conducted. For communities per audience, the Kruskal-

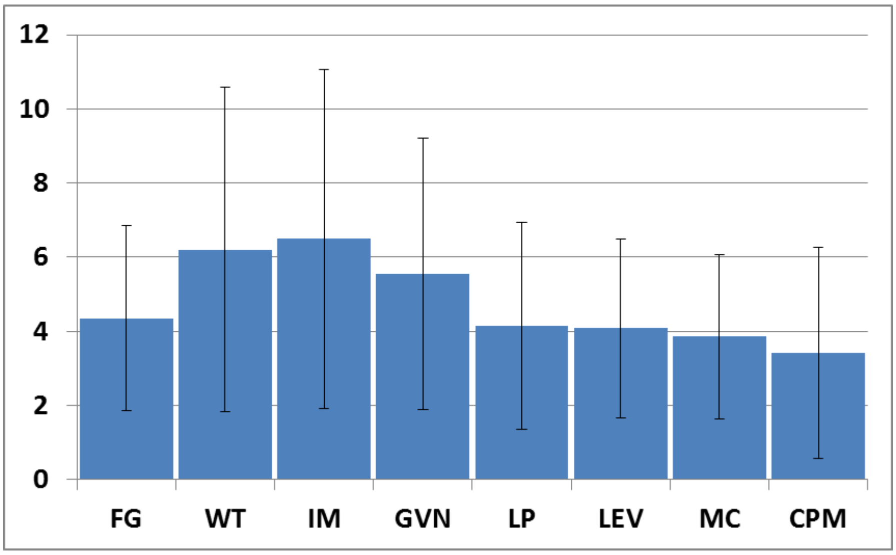


Figure 5.1: Average no. of communities required to create audience for each algorithm

Wallis test showed that there is a significant difference between the algorithms ($p < .05$). To identify the source of difference, a Mann-Whitney U Test [MW47] was conducted and showed that CP performs significantly better than Walktrap, Infomap and Girvan-Newman. There were no other combinations for which the difference between the algorithms was statistically significant.

Ratio in Largest Community

Looking at Fig 5.2, it is clear that CP is the best performer compared to all other algorithms for this metric. On an average, 76.1% of the audience can be selected from a single CP community. All other algorithms produce similar performance to each other where approximately 60% of the audience can be selected from the same community.

The Kruskal-Wallis test for this metric also showed a significant difference between the different algorithms ($p < .05$). The Mann-Whitney U Test revealed that CP had a significant difference in performance with all other algorithms. None of the other 7 algorithms had a statistically significant difference in performance when compared with each other.

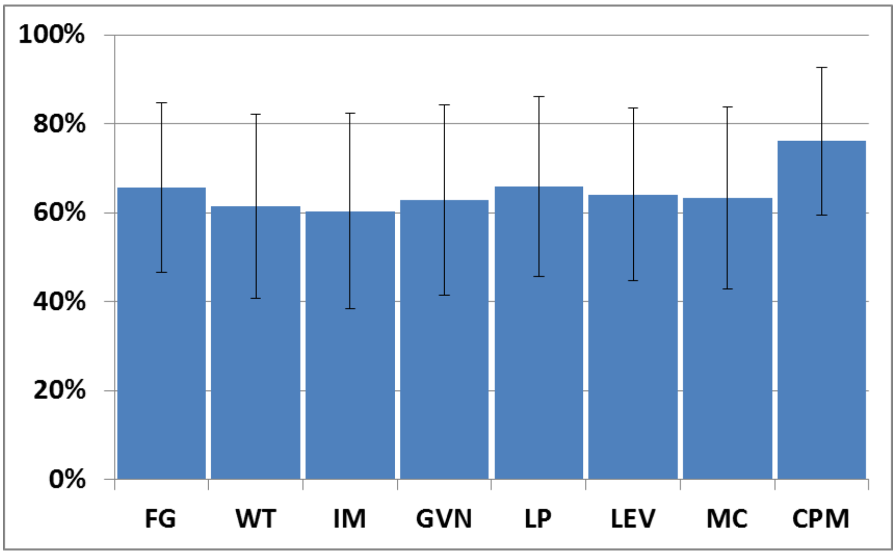


Figure 5.2: Average ratio of largest community in audience for each algorithm

Penalty for exclusion

The penalty values in Fig 5.3 signify the average number of friends the user would have had to exclude to obtain the desired audience from the communities created by the algorithm. As can be seen from the results, Infomap performs better than all other algorithm with a penalty value of nearly 122. The highest average penalty, and hence least desirable performance, is seen for Fastgreedy which produces average penalty value of 159.292.

Kruskal-Wallis test for penalty for exclusion revealed that there was no significant difference between the 8 algorithms for this metric ($p = 0.850$). This can also be anticipated by looking at the descriptive statistics shown in Figure 5.3 where the performance of the algorithms is only marginally different.

5.3.2 Effect of Individual

Having looked at the overall results of our evaluation, aggregated across all users, it was important to examine whether the individual characteristics of the users had any effect on the performance of the algorithms. This was done with the objective of trying to see whether an appropriate algorithm can be chosen with respect to individual users.

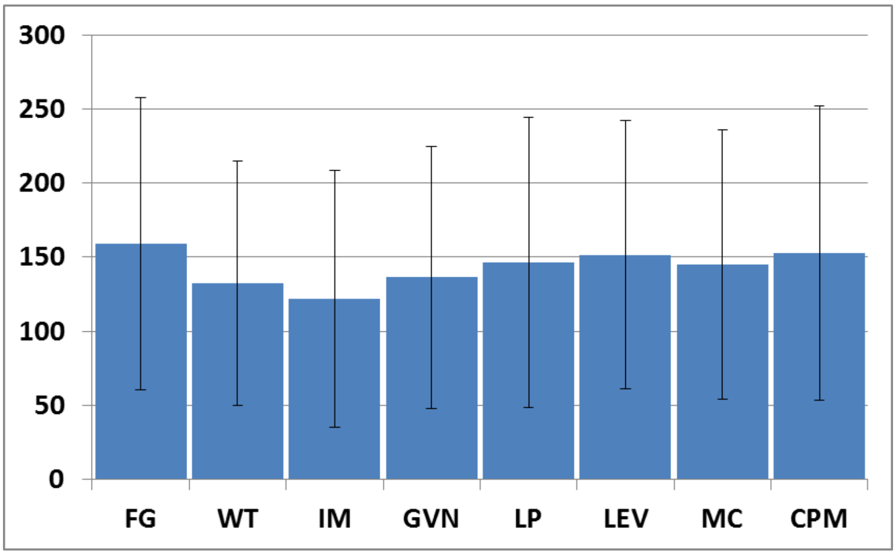


Figure 5.3: Average penalty for exclusion for each algorithm

Personal Characteristics

Table 5.2 shows Pearson correlation for the performance of the algorithms according to the three evaluation metrics with respect to gender and age of the participants as well as their size of friends network and average audience size per photo.

It can be seen from the figure that Communities per Audience and Ratio in Largest Community remain unaffected by Gender, Age and Size of a user's network for all algorithms. The only characteristic that affects these metrics is the Average Audience Size. It positively affects the number of communities per audience as expected (more people, more communities required). On the other hand, it has negative correlation with ratio of audience in largest community. If a user selects larger audiences, it is unlikely that the audience will be constituted from a single community or less number of communities. Looking at penalty for exclusion, age of the user has a weak positive correlation for all algorithms except CP. Males were coded as '0' and females as '1' for the binary correlation analysis and hence it can be concluded that males are more likely to have a higher penalty for exclusion as compared to females. It is also evident that the size of a user's friend network has strong positive correlation with penalty for exclusion for all algorithms. Thus, if a user has a large number of friends in their network, they are more likely to have a higher penalty for exclusion. In terms of particular algorithms,

Table 5.2: Correlation of performance of algorithms with respect to characteristics of individual

	FG	**WT**	**IM**	**GVN**	**LP**	**LEV**	**MC**	**CP**
Communities per Audience								
Gender								
Age								
Size								
Avg. Audience	+ + +	+ + +	+ + +	+ + +	+ + +	+ + +	+ + +	+ +
Ratio in largest community								
Gender								
Age								
Size								
Avg. Audience	−−−	−−−	−−−	−−−	−−−	−−−	−−−	
Penalty for Exclusion								
Gender								−
Age	+	+	+	+	+	+	+	
Size	+++	++	++	++	++	++	++	+++
Avg. Audience								

Strong Correlation (+ + + or − − −) : Coefficient> 0.7
Moderate Correlation (+ + or − −) : Coefficient between 0.5 and 0.7
Weak Correlation (+ or −) : Coefficient between 0.3 and 0.5
Negligible Correlation (no symbol) : Coefficient < 0.3

it is found that CP produces minimal correlation for most factors and hence can be considered the most resilient to variation in individual characteristics of users.

Blacklisting

Having looked at the personal characteristics of the individual users and the effect they had on the performance of the algorithms, it was important to examine whether changing access control decisions with respect to an individual friend has any effect on the usefulness of communities. Particularly, the focus was on the friends for each user for whom the access control policy remained constant across the 10 photos. That is, they were either selected in the audience for all 10 photos by the user or they were excluded from the audience in all photos.

There were only 3 out of the 26 users who had any friends who were *always selected* in the audience for all of the 10 photos. One user had 2 such friends while the other two users had 1 friend each in this category. The other 23 users had no friends who were constantly selected in every photo. This negligible proportion of permanently selected friends rules out the possibility of creating a subset of friends who would always be granted access. The results indicate that such a policy would not affect many friends and hence would not enhance an access control mechanism based on communities sufficiently.

When looking for friends who were *always excluded* from the audience, a larger variation was observed. All the 26 users had at least one friend who was always excluded from the audience of each of the 10 photos. The average ratio of friends who were never selected was found to be *55.4% (s.d = 31.2%)*. This means that the average user excluded more than half of his friend list from each photo.

These always excluded friends could be removed from their respective communities and put into a "blacklist". The effect of such blacklisting on the penalty of exclusion was examined. Although there was no significant difference between the algorithms with respect to this metric, as explained earlier, the algorithm with the best performance for this metric, Infomap, was used for this evaluation. The Infomap communities were reorganized after removing the blacklisted friends.

The average reduction in penalty after removing the blacklisted friends, for the entire set of 26 users, was found to be *44% (s.d = 29.6%)*. This can be regarded as a substantial improvement on the penalty values calculated earlier. The high standard deviation in the reduction suggests that the reduction varies for different users. A cluster analysis was conducted to identify subsets of users based on different levels of penalty reduction using "Two Step Clustering" [SJL10] which generated the following subsets:

1. *High Reduction* : This group of *15 users (5 males & 6 females)* had an average penalty reduction of *66.2%(s.d = 14.56%)*. The average audience across 10 photos for the average user in this subset was 26.98, the median was 12.3 and the standard deviation was 31.34. Thus, these users can be classified as *Consistently Low Selectors*.

2. *Low Reduction* - This group of *11 users (10 males & 5 females)* had an average penalty reduction of *13.8%(s.d = 11.99%)*. The average audience across 10 photos for the average user in this subset was 55.03, the median was 59.9 and the standard deviation was 32.81. Thus, it can be seen that these users select a larger audience on average.

The difference between the average audience of the users in the two clusters was found to be statistically significant using the Mann-Whitney test ($p < 0.05$). The difference in terms of size of friend network was not found to be statistically significant ($p = 0.305$). Moreover, there was a significantly negative correlation (*Pearson Coefficient* $= -0.411$) found between average audience per photo and average reduction in penalty. This means that users who select smaller audiences are more likely to benefit from the blacklisting process.

5.3.3 Effect of Photo

This section examines whether the photo being shared had any effect on the performance of the algorithms. The two attributes to define the photos were the source (whether they were brought by the user or downloaded randomly from the user's Facebook profile during the user study) and the categories (selected by the user during the study).

Source of Photo

Participants chose audiences for 10 photos, 5 randomly chosen from their Facebook account and 5 brought on a USB drive. There was negligible correlation (all coefficients < 0.1) between the source of the photo and the performance of the algorithms for all the 3 metrics.

Photo Categories

There were some cases where a clear difference was evident between the categories users selected for the photos which had a very high audience as compared to the low audience photos. For example, one user selected high audience for photos which they categorized as “Event” and “Animals” but low audience for photos which were categorized as “People(friends or family)”. This observation prompted a more detailed analysis into the effect of photo categories on the audience size of a photo.

As can be seen from Table 5.3, “People(friend or family)”, “Personal” and “Travel” were the most commonly selected categories. On the other hand, “Celebrity”, “Fashion”, “Technology” and “Advertising” were the least selected ones. It can also be seen that “Fashion” has the highest average audience. “Advertising”, “Celebrity” and “Animals” have high average audiences as well. It should be noted here that the total number of photos in the table are more than 260 as the users were able to select multiple categories for each photo.

To give an idea of the privacy implication of each photo category, as perceived by the users, the last two columns of Table 5.3 shows the number of occasions a particular category had the minimum and maximum number of audience members for a particular user. It can be seen that “Personal” photos have the highest number of occurrences when they have the minimum audience for a particular user. However, they also have a substantial number of cases where they have the maximum audience. Similarly, “People (friends or family)” has a high number of minimum as well as maximum audience

Table 5.3: Descriptive statistics of audiences of photos from each photo category

No.	**Categories**	**Photos**	**Audience Size**			**No. of times**	
			Avg.	**S.D**	**Med.**	**Max**	**Min**
1	Landscape	21	53.83	42.05	51	2	1
2	People (Others)	14	54.72	70.81	29	2	3
3	People (friends)	69	52.37	42.73	58	9	5
4	Architecture	20	42.85	62.07	17	2	3
5	Animals	11	86.52	80.58	64	3	2
6	Travel	54	58.51	37.57	60	6	5
7	Fashion	6	143.44	93.85	191	1	1
8	Celebrity	3	84.33	105.10	42	1	0
9	Event	35	62.53	77.4	21	3	2
10	Humor	22	57	60.57	31	3	3
11	Food	13	44.52	47.82	32	2	3
12	Advertising	4	89.33	102.40	57	0	0
13	Entertainment	51	43.68	45.88	25	2	3
14	Personal	60	32.90	30.07	26	5	9
15	Technology	6	49.40	79.68	22	1	1

occurrences. Thus, it can be seen from these results that privacy preferences vary between users and no definitive conclusions can be drawn about privacy implications of categories on their own.

After inspecting this further, it was found that there was negligible correlation between the photo categories and the performance of the algorithms according to the evaluation criteria. It was also observed that photo categories had negligible correlation with audience size of the photo. There was a possibility that the variability in audience selection of the 11 users for whom the reduction in penalty was low was due to photo categories. A linear regression analysis was performed in order to understand the effect of photo category on audience size for these users but found no significant effect. The correlation coefficients were also low with the highest being 0.243 for “Advertising”. Thus, there was negligible or very weak correlation between photo category and audience size even for users selecting variable audiences.

Table 5.3 also shows that the audience size of most categories had a large standard deviation. While this variation could be attributed to different types of users (high selectors or low selectors) selecting the various categories, it was important to examine whether categories being paired with each other had any influence in the different average audiences. There were some cases where combination of different categories influenced the average audience. For example, average audience for “Personal” when taken alone is 32.90 whereas when it is paired with “People(others)”, it goes up to 92.27. Similar variations can be observed across all categories. Categories which have a high average audience, such as “Fashion” or “Advertising” can inflate the average audiences for other categories if selected together for a particular photo. However, even for pairs or combinations of categories, no conclusions could be made about privacy implications. For example, a particular user selected categories “People (friends or family)” and “Travel” for a particular photo and selected 6 audience members while a different photo shared by a different user with the same categories had 191 members. Such variations were commonly observed across all combinations of categories.

It is safe to conclude from the above discussion that there were no noticeable effects of the photo categories on the performance of the algorithms in our evaluations.

5.4 Discussion

This chapter presented a detailed empirical evaluation of 8 network based community detection algorithm with the objective of identifying the most suitable algorithm to be used to represent *Relationship Type* in REACT. The major findings of the evaluation are as follows:

CP best in two out of the three metrics

CP produces the best results for number of communities required to complete a user’s desired audience as well as the ratio of audience in a single community. Its difference with other algorithms in terms of these two metrics is also found to be statistically significant. In terms of penalty for exclusion, the difference between all 8 algorithms is not found to be statistically significant.

Even best penalty not good enough for everyone

Infomap produces the lowest average penalty value of about 122. This means that even leveraging communities created by the best algorithm for this metric, the user would have to remove 122 friends on an average to create their desired access control policies. This value can be considered prohibitively high and suggests that communities cannot be readily leveraged for making access control decisions by all users.

Performance can be enhanced for users who consistently select low audience

We observed that penalty for exclusion could be reduced (avg reduction$\sim$ 44%) by removing "blacklisted" friends from their respective Infomap communities. This suggests that if access control mechanisms can identify frequently excluded friends over time and rearrange communities to exclude these friends during audience selection, it can produce much better results. These friends can be then put into a single community and a default setting of denying access can be implemented for them. The results indicate that such a scenario is most effective for users who select consistently low audiences and their penalty for exclusion can be reduced substantially ($\sim$ 66%).

Size of network has an effect on penalty for exclusion

We observed that number of communities per audience and ratio of audience in largest community have strong correlation with average audience of the individual. These correlations were along expected lines as larger average audience required more communities on average while a smaller average audience produced a higher likelihood that the ratio of audience in a single community would be higher. The penalty for exclusion also had a substantially positive correlation with size of a user's network. A noteworthy finding emerging from this analysis is that CP had the lowest correlation coefficient for most of these factors among all algorithms and can be considered comparatively resilient to these variations in individuals' characteristics as a result.

Nature of content has no significant effect on the performance of the algorithms

Photo category played a negligible role in determining the size of audience. This suggests that participants interpret categories differently and also that category of the photo alone is not enough to determine audience. We also looked at the performance of algorithms according to the photo categories based on the evaluation metrics but no significant conclusions could be made from that analysis as well.

5.5 Conclusion

As a result of the analysis presented in this chapter, *Clique Percolation Method (CP)* emerges as the most suitable network based community detection algorithm which has the best fit with access control decisions made by users during the user study. Therefore, *Relationship Type* would be represented using the CP membership of friends in the implementation of REACT. The high penalty for exclusion values, for algorithms including CP, indicate that using only community detection is not a sufficient solution on its own for the access control problems faced by social media users. This further points in the direction of combining *Relationship Type*, denoted by community membership, with other aspects of the overall context of the disclosure such as *Relationship Strength* and information about the *Content* which, as discussed in Chapter 3, is done in REACT.

Chapter 6

Relationship Strength

The relationship strength component of REACT uses information stored in the social media profiles of the users as was discussed in Chapter 3. The information contained in social media profiles can often be vast and can depend on the particular social media site. Social media sites such as Facebook, for example, have rich user profiles where users are encouraged to provide a lot of personal details such as age, location, workplace, family members, etc. On the other hand, other social media sites such as Twitter allow users to have minimal profiles without even requiring real names. There are numerous ways in which this information can be leveraged to enhance access control mechanisms. One particular way is to use profile information to calculate relationship closeness or "tie-strength" to define relationships between individuals in a social network. Tie-strength can be measured by looking at the amount of communication between individuals or similarity between their profile information [GK09] and can then be used to assist users while making access control decisions [TCS12]. In real life, individuals often share different information with different people based on their perceived closeness with them. For example, a user may choose to disclose something to only his "close friends" but not to "distant" or "weak" relationships [Gra73]. This situation can be reproduced while disclosing information on social media sites where tie-strength information may be provided to the user while making access control decisions [TCS12]. Such interventions may enable the user to share their content with a desired audience in terms of the strength of their relationship with them and preserve its contextual integrity [Nis04].

For relationship strength to be efficiently represented in REACT, it is imperative to identify the most suitable set of profile attributes which are relevant in an access control scenario. This chapter describes the results of an empirical evaluation of 30 profile attributes, which were created from the Facebook profiles of the participants of the user study described in Chapter 4 and their interpersonal communication with their friends, and goes on to find the minimal subset which is deemed to be suitable to represent relationship strength in REACT.

6.1 Challenges

There are several challenges while using social media profile data to enhance an access control mechanism [AFW12]:

- *Fetching Time*: The time taken to fetch the profile information from the social network profiles of users and their friends required for the relevant attributes takes time which slows down the process of access control recommendation. In social media, users cannot be expected to wait for a long time before sharing information. Therefore, the time taken to fetch the relevant information, to facilitate calculation of similarity, should be minimal.

- *Computation*: The profile information should be easily converted into profile vectors which can be analyzed to calculate similarity between users. Attributes which require extensive computation, such as examining the lists of events attended by individual users to identify common attendances, for example, would put a burden on the overall system which can lead to delays in providing the access control recommendation to the user.

- *Availability*: The calculation of profile similarity to represent relationship strength depends on the information in the users' profiles. However, users have often been found to leave profile fields blank (especially for information deemed to be sensitive such as age, religion, sexual orientation, political affiliation, etc.) and this missing information will affect the performance of any resulting mechanism [AFW12].

- *Privacy*: The profile attributes used should not require information which can be considered sensitive or “privacy intrusive” as using such information would defeat the purpose of creating a privacy enhancing mechanism. For example, if an access control mechanism requires access to the content of the personal messages exchanged between users, it can be considered to be privacy intrusive and cannot be considered a suitable solution.

Looking at these challenges, it seems imperative to identify the minimal subset of profile features which can be used to accurately predict access control decisions and make appropriate suggestions to the user while overcoming these challenges. The endeavor is to create an access control mechanism which provides accurate suggestions with minimal profile information from the users or their friends.

6.2 Systematic Feature Engineering

In order to identify the most appropriate profile attributes to represent relationship strength in REACT, a baseline of 30 profile attributes were considered. Table 6.1 shows the full list of all profile attributes which were evaluated. These attributes were created with the objective of including as much available information as possible from Facebook profiles of the users. The table also shows that all attributes considered by previous approaches were included in addition to some extra ones which are generally available in social media infrastructure. We could not compare our list with works which did not explicitly provide the list of profile attributes they used.

6.2.1 Sample

As discussed earlier in Chapter 4, the entire sample size of the user study was 26 users and the total number of access control decisions made by them was 68,840. However, in order to evaluate similarity in profile attributes between users and their friends, access to their

Table 6.1: Comparison of the subsets of profile attributes used to create profile vectors in the different approaches

No.	**Profile Feature**	**Type**	**This Work**	**Social Circles**	**ReGroup**	**BFF**
				[ML12]	[AFW12]	[FSEGF14]
1	Friendship Duration	Derived	✔	✔	✔	✔
2	Recency of Comm.	Derived	✔	✔	✔	✔
3	Amount seen together	Derived	✔	✔	✔	✔
4	Wall Messages	Derived	✔	✔	✔	✔
5	Inbox Messages	Derived	✔	✔	✔	✔
6	Mutual Friends	Direct	✔	✔	✔	✔
7	Gender	Direct	✔	✔	✔	✖
8	Age	Direct	✔	✔	✔	✖
9	Family Membership	Direct	✔	✔	✔	✖
10	Home Town	Direct	✔	✔	✔	✖
11	Home State	Direct	✔	✔	✔	✖
12	Home Country	Direct	✔	✔	✔	✖
13	Current City	Direct	✔	✔	✔	✖
14	Current State	Direct	✔	✔	✔	✖
15	Current Country	Direct	✔	✔	✔	✖
16	Education	Direct	✔	✔	✔	✔
17	Work	Direct	✔	✔	✔	✖
18	Likes	Direct	✔	✔	✖	✔
19	Events	Direct	✔	✔	✖	✖
20	Politics	Direct	✔	✔	✖	✖
21	Religion	Direct	✔	✔	✖	✖
22	Interests	Direct	✔	✔	✖	✖
23	Links Shared	Derived	✔	✖	✖	✔
24	Music	Direct	✔	✖	✖	✖
25	Movies	Direct	✔	✖	✖	✖
26	Languages	Direct	✔	✖	✖	✖
27	Sports	Direct	✔	✖	✖	✖
28	Total Friends	Direct	✔	✖	✖	✔
29	Friend Difference	Direct	✔	✖	✖	✖
30	Age Difference	Direct	✔	✔	✔	✖

profile information was required. Due to the Facebook API[1], only the profile information of users who use an application (and explicitly provide data access permissions) is available for collection. This meant that profile information of only the friends of the participants who were also participating themselves in the study was available. To mitigate this, we particularly encouraged groups of people to participate in the study so that a particular participant would have some Facebook friends also participating in the study which would enable us to get their profile information. The users did not know the information of which of their friends would be accessed and hence made access control decisions with respect to each of their friends in order to avoid biasing their decisions. Nevertheless, this resulted in a reduced sample size for this evaluation as compared to the other analyses presented in this thesis. There were 23 participants who had at least two friends for whom profile information was available. Out of these 23 participants, 14 (61%) were male. The average age of the participants was 32 years (s.d=10, max=61, min=23). Considering each available friend profile of the 23 participants and all photos for which they selected audiences, this dataset contained a total of *689 access control decisions* which was used to evaluate profile attributes to identify the most suitable subset.

6.2.2 Coding Profile Attributes

The profile information, which was collected during the user study, had to be converted into profile vectors for each user and their friends to capture the similarity between them. The aim was to create as many profile attributes as possible using the data that was collected during the user study. Descriptive fields such as “About Me” and “User Bio” were ignored as it would be hard to find similarity between users for such fields. After the study, it was found that none of the participants disclosed “Relationship Status” and “Significant Other” on their profiles so it was automatically discarded. The attributes listed in Table 6.1 are categorized in the following two types:

- *Direct*: These attributes can be directly fetched from the Facebook profile of the user or their friend. No calculation or aggregation is required.

[1] `https://developers.facebook.com/docs/apps/upgrading`

- *Derived*: These attributes are created by aggregating different forms of communication (for eg: posts, messages or photos) between a user and a particular friend.

22 out of the 30 profile attributes shown in Table 6.1 are *Direct* attributes and were fetched directly from the Facebook profiles. Intuitively, the computational cost of creating profile vectors is minimized if a large number of attributes are *Direct* as opposed to *Derived.* Thus, the method of obtaining the attribute is an important factor to consider while evaluating the profile attributes.

The coding of profile attributes followed in this work and described here is most similar to [ML12] as they also looked to create vectors based on similarity of profiles. However, as Table 6.1 shows, this work considers some additional attributes which were not included by them. The various attributes were coded in different ways depending on the type of information they contain in order to capture the similarity between a user and his friends.

- *Friendship Duration* and *Recency of Communication* between a particular user and a specific friend were calculated by counting the number of days since the first and latest communication (such as wall post, message, like, etc.) respectively.

- *Amount seen together, Wall messages* and *Inbox messages* were all coded by simply counting the number of interactions shared by the user and a particular friend. A zero value indicates that no interaction took place between them during the time for which the data was collected.

- *Mutual friends, Links shared* and *Events* denote the common friends, links and events attended by a user and a particular friend.

- *Gender* was coded as a binary variable to capture the similarity between the user and each friend. Here, '1' indicates that the user and the friend had the same gender while a ‘0’ indicates a dissimilarity.

- *Age* and *Total friends* were taken directly from the users’ profiles. In addition to this, the difference (absolute value) between the user’s age and a friend’s age was calculated as a new variable called *Age Difference*. This gives a measure of the gap between the user and a potential audience member in terms of their age. A similar calculation was done for Total Friends to create the variable *Friend Difference*.

- *Family membership* was coded as a binary variable where a ‘1’ indicates that the particular friend was disclosed as a family member of the user while a ‘0’ indicates that no such family relationships was found between friend and the user. This includes cases where the user did not disclose family relationships on Facebook.

- The attributes numbered *10 to 27* in Table 6.1 were coded to represent the number of common entries. For example, if a user and a friend had exactly one common educational institution, a ‘1’ value was put for that variable, if they had two common educational institutions, a ‘2’ was put and so on. A ‘0’ value captures both non-matches and missing entries. Missing entries were coded as 0 to ensure that we maximize the effect of matches between the values while capturing the similarity between profiles.

6.2.3 Metrics

In order to identify the most suitable subset of attributes, a classifier was created with all 30 attributes as an initial configuration. The objective was to identify attributes which enable accurate prediction of access control decisions. The accuracy was evaluated using the dataset of 689 ground truth access control decisions. The performance of the classifier was evaluated using several well-established metrics for machine learning

Table 6.2: Confusion matrix for evaluating performance

		Recommendation	
		Allow	**Deny**
Access Control Decisions	**Allow**	**TP**	**FN**
	Deny	**FP**	**TN**

classifiers [CLB+16] to give a broad picture of the performance. The metrics used are as follows:

- *Specificity*: It is measured as a "true-negative rate" which is a proportion of "deny" instances (from ground truth) that are correctly predicted as such.
- *Sensitivity*: This is the "true-positive rate" or "recall" which is the proportion of "allow" instances that are correctly predicted as "allow" by the classifier.
- *Precision*: This is the proportion of "allow" predictions which were actually "allow" in the ground truth access control decisions.
- *F-measure*: This is a harmonic mean of *Precision* and *Sensitivity* and gives a measure of the overall performance of the classifier.
- *Accuracy* : This measures the proportion of correct, both "allow" and "deny", predictions made by the classifier.

Table 6.2 shows the confusion matrix for evaluating the performance of the classifier. The described metrics are calculated using the following equations:

$$Specificity = \frac{TN}{TN + FP} \tag{6.1}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{6.2}$$

$$Precision = \frac{TP}{TP + FP} \tag{6.3}$$

Table 6.3: Classifier considering all 30 profile attributes shown for each algorithm

Algorithm	**Specificity**	**Sensitivity**	**Precision**	**F-measure**	**Accuracy**
Naive-Bayes	93.4%	45.3%	67.3%	0.541	82.3%
SVM	**98.5%**	37.1%	**88.1%**	0.522	84.3%
Random Forest	93.8%	**64.2%**	75.6%	**0.694**	**86.9%**

$$F - measure = 2 * \frac{Precision * Sensitivity}{Precision + Sensitivity} \tag{6.4}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6.5}$$

where, TP = True Positives, FP = False Positives, TN = True Negatives, FN = False Negatives

6.3 Results

The profile information was coded, as has been described, to create profile attributes to be used to ascertain similarity between profiles. The access control decisions made by the participants were coded as '1' (for "allow") and '0' (for "deny"). The classifier was created using three different classification algorithms: *Naive-Bayes* [Mur06], *Support Vector Machines* [HDO+98] and *Random Forest* [LW02]. All algorithms were implemented using Weka [HFH+09] with 10 fold cross validation using the entire dataset of 689 access control decisions. The results for the classifier created with all 30 profile attributes is shown in Table 6.3.

Table 6.3 shows that while *SVM* produces the best specificity and precision for the classifier, using *Random Forest* clearly produces the best overall performance, especially in terms of Sensitivity. Therefore, it can be concluded that *Random Forest* is the most suitable classification algorithm for this dataset.

6.3.1 Class Balancing

The dataset obtained from the user study, which was considered for the analysis of profile attributes, consisted of *689* access control decisions from 23 users as explained earlier. Out of them, *159* were *"allow"* while *530* were *"deny"*. Thus, there is an imbalance observed as the ratio of "deny" decisions is higher than "allow" decisions. Such a "class imbalance" is known to adversely affect classification and can be mitigated by employing class balancing techniques [JS02]. The class balancing techniques used in this evaluation are:

- *Class Balancer*: In this technique, synthetic instances of the rarer class ("allow" in our case) are added in a way such that ground truth of both classes is equal. This leads to a fractional number of instances for the classifier results (shown in results).
- *Spread Subsampling*: This technique does not introduce any synthetic instances but rather redistributes the frequency of both classes in an attempt to balance the dataset.
- *Cost Sensitive Learning*: "Cost sensitive learning" [LS11, Elk01] penalizes any instance of the rarer class ("allow" in our case) classified as "deny". The penalty or cost can be varied to produce desirable results. For these evaluations, the cost was increased starting from 1 (default) systematically to a point where the F-measure and accuracy of the classifier started decreasing when compared to the original, unbalanced, classifier.

It should be noted that for Random Forest and Naive-Bayes classification algorithms, using the default cost of 1 (not penalizing misclassifications of "allow" class) produced better results than increasing the cost. Overall, it can be seen that Random Forest produces the best results for all balancing modes in terms of F-measure and Accuracy. While *Class Balancer* improves F-measure when compared to the unfiltered mode (Table 6.3), due to the improvement in Precision and Sensitivity, it comes at the expense of losing accuracy of classification. This is due to the possibility that the *Class Balancer* filter overbalances the dataset which in turn ha an adverse effect on accuracy of classification.

Table 6.4: Results of classifier created with all 30 attributes when using the class balancing techniques

Technique	**Algorithm**	**Specificity**	**Sensitivity**	**Precision**	**F-measure**	**Accuracy**
Class Balancer	*Naive-Bayes*	81.5%	57.9%	75.8%	0.656	69.7%
	SVM	78.8%	67.2%	76.1%	0.714	73%
	Random Forest	77.2%	73%	76.2%	0.745	75.1%
Spread Subsample	*Naive-Bayes*	93.2%	45.9%	67%	0.545	82.3%
	SVM	97.4%	37.1%	80.8%	0.509	83.5%
	Random Forest	93.2%	60.4%	72.7%	0.660	85.6%
Cost Sensitive	*Naive-Bayes*	93.4%	45.3%	67.3%	0.541	82.3%
	SVM	95.3%	49.1%	75.7%	0.595	84.6%
	Random Forest	93.8%	64.2%	75.6%	0.694	86.9%

6.4 Identification of Minimal Subset of Attributes

The objective of evaluating all these 30 profile attributes was to identify the minimal profile vector to be used for similarity based access control. This minimal profile vector would be used to represent relationship strength in the implementation of REACT. To achieve this objective, a systematic evaluation of each individual profile attribute was required to understand their contribution to the prediction of access control decisions. Two established techniques of evaluating attributes were used to achieve this objective: *Correlation* and *Information Gain*.

6.4.1 Correlation Order

The 30 profile attributes are ordered according to their correlation coefficient in Table 6.5. It can be seen that the absolute value of the correlation coefficient is used to order attributes as even a negative correlation coefficient means that the attribute contributes to the prediction.

It can be seen in the table that *Wall Messages* have the highest correlation coefficient (0.304). It is also evident that attributes such as *Music, Religion, Sports, Politics, Education*, etc., have very low correlation possibly because many users often leave them blank in their profiles [AFW12].

6.4.2 Information Gain Order

Weka [HFH$^+$09] was used to calculate the “information gain” for each profile attribute to understand how they contribute to the classifier. The 30 profile attributes are ranked according to the information gain value in Table 6.6. It can be seen from the table that *Amount Seen Together* has the highest information gain (0.213). It is also interesting to note that there are 10 profile attributes which do not provide any information gain (attributes 21-30 in Table 6.6)

Table 6.5: Profile Attributes ranked according to Correlation with Access Control Prediction

No.	Profile Attribute	Correlation Coefficient
1	Wall Messages	0.304
2	Photos Together	0.272
3	Interests	0.253
4	Home Town	0.232
5	Friend Difference	-0.214
6	Current State	0.212
7	Mutual Friends	0.211
8	Current City	0.207
9	Events	-0.195
10	Friendship Duration	0.194
11	Home State	0.194
12	Gender	-0.193
13	Recency of Communication	0.186
14	Total Likes	-0.175
15	Total Friends	-0.167
16	Movies	0.143
17	Links Shared	-0.141
18	Age Difference	-0.133
19	Work	-0.115
20	Family	0.106
21	Home Country	0.086
22	Current Country	0.086
23	Age	-0.076
24	Music	0.054
25	Languages	-0.046
26	Sports	0.036
27	Religion	0.016
28	Politics	-0.013
29	Messages	-0.013
30	Education	0.002

Table 6.6: Profile Attributes ranked according to Information Gain

No.	Profile Attribute	Information Gain
1	Amount seen together	0.213
2	Total Friends	0.190
3	Mutual Friends	0.182
4	Friendship Duration	0.175
5	Likes	0.174
6	Friend Difference	0.154
7	Age Difference	0.107
8	Movies	0.102
9	Links Shared	0.096
10	Wall Messages	0.071
11	Events	0.067
12	Interests	0.050
13	Recency of Communication	0.045
14	Home Town	0.037
15	Age	0.036
16	Current State	0.034
17	Current City	0.032
18	Home State	0.026
19	Gender	0.025
20	Family	0.008
21	Work	-
22	Home Country	-
23	Current Country	-
24	Music	-
25	Languages	-
26	Sports	-
27	Religion	-
28	Politics	-
29	Inbox Messages	-
30	Education	-

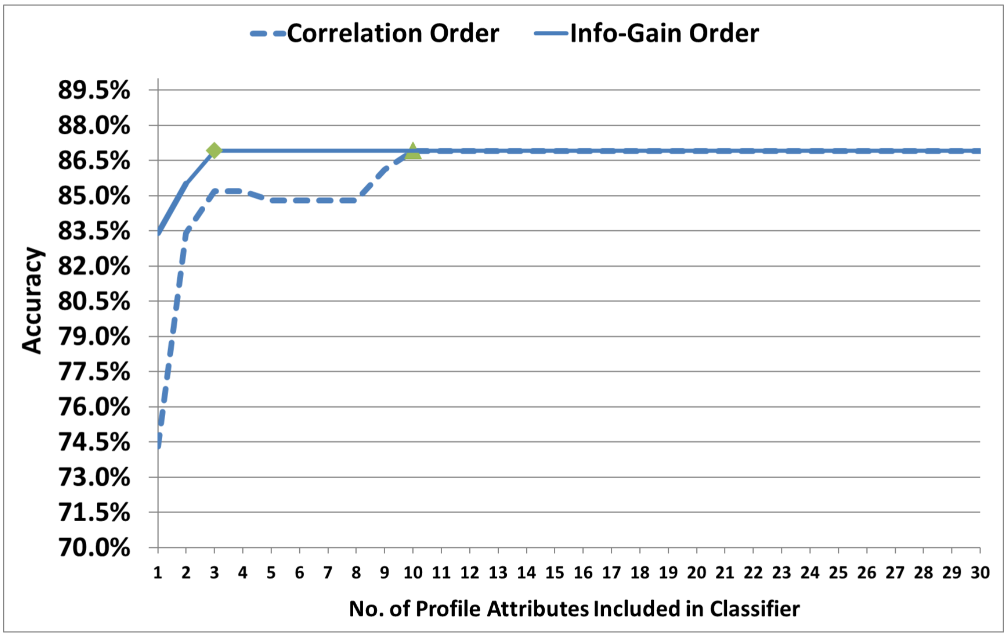


Figure 6.1: Number of attributes required for correlation and information gain ordering to achieve target accuracy of 86.9%

6.4.3 Systematic Identification of Minimal Subset

In order to identify the minimal subset of attributes which are most suited to predicting access control decisions made by users, both correlation and information gain order of profile attributes was used as separate starting points. The objective was to create a classifier using a minimum number of attributes to achieve a target accuracy of 86.9% using *Random Forest* classification algorithm, which was the highest accuracy of prediction found in Table 6.3. No class balancing technique was used for these evaluations.

Figure 6.1 clearly shows that it requires the first 10 attributes of the correlation order to produce an accuracy of 86.9% whereas only the first three attributes in the information gain order produce the same accuracy. Therefore, it is clear that the top three attributes with highest information gain, namely, *Amount Seen Together (0.213), Total Friends (0.190)* and *Mutual Friends (0.182)* (see Table 6.6), are sufficient in producing the target accuracy of prediction.

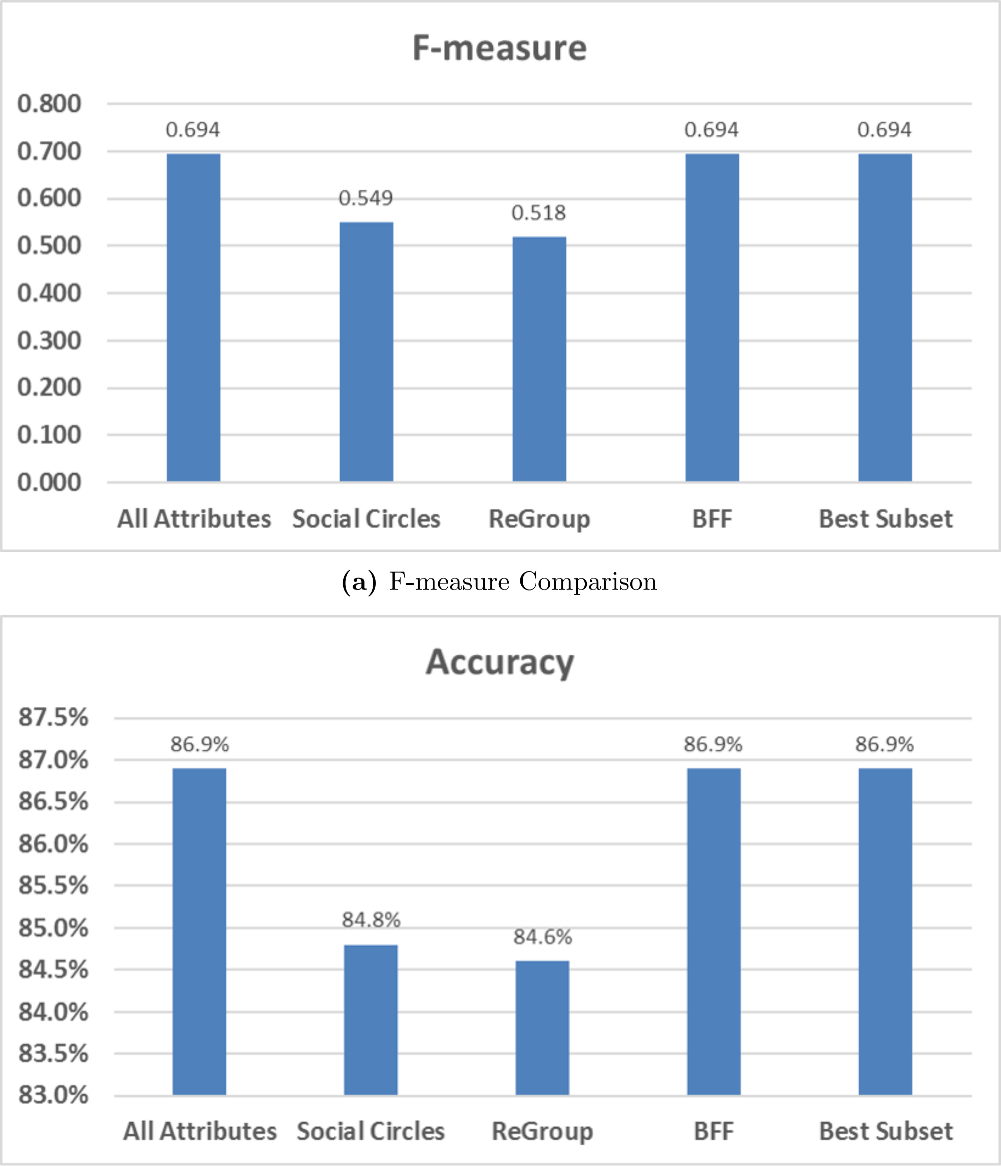
As the objective was to find a minimal subset of attributes, it was important to explore all possibilities of further reducing the subset of these three attributes while keeping the accuracy of prediction intact. Classifiers were created systematically with all possible combinations of these 3 profile attributes to see if any combination, of less than three profile attributes, would be sufficient to produce the required accuracy of prediction. It was found that the only combination of less than three profile attributes which produced the required accuracy of 86.9% was the pair *(Total Friends,Mutual Friends)*. All other combinations of less than three attributes, including using each of the three attributes individually, would reduce the accuracy below this level. Thus, the outcome of the systematic evaluation of 30 profile attributes is the pair *(Total Friends,Mutual Friends)* which is seen to be sufficient in accurately predicting access control decisions.

6.5 Comparison with Previous Approaches

Table 6.1 had shown the sets of profile attributes considered by previous approaches which use profile information to enhance access control mechanisms. After identifying the minimal subset of attributes in the previous section, it is important to compare the results produced by classifiers created by using these two attributes, *Total Friends & Mutual Friends*, with those produced by classifiers created by using subsets of attributes used by previous approaches. This comparison was done on the available dataset of 689 access control decisions described earlier in this chapter.

6.5.1 Performance of Classifier

Figure 6.2 shows the comparison of F-measure and Accuracy produced by classifiers using different subsets of profile attributes used by earlier approaches. The classifier was implemented using *Random Forest* classification algorithm without any class balancing. It can be seen that *BFF* [FSEGF14] produces the same F-measure and Accuracy as the best subset (*Total Friends* and *Mutual Friends*) as it included both these attributes (see Table 6.1). While *Social Circles* [ML12] and *ReGroup* [AFW12] both considered *Mutual Friends*, they failed to include *Total Friends* (or size of network) which results in a slightly lower performance as compared to the best values. Thus, it can be concluded that using these two attributes is sufficient to predict access control decisions with the same accuracy as with using all 30 attributes as well as any attribute subset that has



(a) F-measure Comparison

(b) Accuracy Comparison

Figure 6.2: Comparison of F-measure and Accuracy produced by classifiers created from different subsets of profile attributes

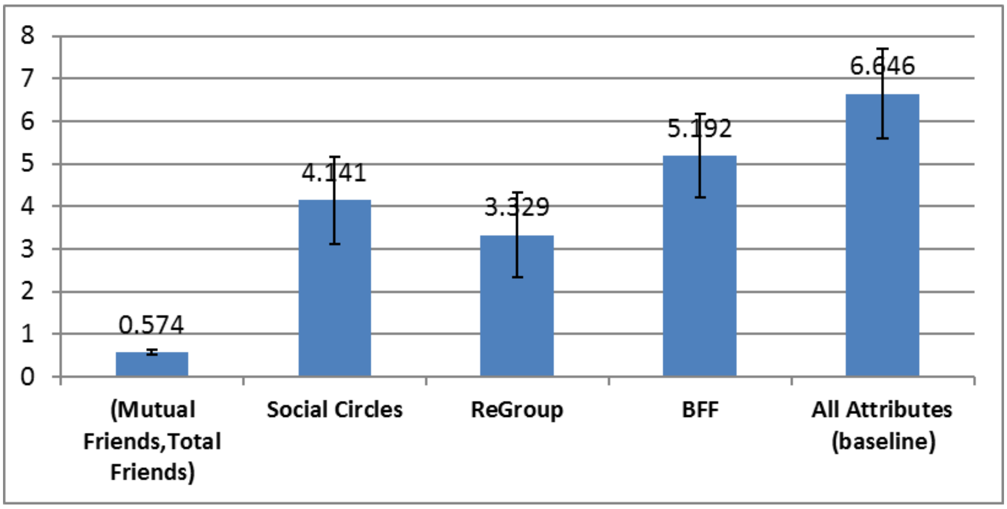


Figure 6.3: Time (in seconds) required to fetch profile information using Facebook API corresponding to each subset of profile attributes

been used in previous approaches.

6.5.2 Temporal Cost

Another important factor to consider when determining the usability of these two identified profile attributes is the time it would take to fetch the necessary profile information to create these attributes. Both *Total Friends* as well as *Mutual Friends* are **Direct** attributes (see Table 6.1) which means that they can be directly fetched from a user's social media profile.

A comparison of the estimated time taken to fetch the profile information necessary to create the profile vectors containing *(Mutual Friends,Total Friends)* and the subset of profile attributes used by *Social Circles* [ML12], *ReGroup* [AFW12] & *BFF* [FSEGF14] is shown in Figure 6.3. This time was calculated by running API calls for a Facebook profile with 320 friends with default settings. The time was calculated by creating 50 API calls to fetch the corresponding subset of profile attributes and taking the average for each proposed mechanism. It can be clearly seen in the figure that using *(Mutual Friends,Total Friends)* takes far less time (about *one-sixth* of the next best) than it would take to fetch attributes corresponding to the other approaches. This is primarily due to the fact that the other approaches required communication information (such as wall messages, inbox messages, etc.) which require more time when compared to "direct" profile attributes such as mutual friends and total friends.

6.6 Discussion

Using profile attributes is often suggested to be an effective method of enhancing access control prediction in existing literature. The evaluation presented in this chapter was dedicated to identify the minimal subset of profile attributes which could be used to represent *Relationship Strength* in REACT. The results of the evaluation enable the identification of *Mutual Friends* and *Total Friends* as the two profile attributes which can be used. The results show that using these two attributes is sufficient to match the accuracy of prediction (86.9%) produced by using all 30 profile attributes which were considered. In addition to providing the required accuracy of prediction, the identification of this minimal subset enhances previous profile similarity based access control models in the following ways:

Less time to fetch required information

It has been shown that the time required to fetch all relevant information from Facebook profiles for the 2 attributes identified in this chapter is about *one-sixth* of the time required for fetching information required for the subset of profile attributes corresponding to the quickest among previous works.

Easy creation of profile vectors

The "direct" attributes *Total Friends* and *Mutual Friends* are both simply numeric values which can be directly fetched from the users' profiles and incorporated in the profile vector without the need of processing a lot of information. There are certain attributes like *Education, Workplace, etc.* that can have different names for the same entity (for e.g. the same organization may have multiple pages which different users may affiliate with while belonging to the same organization) which makes coding more difficult. For both

Total Friends and *Mutual Friends*, coding is unambiguous as these are simply numeric values.

Less intrusive information gathering

The two identified attributes do not rely on communication (messages, wall posts, etc.) between individuals and hence can be considered less intrusive than other models which use such information. The other approaches which are discussed in this chapter all rely on such potentially intrusive information to create the profile vectors.

Cannot be left blank or faked

Both *Total Friends* and *Mutual Friends* are attributes that are automatically calculated and updated by the social media site itself. The users cannot manipulate this data in any way. These attributes do not require access to any identifying information such as *Gender, Address, Age, Workplace, etc.,* which are often left blank by users on their profiles.

Found in most social media infrastructures

The size of the users' network (*Total Friends*) and the number of shared connections (*Mutual Friends*) are attributes which can be found on most social media sites. While the analysis was conducted using information from Facebook profiles as mentioned earlier, the identification of these particular attributes ensures that the resulting mechanism, REACT, can be extended to other social media infrastructures as these particular attributes (or equivalent attributes) can be found in most other social media sites.

6.7 Conclusion

The systematic feature engineering presented in this chapter led to the identification of *Total Friends* and *Mutual Friends* as the two most suitable profile attributes to predict access control decisions. In addition to them being sufficient in matching prediction accuracy of previously proposed approaches, these attributes are also always available in social network infrastructure, are easy to process and require less time to be fetched. These advantages assume a lesser burden on REACT while using these attributes to represent relationship strength as discussed in more detail in Chapter 7 where the implementation of REACT is described.

A limitation of the analysis presented in this chapter is the comparatively reduced size of the dataset of access control decisions which could be considered. This was due to the API restrictions by Facebook[2] (as discussed in Section 6.2.1) which meant that profile information of only the users who participated was available for collection and analyses. We mitigated this by encouraging groups of people to participate in the study so that a particular participant would have some Facebook friends also participating in the study which would enable us to get their profile information. These API restrictions make sense from a privacy perspective as malicious third party applications are not able to access personal information of users who do not use their application and hence do not provide explicit consent to their information being used. However, this makes it challenging for any mechanisms which rely on profile information of friends and evaluation of such mechanisms becomes an extremely challenging task. There is no way, for instance, that profile vectors representing a user's friends can be created if those friends are not using the particular application. In light of this, the profile attributes *Total Friends* and *Mutual Friends* are even more useful as they are mostly publicly disclosed by most users and are hence easily accessible. Therefore, these API restrictions did not impact the implementation and evaluation of REACT shown in Chapter 7 as these profile attributes were available for the entire dataset of users and will also have negligible impact on future implementations of REACT with the design proposed in Chapter 3 of this thesis.

[2] `https://developers.facebook.com/docs/apps/upgrading`

Chapter 7

Implementing and Evaluating REACT

Recommendation of access control decisions has been seen as a solution to ease the burden on social media users and address some of the social privacy problems they face. The primary contribution of this thesis has been an access control recommendation mechanism, REACT, which considers the overall context of the information disclosure by including attributes representing *Relationship Type, Relationship Strength* in conjunction with the information about the *Content.* The detailed design of REACT was presented in Chapter 3. The empirical evaluation of 8 community detection algorithms presented in Chapter 5 aimed to identify the most suitable network best community detection algorithm to be used to represent relationship type in REACT while the evaluation of 30 profile attributes created from Facebook profiles of users presented in Chapter 6 was aimed to identify the most suitable set of attributes to represent relationship strength. Having identified these components of REACT, this chapter presents the detailed evaluation of the performance of REACT implemented using the identified components. The chapter describes how each component of REACT was implemented and then goes on to discuss the results of evaluating the access control recommendations made by REACT using the access control decisions made by the users during the user study described in Chapter 4.

7.1 Instantiating and Implementing REACT

At the heart of REACT is a machine learning algorithm which considers the attributes that constitute the three components, namely, relationship type and strength and information about the content, to provide "allow"/"deny" type access control recommendations to the user. REACT is agnostic to the machine learning algorithm. For the evaluation presented in this chapter, three different classification algorithms, namely, *Naive-Bayes* [Mur06], *Support Vector Machines* [HDO$^+$98] and *Random Forest* [LW02] were implemented using Weka [HFH$^+$09] with 10 fold cross validation with instances generated in the user study made by each user separately. The use of cross-validation is known to minimize over-fitting to the data and it rules out the possible bias associated with division of a dataset into training and test sets.

REACT also relies on blacklisting friends in order to assign a default "deny" recommendation to these friends and hence reduce the number of friends for whom access control recommendations need to be computed for each user. The rest of this section describes how each component of REACT was implemented for the evaluation shown in this chapter. The results of the evaluation are shown later in Section 7.2.

7.1.1 Relationship Type

Relationship type is represented by community membership in REACT and Chapter 5 presented the findings of the detailed empirical evaluation of 8 well-known network based community detection algorithms. The results of the evaluation helped identify *Clique Percolation Method (CP)* [DPV05] as the most suitable community detection algorithm to be used in REACT as it produced the best fit with access control decisions made by users during the user study. CP, like any other network based community detection algorithm in this context, only needs to be executed to update communities when a new friend is added or removed from a user's network, but not every time a user would make an access control decision which happens much more often. Therefore, even though CP has a comparatively higher computational complexity as compared to other similar community detection algorithms (refer to Chapter 5), its impact on the timeliness of access control decisions would be minimal.

The communities were created for each user from their social network (which was downloaded during the user study as an "edgelist" where each "node" represented each friend and an "edge" represents links between them). The edgelist of each user was used as input to the CP algorithm, implemented using the SNAP library [LS14] for this evaluation, to create communities based on identifying "cliques" or connected sub-graphs of the social network [DPV05].

The CP algorithm is based on identifying "k-cliques" depending on a predefined value of the parameter "k". It can produce overlapping communities depending on the value of "k" in the implementation [DPV05], i.e., if the value of "k" is kept at 2, it produces non-overlapping communities, but a value greater than 2 produces increasing number of potential overlapping communities. Given that community detection was used to represent relationship type, it was essential to consider overlapping communities as individuals on social media may share more than one relationship type (for e.g: a user's "co-worker" may also be a "family member"), so the value of "k" was varied systematically, keeping the other components static, to check whether creating overlapping communities had any effect on the overall performance of REACT. The CP membership for each friend of a user, and therefore the *Relationship Type* between the user and that friend, is represented as a binary vector of dimension 'n' (where 'n' is the number of communities CP creates for a user given her social network), in which each element in the vector denotes whether a particular friend belongs to a particular community (denoted by "1")or not (denoted by "0").

7.1.2 Relationship Strength

Strength of relationship or "closeness" between individuals on the social network can be calculated by fetching and analyzing the similarity of profile attributes and information in social networks [GK09, FSEGF14]. Chapter 6 presented a detailed empirical evaluation of 30 profile attributes which could possibly be used to represent relationship type in REACT. The results of the evaluation found a subset of two profile attributes, namely, the size of network (*Total Friends*) and shared contacts (*Mutual Friends*), to be the most suitable to be used in an access control scenario. REACT also uses a third attribute which calculates the difference in network sizes of the user and each of their friends *(Friend Difference)*, in addition to *Total Friends* and *Mutual Friends* to represent relationship strength between a user and each member of their friend network.

Total Friends and *Mutual Friends* are both "direct" attributes and can be directly fetched from the Facebook profiles of the users. *Friend Difference* can be simply calculated as the absolute difference between the *Total Friends* of the user and each of their friends. Thus, none of the attributes required to represent relationship strength in REACT were "derived" and hence no complex computation is necessary. An additional advantage of using these attributes was that the API restrictions introduced in 2015[1], which resulted in a reduced dataset for the evaluation of profile attributes presented in Chapter 6, do not apply to these attributes. Therefore, *Total Friends* (and therefore *Friend Difference*) as well as *Mutual Friends* was available for the entire dataset of 68,840 access control decisions collected during the user study and could be used for evaluating REACT. The values of *Total Friends*, *Mutual Friends* and *Friend Difference* were used as 3-dimension vector to represent *Relationship Strength* between a user and a particular friend.

7.1.3 Content

In addition to considering attributes defining the type and strength of interpersonal relationships, REACT also considers information about the content being shared in order to completely represent the overall context of the information disclosure. Previous research has shown that annotations of content by users employing "tags" can be used to create access control policies and that they are minimally disruptive for the user [YKGS09]. These tags can be about the topic of the content (e.g., Flickr has a list of potential topic categories for the photos users upload) or further information about the social context (e.g., friends being tagged in photos or mentions being made to them in text posts). In the current implementation of REACT, the tags, provided by the users during the

[1] `https://developers.facebook.com/docs/apps/upgrading`

study, about photo categories and friends appearing in the photos were used. The photo categories selected by the users were coded as a binary vector of dimension 15 (which was the total number of photo categories that could be selected by the user for each photo during the user study) which denotes whether the particular category was selected ("1") or not ("0") by the user. Similarly, the tag with respect to friends was also coded as a binary variable for each of the user's friends where "1" represents the case where a particular friend was tagged in a particular photo and a "0" represented case where the particular friend was not tagged by the user.

It is important to note here that the design of REACT is agnostic to the type of content being shared and can be easily extended to other types of data such as text. The design is also agnostic to the method in which the information about the content is supplied and REACT can use any available tags, whether manually provided or generated by automatic analysis of the content. In this implementation, the information about the content was provided by the users in the form of annotations. This was done as users are often found to interpret categories differently [MS16]. Moreover, this can also be considered as a less intrusive method as access to the content itself is not required but only the user's interpretation of the content is necessary. It is also easier to extend REACT for other types of content without too much effort when relying on manual annotation as specialized tools may not be required to be implemented to analyze the content which is an alternative method of categorizing the content and use these classifications to inform REACT about the content [SLSW15]. This method removes the responsibility of having to annotate the content from the user as relying on manual annotation can suffer if users simply avoid annotating the content. It is shown later in this chapter as to how REACT can be personalized and tune itself to the availability and utility of various attributes and still produce accurate recommendations as a possible mitigation to this problem.

7.1.4 Blacklisting

Social media users often have vast social networks consisting of many friends – the users in the sample used for this evaluation had an average of 265 friends (s.d = 121). However,

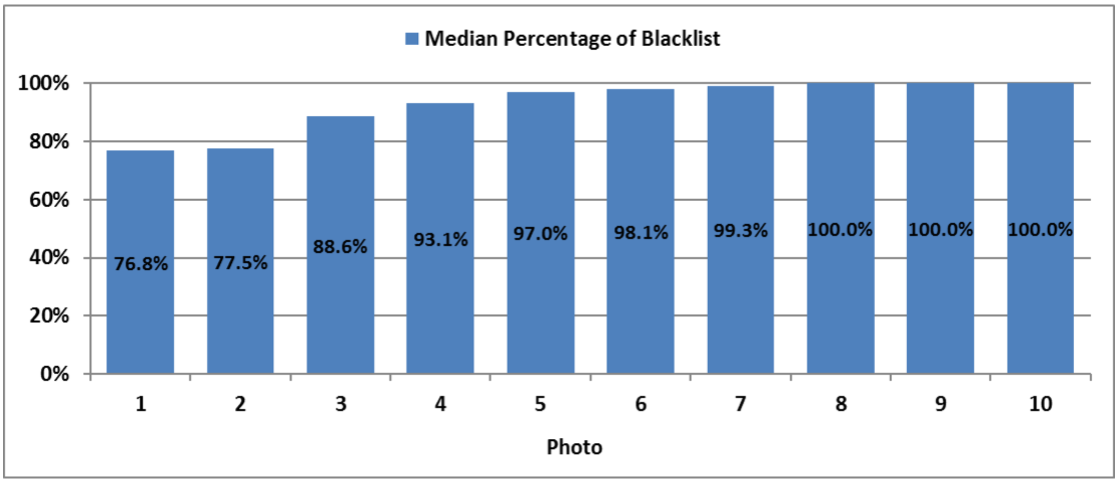


Figure 7.1: Median percentage (across 26 users) of the ratio of blacklist identified after each photo

it has been empirically found that social media users often intend to share their content with a limited subset of their entire friend network [KBHC12, MO11]. REACT leverages this intuition by maintaining a "blacklist" of friends for each user. It does not analyze the relevant relationship information for these blacklisted friends and assigns a default "deny" access control recommendation for them. It is important to note that the specific number of blacklisted friends depends on various factors for individual users such as their access control behavior and their network size. Different social media users use the medium differently and their access control behavior and patterns are different and driven by their privacy preferences and their intentions [MO11]. For example, some users may want to connect to and share with as many people as possible. For these users, the number of blacklisted friends would be minimal. Alternatively, users may choose to interact with only a small subset of their friends and the blacklist for these users may contain a majority of their friends. Thus, the blacklist construction in REACT is a personalized process in which a blacklist is learned based on previous access controls decisions and, hence, it entirely depends on the preferences and access control behavior of individual users.

In the implementation of REACT which is evaluated in this chapter, blacklisted friends were identified by looking at the access control decisions made by users during the user study and all friends who were never granted access by the user ("always excluded") in any of the 10 photos were added to the blacklist. In reality, it may require even less disclosure decisions to identify blacklisted friends for each user. To get a better understanding of how many photos it would require to identify the blacklisted friends for each user, it was necessary to identify the changes in access control decision made by the users with respect to each of their friends. Particularly, it was required to find which was the photo (photo 1 to photo 10) where each friend was granted access for the first time. All friends who were denied access to the first photo are potentially in the blacklist before the access control decisions for the second photo are checked, and so on. The blacklisted friends are the only ones who were never granted access in any of the 10 photos. Thus, if only the blacklisted friends were denied access to all of the photos by the user, which is an extremely static access control policy as the user would be selecting the same audience for each of the 10 photos, then 100% of the blacklist could be identified in the first photo itself. In other words, all friends who were denied access in the first photo would be automatically blacklisted in such a scenario. However, as we find in Figure 7.1, this is not normally the case. The figure shows that the median percentage of blacklisted friends for a user identified after the first photo is 76.8% and reaches 93.1% after only the fourth photo. This essentially means that for most of the users, the entire blacklist can be learned after 4 or 5 photos and anyone not granted access till this time would most likely end up in the blacklist as they would not be granted access in any of the following photos. This finding suggests that the blacklist can be learned by REACT fairly quickly for all users with very few previous access control decisions made by the user.

7.2 Results

The performance of REACT was evaluated using the ground truth access control decisions made by users during the user study described in Chapter 4. The friend networks of all the users was collected during the study in order to create the communities using the CP algorithm. As REACT only requires *Total Friends* and *Mutual Friends* from the Facebook profiles, the evaluation could include all 68,840 access control decisions as this information was present for all friends as it was not restricted by the API changes discussed earlier.

The results shown in this section correspond to the evaluation metrics described earlier in Section 6.2.3 of Chapter 6 which are: *Specificity (i.e., true negative rate)* which is a proportion of “deny” instances (from ground truth) that are correctly recommended as such, *Sensitivity (i.e., true positive rate or recall)* is the proportion of “allow” instances that are correctly recommended as “allow”, *Precision* is the proportion of “allow” recommendations which were actually “allow” in the ground truth access control decisions, *F-measure* is the harmonic mean of precision and sensitivity and finally, *Accuracy* which is the proportion of correct, both “allow” and “deny”, recommendations [CLB^{+}16]. The results were calculated using the confusion matrix shown in Table 6.2 in Chapter 6. The instances for all the users were added up and the metrics were calculated for the entire 68,840 access control decisions for each of the results shown in this section.

7.2.1 Overlapping Communities

It has been discussed earlier in Section 7.1.1 that CP is able to produce overlapping communities. To ensure that REACT was being implemented at its optimum capacity, it was important to identify whether using overlapping communities is better or worse than using non-overlapping communities in terms of overall performance of REACT. While the evaluation of algorithms in Chapter 5 only features CP to produce non-overlapping communities ($k = 2$) in order to facilitate comparison with the other seven algorithms, in this evaluation REACT was implemented using CP of values ranging from $k = 2$ to $k = 5$ to account for overlapping communities while keeping the attributes corresponding to relationship strength and information about the content constant. This could be done as CP is the only algorithm being used and need not be compared to other algorithms here.

Table 7.1 clearly shows that *Random Forest* produces better results than the other two classification algorithms. Looking at the k-values of the CP implementation, it is evident that using $k = 3$ produces slightly better results than the other values. Using any value greater than 2 produces overlapping communities and as 3 is found to be most suitable, it can be said that considering overlapping CP communities benefits REACT. This also

Table 7.1: Performance of REACT for different k values for CP while keeping the attributes corresponding to relationship strength and content constant

K	Algorithm	Specificity	Sensitivity	Precision	F-measure	Accuracy
$k = 2$	*Naive-Bayes*	94.9%	60.7%	58.1%	0.594	91.4%
	SVM	97.2%	55.5%	69.7%	0.618	92.8%
	Random Forest	**97.8%**	64.4%	74.5%	0.691	94.7%
$k = 3$	*Naive-Bayes*	94.3%	60.3%	59.1%	0.597	90.2%
	SVM	96.7%	55.5%	70.1%	0.619	91.6%
	Random Forest	**97.8%**	**65.6%**	**75.3%**	**0.701**	**94.9%**
$k = 4$	*Naive-Bayes*	94.2%	60.3%	58.8%	0.595	90.1%
	SVM	96.7%	54.9%	69.8%	0.615	91.5%
	Random Forest	**97.8%**	64.6%	74.6%	0.692	94.8%
$k = 5$	*Naive-Bayes*	94.1%	60.8%	58.3%	0.596	90.1%
	SVM	96.6%	56.1%	69.9%	0.623	91.6%
	Random Forest	**97.8%**	64.7%	74.4%	0.692	94.7%

seems to match a more accurate representation of a real life scenario as individuals on social media often share more than one community/relationship type (an example is when a user's "family member" is also a "co-worker" or "classmate"). Increasing the value to 4 and 5 however, did not improve the results further. Therefore, the most suitable k-value for the CP implementation is 3. The results produced by REACT, with *Random Forest* algorithm, for $k = 3$ for CP, calculated over the entire dataset of 68,840 access control decisions are: **Specificity = 97.8%, Sensitivity = 65.6%, Precision = 75.3%, F-measure = 0.701** and **Accuracy = 94.9%**.

7.2.2 Class Balancing

It can be seen in Table 7.1 that there is a rather important difference between specificity, sensitivity, and precision; with specificity being much higher than the other two. While specificity is very important in an access control scenario, as the "true negative rate" translates to the accuracy of "deny" classifications and a high specificity would ensure that unintended disclosure is prevented, other metrics are also important as low values for precision, for example, would mean that the user has to spend effort in changing many "deny" recommendations to "allow" to obtain their desired access control policy. Thus, while acknowledging the importance of high specificity and accuracy in an access control scenario, it is important to explore further avenues of achieving high values for all metrics in order to provide a mechanism which minimizes user effort as much as possible.

One of the reasons for specificity being substantially higher than precision and sensitivity may be that, despite the use of a blacklisting approach by REACT, there is still a difference between the number of friends being recommended allow and deny, which is perfectly consistent with the expected access behavior exhibited by users when confronted with individual access control decisions [KBHC12, SGA13]. In cases like this, the machine learning literature recommends the use of class balancing techniques [JS02] to produce more balanced results across all metrics. The class balancing techniques employed are the same as those used in Chapter 6:

- *Class Balancer*: In this technique, synthetic instances of the rarer class ("allow" in

Table 7.2: Overall results produced by REACT for all metrics, for all k-values of CP, and for all learning methods

Mode	**K-value**	**Specificity**	**Sensitivity**	**Precision**	**F-measure**	**Accuracy**
No Balancing	**k=2**	97.8%	64.4%	74.5%	0.691	94.7%
	k=3	97.8%	65.6%	75.3%	0.701	94.9%
	k=4	97.8%	64.6%	74.6%	0.692	94.8%
	k=5	97.8%	64.7%	74.4%	0.692	94.7%
Class Balancer	**k=2**	97.1%	70.1%	84.8%	0.768	92.1%
	k=3	**97.2%**	**72.7%**	**85.7%**	**0.787**	**92.6%**
	k=4	97.2%	70.7%	85.2%	0.773	92.3%
	k=5	97.2%	70.9%	85.2%	0.774	92.3%
Spread Subsample	**k=2**	97.8%	64.5%	74.7%	0.692	94.8%
	k=3	97.8%	64.8%	74.3%	0.692	94.8%
	k=4	97.7%	65.2%	74.2%	0.694	94.8%
	k=5	97.8%	64.5%	74.4%	0.691	94.7%
Cost Sensitive	**k=2**	95.0%	75.9%	60.2%	0.671	93.2%
	k=3	95.1%	76.6%	61.0%	0.679	93.4%
	k=4	95.0%	76.6%	60.6%	0.677	93.3%
	k=5	94.7%	77.5%	59.7%	0.674	93.2%

our case) in a way such that ground truth of both classes is equal. This leads to fractional number of instances for the classifier results (shown in results).

- *Spread Subsampling*: This technique does not introduce any synthetic instances but rather redistributes the frequency of both classes in an attempt to balance the dataset.
- *Cost Sensitive Learning*: "Cost sensitive learning" [LS11, Elk01] penalizes any instance of the rarer class ("allow" in our case) classified as "deny". The penalty or cost can be varied to produce desirable results. For these evaluations, the cost was increased starting from 1 (default) systematically to a point where the F-measure and accuracy of the classifier started decreasing when compared to the original, unbalanced, classifier. The cost was calculated separately for each user in the sample.

Table 7.2 shows the results produced by REACT when using the three class balancing techniques compared with those when no balancing is used. It can be seen from the table that *Cost Sensitive* classification provides highest best sensitivity while no balancing and *Spread Subsample* produce highest accuracy. However, looking at all the metrics, and F-measure in particular, it is clear that *Class Balancer* provides the best overall trade-off. It substantially improves precision of REACT when compared to no balancing. Moreover, an improvement can also been in the values for sensitivity and it is noteworthy that the improvement in both precision and sensitivity do not result in any decrease in the value of specificity. Therefore, for the best configuration of REACT, i.e, using $k = 3$ for CP and class balancer filter, the metrics calculated using the entire dataset of 68,840 access control decision are: **Specificity = 97.2%, Sensitivity = 72.7%, Precision = 85.7%, F-measure = 0.787** and **Accuracy = 92.6%**. Thus, the further results presented in this chapter consider CP implemented with $k = 3$ as well as using the *Class Balancer* filter in the implementation of REACT.

7.2.3 Personalizing REACT

The results presented so far in this chapter show that REACT produces highly accurate access control recommendations according to the various evaluation metrics. It uses the attributes, discussed in detail in Chapter 3, to represent *Relationship Type and Strength* as well as the information about the *Content* itself. However, it is essential to acknowledge that privacy and access control behavior is very personal to individual users [ABL15], so there is the potential that a personalized approach would render even better performance than a“one-size-fits-all” solution. This led to further exploration of possible methods to try and personalize the configuration of REACT such that it uses the most appropriate set of attributes for each individual user.

With these objectives in mind, two established attribute selection techniques were tried to identify subsets of attributes that could be used to configure REACT for individual users:

- *Principal Components*: Principal Components Analysis (PCA) is an established method of attribute selection [LCZT07]. It produces principal components created from an attribute list and the number of these principal components are less than or equal to the number of original attributes. PCA was used to identify the appropriate attributes for each individual user.

- *Information Gain*: Information gain is a good indicator of the contribution of each attribute towards the classifier performance. It was used in Chapter 6 to order the attributes of the classifier. Here, information gain is used to check which of the used attributes actually contribute to the performance of REACT. For each user, the attributes with zero information gain were removed from the configuration.

These attribute selection techniques were tried for each individual user and the results were compared, for each user, with those produced by including all available attributes. The focus was on F-measure as well as accuracy to signify a better trade-off in terms of all metrics for each individual user when selecting the best mechanism. In addition to information gain and PCA, a correlation based approach [Hal99] was also tried. However,

Table 7.3: Number of users for whom each attribute subset provided best results

Attribute Subset	**No. of Users**
Principal Components	8
Information Gain	9
All Attributes	9

Table 7.4: Values for all metrics, calculated over entire dataset, when using best attributes for each user as compared to using all attributes for each user

Attributes	**Specificity**	**Sensitivity**	**Precision**	**F-measure**	**Accuracy**
All	**97.2%**	72.7%	**85.7%**	0.787	92.6%
Best	96.8%	**77.4%**	84.5%	**0.808**	**93.2%**

the results showed that it did not provide the best trade-off for any of the 26 users.

Table 7.3 shows that using PCA to select optimal set of attributes produces best results for 8 users while using information gain is best for 9 users. Additionally, there are 9 users for whom using the entire set of attributes produces the best trade-off. It is worth noting, however, that there were 17 users for whom all three types of attributes (type and strength of relationship as well as content) were represented in the particular attributes that contributed to the classifier (either PCA or Information Gain) and 9 other users for whom at least two types of attributes were represented. This highlights the importance of considering all three types of attributes while designing REACT.

Table 7.4 shows the performance of REACT evaluated over the entire dataset of 68,840 access control decisions made by all users when:

1. All attributes discussed in Section 7.1 were used for each of the 26 users.

2. Only the best subset of attributes was used for each individual user.

It is evident from the table that using the best attribute subset provides a better overall trade-off as compared to using all attributes for all users as it provides a big improvement in terms of sensitivity at the expense of very little change in precision. This results in an improved F-measure and a higher accuracy is also observed.

7.2.3.1 User Characteristics

Having seen the benefits of using the most appropriate subset of attributes, it was important to try and understand whether it is possible to recommend a particular attribute subset (PCA, Information Gain or All Attributes) for an individual user by looking at their characteristics. This would enable early identification of a personalized approach for REACT and mitigate a "cold start" in terms of not requiring a lot of decisions in order to ascertain the most appropriate approach for a user.

Table 7.5 shows the characteristics of users for whom each of the attribute subset produced best performance. There was a statistically significant difference in terms of average audience (across 10 photos per user) ($p < 0.05$) and standard deviation audience (across 10 photos per user) ($p < 0.05$) using the Kruskal-Wallis Test. The Mann-Whitney test found that the 8 users for whom PCA was the most suitable approach had a statistically significantly lower average ($p < 0.05$) and standard deviation audience ($p < 0.05$) than the other 18 users. There were no statistically significant differences between the users for whom using Information Gain was found to be most suitable and for whom using all attributes provided best results. Looking at other characteristics, which do not have statistically significant difference, it is evident that PCA can be used to select attributes, starting from the original configuration of REACT described earlier, for users who have smaller friend networks and who select from a smaller section of their network as shown by their higher blacklist rate (ratio of total friends who were blacklisted), lower allow ratio (percentage of total access control decisions which were "allow") and smaller number of communities used.

7.3 Discussion

This section discusses the major findings from the empirical evaluation of REACT presented in this chapter.

Table 7.5: User characteristics to identify suitable subset of attributes to configure REACT

Attribute Subset		Total Friends	Blacklist Rate	Audience (10 photos)		Communities		Allow Ratio
				Average*	**Stdev***	**Total**	**Used**	
Principal Components (8 users)	**Avg**	214.8	72.1%	13.5	15.2	6.1	5.2	6.1%
	SD	115.2	29%	18.3	19.4	4.3	4.5	6.3%
Information Gain (9 users)	**Avg**	322.4	61.1%	33.4	38.1	9.8	8.7	14%
	SD	150.8	26.8%	22.7	27.9	6.6	6	15.1%
All Attributes (9 users)	**Avg**	251.6	51.1%	42.9	43.1	9.6	8.9	17.6%
	SD	69.1	27%	32.1	28.3	6.1	6.6	15.4%

*Significant difference with Kruskal-Wallis Test ($p < 0.05$)

Overlapping Communities produce better results

REACT was implemented with different values of "k" in order to check whether considering overlapping communities improves quality of the access control recommendations. The results showed that REACT provides best recommendations for $k = 3$ which indicates some overlap between communities. This is also closer to real-life scenarios where a particular member of a user's social network may have more than one type of relationship with them (for e.g. a friend can also be a co-worker).

Class Balancer provides best overall trade-off

It was observed that REACT produced high accuracy (94.9%) and specificity (97.8%) in the baseline mode of operation (with $k = 3$). However, the values of precision and sensitivity were relatively low. Several class balancing techniques were tried to improve the performance of REACT with respect to these metrics and it was observed that using the *Class Balancer* filter improved sensitivity and precision substantially with very little change in accuracy and specificity and hence provided the best overall trade-off.

Personalizing REACT improves performance

It was interesting to examine whether the configuration of REACT could be personalized depending on the access control behavior of various users. Principal Components Analysis (PCA) and Information Gain were utilized as methods to identify the ideal subset of attributes to be considered by REACT for each individual user. The results showed that using the baseline configuration including all attributes produced best trade-off for 9 users while the other 17 benefited from using either PCA (8 users) or Information Gain (9 users). PCA was found to be the most appropriate attribute selection technique for users who have smaller friend networks and who select from a smaller section of their network.

7.4 Conclusion

The empirical evaluation of REACT presented in this chapter shows that it performs very well with respect to all the metrics, namely, specificity, sensitivity, precision, F-measure and accuracy. It is especially important to note that REACT produces very high values for specificity and accuracy in almost all the configurations discussed in this chapter. Specificity can be considered very important in an access control scenario as a high value signifies a high "true negative rate" which means that the members who are originally intended to be denied access by the user are more likely to be recommended a "deny" decision by REACT. Thus, a high specificity shows that REACT is more likely to correctly predict the members of a user's friend network who should not be granted access and thus is capable of mitigating unintended disclosure of information. Nevertheless, for the overall performance of REACT, it is essential to have a high value for other metrics such as precision, sensitivity and especially F-measure which signifies the overall quality of classification. A low F-measure with a high specificity would suggest that while REACT is providing recommendations which enable safeguarding of privacy of the user as no information is shared to unintended audiences, the user would have to overturn a large number of "deny" recommendations to allow access to their desired audience which would require a lot of effort from them. A high F-measure, on the other hand, suggests that REACT is making fewer errors in predicting both types of access control decisions ("allow" and "deny") and hence minimizes the burden on the users of having to modify the recommendations to achieve their desired access control policy.

This chapter also discusses the potential of personalizing REACT by acknowledging the fact that privacy is personal and a "one-size-fits-all" solution may not be feasible to be implemented in a real-life scenario. The design of REACT can be easily adapted to accommodate personalization by implementing established feature selection techniques discussed in this chapter. The results show that using Principal Component Analysis (PCA) to select the most appropriate attribute set for users who grant access to a smaller section of their friend network may improve performance of REACT even more. When the most suitable configuration of REACT is implemented for each of the 26 users in the dataset, the average values for each of the metrics are: **Specificity = 96.8%**, **Sensitivity = 77.4%**, **Precision = 84.5%** & **F-measure = 0.808** and **Accuracy = 93.2%**.

Chapter 8

Conclusions and Future Work

The work presented in this thesis aims to enhance access control mechanisms in social media and ease the burden on social media users by providing them with accurate and contextual access control recommendations. The systematization of literature and social media infrastructure presented in Chapter 2 highlighted the absence of access control mechanisms which consider the overall context of information disclosed by users on social media sites. The primary contribution of the thesis is REACT, an access control recommendation mechanism which leverages interpersonal relationships between the users and each of their friends, by using attributes to define them in terms of type as well as strength in conjunction with information about the content.

8.1 Major Findings

This section highlights the major findings of this thesis in a nutshell.

CP has best fit with access control decisions

Chapter 5 presented detailed results of an empirical evaluation of 8 popular community detection algorithms with the objective of finding the algorithm which produced communities having the best fit with access control decisions made by users during a user study. The results helped identify CP as the algorithm which produced communities which had the best fit with access control decisions made by users. Therefore, this algorithm was chosen to represent relationship type in the implementation of REACT presented in Chapter 7.

Total Friends and Mutual Friends are most suitable profile attributes to predict access control decisions

A detailed evaluation of 30 profile attributes was presented in Chapter 6 which resulted in the identification of *Total Friends* (size of network) and *Mutual Friends* (shared contacts) as the most suitable profile attributes to predict access control decisions. Therefore, these two attributes in addition to *Friend Difference* (difference in size of network, which can be derived from *Total Friends*) were selected to represent relationship strength in the implementation of REACT.

REACT produces highly accurate access control recommendations

The evaluation of the performance of REACT with an implementation where relationship type was represented by CP membership, relationship strength was represented by total friends, mutual friends and friend difference and information about the content was provided in the form of social and topical tags provided by the users, was presented in Chapter 7. The results show that REACT produces highly accurate access control recommendations while also maintaining high values for specificity, which is important in an access control scenario. Chapter 7 also described several enhancements to the baseline implementation of REACT by showing that using class balancer filter improves the precision and sensitivity while maintaining the high accuracy and specificity of the recommendations. The chapter further discussed the ability of REACT to be personalized and presented results when established feature selection techniques, namely, Principal Component Analysis (PCA) and Information Gain were used to select the appropriate set of attributes for each user. The findings showed that using PCA to select attributes for users who grant access to smaller section of their friend network further enhances the performance of REACT.

REACT can be extended to other social networks

REACT was implemented as a Facebook application for the analyses presented in this thesis as it is widely acknowledged to be the most widely used online social network. The analyses led to the identification of attributes to represent the components of REACT, namely, community membership for relationship type (Chapter 5) and size of network (total friends) and common connections (mutual friends) for relationship strength (Chapter 6). REACT can be easily implemented in other social networks as these identified attributes are generally available in most online social networks. It must be noted that REACT will have maximum utility in social networks where the users have greater granularity in terms of creating access control policies and they are able to make access control decisions with respect to individual contacts (or "friends"). REACT can then assist the users by making access control recommendations for each of their contacts, as has been demonstrated in this thesis. However, in some social networks such as Twitter, which primarily have a "follower-broadcaster" model of social engagement, REACT may not be very useful. These social networks do not provide the user with the opportunity of making individual access control decisions with respect to each of their contacts. They can either broadcast the content publicly or just share it with all of their contacts ("protected" tweets are only shared with "followers" on Twitter).

8.2 Limitations

A limitation of the work presented in this thesis is that the collected data and the subsequent analyses is reliant on the activity of users during the user study described in Chapter 4. This is the way in which most related work evaluate similar mechanisms, and in the absence of any publicly available datasets, conducting a user study to obtain ground truth data was deemed to be the most suitable way. The dataset used for the analysis presented in this thesis contains 68,840 access control decisions which is much higher than previous related works [FL10, AFW12, SSLW11, ACF12]. There is always the possibility that fatigue may influence the user study as the users had to make access control decisions and select categories as well as tag any friends for 10 photos sequentially. However, the data suggests that this was not necessarily the case as as there were only 2 users whose number of "allow" decisions decreased after the first photo and many users had audience size of above 50 for the last photo.

A limitation specific to the analysis presented in Chapter 6 is the comparatively reduced size of the dataset of access control decisions which could be considered to evaluate profile attributes of the participants' friends. This was due to the API restrictions imposed by Facebook[1] (as discussed in Section 6.2.1) which meant that profile information of only users who participated in the user study was available. We did mitigate this by encouraging groups of people to participate in the study so that a particular participant would have some Facebook friends also participating in the study which would enable us to get their profile information. The participants, however, were unaware of this limitation and made access control decisions for all their friends normally. They were not informed of this limitation to avoid biasing their access control decisions. The identification of *Total Friends* and *Mutual Friends* to represent relationship strength ensures that future implementations of REACT will not have this limitation as this information is generally available in social media profiles, even with the restrictions, and are shown to accurately predict access control decisions in the analysis presented in Chapter 6.

8.3 Future Work

This section discusses some of the future directions in which research can progress on the basis of the findings of this thesis.

8.3.1 REACT for Mobile Platforms

A possible future work is to implement REACT on mobile platforms as the attributes considered in this thesis do not consider contextual information which can be gathered from mobile activity. It is undeniable that the advent of mobile applications for social media activity have introduced new information which can constitute the context of

[1] `https://developers.facebook.com/docs/apps/upgrading`

information disclosure. Contextual information such as location, for example, assume a far greater role when the user is sharing information via mobile platforms. It has already been shown, for example, that co-location as determined from mobile sensor data of individuals, can influence access control policies [MS12]. A possible future work can be to extend the design of REACT to include information available from mobile phones, such as GPS location, contact lists, etc., which were not relevant in the current evaluation presented in this thesis as REACT was implemented to be used as a desktop Facebook application.

8.3.2 Integrating REACT with Audience Visualization Tools

REACT has been found to recommend highly accurate access control decisions to social media users. It was discussed in Chapter 2 that in addition to access control recommendation, another way of enhancing existing access control mechanisms is to provide better visualization of the audience of the information being shared by the user. Future work may focus on integrating the access control recommendation of REACT with audience visualization tools such as “Audience View” [LBW08] to enhance comprehension of the access control recommendations made by REACT which can be provided to the users in a more actionable way.

8.3.3 Incorporating Multi-party Privacy

Another possible line of future work is to accommodate multi-party privacy into REACT. The design would have to be extended in a way in which the access control preferences of “other parties” (users other than the one who shares the content) related to the content (for e.g. appearing in a photo) are considered in the recommendation process. As outlined in Chapter 2, multi-party privacy is a very active area of research and there are several methods suggested in literature to try and achieve an agreement of access control policy to safeguard multi-party privacy issues, including bidding mechanisms, where the policy with most bids is adopted [SSP09], voting mechanisms either to obtain a majority [CF11] or “veto voting” [TGN10] and game-theoretic approaches [SR16, HAZY14]. One or more of these approaches could be included in the design of REACT to try and safeguard multi-party privacy when providing access control recommendations.

8.3.4 Access Control Recommendation on Social Media Aggregators (SMAs)

Social Media Aggregators (SMAs) are applications which enable the users to access multiple social media accounts from one mobile application. While this has many advantages in terms of ease of use (not having to install separate applications for different social media accounts) and management of resources such as battery, phone memory (one application replacing many), it has several obvious privacy implications. We performed an initial privacy assessment of SMAs which highlighted the potential privacy implications arising from the unique combination of phone data (such as contacts, sms, call records, etc.) with the users' activity on multiple social media accounts being used via the SMA [MSG17]. An interesting future work might be to understand how people manage audiences on SMAs, as there is a much larger scope of a "context collapse" due to the possibility of sharing information on multiple social networks simultaneously, and subsequently evaluate the possibility of access control recommendation for these applications.

8.4 Concluding Remarks

The work presented in this thesis tackles the problem of "unintended disclosure" which leads to privacy breaches for social media users. The main contribution of this thesis is REACT, which provides accurate access control recommendations to social media users using very little information about their friend network and without requiring any personal information. REACT can be implemented according to the requirements of the particular social media infrastructure as it uses information that is generally available in most, if not all, social networks. The evaluation of REACT presented in this thesis shows that the design is implementable and testing its performance with users in the wild would be a natural next step. As part of this implementation process, careful attention shall be paid to ensure the usability of REACT and its recommendations. This is crucial to ensure recommendations are incorporated easily and seamlessly into users' normal use of the social media infrastructure.

Appendix A

Information Sheet

Participant Information Sheet

My name is Gaurav Misra and I am conducting this research as a student in the School of Computing and Communication at Lancaster University, Lancaster, United Kingdom.

What is the study about?

The purpose of this study is to evaluate the factors which are considered while selecting an audience while posting information on social media platforms. I am interested in identifying the most important profile features which can be considered in order to provide intelligent suggestions to social media users when they are required to select an appropriate audience for their content.

Why have I been approached?

You have been approached because the study requires the participation of Facebook users who have more than 100 friends and in excess of 10 photos on the site.

Do I have to take part?

No. It’s completely up to you to decide whether or not you take part

What will I be asked to do if I take part?

If you decide you would like to take part, you would be asked to use a Facebook application, which is developed specifically for this experiment, to access your profile information. The application will download your profile information (which you have already uploaded to Facebook), photos (that you have already uploaded on Facebook), messages (not the content but only the total count) and wall posts (again, only the total count and not the content). The application will also download profile information (such as location, education, workplace, etc.) of your friends who have enabled you to access their profiles. (see “Why is data for my Facebook friends being collected?”).

You will also be asked to bring 5 photos of yourself, which have not been uploaded to Facebook, with you when you appear for the experiment. These photos will be stored on secure servers at Infolab and **no individual images will be published or disclosed under any circumstances**. **NOTHING WILL BE POSTED TO FACEBOOK BEFORE, DURING OR AFTER THE EXPERIMENT**. The environment is completely simulated and isolated and cannot post information to your profile or any other place in Facebook.

After the photos are uploaded, you will be shown 10 photos (5 which you uploaded and 5 randomly selected from the photos you have uploaded on Facebook previously). You will be asked to select the friends (from your Facebook friend list which will be shown to you on the screen) that you would want to share these photos with. You will also be asked to assign a category to the photo (from a list provided by the application) for us to understand the context of the photo.

The entire process should take between **30-45 minutes** in total. The participants would be invited in different batches (according to their preferred time slots as indicated to the researcher as far as possible). The location of the study will be D-53, Infolab, Lancaster University.

Will my data be Identifiable?

The information you provide is confidential. It will, of course be information which is already present in the Facebook servers and will be accessed from there. The data collected for this study will be stored securely and only the researchers conducting this study will have access to this data:

- The databases and files on the computer will be encrypted (that is no-one other than the researcher will be able to access them) and the computer itself password protected.
- After the completion of the data collection (when all participants complete their tasks) the data will be anonymized and aggregated for the analysis. I will ensure that it is impossible to uniquely identify any participants from the data stored on our computers.

Why is data for my Facebook friends being collected?

The major objective of this experiment is to try and identify similarity between a participant and their friends in terms of them being in the same location, going to the same school (education), same workplace, etc. This information is important to analyse whether such similarities influence a participant's decision making when she is required to select an audience for her content. This is why the application will collect profile information (such as education, location, workplace, etc.) from your friends. Importantly, this will only be information that is already visible to you. **No private information from your friends' profiles will be collected such as photos, messages or wall posts that your friends have shared on Facebook.**

What will happen to the results?

The results will be summarised, anonymized and aggregated before they are submitted for publication in an academic or professional conference/journal. They will also be included (in anonymized and aggregated form) in my PhD thesis.

Are there any risks?

There are no risks anticipated with participating in this study.

Will I be able to withdraw from the study?

Yes. You will be able to withdraw from this study. If you withdraw within 2 weeks after completion of the experiment, your data will be taken out of the study and destroyed. If you withdraw later, the data will remain in the study.

Are there any benefits to taking part?

You will receive a cash reward of £10 on completion of your participation.

Who has reviewed the project?

This study has been reviewed by the University Research Ethics Committee (UREC) and the Head of Security Lancaster.

Where can I obtain further information about the study if I need it?

If you have any questions about the study, please contact the main researcher:

Gaurav Misra
PhD Researcher
School of Computing and Communications
g.misra@lancaster.ac.uk

Complaints

If you wish to make a complaint or raise concerns about any aspect of this study and do not want to speak to the researcher, you can contact:

Professor Awais Rashid
Head of Security Lancaster
Email: a.rashid@lancaster.ac.uk
School of Computing and Communications
Lancaster University

Appendix B

Consent Form

Security Lancaster | Lancaster University

Consent Form

Study Title: Automated Relationship based Privacy in Social Media

We are asking if you would like to take part in a research project to understand the factors which are considered to be important by users of social media applications while sharing their information on these platforms.

Before you consent to participating in the study we ask that you read the participant information sheet and mark each box below with your initials if you agree. If you have any questions or queries before signing the consent form please speak to the principal investigator, Gaurav Misra.

Please initial each statement

1. I confirm that I have read the information sheet and fully understand what is expected of me within this study □
2. I confirm that I have had the opportunity to ask any questions and to have them answered. □
3. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason. □
4. I understand that the information collected during the study will be pooled with that of other participants, anonymised and aggregated before being published □
5. I understand that once my data have been anonymised and incorporated into themes it might not be possible for it to be withdrawn, though every attempt will be made to extract my data, up to the point of publication. □
6. I consent to take part in the above study. □

Name of Participant___________________ **Signature**_____________________ **Date** ___________

Name of Researcher ___________________**Signature** _____________________**Date** ___________

Bibliography

[AÁK12] Chaabane Abdelberi, Gergely Ács, and Mohamed Ali Kâafar. You are what you like! information leakage through users' interests. In *19th Annual Network and Distributed System Security Symposium, NDSS 2012, San Diego, California, USA, February 5-8, 2012*, 2012. 18

[ABL15] Alessandro Acquisti, Laura Brandimarte, and George Loewenstein. Privacy and human behavior in the age of information. *Science*, 347(6221):509–514, 2015. 103

[ACD$^+$06] Claudio A Ardagna, Marco Cremonini, Ernesto Damiani, Sabrina De Capitani di Vimercati, and Pierangela Samarati. Supporting location-based conditions in access control policies. In *Proceedings of the 2006 ACM Symposium on Information, computer and communications security*, pages 212–222, 2006. 27

[ACF12] Cuneyt 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. 29, 112

[ADGPV06] Alex Arenas, Albert D\'\iaz-Guilera, and Conrad J Pérez-Vicente. Synchronization reveals topological scales in complex networks. *Physical review letters*, 96(11):114102, 2006. 55

[AEE$^+$14] Gunes Acar, Christian Eubank, Steven Englehardt, Marc Juarez, Arvind Narayanan, and Claudia Diaz. The web never forgets: Persistent tracking mechanisms in the wild. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, CCS '14, pages 674–689, New York, NY, USA, 2014. ACM. 2

[AFW12] Saleema Amershi, James Fogarty, and Daniel Weld. Regroup: Interactive machine learning for on-demand group creation in social networks. pages 21–30, 2012. 4, 27, 29, 30, 32, 34, 42, 71, 72, 73, 81, 85, 87, 112

[AHB04] A. I. Anton, Qingfeng He, and D. L. Baumer. Inside jetblue's privacy policy violations. *IEEE Security Privacy*, 2(6):12–18, Nov 2004. 18

[Ale14] Alexa.com. Top sites in social networking. `http://www.alexa.com/topsites/category/Computers/Internet/On_the_Web/Online_Communities/Social_Networking`, 2014. 21

[ANH16] J. H. Abawajy, M. I. H. Ninggal, and T. Herawan. Privacy preserving social network data publication. *IEEE Communications Surveys Tutorials*, 18(3):1974–1997, thirdquarter 2016. 20

[Ant14] Pauline Anthonysamy. *A Framework to Detect Information Asymmetries between Privacy Policies and Controls of Online Social Networks.* PhD thesis, Security Lancaster, Lancaster University, UK, 2014. 2

[BAAS09] Joseph Bonneau, Jonathan Anderson, Ross Anderson, and Frank Stajano. Eight friends are enough: social graph approximation via public listings. In *Proceedings of the Second ACM EuroSys Workshop on Social Network Systems*, pages 13–18. ACM, 2009. 17

[BAD09] Joseph Bonneau, Jonathan Anderson, and George Danezis. Prying data out of a social network. In *Social Network Analysis and Mining, 2009. ASONAM'09. International Conference on Advances in*, pages 249–254. IEEE, 2009. 17, 19

[BBA75] Ronald L Breiger, Scott A Boorman, and Phipps Arabie. An algorithm for clustering relational data with applications to social network analysis and comparison with multidimensional scaling. *Journal of Mathematical Psychology*, 12(3):328–383, 1975. 54

[BDK07] Lars Backstrom, Cynthia Dwork, and Jon Kleinberg. Wherefore art thou r3579x?: Anonymized social networks, hidden patterns, and structural steganography. In *Proceedings of the 16th International Conference on World Wide Web*, WWW '07, pages 181–190, New York, NY, USA, 2007. ACM. 20

[BFK14] Karl Bringmann, Tobias Friedrich, and Anton Krohmer. *De-anonymization of Heterogeneous Random Graphs in Quasilinear Time*, pages 197–208. Springer Berlin Heidelberg, Berlin, Heidelberg, 2014. 20

[BGLL08] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, 2008. 54

[BK73] Coen Bron and Joep Kerbosch. Algorithm 457: finding all cliques of an undirected graph. *Communications of the ACM*, 16(9):575–577, 1973. 53

[BS08] Justin Brickell and Vitaly Shmatikov. The cost of privacy: Destruction of data-mining utility in anonymized data publishing. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '08, pages 70–78, New York, NY, USA, 2008. ACM. 15, 20

[BZ03] Vladimir Batagelj and Matjaz Zaversnik. An O (m) algorithm for cores decomposition of networks. *arXiv preprint cs/0310049*, 2003. 53

[CF11] Barbara Carminati and Elena Ferrari. Collaborative access control in on-line social networks. In *Collaborative computing: networking, applications and worksharing (CollaborateCom), 2011 7th international conference on*, pages 231–240. IEEE, 2011. 15, 17, 114

[CGNP12] S. Creese, M. Goldsmith, J. R. C. Nurse, and E. Phillips. A data-reachability model for elucidating privacy and security risks related to the use of online social networks. In *2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications*, pages 1124–1131, June 2012. 14

[Cha04] Deepayan Chakrabarti. Autopart: Parameter-free graph partitioning and outlier detection. In *Knowledge Discovery in Databases: PKDD 2004*, pages 112–124. Springer, 2004. 55

[CLB+16] Gul Calikli, Mark Law, Arosha K Bandara, Alessandra Russo, Luke Dickens, Blaine A Price, Avelie Stuart, Mark Levine, and Bashar Nuseibeh. Privacy dynamics: Learning privacy norms for social software. In *Proceedings of the 11th International Symposium on Software Engineering for Adaptive and Self-Managing Systems*, pages 47–56. ACM, 2016. 27, 77, 98

[CN06] Gabor Csardi and Tamas Nepusz. The igraph software package for complex network research. *InterJournal, Complex Systems*, 2006. 58

[CNM04] Aaron Clauset, Mark E J Newman, and Cristopher Moore. Finding community structure in very large networks. *Physical review E*, 70(6):66111, 2004. 54

[Cra03] Lorrie Faith Cranor. P3p: Making privacy policies more useful. *IEEE Security & Privacy*, 99(6):50–55, 2003. 15, 18

[CS14] Gorrell P Cheek and Mohamed Shehab. Human Effects of Enhanced Privacy Management Models. *Dependable and Secure Computing, IEEE Transactions on*, 11(2):142–154, 2014. 28, 32, 52

[DA05] Jordi Duch and Alex Arenas. Community detection in complex networks using extremal optimization. *Physical review E*, 72(2):27104, 2005. 54

[Dan09] George Danezis. Inferring privacy policies for social networking services. pages 5–10, 2009. 28, 32, 52

[DM04] Luca Donetti and Miguel A Munoz. Detecting network communities: a new systematic and efficient algorithm. *Journal of Statistical Mechanics: Theory and Experiment*, 2004(10):P10012, 2004. 53

[DPV05] Imre Derényi, Gergely Palla, and Tamás Vicsek. Clique percolation in random networks. *Physical review letters*, 94(16):160202, 2005. xi, 36, 92, 93

[DTRS12] R. Dey, C. Tang, K. Ross, and N. Saxena. Estimating age privacy leakage in online social networks. In *2012 Proceedings IEEE INFOCOM*, pages 2836–2840, March 2012. 17

[DTS08] Claudia Diaz, Carmela Troncoso, and Andrei Serjantov. *On the Impact of Social Network Profiling on Anonymity*, pages 44–62. Springer Berlin Heidelberg, 2008. 20

[Elk01] Charles Elkan. The foundations of cost-sensitive learning. In *International joint conference on artificial intelligence*, volume 17, pages 973–978, 2001. 79, 102

[EN16] Steven Englehardt and Arvind Narayanan. Online tracking: A 1-million-site measurement and analysis. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, CCS '16, pages 1388–1401, New York, NY, USA, 2016. ACM. 19

[EOK11] Serge Egelman, Andrew Oates, and Shriram Krishnamurthi. Oops, I did it again: Mitigating repeated access control errors on Facebook. pages 2295–2304, 2011. 2, 14

[Fac16] Facebook. Terms as of 2015. `https://www.facebook.com/legal/terms`, Retr. 28/11/2016. 19

[FFB15] Casey Fiesler, Jessica L. Feuston, and Amy S. Bruckman. Understanding copyright law in online creative communities. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work &#38; Social Computing*, CSCW '15, pages 116–129. ACM, 2015. 18

[FL10] Lujun Fang and Kristen LeFevre. Privacy wizards for social networking sites. pages 351–360, 2010. 4, 15, 16, 27, 28, 32, 34, 42, 52, 112

[FLG00] Gary William Flake, Steve Lawrence, and C Lee Giles. Efficient identification of web communities. pages 150–160, 2000. 54

[FLM04] Santo Fortunato, Vito Latora, and Massimo Marchiori. Method to find community structures based on information centrality. *Physical review E*, 70(5):56104, 2004. 54

[Fon11a] Philip W L Fong. Relationship-based access control: protection model and policy language. In *Proceedings of the first ACM conference on Data and application security and privacy*, pages 191–202, 2011. 33

[Fon11b] Philip WL Fong. Preventing sybil attacks by privilege attenuation: A design principle for social network systems. In *Security and privacy (SP), 2011 IEEE symposium on*, pages 263–278. IEEE, 2011. 17

[For10] Santo Fortunato. Community detection in graphs. *Physics Reports*, 486(3), 2010. 55

[FSEGF14] 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. 15, 17, 29, 37, 42, 52, 56, 73, 85, 87, 93

[FSEGF15] Ricard Fogues, Jose M Such, Agustin Espinosa, and Ana Garcia-Fornes. Open challenges in relationship-based privacy mechanisms for social network services. *International Journal of Human-Computer Interaction*, 31(5):350–370, 2015. 4

[GA05] Ralph Gross and Alessandro Acquisti. Information revelation and privacy in online social networks. pages 71–80, 2005. 2

[GK09] 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, 2009. 36, 70, 93

[GL16] Neil Zhenqiang Gong and Bin Liu. You are who you know and how you behave: Attribute inference attacks via users' social friends and behaviors. In *25th USENIX Security Symposium (USENIX Security 16)*, pages 979–995, Austin, TX, 2016. USENIX Association. 17

[GN02] Michelle Girvan and Mark E J Newman. Community structure in social and biological networks. 99(12):7821–7826, 2002. 54

[Gra73] M.S. Granovetter. The strength of weak ties. *American journal of sociology*, pages 1360–1380, 1973. 70

[Gre10] Steve Gregory. Finding overlapping communities in networks by label propagation. *New Journal of Physics*, 12(10):103018, 2010. 55

[GZR$^+$10] Ido Guy, Naama Zwerdling, Inbal Ronen, David Carmel, and Erel Uziel. Social media recommendation based on people and tags. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, pages 194–201. ACM, 2010. 31

[Hal99] Mark A Hall. *Correlation-based feature selection for machine learning.* PhD thesis, The University of Waikato, 1999. 103

[HAZY14] Hongxin Hu, Gail-Joon Ahn, Ziming Zhao, and Dejun Yang. Game theoretic analysis of multiparty access control in online social networks. In *Proceedings of the 19th ACM Symposium on Access Control Models and Technologies*, SACMAT '14, pages 93–102, New York, NY, USA, 2014. ACM. 15, 17, 114

[HCL06] Jianming He, Wesley W. Chu, and Zhenyu (Victor) Liu. *Inferring Privacy Information from Social Networks*, pages 154–165. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006. 17

[HDO$^+$98] Marti A Hearst, Susan T Dumais, Edgar Osman, John Platt, and Bernhard Scholkopf. Support vector machines. *Intelligent Systems and their Applications, IEEE*, 13(4):18–28, 1998. 78, 92

[HFH$^+$09] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H Witten. The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009. 78, 81, 92

[HJ10] 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. 14, 15, 17

[HKT13] Raymond Heatherly, Murat Kantarcioglu, and Bhavani M. Thuraisingham. Preventing private information inference attacks on social networks. *IEEE Trans. on Knowl. and Data Eng.*, 25(8):1849–1862, August 2013. 15, 18

[HLL11] Gordon Hull, Heather Richter Lipford, and Celine Latulipe. Contextual gaps: Privacy issues on Facebook. *Ethics and information technology*, 13(4):289–302, 2011. 3, 14, 20

[Hog16] Michael A Hogg. Social identity theory. In *Understanding Peace and Conflict Through Social Identity Theory*, pages 3–17. Springer, 2016. 27

[HSGH13] Mathias Humbert, Théophile Studer, Matthias Grossglauser, and Jean-Pierre Hubaux. *Nowhere to Hide: Navigating around Privacy in Online Social Networks*, pages 682–699. Springer Berlin Heidelberg, 2013. 17

[IKN05] Hidehiko Ino, Mineichi Kudo, and Atsuyoshi Nakamura. Partitioning of web graphs by community topology. pages 661–669, 2005. 54

[IPA$^+$15] Panagiotis Ilia, Iasonas Polakis, Elias Athanasopoulos, Federico Maggi, and Sotiris Ioannidis. Face/off: Preventing privacy leakage from photos in social networks. In *Proceedings of the 22nd ACM Conference on Computer and Communications Security*, pages 781–792. ACM, 2015. 16

[JBLG05] James B D Joshi, Elisa Bertino, Usman Latif, and Arif Ghafoor. A generalized temporal role-based access control model. *IEEE Transactions on Knowledge and Data Engineering*, 17(1):4–23, 2005. 27

[JEB12] Maritza Johnson, Serge Egelman, and Steven M Bellovin. Facebook and privacy: it's complicated. page 9, 2012. 2, 3, 14, 16, 17, 26, 35

[JO10] Simon Jones and Eamonn O'Neill. Feasibility of structural network clustering for group-based privacy control in social networks. page 9, 2010. 28, 32, 52

[JS02] Nathalie Japkowicz and Shaju Stephen. The class imbalance problem: A systematic study. *Intelligent data analysis*, 2002. 79, 100

[KBHC12] Sanjay Kairam, Mike Brzozowski, David Huffaker, and Ed Chi. Talking in circles: selective sharing in google+. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1065–1074. ACM, 2012. 28, 29, 38, 55, 96, 100

[KLM$^+$12] Peter Klemperer, Yuan Liang, Michelle Mazurek, Manya Sleeper, Blase Ur, Lujo Bauer, Lorrie Faith Cranor, Nitin Gupta, and Michael Reiter. Tag, you can see it!: using tags for access control in photo sharing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 377–386, 2012. 5, 27, 30, 31

[KST+08] Onur Kucuktunc, Sare G Sevil, A Burak Tosun, Hilal Zitouni, Pinar Duygulu, and Fazli Can. Tag suggestr: Automatic photo tag expansion using visual information for photo sharing websites. In *International Conference on Semantic and Digital Media Technologies*, pages 61–73. Springer, 2008. 38

[KVV04] Ravi Kannan, Santosh Vempala, and Adrian Vetta. On clusterings: Good, bad and spectral. *Journal of the ACM (JACM)*, 51(3):497–515, 2004. 54

[KW52] William H Kruskal and W Allen Wallis. Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 47(260):583–621, 1952. 58

[KW08] 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. 17

[KW09] Balachander Krishnamurthy and Craig E Wills. On the leakage of personally identifiable information via online social networks. In *Proceedings of the 2nd ACM workshop on Online social networks*, pages 7–12. ACM, 2009. 19

[LBW08] Heather Richter Lipford, Andrew Besmer, and Jason Watson. Understanding Privacy Settings in Facebook with an Audience View. *UPSEC*, 8:1–8, 2008. 15, 16, 114

[LCZT07] Yijuan Lu, Ira Cohen, Xiang Sean Zhou, and Qi Tian. Feature selection using principal feature analysis. In *Proceedings of the 15th ACM international conference on Multimedia*, pages 301–304. ACM, 2007. 103

[LGKM11] 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, 2011. 14, 16

[LGZ08] Xin Li, Lei Guo, and Yihong Eric Zhao. Tag-based social interest discovery. In *Proceedings of the 17th international conference on World Wide Web*, pages 675–684, 2008. 30

[LHKT09] Jack Lindamood, Raymond Heatherly, Murat Kantarcioglu, and Bhavani Thuraisingham. Inferring private information using social network data. In *Proceedings of the 18th international conference on World wide web*, pages 1145–1146. ACM, 2009. 17, 19

[LHL+09] 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, 2009. 4

[LHLC09] Ian X Y Leung, Pan Hui, Pietro Lio, and Jon Crowcroft. Towards real-time community detection in large networks. *Physical Review E*, 79(6):66107, 2009. 55

[LHY+09] Dong Liu, Xian-Sheng Hua, Linjun Yang, Meng Wang, and Hong-Jiang Zhang. Tag ranking. In *Proceedings of the 18th international conference on World wide web*, pages 351–360. ACM, 2009. 38

[Lil67] Hubert W Lilliefors. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, 62(318):399–402, 1967. 58

[LLLT11] Airi Lampinen, Vilma Lehtinen, Asko Lehmuskallio, and Sakari Tamminen. We're in it together: interpersonal management of disclosure in social network services. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 3217–3226. ACM, 2011. 4

[LLWG11] Qingrui Li, Juan Li, Hui Wang, and Ashok Ginjala. Semantics-enhanced privacy recommendation for social networking sites. In *2011IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications*, pages 226–233, 2011. 29, 32

[LS11] Charles X Ling and Victor S Sheng. Cost-sensitive learning. In *Encyclopedia of Machine Learning*, pages 231–235. Springer, 2011. 79, 102

[LS14] Jure Leskovec and Rok Sosič. SNAP: A general purpose network analysis and graph mining library in C++. http://snap.stanford.edu/snap, jun 2014. 58, 93

[LW02] Andy Liaw and Matthew Wiener. Classification and regression by randomForest. *R news*, 2(3):18–22, 2002. 78, 92

[LWMH13] Sebastian Labitzke, Florian Werling, Jens Mittag, and Hannes Hartenstein. Do online social network friends still threaten my privacy? In *Proceedings of the Third ACM Conference on Data and Application Security and Privacy*, CODASPY '13, pages 13–24, New York, NY, USA, 2013. ACM. 17

[MC10] Aleecia McDonald and Lorrie Faith Cranor. Beliefs and behaviors: Internet users' understanding of behavioral advertising. In *Proceedings of the 2010 Research Conference on Communication, Information and Internet Policy*, October 2010. 14

[ML12] Julian J McAuley and Jure Leskovec. Learning to Discover Social Circles in Ego Networks. In *NIPS*, volume 272, pages 548–556, 2012. 4, 27, 29, 32, 34, 73, 75, 85, 87

[MLA12] Alessandra Mazzia, Kristen LeFevre, and Eytan Adar. No Title. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, page 13, 2012. 15, 16

[MM12] Jonathan R Mayer and John C Mitchell. Third-party web tracking: Policy and technology. In *Security and Privacy (SP), 2012 IEEE Symposium on*, pages 413–427. IEEE, 2012. 2, 19

[MNBD06] Cameron Marlow, Mor Naaman, Danah Boyd, and Marc Davis. Position paper, tagging, taxonomy, flickr, article, toread. In *In Collaborative Web Tagging Workshop at WWW'06*, 2006. 31, 38

[MO11] Alice E Marwick and Others. I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, 13(1):114–133, 2011. 2, 3, 14, 20, 25, 38, 39, 96

[MS12] Pradeep Murukannaiah and Munindar Singh. Platys social: Relating shared places and private social circles. *IEEE Internet Computing*, 16(3):53–59, 2012. 114

[MS16] Gaurav Misra and Jose M Such. How Socially Aware Are Social Media Privacy Controls? *Computer*, 49(3):96–99, 2016. 4, 27, 32, 33, 95

[MSG17] Gaurav Misra, Jose M Such, and Lauren Gill. A privacy assessment of social media aggregators. In *Advances in Social Networks Analysis and Mining (ASONAM), 2017 IEEE/ACM International Conference on.* IEEE, 2017. 115

[MTKC] Brendan Meeder, Jennifer Tam, Patrick Gage Kelley, and Lorrie Faith Cranor. Rt@ iwantprivacy: Widespread violation of privacy settings in the twitter social network. In *Web 2.0 Security & Privacy 2010.* 16

[Mur06] Kevin P Murphy. Naive bayes classifiers. *University of British Columbia*, 2006. 78, 92

[MVC$^+$12] Mainack Mondal, Bimal Viswanath, Allen Clement, Peter Druschel, Krishna P. Gummadi, Alan Mislove, and Ansley Post. Defending against large-scale crawls in online social networks. In *Proceedings of the 8th International Conference on Emerging Networking Experiments and Technologies*, CoNEXT '12, pages 325–336, New York, NY, USA, 2012. ACM. 15, 19

[MVGD10] Alan Mislove, Bimal Viswanath, Krishna P. Gummadi, and Peter Druschel. You are who you know: Inferring user profiles in online social networks. In *Proceedings of the Third ACM International Conference on Web Search and Data Mining*, WSDM '10, pages 251–260, New York, NY, USA, 2010. ACM. 17

[MW47] Henry B Mann and Donald R Whitney. On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, pages 50–60, 1947. 59

[NG04] Mark E J Newman and Michelle Girvan. Finding and evaluating community structure in networks. *Physical review E*, 69(2):26113, 2004. 54

[Nis04] Helen Nissenbaum. Privacy as contextual integrity. *Washington Law Review*, 79:119, 2004. 3, 71

[NS09] Arvind Narayanan and Vitaly Shmatikov. De-anonymizing social networks. In *Security and Privacy, 2009 30th IEEE Symposium on*, pages 173–187. IEEE, 2009. 20

[PAP+15] Iasonas Polakis, George Argyros, Theofilos Petsios, Suphannee Sivakorn, and Angelos D. Keromytis. Where's wally?: Precise user discovery attacks in location proximity services. In *Proceedings of the 22Nd ACM SIGSAC Conference on Computer and Communications Security*, CCS '15, pages 817–828, New York, NY, USA, 2015. ACM. 18

[PCNR10] Rahul Potharaju, Bogdan Carbunar, and Cristina Nita-Rotaru. ifriendu: leveraging 3-cliques to enhance infiltration attacks in online social networks. In *Proceedings of the 17th ACM conference on Computer and communications security*, pages 723–725. ACM, 2010. 17

[PDFV05] Gergely Palla, Imre Derényi, Illés Farkas, and Tamás Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043):814–818, 2005. 53

[PKA+10] Iasonas Polakis, Georgios Kontaxis, Spiros Antonatos, Eleni Gessiou, Thanasis Petsas, and Evangelos P Markatos. Using social networks to harvest email addresses. In *Proceedings of the 9th Annual ACM Workshop on Privacy in the Electronic Society*, pages 11–20. ACM, 2010. 19

[PKVS12] Symeon Papadopoulos, Yiannis Kompatsiaris, Athena Vakali, and Ploutarchos Spyridonos. Community detection in social media. *Data Mining and Knowledge Discovery*, 24(3):515–554, 2012. 35, 36, 53, 55

[PL05] Pascal Pons and Matthieu Latapy. Computing communities in large networks using random walks. pages 284–293. Springer, 2005. 54

[PLZW14] W. Peng, F. Li, X. Zou, and J. Wu. A two-stage deanonymization attack against anonymized social networks. *IEEE Transactions on Computers*, 63(2):290–303, Feb 2014. 20

[PM11] Siani Pearson and Marco Casassa Mont. Sticky policies: An approach for managing privacy across multiple parties. *Computer*, 44(9):60–68, 2011. 15, 16

[RAK07] Usha Nandini Raghavan, Réka Albert, and Soundar Kumara. Near linear time algorithm to detect community structures in large-scale networks. *Physical Review E*, 76(3):36106, 2007. 55

[RB06] Jörg Reichardt and Stefan Bornholdt. Statistical mechanics of community detection. *Physical Review E*, 74(1):16110, 2006. 55

[RB08] Martin Rosvall and Carl T Bergstrom. Maps of random walks on complex networks reveal community structure. 105(4):1118–1123, 2008. 55

[RCC+04] Filippo Radicchi, Claudio Castellano, Federico Cecconi, Vittorio Loreto, and Domenico Parisi. Defining and identifying communities in networks. 101(9):2658–2663, 2004. 54

[RKY06] Indrakshi Ray, Mahendra Kumar, and Lijun Yu. LRBAC: a location-aware role-based access control model. In *International Conference on Information Systems Security*, pages 147–161, 2006. 27

[RSG98] Michael G Reed, Paul F Syverson, and David M Goldschlag. Anonymous connections and onion routing. *IEEE Journal on Selected areas in Communications*, 16(4):482–494, 1998. 15, 19

[S$^+$10] Jesse H Sowell et al. *Deficiencies in online privacy policies: factors and policy recommendations*. PhD thesis, Massachusetts Institute of Technology, 2010. 2, 18

[SC16] Jose M Such and Natalia Criado. Resolving multi-party privacy conflicts in social media. *IEEE Transactions on Knowledge and Data Engineering*, 28(7):1851–1863, 2016. 3

[SGA13] Fred Stutzman, Ralph Gross, and Alessandro Acquisti. Silent listeners: The evolution of privacy and disclosure on facebook. *Journal of privacy and confidentiality*, 4(2):2, 2013. 100

[SHC$^+$09] Norman M. Sadeh, Jason I. Hong, Lorrie Faith Cranor, Ian Fette, Patrick Gage Kelley, Madhu K. Prabaker, and Jinghai Rao. Understanding and capturing people's privacy policies in a mobile social networking application. *Personal and Ubiquitous Computing*, 2009. 15, 18

[SJL10] Ming-Yi Shih, Jar-Wen Jheng, and Lien-Fu Lai. A two-step method for clustering mixed categroical and numeric data. *Tamkang Journal of science and Engineering*, 13(1):11–19, 2010. 64

[SKLD14] Anna Squicciarini, Sushama Karumanchi, Dan Lin, and Nicole DeSisto. Identifying hidden social circles for advanced privacy configuration. *Computers & Security*, 41:40–51, 2014. 29, 32

[SKW15] Florian Schaub, Bastian Könings, and Michael Weber. Context-adaptive privacy: Leveraging context awareness to support privacy decision making. *IEEE Pervasive Computing*, 14(1):34–43, 2015. 27

[SLSW15] Anna Cinzia Squicciarini, Dan Lin, Smitha Sundareswaran, and Joshua Wede. Privacy policy inference of user-uploaded images on content sharing sites. *IEEE transactions on knowledge and data engineering*, 27(1):193–206, 2015. 31, 32, 34, 95

[SM00] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8):888–905, 2000. 54

[Sol06] Daniel J. Solove. A Taxonomy of Privacy. *University of Pennsylvania Law Review*, 154(3):477–560, Januar 2006. 3, 16

[SR16] Jose M Such and Michael Rovatsos. Privacy policy negotiation in social media. *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*, 11(1):4, 2016. 15, 17, 114

[SSLW11] Anna Cinzia Squicciarini, Smitha Sundareswaran, Dan Lin, and Josh Wede. A3p: adaptive policy prediction for shared images over popular content sharing sites. In *Proceedings of the 22nd ACM conference on Hypertext and hypermedia*, pages 261–270, 2011. 5, 15, 16, 27, 30, 31, 32, 38, 42, 112

[SSM11] Ravi Inder Singh, Manasa Sumeeth, and James Miller. A user-centric evaluation of the readability of privacy policies in popular web sites. *Information Systems Frontiers*, 13:501–514, sep 2011. 18

[SSP09] 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, 2009. 15, 17, 114

[TCS12] Antonio Tapiador, Diego Carrera, and Joaqu\'\in Salvachúa. Tie-RBAC: an application of RBAC to Social Networks. *arXiv preprint arXiv:1205.5720*, 2012. 70

[TGN10] Kurt Thomas, Chris Grier, and David M Nicol. unfriendly: Multi-party privacy risks in social networks. In *International Symposium on Privacy Enhancing Technologies Symposium*, pages 236–252. Springer, 2010. 15, 16, 17, 114

[Was94] Stanley Wasserman. *Social network analysis: Methods and applications*, volume 8. Cambridge university press, 1994. 54

[WBIT12] Udi Weinsberg, Smriti Bhagat, Stratis Ioannidis, and Nina Taft. Blurme: Inferring and obfuscating user gender based on ratings. In *Proceedings of the Sixth ACM Conference on Recommender Systems*, RecSys '12, pages 195–202, New York, NY, USA, 2012. ACM. 18

[WHKK10] Gilbert Wondracek, Thorsten Holz, Engin Kirda, and Christopher Kruegel. A practical attack to de-anonymize social network users. In *2010 IEEE Symposium on Security and Privacy (SP)*, pages 223–238. IEEE, 2010. 20

[WKC$^+$11] 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 *UbiComp*, pages 197–206, 2011. 37

[WKR14] Pamela Wisniewski, Bart P Knijnenburg, and H Richter Lipford. Profiling Facebook Users' Privacy Behaviors. In *SOUPS2014 Workshop on Privacy Personas and Segmentation*, 2014. 3, 4, 20, 26

[WLA$^+$14] Yang Wang, Pedro Giovanni Leon, Alessandro Acquisti, Lorrie Faith Cranor, Alain Forget, and Norman Sadeh. A field trial of privacy nudges for facebook. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pages 2367–2376, 2014. 15

[WLW12] Pamela Wisniewski, Heather Lipford, and David Wilson. Fighting for my space: Coping mechanisms for SNS boundary regulation. pages 609–618, 2012. 3

[WNK+11] Yang Wang, Gregory Norcie, Saranga Komanduri, Alessandro Acquisti, Pedro Giovanni Leon, and Lorrie Faith Cranor. I regretted the minute i pressed share: A qualitative study of regrets on facebook. In *Proceedings of the Seventh Symposium on Usable Privacy and Security*, page 10. ACM, 2011. 14

[WSB+10] Christo Wilson, Alessandra Sala, Joseph Bonneau, Robert Zablit, and Ben Y. Zhao. Don't tread on me: Moderating access to osn data with spikestrip. In *Proceedings of the 3rd Wonference on Online Social Networks*, WOSN'10, pages 5–5, Berkeley, CA, USA, 2010. USENIX Association. 15, 19

[XYF+07] Xiaowei Xu, Nurcan Yuruk, Zhidan Feng, Thomas Schweiger, and AJ. Scan: a structural clustering algorithm for networks. pages 824–833, 2007. 53

[YK12] Hakan Yildiz and Christopher Kruegel. Detecting social cliques for automated privacy control in online social networks. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2012 IEEE International Conference on*, pages 353–359, 2012. 32

[YKGS09] Ching-man Au Yeung, Lalana Kagal, Nicholas Gibbins, and Nigel Shadbolt. Providing Access Control to Online Photo Albums Based on Tags and Linked Data. In *AAAI Spring Symposium: Social Semantic Web: Where Web 2.0 Meets Web 3.0*, pages 9–14, 2009. 31, 38, 94

[YLL+09] Ching-man Au Yeung, Ilaria Liccardi, Kanghao Lu, Oshani Seneviratne, and Tim Berners-Lee. Decentralization: The future of online social networking. In *W3C Workshop on the Future of Social Networking Position Papers*, volume 2, pages 2–7, 2009. 15, 19

[ZG09] Elena Zheleva and Lise Getoor. To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles. In *WWW '09: Proceedings of the 18th international conference on World wide web*, pages 531–540, New York, NY, USA, 2009. ACM. 17

[ZSM11] Mark Zuckerberg, Aaron Sittig, and Scott Marlette. Tagging digital media, May 17 2011. US Patent 7,945,653. 38