Analysis of Privacy in Online Social Networks of Runet

Slava Kisilevich
University of Konstanz
Germany
slaks@dbvis.inf.uni-konstanz.de

Florian Mansmann
University of Konstanz
Germany
Florian.Mansmann@uni-konstanz.de

ABSTRACT

In recent years, social networking sites (SNSs) gained high popularity among Internet users as they combine the best of both worlds: befriending people outside real life situations and staying in touch with people already known. An important aspect of any SNS is user profiles, which allow users to virtually publish anything about themselves, including highly personal or sensitive information. With the inception of SNSs, the problem of personal information disclosure and privacy implications has turned into a serious issue. While privacy issues in SNSs have been extensively analyzed in the past five years showcasing flagships of "western" SNSs like Facebook and MySpace, SNSs that target mainly Russian speaking audiences are not yet analyzed and demand investigation. The goals of this paper are twofold: (1) to raise the awareness of the public to the problems of information revelation by studying the amount and type of information disclosed by users of Runet (Russian Segment of the Internet) SNSs (2) to compare our findings to the results of previous studies in the context of "western" SNSs. We investigate different aspects of information revelation of more than 30 million user profiles collected from five Runet SNSs considered in this paper. In addition, we conducted a survey among a Russian speaking population to assess both the level of awareness of the privacy issues and the level of trust, and compared the results to previous studies. While the results indicate that Runet users tend to disclose less information and are more concerned about privacy implications, there is still a substantial gap between western and Runet SNS providers in understanding of privacy implications and implementation of security measures, which leads to exposure of extensive amounts of personal information.

Categories and Subject Descriptors

K.4.1 [**Computer and Society**]: Public Policy Issues—*Privacy*

SIN'10, Sept. 7–11, 2010, Taganrog, Rostov-on-Don, Russian Federation.

General Terms

Human Factors, Measurement, Security

Keywords

Social Networking Sites, Information Revelation, Privacy, Trust, Security

1. INTRODUCTION

The phenomenon of social networking sites introduced to the public about five to six years ago have changed the way people communicate with each other on the Internet. Some social networking sites like Facebook[1] and MySpace[2] attracted millions of users in the first years of their operation and their operators claim to have hundreds of millions of users worldwide. The main benefit of the majority of social networks is to facilitate new friendships, online interaction and communication for which the user profile is the crucial component for establishing connections. In the profile a user can share information such as his name, photos, address, interests, political views, etc. Several studies showed that the members of social networks not only disclose true information about themselves, but provide even more information than they would do in real life [1, 2]. As mentioned in [14], since the success of an SNS depends on the number of members, it attempts to encourage new users to register by improving the design of the website while security and privacy considerations are often left behind. As a consequence, third parties are using this information in different ways to undermine the privacy of SNS users. The privacy risks of this undesired access to profile information can vary from the creation of digital dossiers to stalking, identity theft, spam, cyber-bullying, etc. As a response to different security and privacy related risks as well as to questions related to understanding of users' behavior and reasons for information disclosure, SNSs have been studied from different perspectives such as inference attacks [27, 13, 26], privacy inference [19], sociology [21], psychological studies [3], law [11, 9], privacy policies [4], solutions to privacy protection [17, 8], security issues and recommendations [14], social network structure [18], privacy threats, trust, analysis of amount of disclosed information and user strategies for keeping their privacy [15, 12, 10, 1, 23, 7, 17, 22, 25, 6, 20].

While, most of the listed studies focus on Facebook and few analyze other SNSs such as MySpace[21, 6], the SNSs

[1]http://facebook.com
[2]http://myspace.com

that operate in Runet targeting mainly a Russian speaking audience are not yet analyzed and demand investigation.

The goals of this paper are twofold: (1) to raise the awareness of the public to the problems of information revelation by studying the amount and type of disclosed information by users of social networking sites in Runet (2) to compare our findings to the results made in previous studies. We investigate different aspects of information revelation using about 30 million user profiles collected from five Runet SNSs considered in this paper (*Vkontakte*[3], *MoyMir*[4], *MirTesen*[5], *Love.Mail.Ru*[6], *Loveplanet*[7]). In addition, we conducted a survey among the Russian speaking population to assess both the level of their awareness of privacy issues and their level of trust, and compared the results to previous studies. While the results indicate that Runet users tend to disclose less information and are more concerned about privacy implications, there is still a big gap between western and Runet SNS providers in understanding of privacy implications and implementation of security measures, which leads to exposure of extensive amounts of personal information.

The contributions of the paper are as follows:
(1) We analyze privacy issues in Runet, the Russian segment of the Internet, using data collected from the five most popular social networks where the common unifying characteristic of users is the Russian language.
(2) In contrast to past research, which mostly considered students in their studies, we analyze users from different social groups.
(3) Compared to previous studies, this paper provides the first large-scale study and high levels of detail. For example, [5] reported that 11 million auction users were crawled using parallel machines extracting complete profile information from only 66,000 users, while [18] used 58 computers to download only the graph structure of 4 networks (Flickr, LiveJournal, YouTube, Orkut) totaling in 11.3 million users of which only about $347,000$ users belong to the "pure" social network Orkut. Three other networks provide public APIs to download this information.
(4) We conducted a survey using Russian speaking users of various SNSs including persons from different social groups and countries, statistically analyzed privacy awareness and trust among users of Runet SNSs, and compared the results of past studies performed on "western" SNSs (mainly Facebook). In comparison to these previous studies, that are limited and mainly generalize non-representative samples collected by interviewing university students, our survey results are significantly strengthened through the use of real data extracted from large population groups.

2. RELATED WORK

The type of personal information students disclose on their Facebook profiles, the reasons for joining the social network, and the effects of the survey on changing privacy settings and privacy implications, was shown in a survey among Carnegie Mellon University students [10, 12]. Thu study [10] showed that the amount of personal information disclosed and the awareness of the privacy are very high. However, the survey had no effect on minimizing the amount of disclosed personal information.

Another survey [23] among 38 undergraduate and graduate students was conducted to analyze which social networking sites are popular among students, what personal information they usually disclose, how many members of SNS disclose personal information, and what are the opinions of users regarding privacy issues. The results indicated that the majority (90%) of undergraduates use Facebook. Name, gender, email, birthday, pictures, address, relationship are the most disclosed information. User opinions indicated that the users are not concerned with the implications of personal disclosure, even if in addition to the information exchange with their friends, family, and classmates, their profiles are accessed by strangers.

294 students (mostly the Facebook users) were interviewed [1] to investigate users' privacy concerns, awareness of the privacy issues, attitudes towards Facebook, and level of trust to different community entities like friends, friends of friends, and strangers. Different relations were tested, such as the influence of gender or age on privacy concerns, which revealed that neither gender nor age are significant for those who are members of the social networks, while age may be a restraining factor for older people to join social networks. The statistical results from surveys were compared to the actual data the respondents provided in their profiles. The data revealed that 77.84% of anwers were exactly true, 8% disclosed more than they specified in the survey, and 11% disclosed less then what was reported in the survey.

Influence of trust on information disclosure was studied in [7] on a population of 69 Facebook and 48 MySpace users. The results suggested that Facebook users express more trust than MySpace users and disclose more private information.

Facebook users' strategies for maintaining their privacy was studied in [22] to answer the questions about how and why users share and protect their personal information. The study was conducted by interviewing 18 undergraduate students at the UNC university.

Tufekci [24] compared the level of disclosed information, privacy awareness and gender differences between users of Facebook and MySpace.

In a similar study [6], issues related to personal information disclosure by children in MySpace were investigated. Three reasons for information disclosure were identified: peer pressure, website user interface and signaling.

A recent study [25] extended past research by investigating several privacy issues like privacy concerns among users and strategies to maintain privacy. The study was conducted by interviewing 77 undergraduate university students who have profiles in Facebook. The results indicated that the majority of users do care about privacy issues and develop different strategies to minimize the damage resulting from undesired access to their profles. For example, they changed default settings to control access to their profile, restricting accessibility to certain users or the send private messages instead of making them public. A few users fake their personal data to restrict strangers from collecting true data about them.

[3]http://vkontakte.ru
[4]http://my.mail.ru
[5]http://mirtesen.ru/
[6]http://www.love.mail.ru
[7]http://loveplanet.ru

3. ONLINE SOCIAL NETWORKS IN RUNET

3.1 Overview

In this section we provide an overview of the five most popular social networks in Runet.

3.1.1 Vkontakte

Vkontakte was launched in 2006 and can be regarded as a clone of the early version of Facebook. According to their "users agreements", it is an *internet project that joins people on the basis of their place of study or work.* It is very popular among teenagers and students. Its user interface is similar to the one of Facebook, and many of the privacy settings are built around the early privacy settings of Facebook. It is one of the most visited websites in Russia according to Alexa[8] and claims to have about 65 million users.

3.1.2 MoyMir

MoyMir was launched in 2007 as a social network that combines all content of users on the Mail.ru email portal. In order to register with MoyMir, the user is required to open an email account on Mail.ru first, providing the following mandatory fields: *First Name, Last Name, Birthday* and *Gender*. MoyMir allows for creating and joining "societies", sharing photos, videos, music, and managing a list of friends. According to statistics[9], it has about 40 million registered users.

3.1.3 MirTesen

MirTesen was launched in 2007 as a social network with an important geographical feature for its users. Here he/she can locate other users on a map. Additionally, the users can assign themselves to different places under categories such as *Place of Birth, Place of Work, Place of Residence*, etc., and in the same manner geo-locate other users on the map. The registration requires *First Name, Birthday, Gender* and *Location* as the necessary information. Furthermore, users are required to upload their photo in order to complete the registration. According to the information provided on the main page (April 2010), there are about 11.6 million registered users.

3.1.4 Love.Mail.Ru

Love.Mail.Ru is a dating site launched in 2004. It lacks many features of the classical social networking sites such as creation of groups of interest and searching people according to interests (search includes only gender, location or aim of the acquaintance). The friends list is explicitly created when a user sends a message to another user. There is no way to prevent receiving messages from unknown people. Only after the message is received, the sender can be added to the black list. A special paid VIP-status was introduced, which allows receiving messages solely from other VIP members. A new commercial feature was recently added where the user can confirm that his profile is real by sending a paid SMS message. According to the statistics presented on the main page (March, 10 2010), the total number of profiles is about 12.6 million, of which 7.5 million profiles are searchable. The registration takes place via portal Mail.Ru.

[8]http://www.alexa.com/topsites/countries/RU, April 2010
[9]http://www.corp.mail.ru/about.html, April 2010

3.1.5 Loveplanet

Loveplanet is a dating site competing with Love.Mai.Ru. According to the information presented on its main page, there are about 16 million users registered. For the registration the following fields are mandatory: *Name* or *Nickname* as one field, *Birthday, Gender, Sexual Orientation, Country, Region* and *City.* It allows sharing videos and photos, managing a diary and maintaining a list of friends.

3.2 Security, Privacy and Data Collection

In this section we describe four security and privacy-related threats and data collection methods.

3.2.1 Security and Privacy considerations

As was stated in the Section 1, one of the goals of this paper is to raise the awareness to the problem of information disclosure where both the providers of SNSs as well as the users contribute to this issue. The study in [14] outlined fifteen privacy and security risks that the users or SNSs providers may face. Among the fifteen outlined threats, we would like to mention the four most closely related to our study with respect to SNS: (Threat-1) *digital dossier aggregation* - downloadable SNS profiles, (Threat-2) *secondary data collection* - the data is disclosed by the network operator and visible on the profile (users who visited the profile, time-related activity statistics), (Threat-3) *difficulty of complete account deletion*, (Threat-4) *infiltration of networks* - restricting the information only to friends or sub-networks. The last threat is possible when someone becomes a friend or member of the network (gaining access to the restricted information) using false claims or personal details.

3.2.2 Data collection overview

We used a website crawling application developed in-house using C# programming language to extract profile information from a particular user. The data from four websites (excluding *Vkontakte*) were collected using three network computers during a two month period (with several delays for database maintenance and tuning) beginning on February 10th and finishing on April 10th 2010 (except for *Love.Mail.Ru* that was stopped on March 10th due to statistical analysis). In the following subsections particular methods and limitations will be described in detail for every SNS separately. Table 1 shows what information was available for download (labeled as "•"), was downloaded by our crawlers ("✓"), and was not provided by the SNS ("-").

3.2.3 Vkontakte

Vkontakte provides rich privacy settings. It allows controlling accessibility to the profile, photos, and messages separately. It implements throttling mechanisms to slow down profile download (by returning no results for a specific time interval when the page request exceeded some threshold). All these measures reduce the possibility of building digital dossiers. However, the SNS provider forces the users to reveal more than 30% of personal information before the users can communicate with others. The activity of other users is publicly visible (*Threat-2*). Vkontakte does not implement sub-networks like Facebook and access restriction is possible between groups of friends and non-friends only (*Threat-4*).

Our goal was to check how many *Vkontakte* users change default settings and make their profile and friend network invisible to unknown users. In order not to violate the terms

of use, we collected this information manually from 1,000 users by following their user profiles and friends list. When the profile is not visible to a non-friend user or when the friends list is hidden, the appropriate message is shown on the screen, otherwise the complete information about the person or his/her friends is shown.

3.2.4 MoyMir

Page accessibility is controlled by visibility settings and includes such options like *visible to all, visible to friends, visible to unauthorized users*, whereby the first and third options are checked by default. However, these settings do not include the visibility of personal information, which stays open to everyone. This allows downloading complete profile information as registered or unregistered user (*Threat-1*). Moreover, MoyMir forces the users to reveal some information by restricting accessibility to personal pages of other users, which includes messages, guest books, photo albums, blogs, etc. until the users uploads their photos. Access to profile information, however, stays available. The SNS provider does not implement any throttling mechanisms to slow down page requests and reveals the status of the users as online or the last time they were online (*Threat-2*). However, there is a paid "invisibility" service to temporarily hide one's own online activity. While access to personal items like photos or blogs can be restricted, there is no restriction to access user profiles (*Threat-4*). In addition to all the described threats, we would also like to mention one major flaw. The user's profile URL is composed of his/her personal email that he/she used during registration. Together with the lack of mechanisms to prevent profile downloading, this site comprises a real threat of digital dossier aggregation, which can be used for spam or stalking [14].

Data collection was divided into three steps. In the first step we crawled the list of so called "societies" (interest groups) totaling in 583,252. In the second step, for every society, we acquired the list of members. The website maintains statistics on the number of members in a particular society, but in reality the number of returned users was by far fewer. We ran the crawler several times on all societies and collected 8,617,530 million users. In the third step, we ran the crawler on every user, extracting his/her personal information. In total, we crawled 17,582,267 users and obtained detailed information from 14,575,806 users.

3.2.5 MirTesen

In MirTesen, the users can control the visibility of their profiles in two ways: visible to unregistered or registered users. Both options allow downloading complete user profiles (*Threat-1*). The SNS provider implements a basic throttling mechanism to prevent fast page access requests. We found out that requests made faster than one second return empty results, while requests made with random delays between one second to one minute allow running three parallel crawlers on one machine. According to *Threat-2*, the last time a user was online is visible on his profile. The details of a deleted profile including messages, profile id and photo remain in profiles and in friend lists of other users (*Threat-3*). In addition, the SNS provider retains the email of the deleted profile. This suggests, that the information of the deleted account is not deleted by the SNS provider. There are no means of restricting access to the profile and the complete information is visible (*Threat-4*). The difference between access to the user profile as an unregistered and registered user is only in the ability to see the geolocation status of the user (home, work coordinates) and friends list (changed through privacy settings). By crawling the website as an unregistered user we could infer how many users change their privacy settings by counting the number of users who changed their friend visibility. A two step approach was applied: In the first step, we downloaded initial user information to seed the crawler, using the website feature to get 10 random people. After a sufficient number of people were extracted, we applied step two in which we collected some of the available personal information. We crawled 2,544,833 users and acquired detailed information from 639,649 users. In addition, we could also infer how many users from the whole population of crawled users provide images of themselves in their profiles.

3.2.6 Love.Mail.Ru

The *Love.Mail.Ru* profiles are publicly available (*Threat-1*) and the SNS provider does not implement any throttling mechanisms to slow down the download rate. Moreover, the last time the user was online is disclosed (*Threat-2*). The paid service called VIP applies temporal invisibility and hides the activity status from other users. According to the the site's own statistics[10] only 0.24% of users made use of this service. The deleted profile leaves messages to other users (*Threat-3*). There is no notion of sub-networks and friends, but the complete visibility of the profile and lack of message sending control leaves *Threat-4* valid.

It is possible to navigate sequentially through all the profiles. We used this option to download the whole population of this SNS aiming to provide insight into the demographic statistics and level of sensitive (personal, intimate) information revealed by people from different countries by gender. It is interesting to note, that on March 10, the main page stated that there are 12,503,630 profiles and 7,383,483 are searchable, whereas we managed to download 10,452,992 profiles, of which 932,884 profiles were removed or blocked, leaving us with 9,520,108 valid user profiles. This is almost 29% more than the amount advertised. In the next sections, we consider only the valid profiles when referring to the analysis of this data.

3.2.7 Loveplanet

Loveplanet user profiles are publicly available (*Threat-1*), which allows an unregistered user to perform a search using country, region, city, age, and other parameters. Moreover, the SNS provider does not implement any throttling mechanisms to slow down the download rate. In addition, the activity information of the user (*Threat-2*) is disclosed: the last time the user was online and other users who saw his profile. The user's activity can be hidden using SMS paid service. In this case, the profile will not be located by a search query and allows visiting other profiles without revealing the identity of the user performing the search. The deleted profile leaves messages to other users, the user name and email of the deleted profile is retained by the SNS provider (*Threat-3*). There is no notion of sub-networks while friends list is always accessible (*Threat-4*).

The profile search option returns no more than 110 pages with 10 users on each, and it can vary according to the specificity of the searching query. We used three crawlers

[10] http://love.mail.ru/support.phtml?qid=293, April 2010

and provided them with a separate list of country codes obtained from the HTML page of the search form, popular regions like US states and Canadian provinces as well as several Russian regions and a list of popular cities. We also observed that the website returns new users for the same search query after some period of time and used this observation to run three crawlers in an infinite loop iterating through the country, region or city list and starting again. Using this approach we managed to collect a $1,039,154$ user profiles containing personal and intimate information.

Table 1: Personal information collected from four SNSs, "•": available, "✓": downloaded, "-": not provided.

Fields	Love.Mail.Ru	Loveplanet	MoyMir	MirTesen
Name	✓	✓	✓	✓
Gender	✓	✓	✓	•
Location	✓	✓	✓	•
Age	✓	✓	✓	✓
Aim	✓	✓	-	-
Zodiac	✓	✓	✓	✓
Height	✓	✓	•	-
Weight	✓	✓	•	-
Day regime	✓	✓	•	-
Religion	✓	✓	•	-
Body prop.	✓	✓	•	-
Wealth	✓	•	-	-
Dwelling prop.	✓	•	-	-
Smoking	✓	✓	-	-
Alcohol	✓	✓	-	-
Drugs	✓	✓	-	-
Car	•	✓	•	•
Kids	✓	✓	•	-
Marital st.	✓	✓	•	•
Intimate inf.	✓	✓	-	-
Eye color	•	✓	•	•
Skin color	•	✓	•	•
Hair color	•	✓	•	•
Interests	•	✓	•	•
Languages	✓	✓	•	-

4. RESULTS

The aim of this section is: (1) To show the general demographic statistics of the whole population of the social network site, age and gender distribution as well as to present the amount of disclosed information by age and gender. (2) To introduce survey results and compare to the statistics mined from the data collected and to the results of past studies.

4.1 Demographic statistics

In this section we provide demographic statistics of Russian speaking members of Love.Mail.Ru SNS and the level and type of disclosed information by gender and country. Table 2 presents the 20 most active countries and the percentage of males and females according to the country specified in their profiles. Some interesting patterns can be found. For example, Russia is the only country where females outnumber males. Azerbaijan and Turkey are the countries with a significantly high number of male users.

The age distribution by country and gender is presented in Figure 1 using box-and-whisker plots.

A profile in Love.Mail.Ru contains more categories than other general purpose Runet SNSs: self-description, dating and intimate info, character descriptions and interests. We divided these categories into three broad categories: (1) intimate - includes such fields like *sexual orientation, sex frequency* and *preferences in sex*. (2) type - personal information like *weight, height, smoking habits, alcohol, drugs, body characteristics* and *knowledge of language*. (3) status - *kids, material status, dwelling type, marital status*, and *religion*. Table 3 summarizes the percentage of disclosed information by country and age. According to the results, all genders provide *type* information. Only users from Georgia reveal significantly less information about themselves (82.5% males and 83% females). Males in all countries disclose more intimate information about themselves than females. However, the differences between males and females is the highest in Russia with 20.67% and the lowest in Spain with 5.59%. Females from 17 countries reveal more status information. The exceptions are Russia, Israel and England.

Table 2: User population in 20 most active countries distributed by gender

Country	Total	Male (abs)	Male (%)	Female (abs)	Female (%)
Russia	6,401,680	3,116,796	49	3,284,884	51
Ukraine	1,230,374	668,206	54	562,168	46
Kazakhstan	403,037	227,552	56	175,485	44
Belarus	339,498	185,235	55	154,263	45
Germany	133,734	80,193	60	53,541	40
Azerbaijan	114,521	92,393	81	22,128	19
Uzbekistan	97,359	76,050	78	21,309	22
Moldova	92,753	54,558	59	38,195	41
Georgia	74,147	58,229	79	15,918	21
Armenia	59,665	45,211	76	14,454	24
USA	54,909	33,692	61	21,217	39
Israel	49,304	29,947	61	19,357	39
Estonia	47,095	24,532	52	22,563	48
Latvia	46,231	23,037	50	23,194	50
Turkey	37,458	31,592	84	5,866	16
Lithuania	28,732	15,836	55	12,896	45
England	24,137	15,456	64	8,681	36
Italy	20,694	12,158	59	8,536	41
Spain	19,276	12,305	64	6,971	36
France	13,059	8,736	67	4,323	33

4.2 Survey

The goal of the survey is to examine privacy awareness among different social groups and to compare the statistical results to studies made in the past. Specifically, we adapted four surveys presented in [10, 1, 23, 25] and compiled them into one questionnaire. Several Facebook-related questions were changed to address arbitrary social networks used by respondents. The respondents were also asked to mark what social networks they use which included Odnoklassniki[11], Facebook and MySpace in addition to the five social networks considered in this paper. The survey was conducted in the Russian language on-line and at the faculty of Biology in the Tula State Pedagogical University during one week period on March 2010. The survey was taken only by persons who are members of at least one social network.

50 people (20 men and 30 women) of different age (Mean 30.39; Std. 9.48) participated in the on-line survey (hereafter referred to as *Global*) in which we aimed to examine

[11]http://www.odnoklassniki.ru

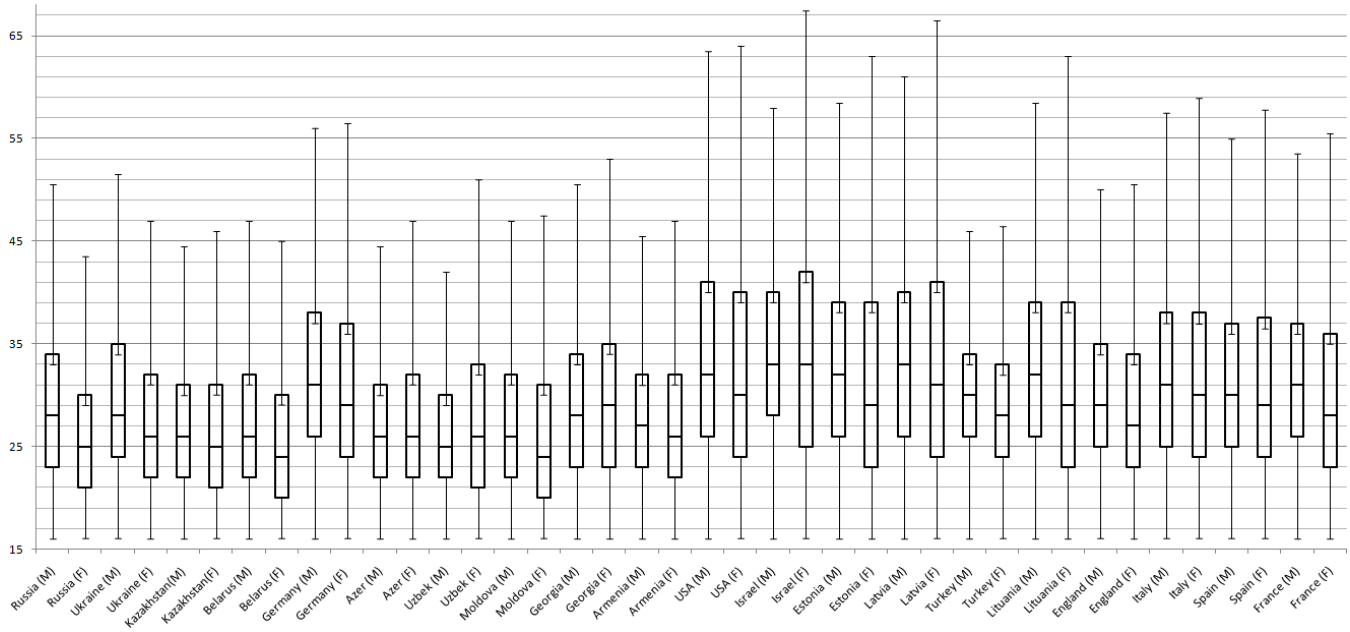

Figure 1: Age distribution

Table 3: Percentage of disclosed information distributed by gender among the 20 most active countries

Country	Male (intimate) (%)	Female (intimate) (%)	Male (type) (%)	Female (type) (%)	Male (status) (%)	Female (status) (%)
Russia	46.9	26.2	97.3	97.5	57.0	46.9
Ukraine	47.0	36.3	96.7	96.2	58.5	64.1
Kazakhstan	39.8	26.8	97.4	97.1	52.6	60.5
Belarus	51.0	38.8	97.4	97.0	63.9	69.5
Germany	45.1	32.6	96.1	95.3	66.2	69.5
Azerbaijan	25.7	15.7	97.6	96.5	35.2	43.2
Uzbekistan	29.4	23.6	98.0	97.6	41.8	56.1
Moldova	38.7	28.6	97.0	96.4	51.9	61.7
Georgia	27.2	17.4	82.5	83.0	36.2	44.4
Armenia	29.5	15.5	97.1	96.6	40.6	40.9
USA	42.7	29.8	96.6	96.3	61.0	62.1
Israel	55.5	36.7	97.3	97.2	71.9	67.5
Estonia	47.7	35.4	95.1	95.3	60.6	62.9
Latvia	45.6	33.8	96.4	95.7	59.4	64.8
Turkey	34.2	21.9	97.0	95.7	46.1	49.0
Lithuania	44.0	32.3	96.5	96.1	60.3	64.9
England	44.3	29.3	96.7	97.0	62.2	57.8
Italy	46.0	31.8	96.8	96.3	59.6	64.0
Spain	40.6	35.0	96.3	96.5	54.1	64.1
France	37.8	26.0	96.0	96.2	55.6	54.3

people from different sectors of society, country of residents (USA, Russia, Ukraine, Germany, Israel), ages and profession (housewives, doctors, students, teachers, programmers, etc.) having only the Russian language as the common denominator. 36 female students (Mean 19.64; Std. 1.58) from the Biology department (hereafter referred to as *Local*) participated in the survey conducted in classes where the aim was to examine students of the same gender and small deviation in age without solid background in computers.

Figure 2 shows the proportion of respondents using social network sites. Vkontakte is used by 91.67% of the Local group compared to only 52.78% of the Global group, which is not surprising since, as mentioned above, Vkontakte is a Russian version of Facebook targeting the student population. However, Facebook is used by 18% of the Global group whereas only by 2.78% of the Local group. This difference is explained by the diversity of respondents' place of residence or profession (in case of the Global group). Odnoklassniki (76% Global, 52.78% Local) and MoyMir (40% Global and 50% Local) are popular among two groups while the dating site Love.Mail.Ru attracts more people from the Global group (18%).

Respondents were asked to mark the personal information they disclose on their profiles. Figure 3 shows the percentage of disclosed information by the Global and Local groups.

Birthday is the most often disclosed piece of information among two groups: 98.00% (Global) and 97.22% (Local). Previous research showed a smaller percentage of disclosed birthdays: about 87% in [10], about 72% in [23], 84% in [1] and 92.2% in [25]. The statistics from the real data collected by us is as follows: 39.39% of MoyMir users, 77.5% of MirTesen users, 99.97% of Loveplanet users and 100% of Love.Mail.Ru users reveal their birthday information.

86% from the Global group and 97.2% of the Local group reveal their true *first* and/or *last names* in contrast to about 87% [23], 94.9% (Facebook), 62.8% (MySpace) [24] and 99.35% [25]. From the data collected we could not clearly determine how many people reveal their true names. However, we observed that many users reveal their names instead of nicknames. We tried to estimate the upper boundary of possibly true names that contain first and last names by counting how many entries are composed from two parts separated by space and do not include non-ascii characters: 27.8% (Love.Mail.Ru), 79.56% (MoyMir), 71.95% (MirTesen), 0% (Loveplanet). The number of only first names is an order of magnitude higher.

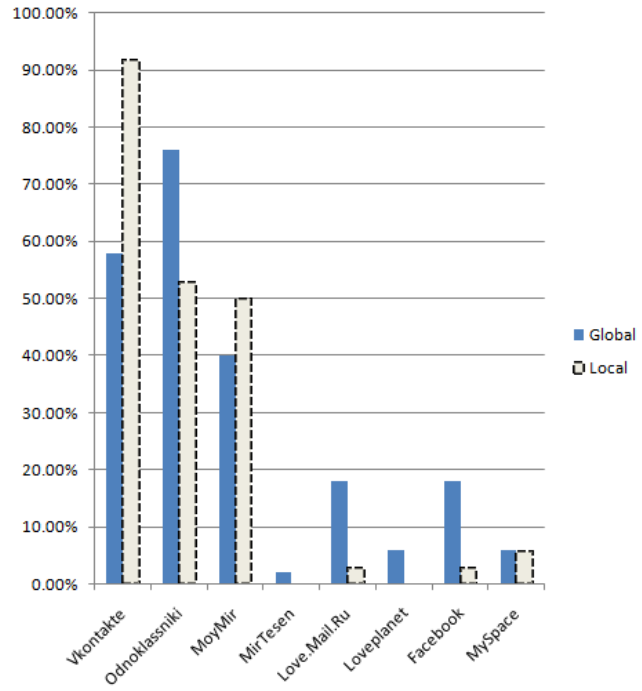

Figure 2: SNS usage. Global (filled), Local (dashed)

Images of the users themselves are posted by 82% (Global) and 94.4% (Local) in contrast to 75% [23], 86.8% [16] and 98.7% [25]. 61.1% of MirTesen users provide their images in their profiles.

Current address is revealed by 12% (Global) and 22.22% (Local) while about 10% in [10], about 65% in [23], 24% in [1], 7.9% females and about 20% males in [25]. The place of residence (country and city) is revealed by 35.35% of MoyMir users and 100% of Love.Mail.Ru users.

The following is a brief comparison of percentage of disclosure:
political views: 36% Global, 41.67% Local, approx. 65% [10], approx. 60% [23], 15% females and 55% males [25].
sexual orientation: 12.00% Global, 2.78% Local, approx. 38% [23], 59% [1], approx. 75% [25]. 31.47% of Love.Mail.Ru users and 100% of Loveplanet[12] users reveal their sexual orientation.
email: 38.00% Global, 36.11% Local, approx. 90% [10], approx. 82% [23], approx. 85% [25].
mobile phone: 12.00% Global, 33.33% Local, approx. 27% [10], 39% [1], approx. 10% [25].
interests: 60.00% Global, 75.00% Local, approx. 83% [10], appox. 63% [23], approx. 70% [25].

In addition, [25] reports that 64% of respondents adjusted a visibility of their profiles such that only friends can access it. We provide the statistics mined from the real data: 15% of Vkontakte users leave their profiles and friends list open, 3.21% adjust their profile visibility but leave friends list open, and 60% leave their profiles visible, while 0.49% of MirTesen users adjust visibility of their friends list in such a way that only friends can see it, 0.95% hide their friend list completely and 98.56% leave the friends list open.

[12]Sexual orientation is a mandatory field for registration at Lovelplanet

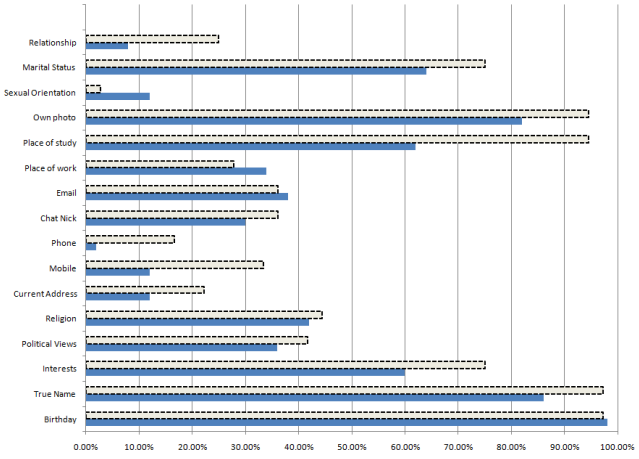

Figure 3: Personal information disclosed by respondents. Global (filled), Local (dashed)

Except for some inconsistencies in the results of past studies, the Russian audience seems to reveal less sensitive information than their western counterparts, which is supported by the survey results and statistics from the real data. Some results are indeed very surprising. For example, the number of Facebook students who report that they reveal their sexual orientation is almost twice as high as the percentage of revealed information on the dedicated dating sites. MoyMir and MirTesen does not contain fields for entering intimate information at all.

To understand why people disclose so much information on their profile, for the study in [23] Facebook students were interviewed about their opinions on different privacy aspects. Table 4 summarizes the results obtained by our survey and provides comparison to the average responses obtained by [23]. In line with [23] we used the five point Likert scale. We applied Mann-Whitney U test to check if there are differences between Global and Local responses. We found a difference between the Global and Local group in the case of profile accessibility by strangers. The Local group is strongly negative about it, while the Global group is more positive (above the average response). However, there is an agreement between both groups in all other questions. In particular, both groups would like to share less information in the future and do not believe that their information is well protected while respondents in [23] are more positive about it.

In past studies it was confirmed that there is no relation between the level of concern in general among the users of social networks and the amount and type of the information the users disclose [10, 1]. High level of trust of users in the social network or their friends was another possible reason for such behavior. Table 5 summarizes average responses on trust attitudes using the Likert seven point scale and compares them to the average responses obtained by [1]. It should be noted that questions asked in [1] were related to Facebook, while we asked about trust in social networks in general. We applied Mann-Whitney U test to check if there is difference between Global and Local responses and obtained no statistical significance at 5% significance level between responses of the Global and Local groups. Accord-

Table 4: Statements related to information disclosure. Average response using Likert five point scale (1 - strong disagreement, 5 - strong agreement). Comparison of Global and Local survey results to Stutzman [23]. "*" indicates statistical significance between Global and Local at the 0.05 level

Statement	Global (Avg.Resp)	Local (Avg.Resp)	Stutzman
I am ok with friends accessing my profile	4.39	4.64	4.55
I am ok with family accessing my profile	4.16	4.05	3.78
I am ok with classmates accessing my profile	4.16	4.33	3.76
I am ok with strangers accessing my profile	*2.86	2.05	3.15
It is important to me to protect my identity information	3.82	4.11	4.21
I an concerned with the consequences of sharing identity information	3.61	3.83	3.29
I am likely to share my identity information online in the future	2.37	2.33	3.34
I believe my identity information is well-protected online	1.98	2.33	2.66

ing to the results, three groups have a comparable level of trust for the social network they use and their friends while they trust less in users not connected to them. Different hypotheses were raised to explain the discrepancy between privacy concerns and information revelation such as peer pressure[1, 6], unawareness of the true visibility of the users' profiles, website interface design [6] or by drawing boundary lines between virtual and real world. It seems that the answer is more prosaic and can be explained by the following metaphoric example:
We are aware of dangers related to driving cars, but we continue to do it since we do not have any other choice and because of the assumption that "nothing bad will happen to me".

Table 5: Trust statements. Average response using Likert seven point scale (1 - strong disagreement, 7 - strong agreement). Comparison of Global and Local average response to Acquisti & Gross [1]

Statement	Global (Avg.Resp)	Local (Avg.Resp)	Acquisti & Gross
The Social Network Site that I am using	4.33	4.94	4.20
Your own friends on SNS	6.02	6.55	5.62
Friends of your friends	4.18	4.47	4.35
On average, SNS users not connected to you	3.14	3.28	3.29

Comparison with the results obtained by [25] reveal interesting patterns of concern about unwanted audiences (see Table 6). According to [25], unwanted audiences are those individuals who are not linked to the SNS user, but who may gain access to the user's profile without his/her knowledge or consent. Both Global and Local groups do not agree or believe that future employers will use the personal information to assess the user's suitability with the company. The results of [25] show that there is concern about this problem. Likewise, it seems unrealistic for both groups that universities or political parties can monitor their profiles and use it for identification of possible illegal activities (universities) or political reasons (parties). This concern is much higher in [25]. There is statistical significance between Local and Global groups on the issue of sexual predators. The Local group is concerned very much and believes that sexual predators use SNS to track potential victims, while the Global group is more conservative on this issue. However, the Global group is highly concerned about being monitored by police and there is a statistical significance between the Global and Local groups on this issue.

Table 6: Concern about access by unwanted audiences. Average response using Likert five point scale (1 - strong disagreement, 5 - strong agreement). Comparison of Global and Local average response to Young & Quan-Haase [25]. "*" indicates statistical significance between Global and Local at the 0.05 level

Statement	Global (Avg.Resp)	Local (Avg.Resp)	Young and Quan-Haase
Future employers will use the personal information contained in my profile to assess my suitability with their company	2.39	2.38	3.15
Universities are monitoring SNS postings, personal info and images to identify involvement in illegal activities	2.12	1.89	3.05
Police/Militia are using SNSs to track illegal activities	*3.49	2.72	2.98
Sexual predators use SNSs to locate potential victims	2.76	*3.44	3.57
Political parties have begun using SNSs to target young professionals and students through the use of advertisements and data mining	2.25	2.72	3.66

The study of privacy concern among students who use Facebook revealed three key types of concerns [25]: (1) the information from the profile can be used by unknown users (2) concerns of data mining and the fact that the company owns every piece of data about the user (3) providing too much information (through wall messages or photos) to unknown people outside of the network. According to [25] the users' privacy concern was based mainly on the bad media coverage and not on their own experience. In our survey, we revealed that most users who have concerns about privacy revelation have experienced something before. The typical experiences in two groups are: (1) receiving spam, (2) receiving unwanted messages originated from friends, but sent by the intruder (3) Internet bullying by unknown people. Only one respondent gave an example from the media that happened in 2008 to Yevgeny Chichvarkin, the founder of the largest Russian mobile phone company Yevroset, who considered to sue Odnoklassniki for hosting several fake accounts bearing his name, photos and sending messages on his behalf[13].

As a response to privacy concerns, users developed their own protection strategies such as blocking non-friends from

[13]http://www.polit.ru/news/2008/01/10/odnoklassniki.html

contacting the user, falsifying parts of the personal information or removing personal images from their profile. Table 7 shows several protection strategies and comparison to the results obtained by [25]. The Global group resorts more to information falsification than the Local group or Facebook users. This can be explained by the fact that most of the Runet SNSs do not provide adequate means of access control or lack these features completely. There is no statistical difference between the Global and Local groups when it comes to the question of excluding personal information to restrict information collection by unknown people. This strategy is more popular among Facebook users. A popular strategy in the Local group is to limit access to certain contacts by changing default privacy settings. According to the average scale, we can say, that all users who change default privacy settings also change the accessibility level. This is valid for both Local and Global groups, although the average response in the Global group is statistically lower, which again can be explained by the fact that most of the SNSs used by the members of the Global group do not have fine grained control on the level of accessibility. According to the average response of Facebook users in [25], there are more users who change default settings rather than users who limit accessibility of their profiles.

Table 7: Privacy protection strategies. Average response using Likert five point scale (1 - strong disagreement, 5 - strong agreement). Comparison of Global and Local survey results to Young & Quan-Haase [25]. "*" indicates statistical significance between Global and Local at 0.05 level

Statement	**Global** (Avg.Resp)	**Local** (Avg.Resp)	**Young** and **Quan-Haase**
I provided fake information to restrict unknown people from gaining information about me	2.18	1.97	1.66
I excluded personal information to restrict unknown people from gaining information about me	2.80	3.0	4.08
Certain contacts only have access to my limited profile	2.74	*4.03	3.47
I changed default privacy settings activated by SNS	2.76	*4.05	4.33

5. CONCLUSIONS

In this paper, we analyzed privacy issues in five Runet social networks: *Vkontakte, MoyMir, MirTesen, Love.Mail.Ru*, *Loveplanet*. The limitation of past studies about social networks is due to several aspects: (1) Statistics were acquired solely using surveys, (2) a lack of real data, and (3) a bias towards Facebook and scarce studies on other social networks.

First, we discussed security and privacy issues of SNSs in Runet that allowed us to download 30 million profiles including almost the complete population from Love.Mail.Ru SNS. The latter facilitated us to gain insight into demographic statistics and understand some trends in revelation of intimate information according to age, gender and country. Second, we conducted a survey among Russian speaking users of social networks from different social groups, professions and gender, and evaluated their privacy concerns, level of trust and amount of personal information disclosed. We performed comparison of our results to the results of previous studies. Although, the surveys have limitation in the generalization, our statistical results are supported by the statistics mined from the data, whereas the limitation of previous studies was due to non-representable population samples (usually limited amount of students at one university) and a lack of real representative data. Our findings suggest that the Runet audience is aware of privacy issues and more concerned about privacy implications than previous studies about western social networking sites suggest. We also showed that the Runet audience discloses less private information than their western counterparts. However, by observing security and privacy measures of SNSs, we can conclude that most of Runet social networking sites do not implement adequate security measures that can prevent automatic profile crawling. All social networking sites except for Vkontakte do not implement adequate privacy measures (and probably do not make users aware of these measures) to restrict access to their profiles by unwanted audiences. In addition, several sites even force their users to uncover some personal information before they can communicate with other peers. Moreover, social networking sites have different number of categories for information disclosure which can lead to disclosure of more personal information on the web sites where more fields are available for filling.

One of the possible recommendations would be to implement some privacy regulations similar to those already established in other critical areas like medicine (HIPAA regulations[14]) to impose the restriction of types of information SNSs may request from their members. Thus, the combination of the amount of disclosed sensitive information and inadequate security measures provide an easy way to collect vast information about users, which can lead to different privacy implications like stalking, identity theft, creation of digital dossiers, spam, etc. Overall, the current privacy issues in Runet social networking sites can be considered as catastrophic.

Due to the space limitation and scope of the paper, we could not conduct all-embracing analysis of privacy issues in Runet. Nevertheless, this paper can be considered as a starting point towards a better understanding of the interrelation between social networks, users, their willingness to share information and the implications on privacy in Russian segment of the Internet.

6. ACKNOWLEDGMENTS

This work was partially supported by DFG Research Training Group GK-1042 "Explorative Analysis and Visualization of Large Information Spaces".
The authors would like to thank Prof. Anna Korotkova from Tula State Pedagogical University for help in conducting the survey.

7. REFERENCES

[1] A. Acquisti and R. Gross. Imagined communities: Awareness, information sharing, and privacy on the Facebook. *Lecture notes in computer science*, 4258:36–58, 2006.

[2] N. Awad and M. Krishnan. The personalization privacy paradox: An empirical evaluation of

[14]Health Insurance Portability and Accountability Act

information transparency and the willingness to be profiled online for personalization. *MIS Quarterly*, 30(1):13–28, 2006.

[3] M. Back, J. Stopfer, S. Vazire, S. Gaddis, S. Schmukle, B. Egloff, and S. Gosling. Facebook Profiles Reflect Actual Personality, Not Self-Idealization. *Psychological Science*, 21(3):372, 2010.

[4] J. Bonneau and S. Preibusch. The Privacy Jungle: On the Market for Data Protection in Social Networks. In *The Eighth Workshop on the Economics of Information Security (WEIS 2009)*, 2009.

[5] D. Chau, S. Pandit, S. Wang, and C. Faloutsos. Parallel crawling for online social networks. In *Proceedings of the 16th international conference on World Wide Web*, pages 1283–1284. ACM, 2007.

[6] Z. De Souza and G. Dick. Disclosure of information by children in social networking – Not just a case of "you show me yours and I'll show you mine". *International Journal of Information Management*, 2009.

[7] C. Dwyer, S. Hiltz, and K. Passerini. Trust and privacy concern within social networking sites: A comparison of Facebook and MySpace. In *Proceedings of AMCIS*, 2007.

[8] A. Felt and D. Evans. Privacy protection for social networking platforms. In *Web 2.0 Security and Privacy Workshop*, 2008.

[9] L. Gelman. Privacy, Free Speech, and "Blurry-Edged" Social Networks. *Boston College Law Review*, 50(5), 2009.

[10] T. Govani and H. Pashley. Student awareness of the privacy implications when using facebook. (unpublished), 2005.

[11] J. Grimmelmann. Saving Facebook. *Iowa Law Review*, 94:1137, 2008.

[12] R. Gross, A. Acquisti, and H. J. Heinz, III. Information revelation and privacy in online social networks. In *WPES '05: Proceedings of the 2005 ACM workshop on Privacy in the electronic society*, pages 71–80, New York, NY, USA, 2005. ACM.

[13] R. Heatherly, M. Kantarcioglu, B. Thuraisingham, and J. Lindamood. Preventing private information inference attacks on social networks. *University of Texas at Dallas, Tech. Rep. UTDCS-03-09*, 2009.

[14] G. Hogben. Security issues and recommendations for online social networks. *Position Paper. ENISA, European Network and Information Security Agency*, 2007.

[15] H. Jones and J. Soltren. Facebook: Threats to privacy. *Project MAC: MIT Project on Mathematics and Computing*, 2005.

[16] E. Kolek and D. Saunders. Online disclosure: An empirical examination of undergraduate Facebook profiles. *NASPA Journal*, 45(1):1–25, 2008.

[17] B. Krishnamurthy and C. Wills. Characterizing privacy in online social networks. In *Proceedings of the first workshop on Online social networks*, pages 37–42. ACM, 2008.

[18] A. Mislove, M. Marcon, K. Gummadi, P. Druschel, and B. Bhattacharjee. Measurement and analysis of online social networks. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, pages 29–42. ACM, 2007.

[19] A. Mislove, B. Viswanath, K. Gummadi, and P. Druschel. You are who you know: inferring user profiles in online social networks. In *Proceedings of the third ACM international conference on Web search and data mining*, pages 251–260. ACM, 2010.

[20] F. Nagle and L. Singh. Can friends be trusted? Exploring privacy in online social networks. In *Proceedings of the 2009 International Conference on Advances in Social Network Analysis and Mining*, pages 312–315. IEEE Computer Society, 2009.

[21] U. Pfeil, A. Raj, and Z. Panayiotis. Age differences in online social networking - a study of user profiles and the social capital divide among teenagers and older users in myspace. *Computers in Human Behavior*, 25:643–654, 2009.

[22] K. Strater and H. R. Lipford. Strategies and struggles with privacy in an online social networking community. In *BCS-HCI '08: Proceedings of the 22nd British HCI Group Annual Conference on HCI 2008*, pages 111–119, 2008.

[23] F. Stutzman. An evaluation of identity-sharing behavior in social network communities. *International Digital and Media Arts Journal*, 3(1):10–18, 2006.

[24] Z. Tufekci. Can you see me now? Audience and disclosure regulation in online social network sites. *Bulletin of Science, Technology & Society*, 28(1):20–36, 2008.

[25] A. Young and A. Quan-Haase. Information revelation and internet privacy concerns on social network sites: a case study of facebook. In *Proceedings of the fourth international conference on Communities and technologies*, pages 265–274. ACM New York, NY, USA, 2009.

[26] E. Zheleva and L. Getoor. To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles. In *Proceedings of the 18th international conference on World wide web*, pages 531–540, 2009.

[27] B. Zhou and J. Pei. Preserving privacy in social networks against neighborhood attacks. In *Proceedings of the 24th IEEE International Conference on Data Engineering (ICDEŠ08)*, pages 506–515, 2008.