PERCEPÇÃO DE PRIVACIDADE EM REDES SOCIAIS

GUSTAVO COSTA RAUBER

PERCEPÇÃO DE PRIVACIDADE EM REDES SOCIAIS

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais como requisito parcial para a obtenção do grau de Mestre em Ciência da Computação.

Orientador: Virgílio Augusto Fernandes de Almeida

Belo Horizonte

Setembro de 2012

GUSTAVO COSTA RAUBER

PRIVACY AWARENESS IN SOCIAL NETWORKS

Dissertation presented to the Graduate Program in Computer Science of the Universidade Federal de Minas Gerais in partial fulfillment of the requirements for the degree of Master in Computer Science.

Advisor: Virgílio Augusto Fernandes de Almeida

Belo Horizonte

September 2012

Rauber, Gustavo Costa
R123p Percepção de Privacidade em Redes Sociais / Gustavo Costa Rauber. — Belo Horizonte, 2012
xxiv, 54 f. : il. ; 29cm

Dissertação (mestrado) — Universidade Federal de Minas Gerais
Orientador: Virgílio Augusto Fernandes de Almeida

1. Computação - Teses. 2. Redes sociais on-line - Teses. 3. Privacidade - Teses. I. Orientador. II. Título.

CDU 519.6*04(043)

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

FOLHA DE APROVAÇÃO

Percepção de privacidade em redes sociais

GUSTAVO COSTA RAUBER

Dissertação defendida e aprovada pela banca examinadora constituída pelos Senhores:

PROF. VIRGÍLIO AUGUSTO FERNANDES ALMEIDA - Orientador
Departamento de Ciência da Computação - UFMG

PROFA. OLGA NIKOLAEVNA GOUSSEVSKAIA
Universidade Federal de Minas Gerais - Brasil

PROFA. RAQUEL OLIVEIRA PRATES
Departamento de Ciência da Computação - UFMG

Belo Horizonte, 06 de setembro de 2012.

This work is dedicated to Professor Christiano Gonçalves Becker (in memoriam).

Acknowledgments

I owe my deepest gratitude to my whole family for their love and paramount support throughout this dissertation. I am also thankful to my wife, Mariana Caetano, for her loving support.

I am grateful to have worked with my advisor, Virgílio Augusto Fernandes de Almeida, and thankful to him for sharing his brightest thoughts and research insights and representing so well our country.

I would like to specially thank my overseas guest advisor, Ponnurangam Kumaraguru (PK), for his enthusiasm, support and hard work.

I am also thankful to my comrades at III-T, Delhi: Denzil Correa, Paridhi Jain and Aditi Gupta. It was just great spending time with you.

To my lab mates at the Center of Analysis and Modeling of System Performance (CAMPS) thank you so much for making it such a great place to study. In particular, I would like to thank Emanuel Vianna, Giovanni Comarela, Tiago Rodrigues de Magalhães, Tatiana Pontes, Marisa Vaz, Gabriel Magno, Saulo Ricci, César Fernandes, Matheus Caldas and Fabrício Benevenuto, you made it all much more fun. I could not forget to thank our greatest foreign PhD visitor, Diego Saez Trumper, *a la salud*!

I am thankful to have met again at the university during the course of this dissertation my undergrad friends Cristiano Arbex Valle, Pedro de Carvalho Gomes, Carla Bechelane, André Cavatoni and Yolanda Vieira.

Last but not least, I would like to thank all my high school friends, specially Mateusão, who made my schedule during the course of my master's degree so much more complicated.

The present study was supported by the Indo-Brazil Science Council, CNPq and CAPES.

"You must be the change you want to see in the world."
(Mahatma Gandhi)

Resumo

Redes Sociais como Facebook, Twitter e LinkedIn experimentaram um crescimento exponencial nos últimos anos. Usuários gastam mais tempo em sítios de rede social do que em qualquer outro tipo de sítio ou serviço na Internet. Os usuários dessas redes publicam e compartilham uma grande quantidade de informações pessoais sem estar cientes muitas vezes das suas implicações na vida privada. As informações pessoais publicadas nesses sítios podem se tornar uma mina de ouro para empresas de marketing e também para criminosos virtuais. A caracterização da percepção de privacidade dos usuários é importante para se definir soluções tecnológicas e jurídicas. Os usuários de uma rede social esperam que a mesma forneça uma boa proteção de privacidade ou que forneça os mecanismos apropriados que lhes permitam tomar decisões sobre o controle de privacidade dos seus dados. Esta dissertação investiga a percepção de privacidade de usuários do Facebook, a maior rede social da atualidade. O presente estudo é um dos primeiros a caracterizar a percepção de privacidade em uma rede social através de um experimento no mundo real e não através de entrevistas. As principais descobertas são: apenas uma pequena parcela de usuários troca as configurações padrão de privacidade; a maior parte dos usuários exibe publicamente o gênero e a lista de amigos; a maioria dos usuários que fornece informações sobre localização as exibe publicamente; os usuários exercem maior controle sobre conteúdos potencialmente mais perigosos à reputação; as pessoas marcadas em fotos por um indivíduo formam redes egocêntricas de alto aglutinamento; uma importante parcela dos usuários exibe a data de nascimento à sua rede de contatos; a agregação de dados isolados pode vir a revelar informações outrora privadas.

Palavras-chave: Privacidade, Redes Sociais, Experimentos no Mundo Real.

Abstract

Online social networks such as Facebook, Twitter and LinkedIn have experienced exponential growth in recent years. Users are spending more time on Online Social Networking (OSN) sites than on any other sites and services on the Internet. Users post and share a lot of personal information on these sites without being aware of their privacy implications. Personal information posted on these OSNs can be a treasure for marketing companies and cyber criminals. Characterizing the privacy awareness of users is important to design technologies and policy solutions. Users expect the OSN site to provide good privacy protection or provide controls so they can make informed decisions about their privacy. This dissertation investigates the privacy awareness of users on Facebook, the largest OSN. The present study is one of the first to characterize the privacy awareness on OSN through a real world experiment, not self-reported data. The main findings are: only a low percentage of users change the default privacy settings; most users expose their gender and friends list publicly; most users who have commended their location information to Facebook expose it publicly; users exercise more control over content with more potential to endanger their reputation; people tagged by an individual form strong-tie egocentric networks; an important share of users expose their full date of birth to their network; the aggregation of individual bits can reveal once private information.

Keywords: Privacy, Online social networks, Real-world experiments.

List of Figures

List of Tables

Contents

Chapter 1

Introduction

Popular destinations on the Web such as search engines, news media, social networking, video or photo sharing and online games attract hundreds of millions of users every day, where they interact with different kinds of services. On one hand, these interactions yield valuable data that can be used to personalize the user's web experience. On the other hand, these interactions always leave data crumbs that can be used to breach user's privacy. Also, these destinations can share sensitive or personal user information with other users or third parties without proper user consent. Search engines can consciously or inadvertently also build user profiles, store user IP addresses, or collect any other information that could ever tie a particular search to a specific user [Krishnamurthy and Wills, 2008].

Social networking sites offer attractive means of online social interactions and communications, but also raise privacy and security concerns. Facebook is the number one network in the world, except for a few countries, like Japan and China [Brentcsutoras, 2010]. Web surfers now spend more time socializing on Facebook than searching with Google [New York Post, 2010]. Facebook has more than 800 million active users at any given point in time and 30 billion pieces of content (hyperlinks, notes, photos, etc.) are shared by its users each month. Facebook supports more than 70 languages, what makes it a huge global digital space. Like the Web itself, Facebook is a powerful technology to increase connection between people separated by borders of nation, language, religion and culture. With an estimated 65 billion friendships, it is important to study how this crucial technology is perceived across different cultures and understand user's privacy awareness [Facebook Statistics, 2010; Huffington Post, 2010]. Facebook has also been revising its privacy policy and settings from the day of inception, what directly affects a large population in the world. The focus of this research is to study Facebook users' privacy awareness / carelessness around the globe, and in particular,

in Brazil and India.

To the best knowledge of the author, this is the first study to analyze and compare the privacy awareness of Facebook users in Brazil and India through a real world experiment. It has used real-world data (not self reported) for studying privacy preferences. Understanding users' behavior in real world settings is critical to develop any technological or policy solutions [Kumaraguru et al., 2008]. The findings from this dissertation can be useful for other Online Social Networks and not just Facebook.

The main contributions of this dissertation are:

- Investigate privacy awareness of Facebook users using real world data.
- Show that the majority of the users are oblivious to privacy and reveal a lot of personal information on Facebook.
- Show that users from two different countries have different perceptions about the desired levels of privacy.
- Compare commended information to Facebook with what is made publicly available.

This dissertation is organized as follows. Chapter 2 introduces a notion about privacy and the idea of the right to privacy. It also portrays a great number of studies that were conducted about privacy related to the usage of Facebook and other Online Social Networks and many other related works. The growth of Online Social Networking experienced in Brazil and India is also a topic of discussion. Chapter 3 focuses on the study methodology, the application which was developed and the data sets that were acquired. Chapter 4 presents some hypotheses about how users perceive or ignore privacy on Facebook. Chapter 5 reports the results that support the hypotheses made and a couple further analyses. Chapter 6 discusses the results found, presents some conclusions and future work directions.

Chapter 2

Background

In this chapter is presented a brief background on various studies (in particular, privacy) that have been done on Facebook. It also describes some results from studies which have analyzed cultural aspects of privacy to provide a background on the comparison that is aftwerwards made between users from Brazil and India.

2.1 Privacy Concept and the Right to Privacy

The concept of privacy is not something new. A panoply of definitions can be found and consent might be hard to achieve although everyone has a certain understanding of its concept.

The New Oxford American Dictionary defines privacy as:

> 1) the state or condition of being free from being observed or disturbed by other people; 2) the state of being free from public attention - a law to restrict newspaper's freedom to invade people's privacy. [Stevenson and Lindberg, 2010]

One of the first signs of the privacy principle on contemporary law appears on the fourth amendment of the United States Constitution, which states:

> The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no warrants shall issue, but upon probable cause, supported by oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized. [Madison, 1791]

Warren and Brandeis [1890] developed "the right to privacy" concept based upon the right "to be let alone", an expression coined a couple years earlier by Judge Thomas Cooley. They were responsible for important developments of people's privacy law and their Harvard Law Review article still seems fresh.

Nowadays, Solove [2010] argues that the fourth amendment should be abandoned as a test of privacy invasion in exchange for a more pragmatic approach to face the issue of government information gathering.

For a broad review of international privacy laws, the interested reader is directed to Ishitani et al. [2003].

2.2 Privacy on Facebook

Due to its immense popularity (over 800 million active users at any given point in time), various research studies have been conducted on Facebook. Researchers have: analyzed the social network of Facebook users to find different patterns [Lewis et al., 2008]; analyzed the impact of Facebook applications and games [Wei et al., 2010]; used it to study statistical sampling of participants on the Internet to generalize the result from the analysis [Gjoka et al., 2009].

One key topic that has been studied with respect to Facebook is privacy. Thomas et al. [2010] examined how the lack of joint privacy controls can inadvertently reveal sensitive information about a user. Besmer and Lipford [2010] identified the social tensions that photo tagging generates and the needs of privacy tools to address the social implications of photo privacy.

Liu et al. [2011] revealed the disparity between the desired and actual privacy settings of users on Facebook. They found that 36% of content remains shared with the default privacy settings. They also found that privacy settings matched users' expectations only 37% of the time and, when incorrect, they almost always tend to expose content to more users than expected. Nevertheless, they noted that photos have the most privacy-conscious setting among content shared. Some possible reasons pointed by the authors for such gap between the desired and actual privacy settings include: poor human-computer interaction mechanisms, the static nature of privacy settings and the significant amount of work forced on the user to maintain the privacy of their content.

Also, in general, privacy has been an important topic of study on Online Social Networks [Gross and Acquisti, 2005; Zhou and Pei, 2008; Beye et al., 2010].

There have been many instances where users have consciously or inadvertently

shared personal information on Facebook that has later become an embarassement for the users involved [Mail Online, 2009]. It has also been found that government and council employees in the U.K. are using social networking sites from office, where they are also exposing private or classified information [Yahoo! News India, 2010]. For a profound account of gossips, slanders and a series of rumors and disputes on the Internet over personal reputation the interested reader is referred to the work of Solove [2004, 2007].

Privacy settings on Facebook have been on scrutiny for some time and various factors related to privacy settings on Facebook have been studied [The Guardian, 2010]. Privacy settings of Facebook have evolved over the time [The Economist, 2010]. Boyd and Hargittai [2010] showed that both frequency and type of Facebook users as well as Internet skill are correlated with making proper modifications to privacy settings. They have also observed a few gender differences in how young adults approach their privacy configurations, which is notable, given that gender differences exist in so many other online domains.

Facebook has been used to study the Personally Identifiable Information leakages online. Krishnamurthy and Wills [2008] analyzed Facebook for profile and friends to be viewed by others and found varying levels of public exposure ranging from 76% to 99% of the users among different regional networks worldwide. Gjoka et al. [2009], while analyzing Facebook for unbiased sampling, showed that the majority of users (84%) did not change their default privacy settings and only 7% of global users hid their friends from strangers.

Many factors seem to influence the privacy awareness of users – geographical location, ethnicity, node degree and even the privacy awareness of friends. Gjoka et al. [2009] found that users around the globe were split between two extremes of privacy settings, which is inline with literature (Individualist and Collectivist society) [Hofstede et al., 2010]. Chang et al. [2010], using the Facebook data from the U.S. users, showed that ethnicity of users impacted their privacy preferences. For instance, Hispanic users share more photos than the average U.S. citizen user. Gjoka et al. [2009] showed that users with low degree nodes tend to have stringent privacy settings while users with high degree nodes tend to be liberal in their privacy settings. This is counterintutive, as one would imagine that users with high degree nodes would be more aware of privacy settings and therefore would have changed it to being stringent. They also showed a positive correlation between one's privacy awareness and their friends' privacy awareness. The power of one's connections to influence every aspect of social behaviour has been the subject of many recent studies [Barabási, 2003; Christakis and Fowler, 2010].

2.3 Privacy Studies in Brazil and India

Very little research work has been done in studying privacy perception or awareness in Brazil and India. Studying privacy awareness of users in these countries will help in decision making of technologies and policies for the use of the Internet. Countries like Brazil and India are expected to play a central role in the world of 21^{st} century.[1]

A large amount of research is conducted in the U.S. [Kumaraguru and Cranor, 2005b] and Europe on various aspects of privacy. Due to cultural background, there is a large difference in privacy perceptions among different parts of the world [Bellman et al., 2004]. Hofstede et al. [2010] has classified societies around the world into many categories and the two extremes are individualist and collectivist. According to Hofstede both Brazil and India are collectivist societies. People in Brazil and India are unaware of various privacy issues both in the online and offline worlds [Diller et al., 2003; Kumaraguru and Cranor, 2005a; Kumaraguru et al., 2005].

2.4 Facebook Photos

The ability to upload and share photo albums on Facebook was launched on October 2005, when the OSN accounted about 5 million users [Facebook Timeline, 2010]. By then, photo hosting was already exploding on the Internet and other sites which offered photo hosting services were already quite popular, like MySpace and Flickr; the latter was by that time in the hands of Yahoo. Nonetheless, the simplistic interface design and the leverage of social features allowed the service to become the most popular feature of Facebook and the number one online photo service by late 2009, with more than 30 billion photos [Kirkpatrick, 2010].

Since its start, the service allowed users to tag their friends and comment on the photos. Friends received in return e-mail alerts when they were tagged, driving lots of traffic to the website. With one month of its launch, 85% of the subscribed users were tagged at least once [Kirkpatrick, 2010]. It is not hard to imagine the pervasive privacy consequences made possible by the service advent.

Nowadays, people upload more than 3 billion photos each month and add more than 100 million tags to photos on Facebook every day [Inside Facebook, 2010; Facebook Blog, 2011]. Facebook also reached a new record with 750 million photos uploaded over the first weekend of 2011 [TechCrunch, 2011].

[1] The combined BRIC (Brazil, Russia, India and China) economies by 2050 is expected to be more than the combined economies of the richest countries in the world [Wilson and Purushothaman, 2003].

2.5 OSN Growth in Brazil and India

In Brazil, traffic to social networking sites grew 51% in 2009, reaching more than 36 million visitors aged 15 and older in August 2010. Facebook experienced triple-digit growth, increasing its audience 479% in 2009 [ComScore, 2010b]. From December 2010 to December 2011, Facebook experienced once more triple-digit growth in Brazil, increasing its audience by 192% and reclaiming the top position of social networking from Orkut. Windows Live Profile ranked third, followed closely by Twitter. More surprisingly, the average time spent per visitor during the month of December on Facebook in Brazil grew from 37.2 minutes in 2010 to 285.2 minutes in 2011, a change of 667% [ComScore, 2012].

LinkedIn also experienced triple-digit growth of 428% in Brazil year-over-year in March 2010, while its audience in India grew 76% during the same period [LinkedIn Blog, 2011].

In India, more than 33 million Internet users aged 15 and older visited social networking sites in July 2010, representing 84% of its total Internet audience. India ranked then as the seventh largest market worldwide for social networking, after the U.S., China, Germany, Russia, Brazil and the U.K. The total Indian social networking audience grew 43% year-over-year, more than tripling the rate of growth of the total Internet audience. Facebook already had the top spot among social networking sites, with 20.9 million visitors. Orkut ranked second with 19.9 million visitors (up 16% from the year before), followed by BharatStudent.com with 4.4 million visitors [ComScore, 2010a].

Facebook achieved an astonishing growth on its active user base in the past couple years in Brazil and India. By keeping track of the data provided through Facebook Ads [2012], it was seen that Brazil registered a growth of 860% on its reported active users base, while India registered a growth of 655%, both during the period comprehended between October 2010 and September 2012, as depicted in Figure 2.1. During the same period, Facebook also jumped from the 15^{th} to the 1^{st} spot in Brazil top sites by audience provided by Alexa [2012], and also consolidated the 3^{rd} spot in India.

As of September 2012, Brazil and India represented the 2^{nd} and 3^{rd} largest geographical communities on Facebook, respectively, as shown in Table 2.2. This scenario tends to continue or improve influenced by the economical development of these countries and the relatively low penetration rates they present. In Brazil, the penetration rate considering the whole population is of 29.48%, while in India it only attains 4.43% of the inhabitants. Both are quite far from the highest occurrent rates, summarized in Table 2.1. For instance, Facebook presents a penetration rate of 68.14% in Iceland, the

highest among countries with more than 100,000 inhabitants. All population figures used were the most current available and were extracted from Wikipedia [2011].

For a global panorama of Facebook penetration rates it was drawn a world map in Figure 2.2. It is noticeable how low is the presence of Facebook on Central Africa and also on other countries ruled by authoritative regimes.

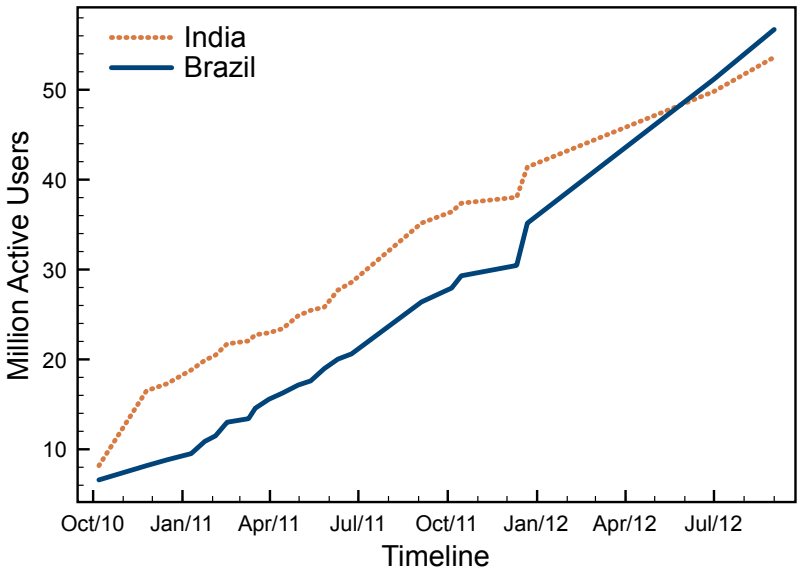

Figure 2.1. Reported growth of Facebook active users in Brazil and India [Facebook Ads, 2012].

Table 2.1. Countries populated by more than 100,000 people with the highest penetration rates on Facebook. As of September, 2012.

Rank	Country	# of Users	% of the Population
1	Iceland	216,980	68.14
2	United Kingdom	40,084,660	64.38
3	Chile	9,639,260	57.08
4	Hong Kong	3,939,960	55.85
5	Taiwan	12,792,580	55.20
6	Singapore	2,800,400	54.02
7	Norway	2,696,920	54.01
8	United States	163,442,860	52.24
9	Denmark	2,915,740	51.41
10	Canada	17,782,180	51.23

Table 2.2. Countries with most users on Facebook and the corresponding percentage of the population. As of September, 2012.

Rank	**Country**	**# of Users**	**% of the Population**
1	United States	163,442,860	52.24
2	Brazil	56,704,840	29.48
3	India	53,622,460	4.43
4	United Kingdom	40,084,660	64.38
5	Indonesia	39,769,280	16.75
6	Mexico	37,560,560	33.44
7	Turkey	31,673,840	42.96
8	Philippines	29,273,380	31.14
9	France	24,631,920	37.42
10	Germany	24,215,200	29.60
11	Italy	21,938,480	36.15
12	Argentina	19,811,420	49.38
13	Canada	17,782,180	51.23
14	Colombia	17,174,540	37.40
15	Thailand	16,411,860	24.60
16	Spain	16,090,160	34.96
17	Japan	14,006,740	10.95
18	Malaysia	12,822,900	45.26
19	Taiwan	12,792,580	55.20
20	Egypt	11,592,340	15.10

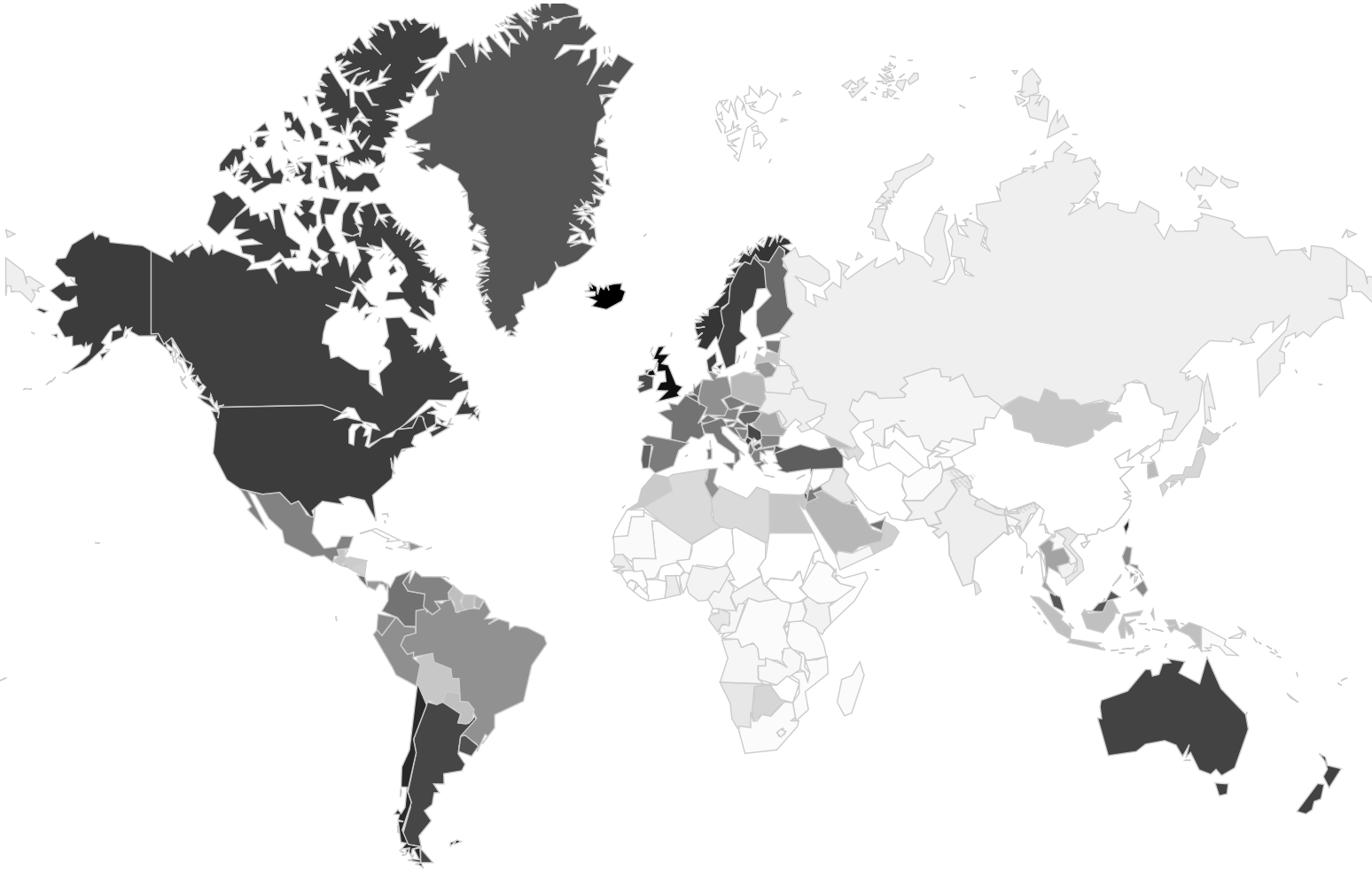

Figure 2.2. Facebook worldwide presence. The darker the region, the higher is the population penetration rate. As of September, 2012

Chapter 3

Methodology

In this chapter is presented how participant recruitment was promoted for the Facebook application developed for this particular study. The application itself is also depicted. The data collected is described in detail along with the demographics of participants and their friends.

3.1 Recruitment

To recruit participants to the study, emails and Facebook messages were sent, posters like the one in Figure 3.4 were affixed on university notice boards and campuses around the globe (Figure 3.1), much word-of-mouth was made and also a couple of media posts were published. Most of the campaigning about the study was done in Brazil and India. A domain was registered `http://www.theprivacystudy.org/` to host all information (link to the Facebook application, details about prizes, etc.) about the study. The website allowed participants to spread the word about the Facebook application through several online channels such as *Twitter* and *Google Buzz*. In the final count of the data that is hereby analyzed, the application received 416 likes on Facebook out of 605 participant users; 75.4% of the participants had at least one friend who also joined the study. Prizes were offered through raffle to participants, comprising one high-end mp3 player of 32 GBs and five games for PC/Mac. It was also noted "Get your friends to install the application and increase your chances to win a prize!" on the website to help attract more participation in the study.

The recruitment of participants and the spread of the application adoption is acknowledged as the most difficult part of the present study. Some other possible approaches that might be explored in future studies is to recruit participants through Mechanical Turk [Amazon, 2005] or to develop more engaging applications like games.

3.2 Study Setup

Six hundred and five participants installed the application in their Facebook account. When participants installed the *Privacy Study* application[1], they were presented with the study privacy policy which explained what kind of data would be collected from them and for what purpose. The application worked in conformity with the Terms of Service of Facebook and ethical ways of studying social media [Fisher et al., 2010]. Participants were then invited to authorize the application access to their Facebook data and some pieces of information from their friends as well, as depicted in Figure 3.2. The following data items were collected from participants who installed the application:

- User and friends basic information (sex, birthdate, timezone, current city, hometown)
- User content privacy settings
- User content meta information (eg: album size, creation date, type, tags)

[1] `http://www.facebook.com/apps/application.php?id=144781782200917`

Figure 3.1. Poster affixed at a university bus stop in Stockholm, Sweden.

After installing the application, users were presented with a breakdown of the visibility of the content (photo albums, videos and links) that they have shared on Facebook (as shown in Figure 3.3). A pie chart shows the percentage of what is visible to *Everyone*, *Friends and Networks*, *Friends of Friends*, *Friends Only*, what has *Custom* visibility and what is available to *Self* only. Users can see their personal breakdown for photo albums, videos and links by clicking on the radio buttons placed beneath the pie chart. The application also displays the number of friends who have currently installed the application and the quantity of coupons for the prizes draw earned so far.

Figure 3.2. Permission dialog box presented to participants while installing the *Privacy Study* application. It is requesting the participant for accessing the account information. Blurred the user ID in the figure.

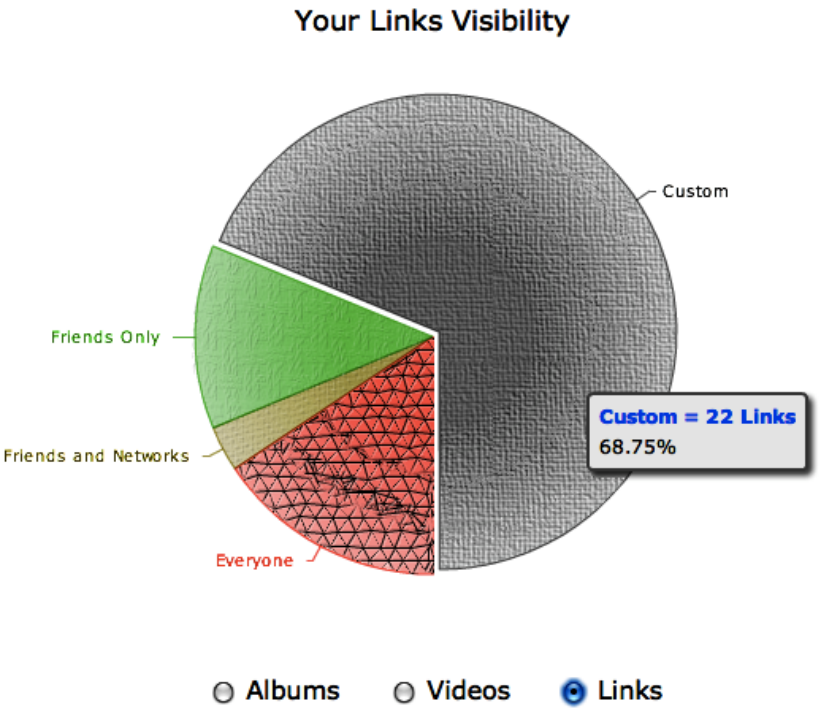

Figure 3.3. The *Privacy Study* application overview. It presents a breakdown of the visibility of the contents (photo albums, videos and links) that the participant has shared on Facebook. The application also portrays the number of friends who have installed the application and the quantity of coupons for the prizes draw earned (cropped from image, as they are normally presented beneath the chart).

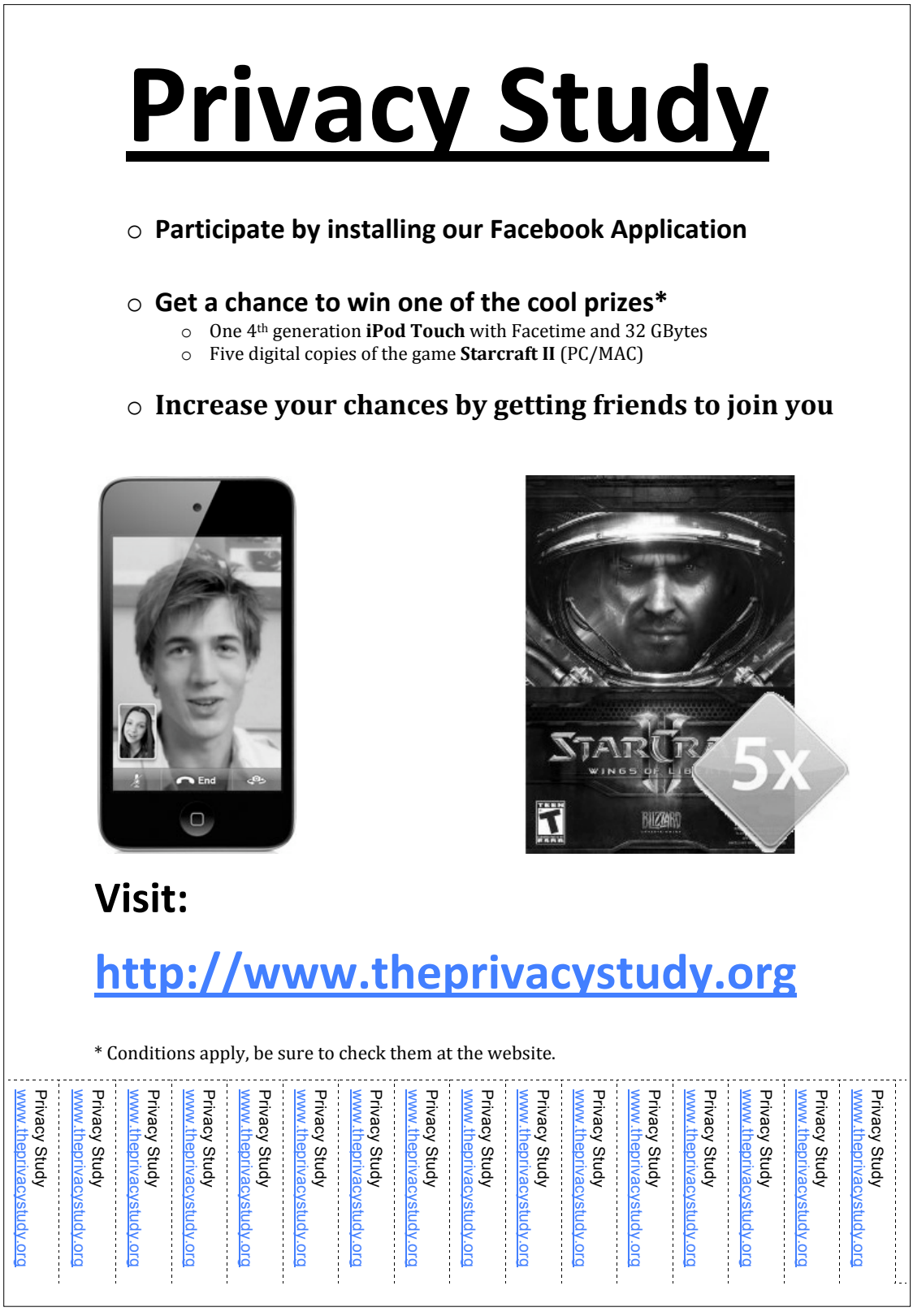

Figure 3.4. Sample of the study poster that was affixed on university boards.

3.3 Data Set and Demographics

Two sets of data were collected through the Facebook API, one about the participants who installed the application and another one about the friends of participants. Data set P represents the users who installed the study application and is summarized in Table 3.1. Both in Brazil and India male participants were more prevalent than female participants, which is opposite to the data that Boyd and Hargittai [2010] used in their study. According to Facebook Ads [2012], there are far more male users (72.5%) in India compared to female users (27.5%), whereas there are more female users (54.5%) in Brazil than male (44.5%). An important percentage of participants (53%) in the study belonged to the age group 18 - 25, which is likewise the most active age group on Facebook. Young adults in this specific age range are the most prevalent users of most Online Social Networks on the Internet [Social Media Optimization, 2008]. Data set PF represents the participants along with their friends and its details are given in Table 3.2.

A third data set PP was collected by distributedly crawling the public profile and the public friends list of participants and their friends. A total of 89825 public profiles were obtained this way, which represents 92% of the users present in PF data set. The missing profiles represent a share of no longer active users. Despite the fact that these users were no longer active on Facebook, their data remained available to applications through the Facebook API, their names notably. This is not the first time that Facebook is pointed to keep removed data for unknown reasons and for unlimited time [The Washington Times, 2011]. It was also acquired 30 million friendship connections through the crawling process.

Table 3.1. Demographics of the study participants (data set P). The data set comprehends participants from 21 countries. Data was not available for some users, presented as N/A in the table.

	Total N = 605	**Brazil** N = 301	**India** N = 225	**Others** N = 79
Gender (%)				
Female	25.45	29.57	19.11	27.85
Male	65.45	61.46	71.11	64.56
N / A	9.09	8.97	9.78	7.59
Age (%)				
Under 18	0.66	0.33	1.33	0.0
18 - 25	52.56	43.19	69.78	39.24
26 - 35	23.97	30.23	12.89	31.65
36 - 45	5.29	5.98	3.56	7.59
46 - 55	1.49	2.33	0.89	0.0
Over 55	0.5	0.33	0.89	0.0
N / A	16.03	17.94	11.56	21.52
# of friends				
Average	205.08	149.7	257.88	265.58
Median	159	118	207	235
User content				
#Albums	3168	1173	1231	764
#Photos	67209	19240	20184	27785
#Links	17020	5303	7717	4000
#Videos	275	59	111	105

Table 3.2. Demographics of the study participants and their friends (data set *PF*). The data set comprehends users from 127 countries. 63185 users did not reveal their country and are accounted in *Total* column only. Data was not available for some users, presented as N/A in the table.

	Total N=97687	**Brazil** N=8376	**India** N=15896	**Others** N=10230
Gender (%)				
Female	32.07	39.68	22.74	30.76
Male	55.38	51.04	69.38	53.02
N/A	12.54	9.28	7.89	16.22
Age (%)				
Under 18	2.09	1.06	2.83	1.51
18 - 25	37.01	27.48	50.25	26.0
26 - 35	18.07	26.34	6.5	16.96
36 - 45	4.12	5.89	1.14	4.0
46 - 55	1.93	2.36	0.74	1.47
Over 55	1.18	1.23	0.23	1.03
N/A	36.79	36.87	38.54	50.07

Chapter 4

Hypotheses

In this chapter is presented the study hypotheses.

4.1 H1: Default Privacy Settings

Privacy settings is a means by which users set their preferences about how their profile or other information should be handled by the organization who is collecting the information (e.g. Facebook). Default settings are supposed to capture the most acceptable preferences so that users do not have to keep changing the default settings. It has been shown that 84% of the Facebook users did not change their default privacy settings [Gjoka et al., 2009]. It has been also shown that presenting information that is easily accessible to users can significantly aid the user to change their privacy settings on Facebook [Lipford et al., 2008]. Acquisti et al. found that participants in their study had misconceptions about privacy on Facebook [Acquisti and Gross, 2006; Gross and Acquisti, 2005].

Hypothesis 1: *The privacy settings of users are significantly different from the default privacy settings on Facebook.*

4.2 H2: Gender Exposure

From an individual standpoint, gender represents the least perilous personal information considered in the present study. Nonetheless, it represents an important piece for total information systems being built by marketeers, government agencies and criminal organizations. Gender is one piece of information (along with birth date and zip code)

which can be used to identify a large percentage of U.S. citizens uniquely [Sweeney, 2002]. Gender can be fairly estimated using first name databases[1] and can even be predicted by image-based classifiers [Gallagher, 2008]. It has a specific privacy setting, which is governed by a checkbox entitled "Show my sex in my profile" on the profile edit page. As one can guess, the default is set to visible. Users cannot control who can see the information, being it public available when disclosed.

Hypothesis 2: *A high percentage of Facebook users conceal their gender information.*

4.3 H3: Location Exposure

Location-aware systems are rapidly becoming paramount on modern societies. From communication to recommendation systems, from mobile phones to GPS-enabled devices, individuals' location is no longer a secret these days. Famished for the latest technology novelty, people forget to reason about the underlying privacy consequences. Large amount of research is being done in the space of *location based privacy systems* [Toch et al., comp; Tsai et al., 2009]. For Facebook, the story could not be any different. The default privacy setting for the current city and hometown attributes is configured so everyone in the social network can access them when disclosed.

Hypothesis 3: *A large percentage of users hide their current location on Facebook.*

4.4 H4: Content Exposure

No day goes by without media coverage about someone having her reputation disputed on the Internet. The video shooted or the picture taken of someone's private life might become public and make the headline or reach an unattended audience one day. Watching this recurrent situation serves as an alarm for people to try to control what they expose and share online and specially with whom. In a decreasing scale of reputation endangerment, certain contents can be shared on Facebook: videos, photo albums and links. Facebook allows users to fine tune their privacy settings for every piece of content shared over its network. Nevertheless, to achieve broader audience and network growth, in despite of user reputation preservation, the default visibility

[1] `http://www.socialsecurity.gov/OACT/babynames/`

is configured so *everyone* in the social network can access these contents. Even so, users are given the option to customize the visibility of their content with the following options: *Everyone*, *Friends and Networks*, *Friends of Friends*, *Friends Only*, *Custom* or *Self*.

Hypothesis 4: *User privacy control over distinct content types is not significantly different.*

4.5 H5: Friends List Exposure

How many friends publicly disclosing the city they live in does it take to infer yours? How many others would it take to infer where you have studied? How many openly gay friends must you have on a social network before you are outed by implication? [Grimmelmann, 2009]. These and other questions can be attempted to be answered by the disclosure of your friends list and some other pieces of information by your friends. As a personal network expands, privacy protection goes beyond personal settings and become a social networking problem [Gundecha et al., 2011]. Facebook allows access to the friends list to everyone who has joined the OSN by default, although one can customize its visibility at will.

Hypothesis 5: *Most users conceal their friends list from the public.*

4.6 H6: Date of Birth Exposure

Date of birth is one of the crucial information among others that criminals look for in order to perform identity theft or impersonation. Researchers have shown that the date and place of birth combined can be exploited to predict Social Security Numbers (SSN) of U.S. citizens [Gross and Acquisti, 2005]. Date of birth is also often used for marketing purposes. In Facebook, there is a specific privacy setting from where users can choose: "Show my full birthday in my profile"., "Show only month & day in my profile". or "Don't show my birthday in my profile". The display of the complete date of birth is the default setting. Users also have a separate control to select the visibility of the information, where the default is set to *Friends of Friends*.

Hypothesis 6: *An important share of Facebook users hide their full date of birth (year, month and day) from their network.*

Chapter 5

Results

The results presented in this chapter refute Hypotheses from 1 through 6. Section 5.7 presents further analyses done with the collected data.

5.1 H1: Default Privacy Settings

For this analysis, the privacy settings S_u of each user u of PF data set were converted to a 3-bit word, where $S_u = 111$ represents public disclosure of the studied attributes, according to the encoding presented in Table 5.1. This approach gives an overall idea of how users changed the privacy settings and it has been used in the past [Gjoka et al., 2009].

Table 5.1. Encoding for basic privacy settings S_u of a user u. Bit 1 denotes the leftmost bit in the representation.

Bit	**Attribute**	**Description**
1	Date of Birth	=1 if full date of birth is visible
2	Location	=1 if current location is informed
3	Gender	=1 if sex is revealed

Facebook has two possible default settings: $S_u = 101$, obtained after filling out the initial registration form (i.e. date of birth and gender visible to everyone), and $S_u = 111$, if the user informs his current location afterwards. As depicted in Figure 5.1, the majority (58.1%) of users keep the two aforementioned default settings ($S_u = 1*1$), where 39% of users remain with the initial registration settings ($S_u = 101$). Only 4.1%

of all users appear to be really concerned of exposing their information and do take the time to conceal all the three pieces under analysis ($S_u = 000$).

In Brazil, the scenario exhibits 67.1% of users with one of the two possible default settings. In India, 22.1% of users disclose all the three attributes, while the equivalent statistic in Brazil is somewhat lower, 14.3%. The precedent difference is statiscally significant (Proportion Test, p-value < 0.05). This contradicts Hypothesis 1 and reinforces the fact that simply providing a set of customization features does not ensure that users will take advantage of them [Mackay, 1991].

Taking the gender influence into consideration, the encoding possibilities are reduced by half, as the third bit becomes necessarily 1. Among the remaining configurations it is considered the two extremes: $S_u = 001$, which stands for total concealment, and $S_u = 111$, which stands for total revelation. For all cases, women are significantly more conservative than men, as shown in Table 5.2. For instance, 24.1% of all the women do not disclose their full date of birth neither their current location ($S_u = 001$). The same statistic reaches a lower portion of the men, 17.2% overall.

These results help to refute Hypothesis 1.

Table 5.2. Reach across different genders for both extreme privacy settings S_u of users in the PF data set. All figures with 95% confidence level.

	S_u	**Female** Avg(%) ± Std. Error	**Male** Avg(%) ± Std. Error
Brazil	111	13.89 ± 0.53	18.14 ± 0.55
	001	23.78 ± 0.65	15.19 ± 0.51
India	111	19.86 ± 0.76	26.54 ± 0.52
	001	25.30 ± 0.82	17.53 ± 0.45
All	111	18.15 ± 0.40	23.99 ± 0.34
	001	24.09 ± 0.44	17.19 ± 0.30

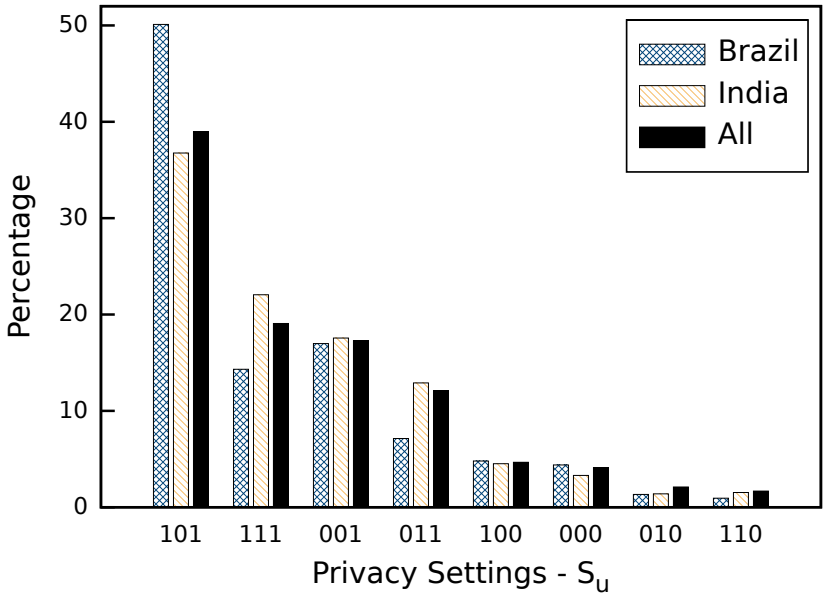

Figure 5.1. Distribution of the privacy settings S_u of users in the PF data set. The encoding hereby presented is in accordance with Table 5.1. Only 4% of the users conceal all the three pieces of information under analysis ($S_u = 000$).

5.2 H2: Gender Exposure

Using the PF data set, it was found that 87.5% of users reveal their gender information to their network, the highest exposure level among factors studied in this research. Figure 5.2 presents the gender breakdown and exposure of participants and their friends. Concealment or non-availability of gender information is low in all regions considered ($< 16\%$).

It was also found that 75.83% of the users in the PF data set reveal their sex publicly on their profiles. Moreover, 86.39% of the users who are known to have confided their sex information to Facebook reveal it on their profiles. Such high levels of exposure obtained reinforce the assumption that users are diving in a sea of obliviousness towards privacy. Moreover, the information is publicly available on the profile page when disclosed, what corroborates the importance of default settings. Research has shown that most people rarely change them [Boyd and Hargittai, 2010]. Public disclosure rates are summarized in Table 5.3.

On every considered region women tended to reveal their gender on their profiles more often than men. All the differences between genders were statiscally significant (Proportion Test, p-value < 0.05).

Gundecha et al. [2011] report a public gender exposure rate of 81.77%, out of more than 2 million profiles collected, in accordance with the results hereby presented.

Nevertheless, it was shown that the rate can be even higher if only users who confide the information are considered.

These results refute Hypothesis 2.

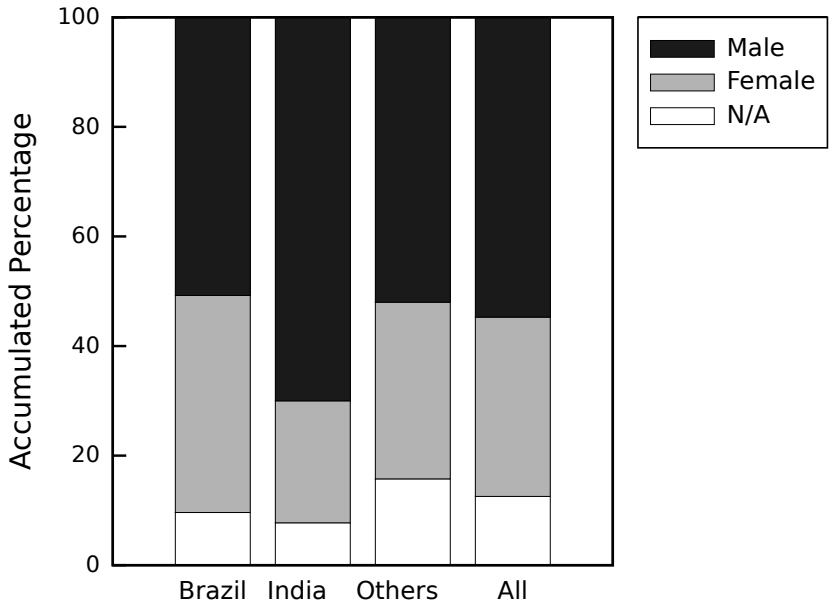

Figure 5.2. Gender breakdown and exposure of participants and their friends. About 87% of users expose their gender to their network. Used PF data set for the analysis. Data was not available for some users, presented as N/A.

Table 5.3. Public disclosure rates of gender information for those in the PF data set. Particular gender rates are discrepant because they acknowledge that users have informed their sex. Figures with 95% confidence level.

	Gender Public Disclosure Rate Avg(%) ± Std. Error			
	All	Informed Gender	Female	Male
Brazil	76.53 ± 0.85	84.59 ± 0.76	86.70 ± 1.08	82.94 ± 1.05
India	79.84 ± 0.61	86.18 ± 0.55	90.27 ± 0.96	84.88 ± 0.65
All	75.83 ± 0.25	86.39 ± 0.22	88.61 ± 0.33	85.06 ± 0.29

5.3 H3: Location Exposure

Location is an important feature for applications built on top of OSNs. It is used to locate old friends, provide customized recommendations such as concerts and restau-

rants, provide metropolitan transport routes and so forth and so on. There are many benefits from the users' standpoint which can be brought to them by the correct use of their location. Nonetheless, the misuse of such information could lead to many other less fortunate events, such as the robbery of his house when on vacation or the inconvenient of being approached online by close sexual predators.

To derive the regional statistics presented in this section it was considered that participants and their friends belonged to the same region. Otherwise, it would not be possible to estimate regional exposure levels, but only the overall scenario. To help justify the fairness of the aforementioned approach it was derived the percentage of friends who belonged to the same region of the participants, from those who happen to reveal that information. It was found that 78.2% of the friends who reveal their location share the same country of participants from Brazil and the equivalent statistic reaches an even higher bar for India, 87.8%. When the region considered is restricted to the city level, these rates drop to 49.5% of the friendships in Brazil and 42.3% in India. Previous research has shown that people tend to interact online more with those who share their same age, language and location [Leskovec and Horvitz, 2008]. Ugander et al. [2011] using the entire Facebook graph have also found that 84.2% of the friendship edges are within countries.

Using the PF data set collected by the application it was found that 32.7% of the users disclose their current city on Facebook overall. The statistic exposes a significant contrast between Brazil and India, reaching 22% of the users in the former country and 36.4% in the latter. One can conjecture that the difference stem from high crime rates in the main urban centers in Brazil. As a consequence, Brazilian user networks try to hide that piece of information from the scrutiny of strangers, to protect themselves. Also contrasting was the exposure levels across genders, where a trend towards less revelation by women was observed on every considered region (Figure 5.3). The difference found between genders was of 2.3% in Brazil, 7% in India, 2.5% elsewhere and 6.1% overall. All the differences were statiscally significant (Proportion Test, p-value < 0.05). It is believed that women are more concerned about disclosing personal information on online social networks as they are often targeted by sexual predators.

For those who are known to inform their current city it was found that 61.6% of them disclose it publicly on their profiles overall, as shown in Table 5.4. In Brazil, the same statistic reaches exactly the same bar of 61.6%, while in India it goes a little higher and attains 63.2%. These are extremely high rates for such a sensitive information and reveal how a great portion of users ignore the implications of its exhibition or even the existence of particular privacy controls that govern its visibility reach. Taking

the gender into consideration, once more women were less likely to reveal a piece of information about themselves on every considered region, except for Brazil.

Taking the same approach for the hometown information it was found higher public explosure rates for every considered subset when contrasted to the current city information, except for male in Brazil. This seems to be intuitive as the hometown information poses a less direct threat to users. Also, 52.44% of the users who reveal the current city or the hometown publicly, reveal them both.

Moreover, contrasting the public profiles collected with the data obtained through the application it was found that 8.8% of all the users from PF data set do not let their friends disclose their current city through third party applications by a particular opt-out mechanism, but do it so publicly on their profiles. This can be understood as an attempt to avoid automatic data collection or the annoyance by third party applications.

A corresponding study performed by Gundecha et al. [2011] found that 30.17% of the users reveal their current city publicly on Facebook and 35.38% do so for their hometown information, out of more than 2 million public profiles. These figures are close from those found for overall disclosure but hide the more sinister aspect of the public disclosure rates over those who are known to have commended the information to Facebook.

These results refute Hypothesis 3.

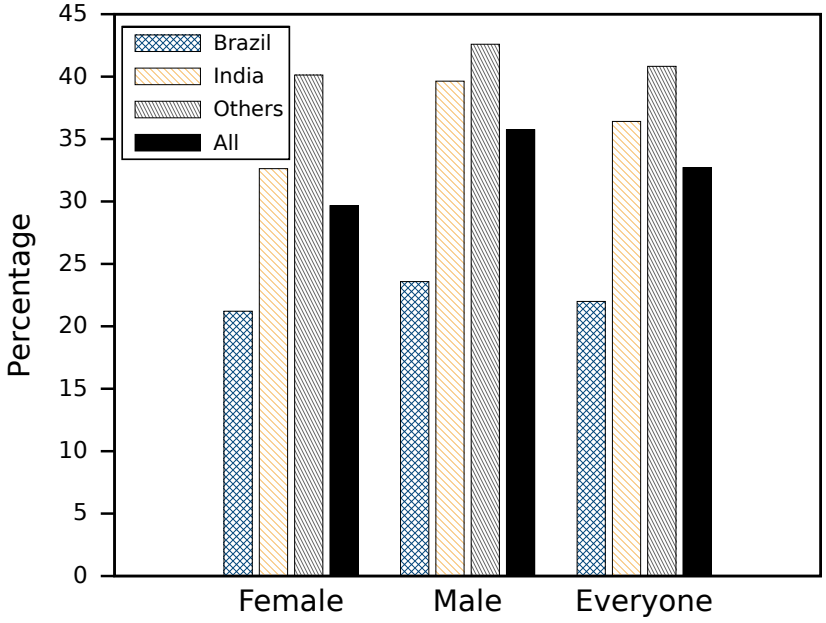

Figure 5.3. Distribution of current city disclosure from the PF data set. Women are less likely to reveal their location.

Table 5.4. Current city and hometown public disclosure rates for those in the *PF* data set that are known to have informed their current city / hometown on Facebook. Rates are high for all cases considered. Figures with 95% confidence level.

	Gender	**Current City Public Disclosure Rate** Average(%) ± Std. Error	**Hometown Public Disclosure Rate** Average(%) ± Std. Error
Brazil	All	61.64 ± 0.01	61.70 ± 0.10
	Female	62.91 ± 0.02	72.34 ± 0.15
	Male	60.80 ± 0.02	56.47 ± 0.14
India	All	63.21 ± 0.01	77.27 ± 0.09
	Female	59.72 ± 0.02	81.82 ± 0.25
	Male	64.49 ± 0.01	81.32 ± 0.09
Others	All	59.46 ± 0.01	69.49 ± 0.14
	Female	57.38 ± 0.02	69.23 ± 0.30
	Male	61.19 ± 0.02	73.17 ± 0.16
All	All	61.64 ± 0.01	70.03 ± 0.06
	Female	59.92 ± 0.01	74.29 ± 0.12
	Male	62.84 ± 0.01	71.15 ± 0.07

5.4 H4: Content Exposure

Facebook allows users to fine tune their privacy settings for every piece of content (e.g. photo albums, links, videos) shared over its network. In order to get this content to broader audiences and increase network value, the default visibility is configured to *Everyone*. Users are given the option to customize the visibility of the content with the following options: *Everyone*, *Friends and Networks*, *Friends of Friends*, *Friends Only*, *Custom* or *Self*. For the present analysis, user content privacy settings were obtained from the data set P . As summarized in Table 3.1, only a few hundred videos were uploaded and shared on Facebook by the study participants, demonstrating that the feature is not so popular as in other video specific platforms (e.g. YouTube or Vimeo). For this reason, videos are left out of the present analysis and focus is given to photo albums and links.

Users have four different kinds of photo albums on Facebook: *profile pictures*, *wall photos*, *mobile uploads* and *normal ones*. The first three kinds are unique per user, while the latter serves for the general purpose and can be created at will. Although the visibility of the *profile pictures* album can be configured in the same way as the others, the privacy data obtained through the Facebook API does not reflect that - being it always set to *Self*, and they are thus left out of the investigation.

The examination commences by taking an overall look on content visibility. For that, the entire content set was broken down by their current visibility setting, on a per content basis. For photo albums, the only regional statistically significant difference found was for the *Custom* setting (Proportion Test, p-value < 0.05), where in India it reaches 10% and in Brazil it stays around 5%. More important are the contrasts found between visibility levels, where the *Friends Only* setting appears as the first choice among all participants (Proportion Test, p-value < 0.05), showing a clear discontentment with the default exposure to *Everyone*.

As depicted in the upper half of Figure 5.4, the album visibility breakdown clearly diverges from its link counterpart. While the exposure of photo albums to *Everyone* reaches 35% overall, the equivalent statistic for links notches 56%. The situation is more accentuated in Brazil, where the gap attains 27% (42% for albums and 69% for links).

Another interesting way to look at the content exposure scenario is to analyze the visibility usage reach. For that examination a bucket is created for each possible setting and each user is placed once in every bucket for which he has a content shared with that particular setting. So, for instance, if user u_1 has two albums shared with *Everyone* and two other albums shared with his *Friends Only*, he is placed in buckets *Everyone*

and *Friends Only* a single time. The results for visibility usage reach are depicted in the lower half of Figure 5.4. The visibility for *Everyone* reaches 82.5% of all users for links and 55.8% for photo albums. Another clear display of preoccupation towards photo albums exposure is given by the *Friends Only* setting, where the usage reaches 55.5% for albums and 45% for links. The gap is even more prominent in Brazil and India, where it reaches 19.9% for the former and 29.6% for the latter. The precedent differences are all statiscally significant (Proportion Test, p-value < 0.05).

Liu et al. [2011] while conducting a survey of U.S. users on Facebook found that 22% of the photos albums in their sample were shared with the default visibility setting and 36% of the content overall. They also noted that photo albums had the most privacy-conscious setting across all kinds of content, what reinforces the results hereby presented.

These results refute Hypothesis 4.

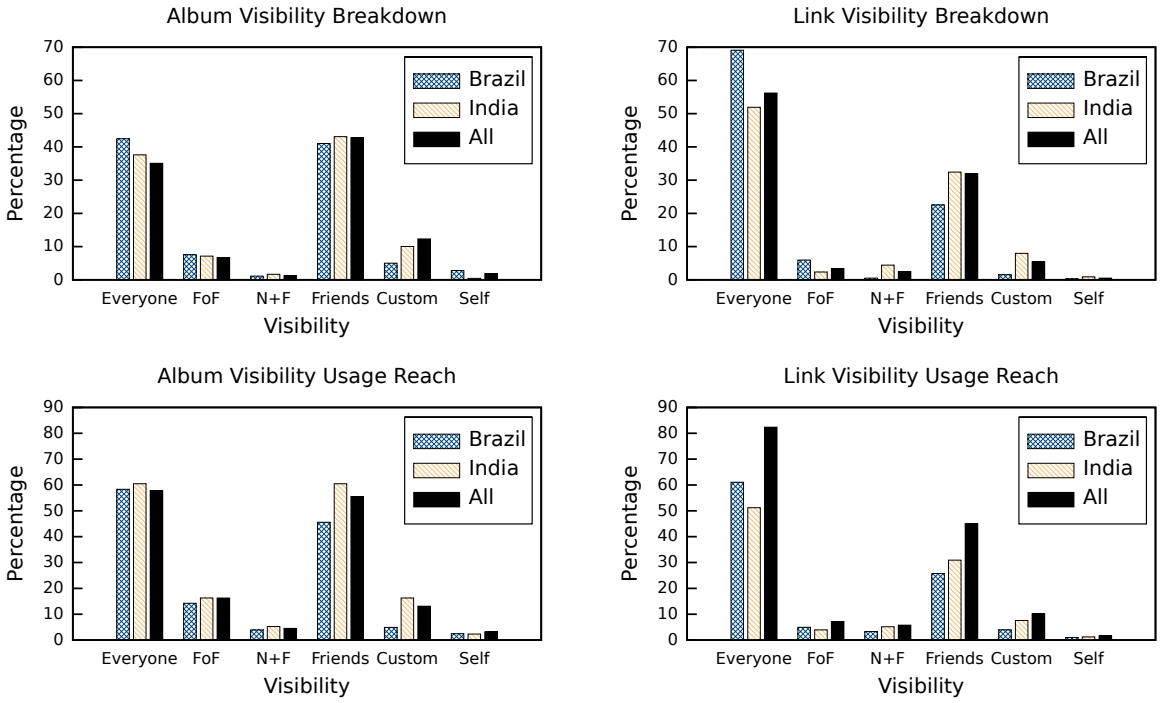

Figure 5.4. On all charts *FoF* is the acronym for *"Friends of Friends"*, *N+F* for *"Networks and Friends"* and *Friends* stands for *"Friends Only"*. The two charts at the top contrast the visibility breakdown between photo albums and links. As it can be seen, users are overall more concerned about their photo albums exposure, what is shown by higher peaks for *Friends*. This message is reinforced by the charts at the bottom, where is displayed the usage reach for every possible setting. For instance, the visibility to *Everyone* reaches more than 82% of the participants for links and 58% for albums.

5.5 H5: Friends List Exposure

Using the public profiles collected for the PF data set, it was found that 67% of the users reveal their friends list publicly. In Brazil, the exposure level reaches 74% of the users, while in India the equivalent statistic is at 69%. The difference reported between the two countries is statiscally significant (Proportion Test, p-value < 0.05). Other representative countries in the sample were analyzed and are resumed in Table 5.5. One can conjecture that the higher rates presented by Brazil and India are due to the recency of the expansion of Facebook on these countries.

Except for Brazil, women were less likely to reveal their friends list on every other region analyzed. Overall, 65.9% of female individuals revealed their friends list publicly, while the statistic reaches 69.6% of male persons. The difference is statiscally significant (Proportion Test, p-value < 0.05).

A parallel work performed by Gundecha et al. [2011] found that 72.03% out of more than 2 million users revealed their friends list publicly on Facebook, in accordance with the results hereby presented.

Table 5.5. Public disclosure rates of the friends list by those in the PF data set. Figures with 95% confidence level.

	All	**Female**	**Male**
	Avg(%) ± Std. Error	Avg(%) ± Std. Error	Avg(%) ± Std. Error
Brazil	73.95 ± 0.88	75.53 ± 1.37	74.27 ± 1.23
India	69.32 ± 0.70	67.22 ± 1.52	70.49 ± 0.83
U.K.	65.59 ± 2.62	63.81 ± 4.95	67.93 ± 3.56
U.S.	62.93 ± 1.21	61.41 ± 2.17	64.67 ± 1.64
All	67.04 ± 0.28	65.89 ± 0.49	69.55 ± 0.37

Another remark regarding the friends list exposure is its access through third party applications. It might be hard to imagine but almost every third party application has access to the entire friends list of a user once it is properly installed. Facebook has an elaborated mechanism where users are asked to grant permissions every time an application wants to make use of a new piece of information (e.g: user birthday or his friends birthdays, etc). Nonetheless, when it comes down to the use of the friends list there is no such exclusive permission (refer to Figure 5.5). As so, unless the application does not ask for any permission, it will have access to the user friends list. It is understandable that applications can be more valuable to a user with the use

of his friends list, but not every application needs it or should have access to it. It is possible to hide the friends list from particular friends on the user profile page but not from applications, which are made most of the time by complete strangers and most of whom are rarely full of good intentions.

These results refute Hypothesis 5.

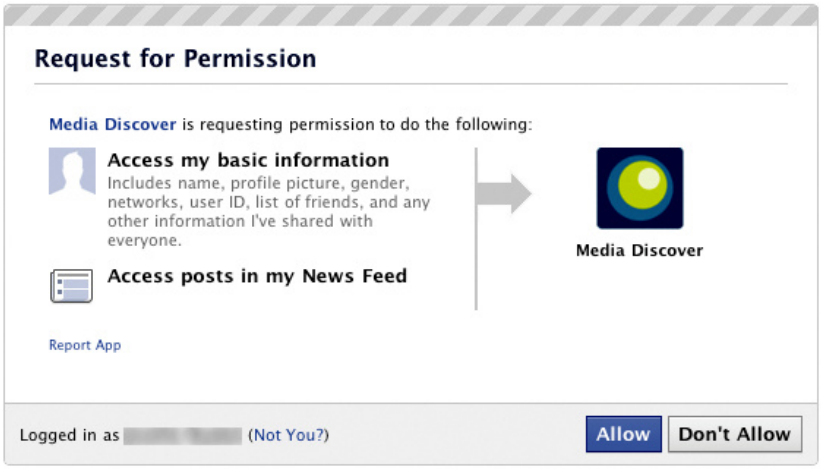

Figure 5.5. Example of a simple permission request dialog on Facebook. The friends list is treated as a basic information such as the name or the gender of the user and does not require a specific permission.

5.6 H6: Date of Birth Exposure

Using the PF data set, no huge contrast was found between Brazil and India with respect to date of birth protection. Nonetheless, the exposure levels are disturbingly high for such a personal information (refer to Figure 5.6). In Brazil, about 64.4% of the participants and their friends expose their full date of birth to their network, while in India the equivalent statistic is a bit lower, at 62.3%.

An important contrast was found between genders concerning the date of birth exposure. Although still high, women seem to be more concerned than men regarding its exposure, wheresoever. For instance, in Brazil 60.7% of the women expose their full date of birth while the same statistic reaches 71.6% of the men. The full results are presented in Table 5.6.

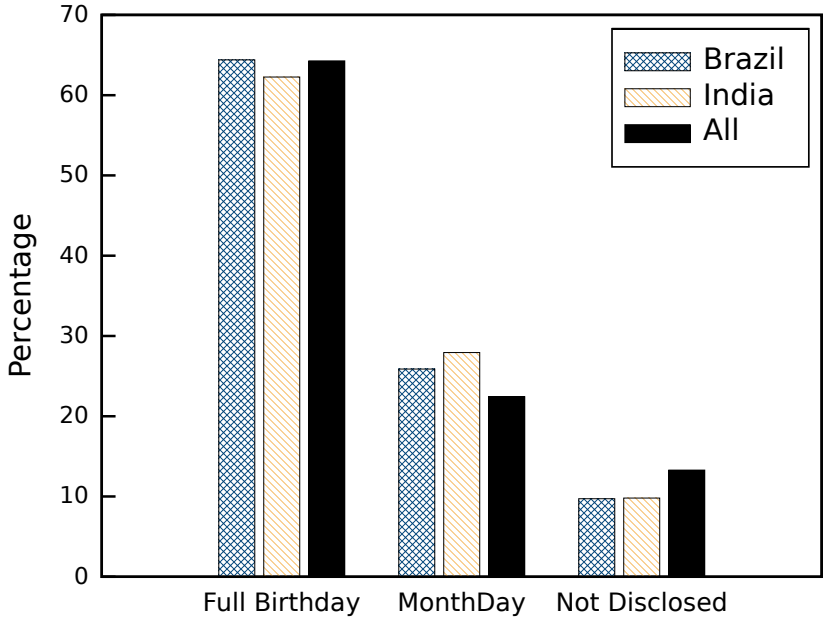

Figure 5.6. Date of birth exposure distribution from the data set PF. The majority of users expose their full birthday.

The age distribution of users could be derived from those who disclose their date of birth (refer to Figure 5.7). It is observed that participants and their networks of friends are concentrated in young age strips (< 36 years). One might think that there is an over representation from these age strips but, according to Facebook Ads [2012], they represent 76% and 90% of users in Brazil and India, respectively.

These results refute Hypothesis 6.

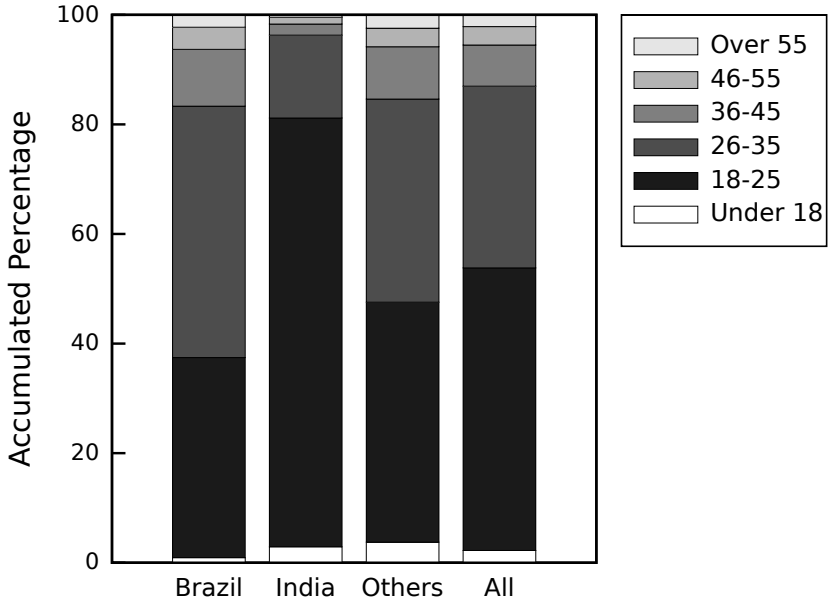

Figure 5.7. Age breakdown from those in the PF data set who revealed their date of birth. The majority of users belong to 18 – 35 years age group.

Table 5.6. Full date of birth exposure reach across different genders in the PF data set. A high percentage of users share their birthdate with their network, specially men. Figures with 95% confidence level.

	Female	**Male**
	Avg(%) ± Std. Error	Avg(%) ± Std. Error
Brazil	60.70 ± 0.02	71.61 ± 0.01
India	58.11 ± 0.02	64.60 ± 0.01
Others	52.54 ± 0.02	54.81 ± 0.01
All	62.07 ± 0.01	68.76 ± 0.01

5.7 Further Analyses

5.7.1 Egocentric Networks

Even though most people have hundreds of friends on OSNs like Facebook, it is hard to maintain a stable social relationship with all of them. In fact, this number of average stable social relationships one can maintain (known as Dunbar's number) has been calculated and lies between 100 and 230, although 150 is a commonly accepted value [Hernando, A. et al., 2010]. Features like wall posts, personal messages and comments on shared items allow users to interact virtually to a great number of friends asynchronously. One feature, though, reflects a kind of interaction that is only possible in real life: photo albums, since you can only take pictures with people you have met personally. These albums portray a part of one's social life interactions and, using information contained on their photo tags, it is possible to create a parallel network of friends [Golder, 2008]. It is only natural that this new network form a much stronger graph than the usual OSN friends network.

The average clustering coefficient found for the participant ego networks was 0.10, which is lower than values found by previous studies, ~ 0.16 [Gjoka et al., 2009; Wilson et al., 2009]. This might be explained by the fact that the set of friendships available is incomplete due to the concealment of the friends list by a portion of the users, as exposed in Section 5.5. When considered only nodes with full information, the average clustering coefficient is raised to 0.13.

The clustering coefficient found for ego networks of participants in India (0.14) is significantly higher then that of participants in Brazil (0.09) (Proportion Test, p-value < 0.05). One might conjecture that this is due to the very high presence on Facebook of youngsters in India, what culminates in a misrepresentation of circles of friendship from other age strips (refer to Figure 5.7). The results are summarized in Table 5.7.

The act of posting a picture and tagging someone exposes a certain level of intimacy between comrades. Based on that fact, it was drawn the ego networks of participants yet this time with only the tagged friends, like the one depicted in Figure 5.8. It was found that the average clustering coefficient for these networks was of 0.52, what demonstrates not only a level of intimacy betweens peers but how ego tends to tag only certain tightly connected circles of companionship. Once more, the average clustering coefficient found for ego networks of participants in India was significantly higher than that found for their peers in Brazil (Proportion Test, p-value < 0.05).

It was also verified that users tend to tag twice as many friends than being tagged by them, as shown in Figure 5.9, a similar result found by Lewis et al. [2008].

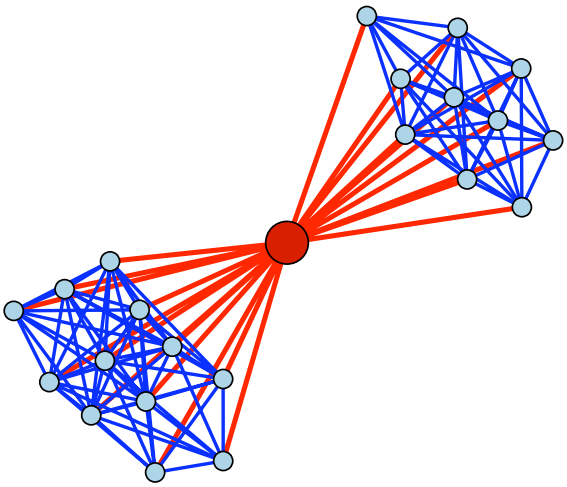

Figure 5.8. Example of egocentric network of people tagged on photos by a user. The user who tags is represented by the larger node at the center, while the friends tagged are shown at the two extremes. A lot of friendship links between tagged friends can be seen. The actual clustering coefficient for this instance is 0.42.

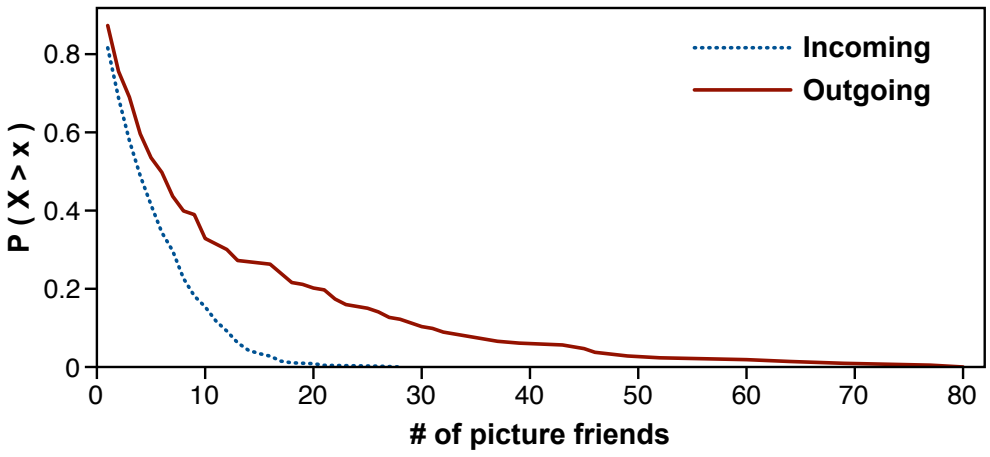

Figure 5.9. Complementary CDF of friends tagged by users in the data set P. The majority is tagged by less than 10 friends but does tag twice as much friends. ($n = 205$)

Table 5.7. Clustering coefficient (CC) for egocentric networks of participants friends and participants tagged friends. Tagged friends form denser egocentric networks. Figures with 95% confidence level.

	Friends CC Avg(%) ± Std. Error	**Tagged Friends CC** Avg(%) ± Std. Error
Brazil	0.09 ± 0.01	0.48 ± 0.09
India	0.14 ± 0.01	0.56 ± 0.07
Others	0.06 ± 0.01	0.45 ± 0.12
All	0.10 ± 0.01	0.52 ± 0.05

5.7.2 Aggregation Erodes Privacy

The revelation of a particular piece of information might seem innocuous for an individual or a community from their particular point of view. For instance, assume that you have submitted a query to a search engine looking for a book about lung cancer. Thus far not a big deal, right? Perharps so, perharps you are a student or maybe a physician or even just a curious hypochondriac. Now suppose that you have purchased a wig on a web store or even on a physical one which offers a good bonus program that in turn track consumer purchases. Far from being a problem, correct? People buy wigs all the time to attend costume parties or possibly for other reasons. Still following the thought? What if an aggregator could collect both data crumbs? Worse, what if an aggregator could collect these and some other particular bits that could lead back to you? Are you concerned by now about sharing these facts? The inference that you have or have had lung cancer and are undergoing or underwent chemotherapy might be something that you would like to keep private or restrict to some of your contacts. This is also probably the case for most of your medical records.[1] Krishnamurthy and Wills [2009] report how user-related data aggregation has been increasing by a decreasing number of entities.

The goal of the present study case is to derive a particular hidden user attribute based on a multitude of other publicly available bits of information and thus show that the aggregation of individual data bits can reveal once private information. Using the public profiles collected for the PF data set it was possible to estimate the age of 79.61% of the users with an error of less than 4 years, for those that age verification was possible. The developed approach was built upon the work of Dey et al. [2012].

[1] Aggregation example adapted from Solove [2011].

By making use of the PP data set it was found that only 3.33% of the users in the PF data set reveal their birthday publicly on their profiles. Also, the exposure rate in India is signifcantly higher than in Brazil, 5.62% in the former and 3.98% in the latter (Proportion Test, p-value < 0.05), as shown in Table 5.8. Moreover, only 1.66% of the users reveal their full date of birth publicly, what indicates how sensitive this particular information is considered by users.

The ground truth set GT consists then of 59916 users of the PF data set who have revealed their full date of birth to their network of friends or have confided it to the *Privacy Study* application directly and also report a birth year later than 1930. The last restriction is meant to exclude users who blatantly lie about their ages on their profiles.

Gundecha et al. [2011] report a birthday disclosure rate of 3.30% out of more than 2 million public profiles, in line with the results hereby presented.

Table 5.8. Birthday public disclosure rates in the PP data set. All figures with 95% confidence level.

	All	**Female**	**Male**
	Avg(%) ± Std. Error	Avg(%) ± Std. Error	Avg(%) ± Std. Error
Brazil	3.98 ± 0.39	4.43 ± 0.65	3.72 ± 0.53
India	5.62 ± 0.35	5.42 ± 0.73	5.73 ± 0.42
All	3.33 ± 0.11	3.19 ± 0.18	3.54 ± 0.15

The several steps of the developed procedure to derive the age of users are described in the following subsections. All of them use publicly available information only.

5.7.2.1 Profile Birth Year

From the 1824 users who reveal their birth year publicly, 1804 of them fulfill the proposed requirements, accounting for 3.01% of the GT data set. No math to be done during this step.

5.7.2.2 High School Graduation Year

As it was stated previously, many users hide their age on their public profiles. Nonetheless, many of them make their high school graduation year publicly available. Both

years are clearly correlated because a person normally graduates from high school around 18 years old.

The approach is quite straightforward. It was performed a linear regression using the high school year and birth year of users who provided them both on their profiles. This particular subset accounted for 629 users or 1.05% of GT.

The linear regression yielded:

$$\text{Birth Year} = 0.97757902 \times \text{High School Graduation Year} + 27.3272185 \quad (5.1)$$

By making use of the high school graduation year provided in public profiles it was possible to estimate the age of 11264 users or 18.80% of GT.

5.7.2.3 High School Network With Graduation Year

Several profiles on Facebook contain high school network affiliation with year information. For instance: *Smithtown High School East '08*. At the time of the collection of the PP data set these were publicly available information only.

Using the graduation year of a high school network affiliation and the Equation 5.1 it was possible to estimate the age of 390 users or 0.65% of GT.

5.7.2.4 High School Only

At this point there is left the high school information on some public profiles yet without the year information. By harvesting the friends list of such users for friends with equivalent high school with year information it is also possible to predict their age. The assumption is simple, people tend to have friends from their high school.

Using this approach it was possible to predict the age of 1840 users or 3.07% of GT.

5.7.2.5 College Graduation Year

Using the least college graduation year provided in public profiles of the PP data set it was taken a similar approach to the one described in Section 5.7.2.2. It was performed a linear regression using the users who provided both their college year and birth year on their profiles. This subset accounted for 580 users or 0.97% of GT.

The linear regression yielded:

$$\text{Birth Year} = 0.932729994 \times \text{College Graduation Year} + 113.0652051 \quad (5.2)$$

The least college graduation year approach could predict the age of 3868 users or 6.47% of GT.

5.7.2.6 College Network with Graduation Year

Similarly to the approach described in Section 5.7.2.3, it was extracted the college affiliation networks with year from the public profiles and then it was applied the Equation 5.2, making possible the age estimation of 939 users or 1.57% of GT.

5.7.2.7 College

The final step with college information is similar to the one described in Section 5.7.2.4 for high school information. By checking the friends list for similar colleges with year information it was possible to predict the age of 148 users or 0.25% of GT.

5.7.2.8 Friends

For users without any of the previous information available it was taken the median birth year of their friends as the estimated age. This approach can be iterated several times until no new prediction is possible. It was performed two iterations of the present step in order to avoid the degradation of the results. The first one could predict the age of 33589 users or 56.06% of GT. The second iteration was able to predict the age of other 3577 users or 5.97% of GT.

5.7.2.9 Likes and Interests

35.5% of the profiles collected in the PP data set displayed some kind of interest or user taste preference. These signals span many categories like music, books, movies, television shows, daily activities and other life interests. All of them are textual information. Many of these signals present a high concentration of users in certain age strips, like the example given in Figure 5.10, where more than 60% of the users who like the singer "Kesha" were born around 1988.

To predict user age using this information it was used a Top-N nearest neighbors algorithm using a Vector Space Model (VSM) retrieval method with a basic TF-IDF (Term Frequency - Inverse Document Frequency) weighting scheme. For each kind of interest it was used a different dimension and the best experimental N found was 25. The details to implement such an algorithm can be found in [Croft et al., 2010; Ricci et al., 2010].

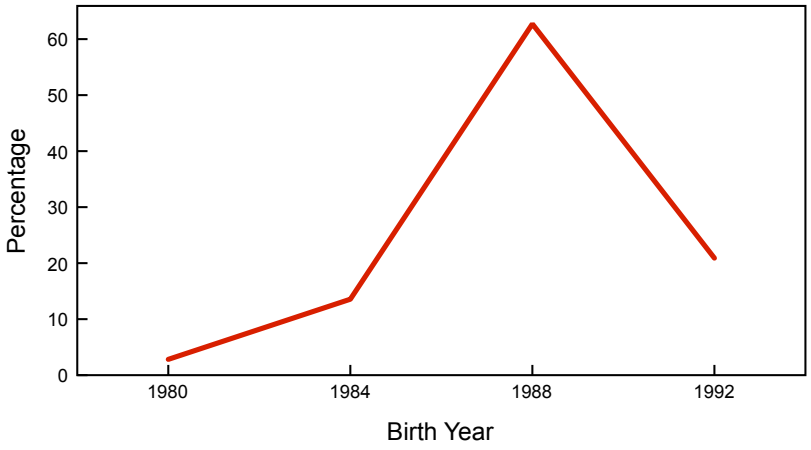

Figure 5.10. Birth year distribution of users in the PF data set that acknowledge interest in artist "Kesha". (bin size = 4 years, $n = 184$)

The age of the nearest neighbors were weighted according to their score, resulting in the following equation:

$$\text{Predicted Age} = \frac{\sum_{n \in N} Score(n) \times Age(n)}{\sum_{n \in N} Score(n)} \tag{5.3}$$

Using this approach it was possible to predict the age of 1080 users or 1.8% of GT.

5.7.2.10 Constant Age

For the remaining users it was simply used a constant age prediction. At this stage, the best result was achieved with a 25 years guess and it was used to predict the age of the last 1410 users, representing 2.35% of GT.

5.7.2.11 Summary

The results of each step of the procedure are summarized in Table 5.9. As can be seen, the "High School" information served as the best predictor, accurately guessing the age of 96.38% of the users with such information available.

For a different perspective of the results it was drawn the complementary cumulative distributed function of the cumulative score in Figure 5.11. As can be noticed, almost 50% of the users had their age predicted within one year interval of error.

In order to compare the results obtained with other possible approaches, it was performed the age prediction of the users in GT with other two simpler algorithms. Both of them are based on the probability mass function of the age

distribution of Facebook users in Brazil and India, drawn in Figure 5.12. The first one is a random guesser that follows the average age distribution of users in Brazil and India and the second one is just a constant age predictor like the step described in Section 5.7.2.10. The former obtained the worst result and could predict the age of only 30.87% of the users within 4 years of error, while the latter could reasonably predict the age of 62.59% of the users in the *GT* data set with a constant guess of 23 years. The results of the comparison are summarized in Table 5.10.

Table 5.9. Summary of the results from the several steps of the the *Privacy Study* proposal age prediction procedure. MAE stands for Mean Absolute Error. CS stands for Cumulative Score (% < years of error).

Step	**# of Users**	**% of *GT***	**MAE**	**CS(4) - %**
Profile Birth Year	1804	3.01	0	100
HS Graduation Year	11264	18.80	1.41	96.52
HS Network with Grad. Year	390	0.65	0.69	99.49
High School (HS)	1840	3.07	1.5	94.84
Subtotal for High School	13494	22.52	1.4	96.38
College Graduation Year	3868	6.47	2.95	83.66
College Network with Grad. Year	939	1.57	2.49	83.49
College	148	0.25	2.94	77.70
Subtotal for College	4955	8.27	2.87	83.45
Friends - 1^{st} iteration	33589	56.06	3.96	76.61
Friends - 2^{nd} iteration	3577	5.97	6.98	52.81
Subtotal for Friends	37166	62.03	4.25	74.32
Likes and Interests	1080	1.80	6.62	56.48
Constant Age (25 years)	1410	2.35	9.1	36.81
Total	59916	100	3.52	79.61

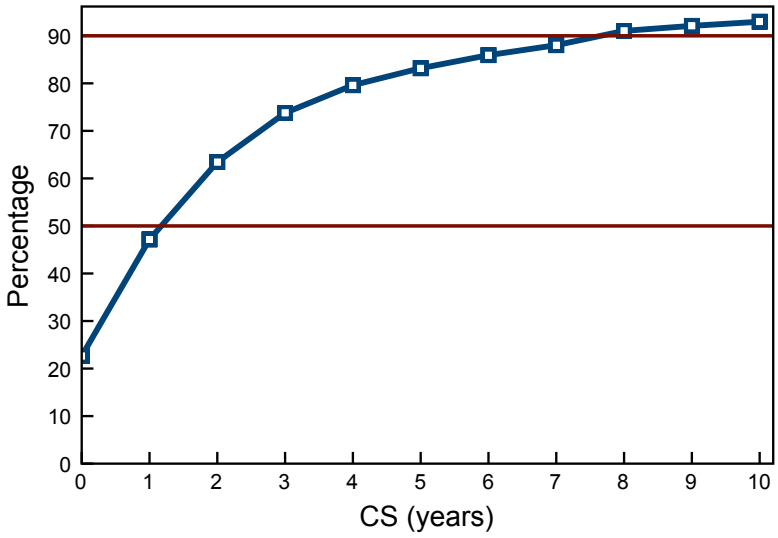

Figure 5.11. Complementary Cumulative Distribution Function (CCDF) of the Cumulative Score (CS) for the *Privacy Study* proposal age prediction procedure. Almost 50% of the users had their age predicted within one year of error.

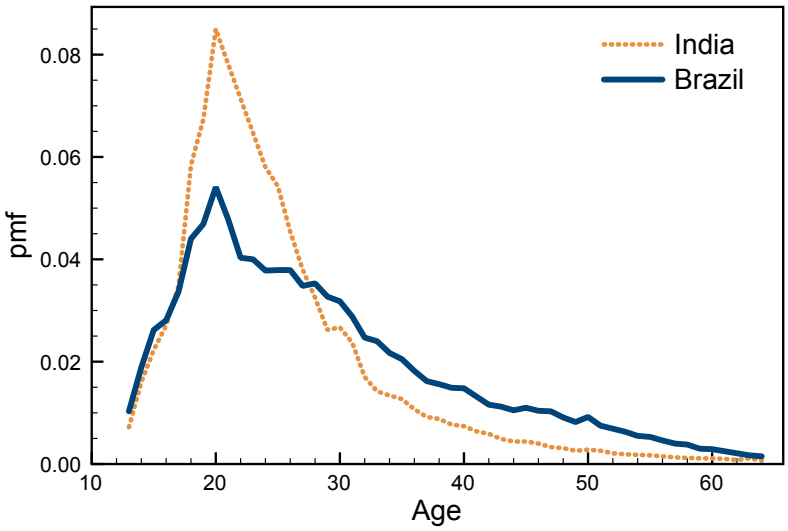

Figure 5.12. Probability mass function of the age distribution of Facebook users in Brazil and India [Facebook Ads, 2012].

Table 5.10. Summary of the results for different age prediction algorithms. MAE stands for Mean Absolute Error. CS stands for Cumulative Score (% < years of error).

Algorithm	**MAE**	**CS(4) - %**
Random Age Distribution	10.61	30.87
Constant Age (23 years)	5.67	62.59
Privacy Study Proposal	3.52	79.61

Chapter 6

Conclusion

Privacy can be understood as one person's ability to control the access to information about herself. In most cases portrayed throughout this study these controls are present but go unnoticed or are often misunderstood by end users. People want freedom to express themselves in the digital era without having to deal with the enormous complexity of current privacy control designs. The default privacy settings of an Online Social Network plays a key role to preserve its users' reputation and shall not be subject to commercial interests solely. It was found that very few users change their default privacy settings. Thereupon, Online Social Networks need to pay more attention when designing their defaults to best serve their users' privacy protection.

Online Social Networks might serve as a key tool to empower people in the digital democracy era. It is paramount that the rollout and spread of this new important technology be supported by strong legal and technical privacy solutions in order not to let it become the materialization of once fictional novels like George Orwell's "1984" or Franz Kafka's "The Trial".

As it was demonstrated, only a small part of users appear to be really aware of the consequences of permissive settings and their pervasive consequences and therefore do not reveal any of the personal identifiable information under study. It was also shown that the majority of the users reveal publicly their gender on Facebook.

Is was revelead that is preponderant the public visibility setting of locational information among users who are known to have commended the information to Facebook, exposing a darker scenario than other previous studies which have shown only overall crude statistics.

It was also shown that users exercise more control over content with more potential to endanger their reputation. For instance, the exposure of photo albums was oftenly configured to reach a narrower audience than links. An implication of that re-

sult would be to modify the Online Social Network default visibility setting according to the content being shared, instead of having a single rule for every content kind. In other words, photo albums should not be made visible to *everyone* by default.

Afterwards, it was elucidated how most part of the users reveal their friends list publicly, what potentially allows the automatic collection of users' networks and the indirectly inference of users' attributes. The list of friends constitutes perharps the most crucial part of one's privacy and nonetheless is treated as a basic information by Facebook, being pratically impossible to keep it hidden from third party applications. A simple solution to this matter could be the adoption of a second privacy control, where one could opt-out from being displayed in everyone else's list of friends.

Using the friends list and photo tagging information from users, it was pointed out how tagged friends form tightly connected circles of friendship. This information can be used to derive not only closest friends but also to construct community detection algorithms.

Lastly, it was demonstrated how aggregation can erode user privacy. By starting from the birthdate information of only 3% of the users of a particular subset it was possible to derive the age of almost 80% of the users within a reasonable margin error. The approach can also be extended to reveal other private attributes of users. For instance, if a user likes several local businesses from a particular town there is a good chance that he or she lives in that city, and so forth and so on.

As in any real world experiment, a few limitations were present. The data was collected mainly through acquaintances and people whom it was able to reach through emails and fliers. Therefore, the sample obtained is a convenient sample, so the results may not be generalizable to all Facebook users. It is also understood that conducting such a study where the users are statistically representative of the country or group of interest is difficult to achieve. Furthermore, the possible issue of bias is commonly present in user studies, such as psychology or sociology surveys among college students, and the present dissertation has no aim to be different.

An interesting future direction of the work hereby presented would be to study the longitudinal effects on the privacy settings on Facebook over the time.

The present study has generated one full article during its course [Rauber et al., 2011].

Bibliography

Acquisti, A. and Gross, R. (2006). Imagined communities: Awareness, information sharing, and privacy on the facebook. In *In 6th Workshop on Privacy Enhancing Technologies*, pages 36--58.

Alexa (2012). Alexa, the web information company. Website, Retrieved in January, 2012. `http:///www.alexa.com`.

Amazon (2005). Mechanical turk. Website. `http://www.mturk.com/`.

Barabási, A. (2003). *Linked: how everything is connected to everything else and what it means for business, science, and everyday life.* A Plume book. Plume.

Bellman, S., Johnson, E. J., Kobrin, S. J., and Lohse, G. L. (2004). International Differences in Information privacy concerns: A global survey of consumers. *The Information Society*, 20:313 – 324.

Besmer, A. and Lipford, H. R. (2010). Moving beyond untagging: photo privacy in a tagged world. *Conference on Human Factors in Computing Systems*, pages 1563--1572.

Beye, M., Jeckmans, A., Erkin, Z., Hartel, P., Lagendijk, R., and Tang, Q. (2010). Literature Overview - Privacy in Online Social Networks.

Boyd, D. and Hargittai, E. (2010). Facebook privacy settings: Who cares? *Journal on the Internet*, 15(8).

Brentcsutoras (2010). Top social networks from the top internet countries. News Article, Retrieved in September, 2010. `http://www.brentcsutoras.com/2010/09/02/top-social-networks-top-internet-countries`.

Chang, J., Roseen, I., Backstrom, L., and Marlow, C. (2010). ePluribus: Ethnicity on Social Networks. In *Proceedings of the Fourth International Conference on Weblogs and Social Media*.

Christakis, N. and Fowler, J. (2010). *Connected: The Amazing Power of Social Networks and How They Shape Our Lives*. HarperCollins Publishers.

ComScore (2010a). Facebook captures top spot among social networking sites in india. News Article, Retrieved in October, 2010. http://www.comscore.com/Press_Events/Press_Releases/2010/8/Facebook_Captures_Top_Spot_among_Social_Networking_Sites_in_India.

ComScore (2010b). Orkut continues to lead brazil's social networking market, facebook audience grows fivefold. News Article, Retrieved in October, 2010. http://www.comscore.com/Press_Events/Press_Releases/2010/10/Orkut_Continues_to_Lead_Brazil_s_Social_Networking_Market_Facebook_Audience_Grows_Fivefold.

ComScore (2012). Facebook blasts into top position in brazilian social networking market following year of tremendous growth. News Article, Retrieved in January, 2012. http://www.comscore.com/Press_Events/Press_Releases/2012/1/Facebook_Blasts_into_Top_Position_in_Brazilian_Social_Networking_Market.

Croft, W., Metzler, D., and Strohman, T. (2010). *Search engines: information retrieval in practice*. Alternative Etext Formats. Addison-Wesley.

Dey, R., Tang, C., Ross, K., and Saxena, N. (2012). Estimating age privacy leakage in online social networks. *Infocom Mini-Conference*.

Diller, S., Lin, L., and Tashjian, V. (2003). The human-computer interaction handbook. chapter The evolving role of security, privacy, and trust in a digitized world, pages 1213--1225. L. Erlbaum Associates Inc., Hillsdale, NJ, USA.

Facebook Ads (2012). Website, retrieved in January, 2012. http://www.facebook.com/ads.

Facebook Blog (2011). Making photo tagging easier. News Article, Retrieved in June 2011, http://blog.facebook.com/blog.php?post=467145887130.

Facebook Statistics (2010). Website, retrieved in November, 2010. http://www.facebook.com/press/info.php?statistics.

Facebook Timeline (2010). Timeline. Website, Retrieved in January, 2010. http://www.facebook.com/press/info.php?timeline.

Fisher, D., McDonald, D. W., Brooks, A. L., and Churchill, E. F. (2010). Terms of service, ethics, and bias: Tapping the social web for cscw research. *Computer Supported Cooperative Work (CSCW), Panel discussion.*

Gallagher, A. C. (2008). Estimating age, gender, and identity using first name priors. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1--8.

Gjoka, M., Kurant, M., Butts, C. T., and Markopoulou, A. (2009). A Walk in Facebook: Uniform Sampling of Users in Online Social Networks. Technical report, arXiv.org.

Golder, S. (2008). Measuring social networks with digital photograph collections. In *Proceedings of the nineteenth ACM conference on Hypertext and hypermedia*, HT '08, pages 43--48, New York, NY, USA. ACM.

Grimmelmann, J. (2009). Saving facebook. *Iowa Law Review*, 94:1137 – 1206.

Gross, R. and Acquisti, A. (2005). Information revelation and privacy in online social networks. *Proceedings of the 2005 ACM workshop on Privacy in the electronic society*, pages 71--80.

Gundecha, P., Avenue, S. M., and Barbier, G. (2011). Exploiting Vulnerability to Secure User Privacy on Social Networking Site. *KDD11.*

Hernando, A., Villuendas, D., Vesperinas, C., Abad, M., and Plastino, A. (2010). Unravelling the size distribution of social groups with information theory in complex networks. *Eur. Phys. J. B*, 76(1):87–97.

Hofstede, G., Hofstede, G., and Minkov, M. (2010). *Cultures and Organizations: Software for the Mind, Third Edition.* McGraw-Hill.

Huffington Post (2010). Facebook changes raise privacy concerns among us senators. News Article, Retrieved in April, 2010. `http://www.huffingtonpost.com/2010/04/27/facebook-changes-raise-pr_n_553129.html`.

Inside Facebook (2010). New facebook statistics show big increase in content sharing, local business pages. News Article, Retrieved in February, 2010. `http://www.insidefacebook.com/2010/02/15/new-facebook-statistics-show-big-increase-in-content-sharing-local-business-pages/`.

Ishitani, L., Almeida, V. A. F., and Júnior, W. M. (2003). *Uma arquitetura para controle de privacidade na web.* PhD thesis, UFMG.

Kirkpatrick, D. (2010). *The Facebook Effect: The Inside Story of the Company That Is Connecting the World*. Simon & Schuster.

Krishnamurthy, B. and Wills, C. (2009). Privacy diffusion on the web: a longitudinal perspective. In *Proceedings of the 18th International Conference on World Wide Web, WWW 2009, Madrid, Spain, April 20-24, 2009*, pages 541–550.

Krishnamurthy, B. and Wills, C. E. (2008). Characterizing privacy in online social networks. In *Proceedings of the first workshop on Online social networks*, WOSN '08, pages 37--42, New York, NY, USA. ACM.

Kumaraguru, P. and Cranor, L. F. (2005a). Privacy in India: Attitudes and Awareness. In *Proceedings of the 2005 Workshop on Privacy Enhancing Technologies (PET2005)*.

Kumaraguru, P. and Cranor, L. F. (2005b). Privacy indexes: A survey of westin's studies. Technical report, Carnegie Mellon University.

Kumaraguru, P., Cranor, L. F., and Newton, E. (2005). Privacy perceptions in india and the united states: An interview study. *In The 33rd Research Conference on Communication, Information and Internet Policy (TPRC)*.

Kumaraguru, P., Sheng, S., Acquisti, A., Cranor, L. F., and Hong, J. (2008). Lessons from a real world evaluation of anti-phishing training. *e-Crime Researchers Summit, Anti-Phishing Working Group*.

Leskovec, J. and Horvitz, E. (2008). Planetary-scale views on a large instant-messaging network. *International World Wide Web Conference*, pages 915--924.

Lewis, K., Kaufman, J., Gonzalez, M., Wimmer, A., and Christakis, N. (2008). Tastes, ties, and time: A new social network dataset using facebook.com. *Social Networks*, 30(4):330 -- 342.

LinkedIn Blog (2011). 100 millions members and counting. News Article, Retrieved in March, 2011. `http://blog.linkedin.com/2011/03/22/linkedin-100-million/`.

Lipford, H. R., Besmer, A., and Watson, J. (2008). Understanding privacy settings in facebook with an audience view. In *UPSEC'08: Proceedings of the 1st Conference on Usability, Psychology, and Security*, pages 1--8, Berkeley, CA, USA.

Liu, Y., Gummadi, K. P., Krishnamurthy, B., and Mislove, A. (2011). Analyzing Facebook Privacy Settings: User Expectations vs. Reality. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference IMC 11*, page 61. ACM Press.

Mackay, W. E. (1991). Triggers and barriers to customizing software. In *CHI '91: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 153--160, New York, NY, USA. ACM.

Madison, J. (1791). The United States Bill of Rights. Website, retrieved in December, 2010. `http://www.law.cornell.edu/constitution/`.

Mail Online (2009). Mi6 chief blows his cover as wife's facebook account reveals family holidays, showbiz friends and links to david irving. News Article, Retrieved in September, 2010. `http://www.dailymail.co.uk/news/article-1197562/MI6-chief-blows-cover-wifes-Facebook-account-reveals-family-holidays-showbiz-friends-links-David-Irving.html`.

New York Post (2010). No. 1 facebook unseats google. News Article, Retrieved in September, 2010. `http://www.nypost.com/p/news/business/no_facebook_unseats_google_aL6kAkw7i9xjA88LdcicHO`.

Rauber, G., Almeida, V. A. F., and Kumaraguru, P. (2011). Privacy albeit late. *XVII Simpósio Brasileiro de Sistemas Multimídia e Web*.

Ricci, F., Rokach, L., Kantor, P., and Shapira, B. (2010). *Recommender Systems Handbook*. Springer.

Social Media Optimization (2008). Social network user demographics. Website. `http://social-media-optimization.com/2008/05/social-network-user-demographics/`.

Solove, D. (2004). *The Digital Person: Technology and Privacy in the Information Age*. Ex machina. NYU Press.

Solove, D. J. (2007). *The future of reputation: gossip, rumor, and privacy on the Internet*. Yale University Press.

Solove, D. J. (2010). Fourth Amendment Pragmatism. *Boston College Law Review Vol 51 p 1 2010*.

Solove, D. J. (2011). *Nothing to Hide: The False Tradeoff between Privacy and Security*. Yale University Press.

Stevenson, A. and Lindberg, C. (2010). *New Oxford American Dictionary*. Oxford University Press.

Sweeney, L. (2002). k-anonymity: a model for protecting privacy. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 10(5):557--570.

TechCrunch (2011). Facebook users uploaded a record 750 million photos over new year's. News Article, Retrieved in January, 20111. `http://techcrunch.com/2011/01/03/facebook-users-uploaded-a-record-750-million-photos-over-new-years/`.

The Economist (2010). Facebook and the hotel california. News Article, retrieved in October, 2010. `http://www.economist.com/blogs/babbage/2010/10/facebook_and_transparency`.

The Guardian (2010). Facebook privacy hole 'lets you see where strangers plan to go'. News Article, Retrieved in October, 2010. `http://www.guardian.co.uk/technology/2010/apr/26/facebook-privacy-hole`.

The Washington Times (2011). Austrian student takes on facebook over privacy. News Article, Retrieved in October, 2011. `http://www.washingtontimes.com/news/2011/oct/26/austrian-student-takes-on-facebook-over-privacy/`.

Thomas, K., Grier, C., and Nicol, D. M. (2010). unFriendly: Multi-party Privacy Risks in Social Networks. *Privacy Enhancing Technologies Symposium*, 6205:236--252-pgccpagedoformat-252.

Toch, E., Cranshaw, J., Drielsma, P., Tsai, J., Kelley, P., Springfield, J., Cranor, L., Hong, J., and Sadeh, N. (ACM International Conference on Ubiquitous Computing (Ubicomp)). Empirical models of privacy in location sharing. *To appear in.*

Tsai, J., Kelley, P., Cranor, L., and Sadeh, N. (2009). Location-sharing technologies: Privacy risks and controls. *The 37th Research Conference on Communication, Information, and Internet Policy (TPRC).*

Ugander, J., Karrer, B., Backstrom, L., and Marlow, C. (2011). The Anatomy of the Facebook Social Graph. *ArXiv e-prints.*

Warren, S. and Brandeis, L. D. (1890). The right to privacy. *Harvard Law Review*, 4(5):193–220.

Wei, X., Yang, J., and Adamic, L. (2010). Diffusion dynamics of games on online social networks. *3rd Workshop on Online Social Networks.*

Wikipedia (2011). English version of wikipedia. Database dump of December 1st, 2011. http://dumps.wikimedia.org/enwiki/.

Wilson, C., Boe, B., Sala, A., Puttaswamy, K. P. N., and Zhao, B. Y. (2009). User interactions in social networks and their implications. *Proceedings of the fourth ACM european conference on Computer systems EuroSys 09*, page 205.

Wilson, D. and Purushothaman, R. (2003). Dreaming with brics: The path to 2050. Technical report, Golman Sachs.

Yahoo! News India (2010). Using too much facebook and twitter may cost you your job. News Article, Retrieved in October, 2010. http://in.news.yahoo.com/139/20101010/882/twl-using-too-much-facebook-and-twitter.html.

Zhou, B. and Pei, J. (2008). Preserving privacy in social networks against neighborhood attacks. In *ICDE '08: Proceedings of the 2008 IEEE 24th International Conference on Data Engineering*, pages 506--515.