Journal of Service Science Research (2012) 4:175-212
DOI 10.1007/s12927-012-0008-z

Fake Identities in Social Media: A Case Study on the Sustainability of the Facebook Business Model

Katharina Krombholz, Dieter Merkl, Edgar Weippl

Received: 26 July 2012 / Accepted: 17 November 2012 / Published: 31 December 2012

ABSTRACT

Social networks such as Facebook, Twitter and Google+ have attracted millions of users in the last years. One of the most widely used social networks, Facebook, recently had an initial public offering (IPO) in May 2012, which was among the biggest in Internet technology. For-profit and nonprofit organizations primarily use such platforms for target-oriented advertising and large-scale marketing campaigns. Social networks have attracted worldwide attention because of their potential to address millions of users and possible future customers. The potential of social networks is often misused by malicious users who extract sensitive private information of unaware users. One of the most common ways of performing a large-scale data harvesting attack is the use of fake profiles, where malicious users present themselves in profiles impersonating fictitious or real persons. The main goal of this research is to evaluate the implications of fake user profiles on Facebook. To do so, we established a comprehensive data harvesting attack, the social engineering experiment, and analyzed the interactions between fake profiles and regular users to eventually undermine the Facebook business model. Furthermore, privacy considerations are analyzed using focus groups. As a result of our work, we provided a set of countermeasures to increase the awareness of users.

KEYWORDS

Social Media, Facebook, Privacy, Social network analysis, Fake profiles, Sustainability, Reliable data

Katharina Krombholz
SBA Research
e-mail: kkrombholz@sba-research.org
Dieter Merkl
Vienna University of Technology
e-mail: dieter.merkl@ec.tuwien.ac.at
Edgar Weippl (✉)
SBA Research
e-mail: eweippl@sba-research.org

1. INTRODUCTION

In recent years, online social networks such as Facebook, Twitter an Google+ have become a global mass phenomenon and one of the fastest emerging e-services according to Gross and Acquisti (2005) and Boyd and Ellison (2007). A study recently published by Facebook (2012) indicates that there were about 901 million monthly active users on the platform at the end of March 2012. Therefore, Facebook is one of the largest online social networks. Not only common users but also celebrities, politicians and other people of public interest use social media to spread content to others. Furthermore, companies and organizations consider social media sites the medium of choice for large-scale marketing and target-oriented advertising campaigns.

The sustainability of the business model relies on several different factors and is usually not publicly disclosed. Nonetheless, we assume that two major aspects are significant for Facebook. First and foremost, Facebook relies on people using their real-life identity and therefore discourages the use of pseudonyms. Verified accounts allow (prominent) users to verify their identity and to continue using pseudonyms, e.g., stage names such as ‘Gaga’. This is considered to be a security mechanism against fake accounts (TechCrunch 2012); moreover, users are asked to identify friends who do not use their real names.

Second, the revenue generated by advertising is substantial and thus protection of the revenue streams is important. Media reports (TechCrunch 2012) have indicated that a large portion of clicks are not genuine clicks by real users. In the short term, Facebook does not suffer from bot-click attacks but instead benefits from increased revenue, so there may be no incentive to prevent such fraud. In the long run, however, advertisers will move away from the platform if the promised targeted advertisements are not delivered correctly.

Amongst users, social media are widely regarded as an opportunity for self-presentation and interaction with other participants around the globe. Due to the wide circulation and growing popularity of social media sites, even for-profit organizations, such as companies, and non-profit organizations have gained interest in presenting themselves and reaching potential customers. A presence on Facebook, Twitter etc. is nearly taken for granted. The social media site operators have a huge amount of personal data and other shared content such as links, photos and videos stored on their servers. Many individuals and organizations use social media sites to access and gather information about other users. Not only has user privacy become an issue due to complex data-accessing models as stated by King et al. (2011), the reliability and sustainability of Facebook user data have also become subjects of interest for for-profit organizations that rely on this data, as examined by Krasnova et al. (2009) and Wang et al. (2011). One of the major issues concerning data acquisition in online social networks is the problem of fake user data or even entirely fake profiles. Reasons for providing fake user data are usually a result of privacy enhancement strategies due to conflicting privacy configurations and data protection policies caused by the platform. While many Facebook users provide partially fake data in their user profiles, some profiles do not even represent a person who exists in real life, as discussed by Gao et al. (2010). Such profiles are widely used for malicious attacks on users' privacy, as Boshmaf et al. (2011) and Bilge et al. (2009) have shown, e.g., data harvesting campaigns conducted by botnets. Sometimes they are even used to create an artificial audience to review products or to make a business grow by making it popular through many profiles, or even to spread opinions and ideologies. According to Facebook, 5% to 6% of registered Facebook accounts are fake accounts. Facebook clearly states in their *Legal Terms* that users are not allowed to provide fake information and that they must keep their information up to date (Legal Terms 2012). This clearly indicates that the accuracy and correctness of Facebook user data is important for Facebook's business model. Inaccurate or false information endangers the sustainability of the Facebook business model. On the other hand, the Facebook platform attracts companies to use the platform for advertising and marketing by offering them a high number of users who are easy to access. Using methods for detecting and eliminating fake profiles would lead to a reduction in registered accounts, raising the risk of eliminating false positives and making it more difficult for regular users to create and maintain a profile. It was also considered helpful for a successful IPO in May 2012 to have as many accounts as possible.

Organizations, the service platform, and the individual users have different interests concerning data access and sharing. Whereas the platform and third parties, such as for-profit organizations that use the platform for business, are interested in gathering as much user data as possible, the users mostly do not want to share their personal data with them. However, as the data-sharing model of Web 2.0 services differs from traditional Web applications, users are often unaware who they are sharing their information with. Facebook provides data access control and privacy regulations to protect its users' privacy. However, in many cases the user is not sufficiently protected and private information is leaked

In our research, we conducted a social engineering experiment on Facebook using fake profiles. Furthermore, we held focus groups and a survey with an overlapping sample. The purpose of this work is to describe and discuss privacy-related issues in the context of social media and discuss the reliability and sustainability of Facebook user data. Our main contributions are the following:

- We created socially attractive fake Facebook profiles and integrated them into existing friendship networks to simulate a data harvesting attack.
- We analyzed the Facebook user data of profiles that interacted with our fake profiles.
- We analyzed human factors to get a deeper understanding of the procedure of successfully integrating a fake profile into an existing friendship network.
- We provided countermeasures by raising user awareness.

The remainder of this paper is structured as follows: In section 2 we provide background information on social media usage and related work on privacy violations and data harvesting. In sections 3 and 4, we discuss our research methods, namely the Facebook social engineering experiment and the focus groups, and furthermore provide a detailed analysis and discussion of the results. Finally, we conclude the paper with a review of our findings and a look at further research.

2. RELATED WORK AND BACKGROUND

Boyd and Ellison (2007) define social network sites as "web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection and (3) view and traverse their list of connections and those made by others within the system."

Furthermore they argue that there are different approaches concerning the network principles on different sites. Some sites focus on the visualization of established real-life social networks, while others try to facilitate the forming of new relationships. Social networks provide not only tools for communication amongst the users, but also functionality for Life Logging. Smith et al. (2011) defined LifeLogging as “the collection of data in order to illustrate a person’s life.”

In general, a social network can also be modeled as a graph as proposed by Asuncion et al. (2010), Narayanan et al. (2009), Puttaswamy et al. (2009) and Zheleva et al. (2009). Hence, we define $G = (V, E)$, where V is a set of n nodes. Each participant of a social network, namely a user account, is represented by a node. E is a set of edges. An edge $e_{i,j}$ connects two nodes v_i, v_j. If two nodes are adjacent, this means that there is a social connection established between the users modeled by the nodes.

In case of Facebook, where user connections are bidirectional, the edges $e_{i,j}$ are undirected. In other social networks, such as, e.g., Twitter or Google+, Graph G can be modeled as a directed graph as the user connections are not necessarily bidirectional. Agichtein et al. (2008) described a paradigm shift from Web users as being consumers of content to producers of content in the early 2005. User-generated content is one of the key characteristics of social media. The content as well as the community and the disclosed personal information are of high value for the company behind the platform, as they can be used for target-based advertising and product placement. Excessive disclosure and privacy are highly correlated, generating an area of conflicts. Privacy in general is hard to measure according to Stutzman et al. (2010, 2011), Liu et al. (2011), Besmer et al. (2010), Strater et al. (2008) and Gross and Acquisti (2005). In their work, they have discussed various aspects and characteristics of user privacy such as the relationship between perceived and actual privacy settings. Stutzman et al. (2010, 2011) discussed factors that influence the correlation between privacy attitudes and disclosure behaviors on Facebook and examined the baseline relationship between privacy engagements and disclosure practices. Liu et al. (2011) have discussed the difficulty of measuring privacy. According to them, boundary regulation and the perception of audiences is hard and expectations do not always represent reality. Besmer et al. (2010) discussed privacy as a tool for boundary management, based on the privacy definition by Altman (1975). Stanton et al. (2003) argue that boundary management always means the restricttion of an audience. Baden et al. (2009) have engaged in designing an online social network with user-defined privacy instead of platform-defined privacy as it is the case on Facebook. Asuncion et al. (2010), Narayanan et al. (2009), Puttaswamy et al. (2009) and Zheleva et al. (2009) proposed and evaluated successful attacks on user privacy in social media. They applied several algorithms on a graph representing a social network. Asuncion et al. (2010) presented an attack scenario based on group testing that had previously been introduced by Dorfman (1943) to efficiently test blood samples. Narayanan et al. (2009) discussed cross-networking identification attacks. Puttaswamy et al. (2009) described social intersection attacks as an effective, low-cost privacy attack to reveal private information. They furthermore proposed anonymization techniques to protect users from attackers based on k-anonymity (Sun 2011) and introduced a new graph structure to provide k-anonymity, edge minimality and latent edges. Zheleva et al. (2009) examined privacy attacks using links and groups within a social network to determine undisclosed attributes.

As social network sites hoard an enormous amount of personal data, they are also a potential target of illegal data collectors and profilers such as socialbots, which are spam bots that exploit social networks to harvest user data. Gao et al. (2010), Bilge et al. (2009) and Boshmaf et al. (2011) show that automated data collection techniques such as socialbots can be used successfully to harvest data on social media platforms. Boshmaf et al. (2011) found that open social networks such as Facebook are highly vulnerable to large-scale infiltration. The architecture of a socialbot network consists of a botherder, a botmaster and a variable number of socialbots. The botherder controls the automation software that manipulates the operation of the socialbot network. The botmaster is connected to the social network through a command and control channel. Not only is the process easy to automate, they also found that users are not careful enough. They point out several vulnerabilities in social networks. Bilge et al. (2009) have performed rather successful profile cloning and cross-site profile cloning attacks and found that these automated attacks are very effective due to the fact that the general profile information was cloned from profiles of real Facebook users. The associated assumption is that people accept friend requests from someone they know more readily. The vulnerable point in profile cloning is that the full name and the profile pictures in Facebook are not unique. The only key value that has to be unique is the e-mail address. As there are several e-mail providers that make it easy to create multiple e-mail addresses (e.g., mail.ru), it is rather easy to just clone the name and the profile picture from another account and create a fake account with another email address. Another cloning social attack they tried is cross-site profile cloning. In cross-site profile cloning, the profile information of participants on one social media platform is copied and a new profile with the cloned information is created in another social network, where the person who's information is being used has no account. They also point out the potential danger for users of social networks arising from the fact that there is not as much awareness for social spam as for e-mail spam, as it is a relatively new form of spam. Furthermore, the available study claims that the already implementted Facebook Immune System (Stein 2011) is not sufficient for detecting large-scale social spam campaigns. Nevertheless, data can be leaked not only through targeted attacks-sometimes even the users themselves leak sensitive information, as discussed by King et al. (2011) and Mao et al. (2011). Third parties, such as applications hosted by companies other than the platform operator also leak private information by violating the Facebook privacy settings configured by the user, as discussed by Gross and Acquisti (2005) and Lipford et al. (2008). An example of such a privacy leak caused by a third-party application has been shown by Wang et al. (2011). They examined a third-party application that discloses a user's friends' birthdays. Even though some users did not disclose their birthday publicly on their profiles, the third-party application made the entire birthdate publicly viewable within the application, which is in conflict with privacy.

3. THE FACEBOOK SOCIAL ENGINEERING EXPERIMENT

3.1 Framework

Researchers have found that socialbots are a successful method for harvesting user data from social network sites as determined by Boshmaf et al. (2011) and Bilge et al. (2009). Furthermore, an enormous amount of privacy leaks is actually caused by the users themselves, according to Mao et al. (2011), Wang et al. (2011), Strater et al. (2008) and Stutzman et al. (2011). Therefore, we conducted a social engineering experiment on Facebook and collected user data as well as qualitative data from observing the inter-user interactions. The basic idea was to construct fake Facebook profiles and establish friendship connections with as many

Facebook users as possible and collect all the data they revealed. To perform large-scale data harvesting attacks in a social network, the task has to be automated. According to Boshmaf et al. (2011) and Bilge et al. (2009), it is easy to automate a large-scale infiltrating socialbot network to harvest data from social networks such as Facebook. As we achieve a deeper understanding of how users react to and interact with such fake users profiles, we want to perform only actions that can be automated.

3.2 Methodology

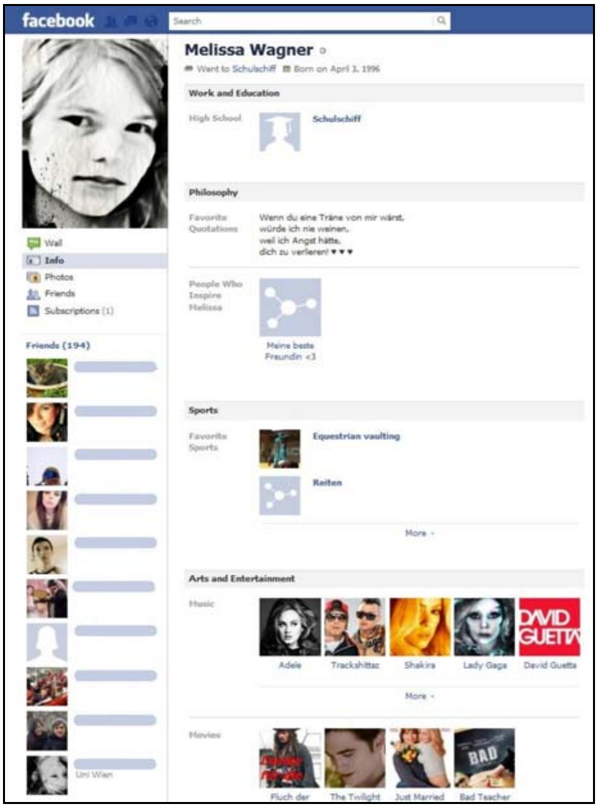
Initially, we created six fake Facebook profiles in three different age groups, half of them female (*Melissa, Laura, Ilse*) and half of them male (*David, Chris, Ferdinand*). Furthermore, we created another profile that showed a cartoon cat (*Mitzi*) instead of a human being and disclosed no gender information at all. To generate social interactions, we introduced another fake female profile representing a teenage girl (*Laura*) that *friended* all of the other fake profiles from our experiment. In contrast to the profiles used in other socialbot experiments (e.g., by Gao et al. 2010, Boshmaf et al. 2011, Bilge et al. 2009), which tried to make the profiles as simple as possible, we created realistic and complex profiles to guarantee a high social attractiveness. To do so, we examined the Barracuda Labs social networking analysis (2011). Their research project aimed to analyze the differences between fake and real Facebook profiles and compared a selection of features to point out the differences. In order to create as socially attractive and unsuspicious profiles as possible, we referred to the findings of the Barracuda Labs social networking analysis. Furthermore, we algorithmically generated profile pictures. The goal was to obtain artificial images that did not represent to a specific real person but that nevertheless had features of real faces. To this end, we applied image transformation algorithms recursively to average faces derived from a set of input faces. To reduce artifacts caused by the image transformation algorithms, we applied filters and various visual effects to mask them. The data was obtained from statistical evaluations of data based on the manually selected birth dates. Figure 1 shows Melissa's Facebook profile as an example. Table 1 outlines the characteristics of the fake profiles we designed to perform the experiment.

Table 1. This Table Outlines the Characteristics of the Facebook Fake Profiles We Used to Conduct the Social Engineering Experiment on Facebook. This Table only Contains Publicly Disclosed Information as It is Viewable for Another Facebook User That is Not Friend with the Respective Profile

Name	Lena	Melissa	Laura	Ilse
Gender	Female	Female	Female	Female
Interested in	Men	-	Men	-
Relationship status	-	Single	Men	-
Birthdate	9/23/1995	4/3/1996	4/24/1987	3/1/1972
Hometown	Salzburg (A)	-	Vienna (A)	Unterhilinglah (A)
Current location	Vienna (A)	Vienna (A)	Vienna (A)	Kalsdorf (A)
School	BG Zaunergasse	-	-	-
High school	Gymn. Sacre-Coeur	Schulschiff	-	HAK Eferding
University	-	-	Univ. of Vienna	-
Workplace	Freelancer	-	-	Graz Airport

Name	David	Chris	Ferdinand	Mitzi
Gender	Male	Male	Male	-
Interested in	-	-	-	Single
Relationship status	-	-	-	Single
Birthdate	3/17/1995	3/8/1982	8/14/1972	11/9/1990
Hometown	-	-	Salzburg (A)	Vienna (A)
Current location	Vienna (A)	-	Vienna (A)	-
School	-	HTL Rennweg	-	-
High school	Schulschiff	Schulschiff	-	-
University	-	-	-	-
Workplace	-	Graphic designer	Lawyer	-

The Facebook Social Engineering Experiment was actively conducted between March 12 and April 11, 2012. However, the fake Facebook profiles remained online for another 4 weeks, during which we performed no interactions on their behalf except for accepting incoming friendship requests. Afterwards, we deactivated the accounts to protect the participants' privacy as they contained friendship connections to other profiles. During the active period of the experiment, we maintained a logbook to monitor the activities. On the one hand, we collected numerical data such as the amount of outgoing and incoming friendship requests as well as the total number of friends at every point in time. On the other hand, we annotated qualitative data to describe the events occurring during the observation. Furthermore, at the end of the active implementation of the experiment, we collected Facebook user data via the Facebook Graph API and the social snapshot tool developed by Huber et al. (2011). In the following subsections, we will provide a detailed analysis of the data and, finally, triangulate it. We defined interaction policies to regulate the behavior of the profiles. As Boshmaf et al. (2011) have shown that social engineering on Facebook is easy to automate, we only performed tasks that could in principle be automated to simulate a large-scale infiltration attack. At the start of the experiment, we sent friendship requests to people suggested in the *People You May Know* section on Facebook. After four weeks of perio-dically sending out friendship requests, we stopped our activities and just observed the incoming actions from other users. As the users who friended our profiles were uninten-tionally participating in our study, we deactivated our Facebook accounts at the end of the observation. We anonymized the collected user data by replacing the names with random numbers to protect the participants' privacy.



Melissa Wagner

Went to Schulschiff Born on April 1, 1996

Work and Education

High School: Schulschiff

Philosophy

Favorite Quotations: Wenn du eine Träne von mir wärst,
würde ich nie weinen,
weil ich Angst hätte,
dich zu verlieren! ♥ ♥ ♥

People Who Inspire Melissa: Meine beste Freundin <3

Sports

Favorite Sports: Equestrian vaulting, Reiten

Arts and Entertainment

Music: Adele, Trackshittaz, Shakira, Lady Gaga, David Guetta

Movies: Fluch der, The Twilight, Just Married, Bad Teacher

Figure 1. The Facebook Profile of Melissa

3.3 Analysis of the Quantitative Data

The quantitative data was collected at least once each day from Monday to Friday. No data was collected on the weekends. For every observation point, we monitored the number of outgoing and incoming friend requests as well as the number of new friends. We also monitored the absolute number of friends. This additional information was useful to estimate the number of friends that *unfriended* our fake profiles. Moreover, we recorded the numbers of likes and messages. In general we observed that over time, the number of friends was increasing, as well as the number of incoming friend requests. We ceased performing actions and monitoring the activities on April 11, 2012. Surprisingly, when we collected the Facebook user data, we logged into Facebook with our fake profiles and observed that the number of incoming friend requests was increasing. Table 2 lists the total numbers of friends at certain moments. As stated in the introduction to this section, we conducted the experiment between March 12 and April 11, 2012, but the number of incoming friend requests increased even though we no longer performed any actions. From this we can determine that at a certain point a fake profile that could potentially be misused for data harvesting begins to yield results without any further action. Table 2 shows that the highest numbers of friends belong to fake profiles representing young women. From this we conclude that female fake profiles in general are more likely to be friended on Facebook and, therefore, more likely to be used for malicious data harvesting attacks. Barracuda Labs (2011) showed that 97% of fake profiles on Facebook are female. This fact raises the question whether the friends of the respective fake profiles are of the same or opposite gender. To answer this, we extracted the gender information from the Facebook user data we had gathered via the Facebook Graph API and analyzed it. All of the analyzed friendship networks had profiles that did not disclose gender information; however, these were only a handful. Mitzi, Laura, Ilse and Ferdinand had mostly male friends, where Laura had the highest number of male friends. Melissa and David had more female friends, and in the case of Lena and Chris, the numbers were almost equal for male and female friends. Figure 2 shows the development of Lena's friendship connections. The data was measured at the times indicated on the x-axis and accumulated to visualize and compare the development of the friendship connections.

Table 2. This Table Shows How Many Friends the Fake Profiles Had at Given Times. The Table is Sorted (from Left to Right) by Gender (Female, Male, Undefined) and Age in Ascending Order. The Top 3 Numbers of Friends are Highlighted in Bold Font

Date	Lena	Melissa	Laura	Ilse	David	Chris	Ferdinand	Mitzi
3/26/2012	94	76	55	56	46	53	11	48
4/11/2012	218	178	200	60	63	100	14	75
6/26/2012	**237**	**204**	**272**	66	74	111	20	99

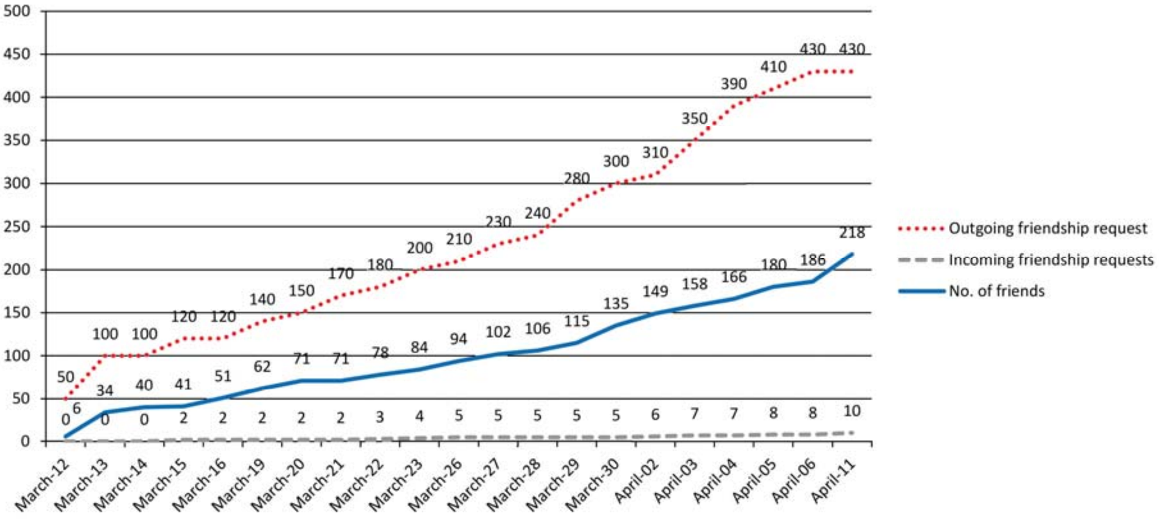


Figure 2. Lena’s Accumulated Incoming and Outgoing Friendship Requests and the Accumulated Number of Facebook Friends Measured Between March 12 and April 11 2012

3.4 Analysis of the Facebook User Data Records

Facebook calls the core structure that represents people and connections between them a social graph. The easiest way to access data from the graph is the Graph API (2012). People, organizations, pages, photos and events are represented as objects with connections between them such as, e.g., friend relationships, shared content and photo tags. Each object has a unique ID and can be accessed through the Open Graph API. All objects are represented as JSON objects (2012). JSON is a slim and hierarchical data format that is easy to parse and process in several programming languages. The social snapshot tool provided by Huber et al. (2011, 2012) uses the Facebook Graph API to collect data. We used the social snapshot tool and our own scripts to retrieve JSON objects from Facebook to analyze the Facebook user data of our fake profiles' friends. We analyzed the datasets in order to learn about the people who became friends with our fake profiles and to find patterns and similarities among them. We wanted to identify factors that supported the establishing of friendships between our fake profiles and other Facebook users. Therefore we analyzed the Facebook user data of our fake profiles' friends with respect to the factors mentioned in the definition of the People You May Know functionality. It must be mentioned that the results obtained by analyzing the Facebook user data is always biased as Facebook users are not forced to disclose all information. Furthermore, it is hard to estimate whether the provided data is correct. This section contains a detailed discussion of the analysis of the collected Facebook user data records based on an experimental observation. To visualize the information of the Facebook user data and support interpretations, we created network graphs for all collected datasets and applied clustering and filtering algorithms for easier discussion of the results. We used the Python programming language (2012) and the NetworkX (2012) package to process and create the graphs. The coloring as well as the layout adjustments were made in Gephi (2012), an open source graph drawing software. The graph layout was arranged according to a force-based algorithm proposed by Fruchterman-Reingold (1991) that comes already implemented in Gephi. We did not label the nodes with names or Facebook IDs as the test subjects were unintentional participants in our study. .

3.4.1 Mutual Friends-Cluster Analysis

The JSON objects retrieved from Facebook include information on friendship connections between the friends of a certain user, who in this case was each of our fake Facebook profiles. For every fake profile, we created a graph to illustrate its friendship network. The nodes represent the profiles of the fake profile's friends and the edges indicate friendship connections between them. We chose to color and size the nodes according to their degree. The degree of a node is determined by the number of edges directly connecting to other nodes within a network. The darker a node is colored and the bigger it is, the higher the degree of the node. In this context, a high degree means higher network density and a higher number of mutual friends. Comparing the resulting graphs, we observed a connection between the size of a friendship network and the network density. Larger graphs appear to have a higher network density and thus contain nodes with a higher node degree. An example of a dense friendship network is shown in Figure 3. Table 3 shows the maximum node degree and number of friends per fake profile obtained from the cluster analysis. The friendship network of Melissa has a high density and the nodes have the highest degree of all our fake profiles' friends, which is 138. In this case, a high node degree is equivalent to the number of mutual friends of the profile represented by the node in question. Laura's friendship network is also dense, with the highest degree within the network being 110. Her friendship network also contains the highest number of satellites, which have a low degree of edges and are therefore isolated from the denser clusters. Within a cluster, Facebook users have a higher probability of knowing each other. Our results suggest that the number of nodes within a graph correlates with the maximum occurring node degree within the graph. Therefore, we calculated a linear model. The slope of the linear function is 0.48. Figure 4 shows the correlation between the number of nodes, which represents the size of a friendship network, and the maximum node degree within the graph, which represents the profile with the most mutual friends.

Table 3. This Table Compares Parameters of the Friendship Networks Based on the Numbers Collected on 6/26/2012

Parameter	Lena	Melissa	Laura	Ilse	David	Chris	Ferdinand	Mitzi
No. of friends	237	204	272	66	74	111	20	99
Max. node degree	110	138	110	47	27	29	4	48

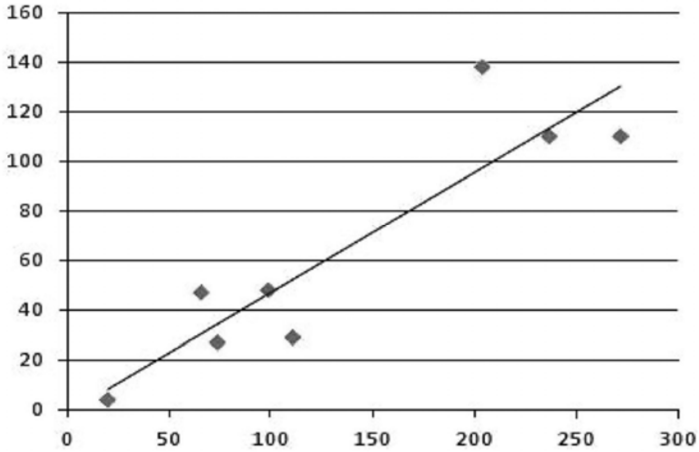


Figure 3. This graph displays the correlation between the number of nodes within a graph and the maximum node degree occurring within the considered graph. The x-axis displays the size of a friendship network, which is the number of friends of the fake profiles used in the social engineering experiment measured on 6/26/2012. The y-axis represents the maximum node degree, which is the highest occurring number of mutual friends within the network. This data was derived during the cluster analysis of the Facebook user data.

The linear regression model derived from this data has a slope of *m = 0,48*.

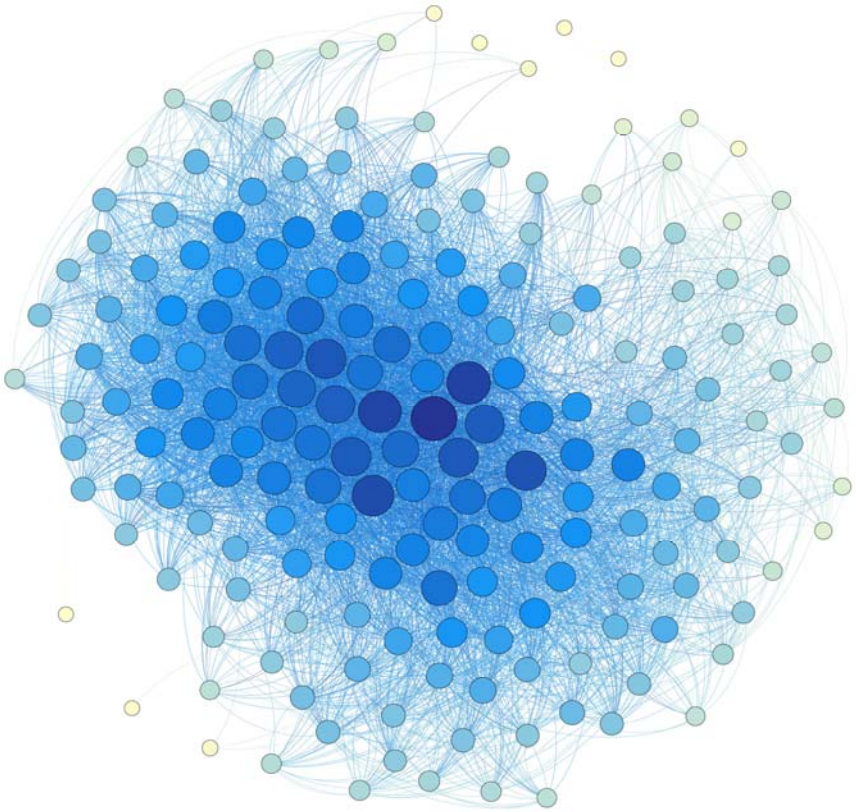


Figure 4. The Friendship Network of Melissa

3.4.2 Work and Education Information

One of the factors mentioned in the description of the People You May Know functionality is work and education information. We analyzed the datasets and labeled the nodes with educational or workplace information depending on the subject's profile information, which, however, was not available for all profiles. In general, we indicated the information by coloring and resizing the respective nodes and drawing the other nodes smaller and grey. If a node is black or blue, a certain educational institution, namely the same as indicated by the fake profile, was listed on the associated profile. All nodes that are not colored black or blue did not list the institution, but may have listed other. We chose to label the nodes if and only if they listed the same educational institute or workplace as the fake profile they were friends with. We performed this education and workplace information analysis on all the friendship networks. An example of such a labeled graph is shown in Figure 5, containing the information obtained from Lena's friendship network. Lena listed two (high-)schools on her profile, namely the Gymnasium Sacre-Coeur in Vienna and the Bundesgymnasium Zaunergasse in Salzburg. We labeled all the nodes respectively. The black coloring indicates that the profiles

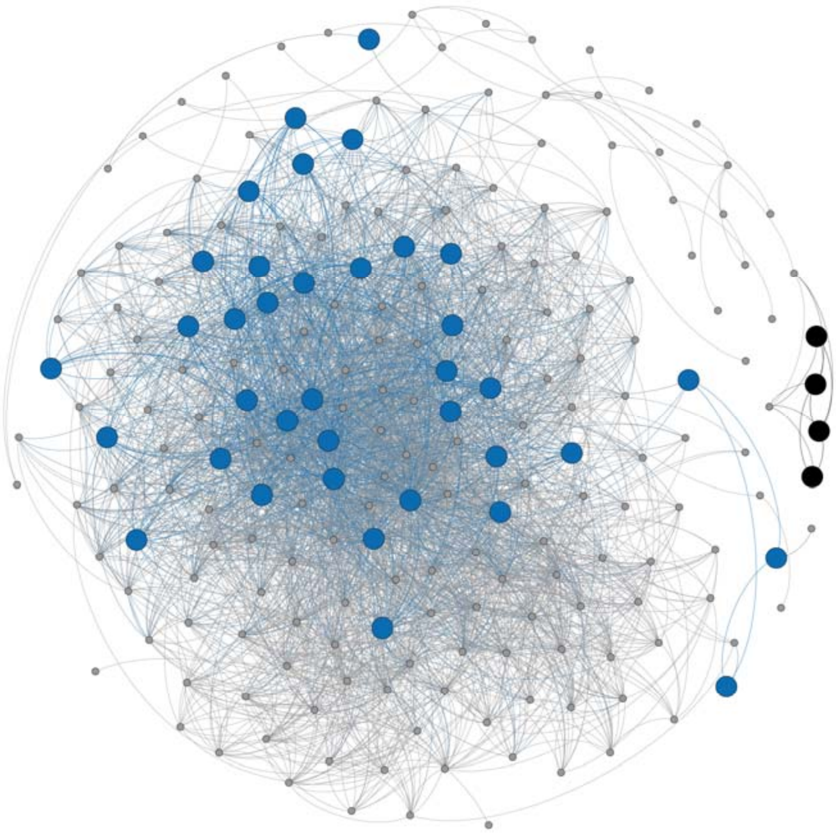


Figure 5. The Friendship Network of Lena, Where Black Nodes Indicate That the Profile Listed Gymnasium Sacre-Coeur and the Blue Nodes Indicate That the Profile Listed Bundesgymnasium Zaunergasse as Educational Institutes in Their Profiles

represented by the nodes listed Gymnasium Sacre-Coeur in their profiles whereas blue indicates Bundesgymnasium Zaunergasse. We took into account that educational institutes may not be named consistently within the network (e.g., the Bundesgymnasium Zaunergasse may also be listed as Gymnasium Zaunergasse or BGZ), and searched for all possible variations referring to the same school in the user data. Furthermore, we found that none of the participants listed both schools in their profile information. Lena's graph shows a typical node distribution for profiles that list more than one educational institute. As can be observed in the example graph showing Lena's labeled friendship network, her friendship circles are clustered by listed educational institutes.

3.4.3 Location Information

Many profiles list location information. In many cases, location information is strongly tied to mutual friends and work and education information. Hence, we assume that location is a factor that highly influences the distribution of a friendship network. Similar to the analysis of education and workplace information, we filtered the locations indicated in the profiles according to the location information disclosed on the fake profiles. Figure 6 shows the friendship network associated with Lena's profile, where all nodes that listed the same hometown as her are marked blue and those indicating the same current location are colored black. During the experiment, we observed that most friends that were suggested to Mitzi Turkish names. Mitzi indicated Vienna as current location in the profile. Therefore, we chose to analyze the friendship network for profiles listing a place in Turkey and Vienna as current location. In this case, the density of friendship connections is also strongly tied to the disclosed location information. Figure 7 shows Mitzi's friendship network, where friends listing a place in Turkey are colored blue and those listing Vienna are black. As mentioned above, we only sent friendship requests to people suggested in the *People You May Know* section. Most of them have mutual friends with the fake profile, therefore we determine that the establishment of a friendship connection with a Turkish profile in the beginning of the experiment caused a snowball effect and integrated the profile into a friendship network consisting of many profiles located in Turkey.

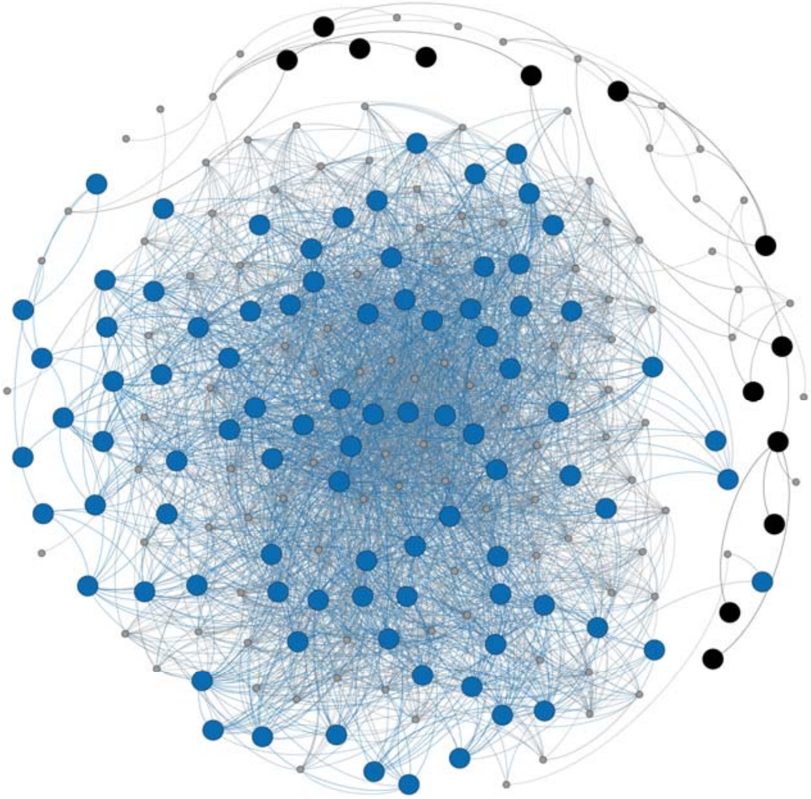


Figure 6. The Friendship Network of Lena, Where Blue Nodes Indicate That the Associated Profiles Listed the Same Hometown as Her and the Black Nodes Represent Profiles with the Same Current Location

3.4.4 Networks and Common Interests

People who share the same interests often join the same networks and like the same Facebook pages. Many Facebook users list arts and entertainment interests or their favorite leisure time activities on their profiles. Among the huge variety of interests, we chose to analyze the favorite TV shows and music listed on the harvested profiles. First, we counted the occurrences of TV shows and music artists listed in the profiles. An example is shown in Figure 8, which illustrates the distribution of the mentioned TV shows among Laura’s friends. Furthermore, we determined the most frequently mentioned TV shows and music artists among the friends of our fake Facebook profiles. The resulting items may not necessarily be listed on our fake profiles. After determining the most frequently listed music artist and TV show for each friendship network, we labeled the nodes respectively and again

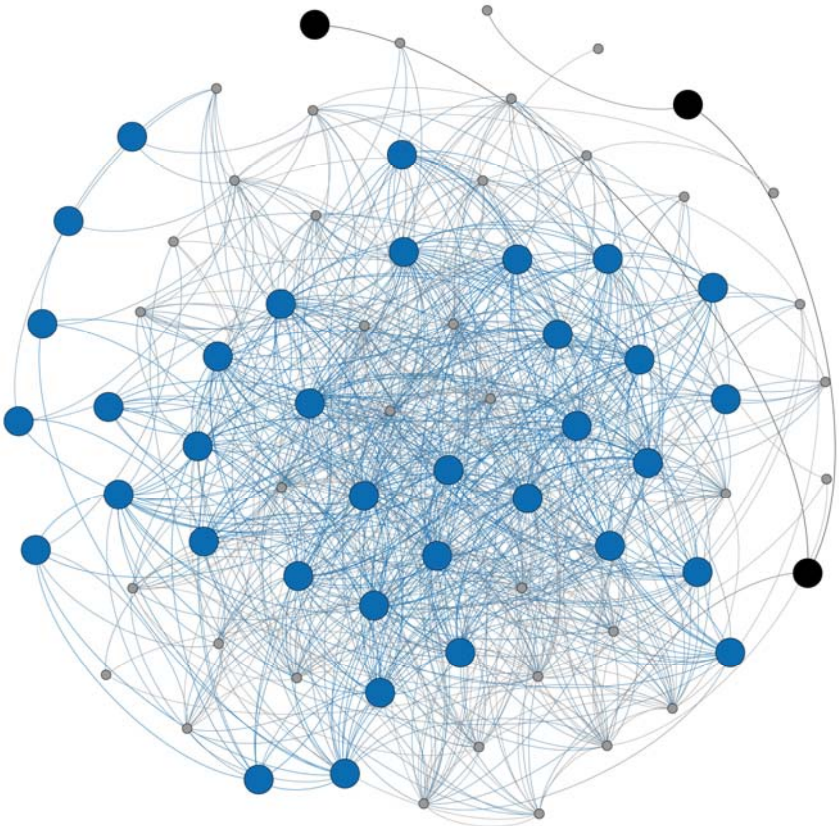


Figure 7. The Friendship Network of Mitzi, Where Blue Nodes Indicate That the Profile Listed a Place in Turkey as Current Location and the Black Nodes Indicate That the Profile Listed Vienna, the Place That Mitzi Herself Actually Mentioned in Her Profile

indicated the occurrence by coloring the nodes accordingly. The most frequently mentioned TV show among Lena's friends was The Simpsons. In Figure 9, the blue labeling indicates that The Simpsons among favorite TV shows. Figure 10 illustrates the occurrence of Rihanna among the lists of favorite artists among David's friends. Taking a closer look the graphs of favorite music and TV shows, we can observe that there is no visible pattern and the distribution among the nodes seems to be random. This is the case not only for the friendship networks shown in Figures 9 and 10 but also for all other analyzed friendship networks. We assume that the reason for this is that the most frequently mentioned items are not specific to a certain community and are therefore found in different communities and peer groups that are not necessarily part of a specific subnetwork.

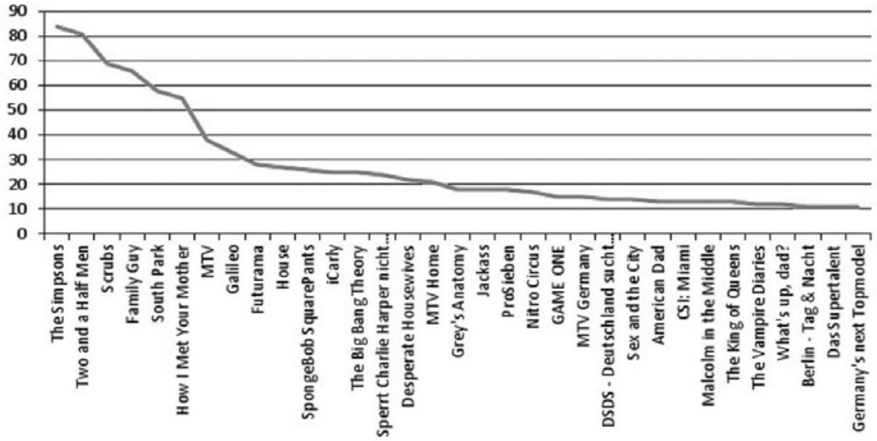


Figure 8. The Distribution of Favorite TV Shows Listed by Laura's Friends. The x-Axis Lists the TV Shows in Descending Order. Only TV Shows Mentioned on More Than 10 Profiles are Show for Better Overview. A TV Show Can Only be Listed Once Per Profile, but More Than one TV Show Can be Listed. The y-Axis Indicates the Total Number of Counts Per TV Show

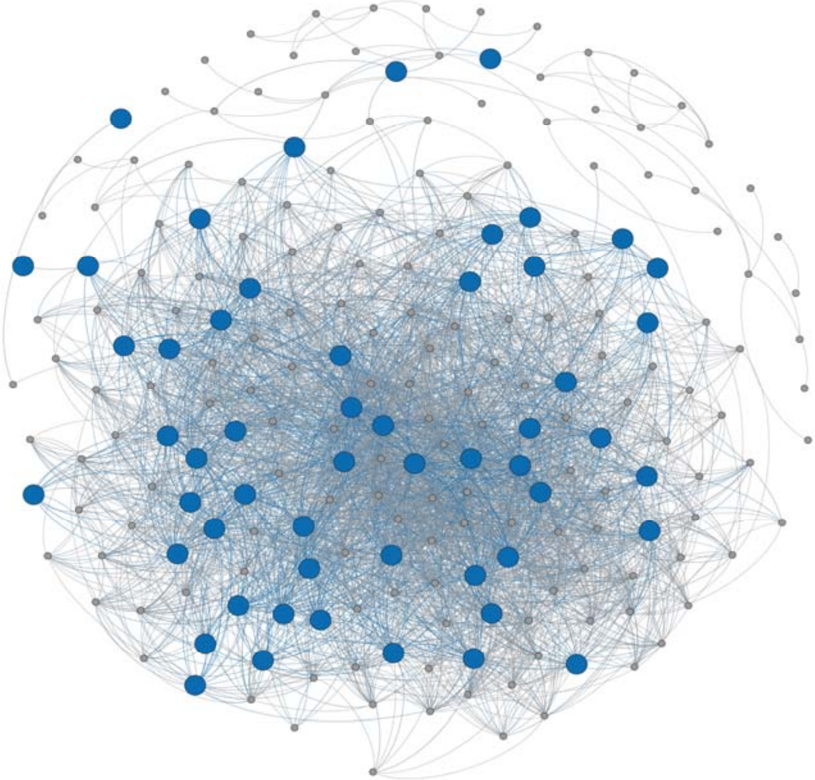


Figure 9. The Friendship Network of Lena, Where Blue Nodes Indicate That *The Simpsons* was Listed among the TV Shows on the Represented Profiles

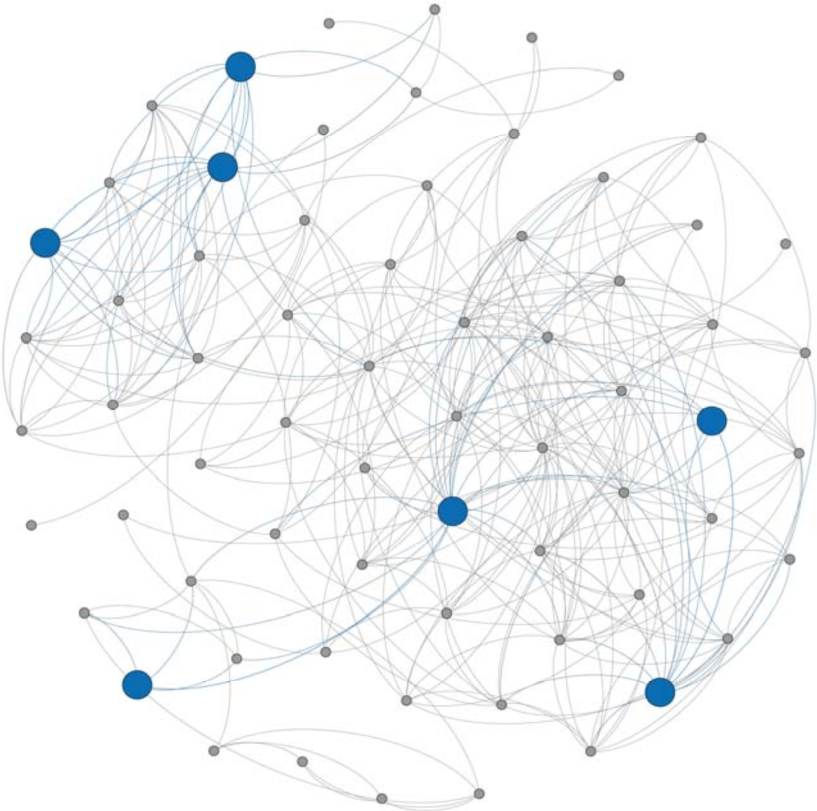


Figure 10. The friendship network of David, where blue nodes indicate that *Rihanna* was listed among the favorite music artists on the represented profiles

3.5 Qualitative Analysis of the Observation

During the experiment, we carefully observed the interactions between the fake profiles and other Facebook users. At every observation point, we collected qualitative data on how we perceived the interactions. We annotated the observations in our logbook together with the numerical data we collected. We observed a strong tendency for Facebook users to perceive a profile as fake if it appears to be new to Facebook, meaning that the profile has few friends and an almost empty wall. When the number of friends increases, the skepticism of the other Facebook users decreases. Having a higher number of friends as well as a lot of social activity makes a profile socially attractive for other users as determined by Boshmaf et al. (2011). At the beginning of the experiment, all our fake profiles received messages from the people we sent friendship requests to. Most of them were very friendly and contained questions such as “Hi, how are you:)”, “From where do I know you?” or “Tell me more about you." Some of them, especially Lena, David and Ferdinand, received messages that contained skeptical statements doubting the authenticity of the profile. On David's wall, one person posted "From where do we know each other?." Some of his friends started com-menting on the post and discussed from where they could all know him, and finally decided that it must be because they went to the same school. Interestingly, almost every Facebook user who sent a message to one of our profiles was already friends with them at the time the messages appeared in the inboxes. Furthermore, we observed that after some time, as the number of friends increased, the number of incoming messages decreased and the skepticism among the other Facebook users gradually disappeared. Laura received many messages containing flirting attempts from men over the whole period of observation. Every fake profile was blocked by Facebook at least once, which means that they were not allowed to send friendship requests to other Facebook users. Most of them were only blocked for one or two days, others were blocked for up to two weeks. The most frequently blocked profile was Ferdinand's profile. The reason for blocking was that Facebook users rejected the friendship requests they received from our fake profiles. In general, we observed that once a certain degree of social attractiveness is reached, the fake profiles start functioning with only minimum interaction (such as accepting incoming friendship requests) required to maintain the profiles as active participants in the social network. This shows that the Facebook Immune System proposed by Stein et al. (2011) is not able to reliably detect fake profiles. Furthermore, this behavior also demonstrates that it is hard for Facebook to detect fake profiles without the users' active support by reporting a specific profile. To ensure a sustainable collection of user profiles, Facebook actively involved the users when they launched a campaign in July 2012 where the platform invited users to report whether the names their friends indicated were real or fake.

3.6 Discussion

The analysis of the fake Facebook profile experiment showed that creating and maintaining a fake profile is an easy task. The Facebook Immune System by Stein et al. (2011) was not able to identify our profiles as fake. Furthermore, when creating the profiles, we did not come across a single CAPTCHA (Completely Automated Public Turing test to tell Computers and

Humans Apart, Ahn 2008). As we created the profiles in March 2012, we suppose that Facebook at that time preferred to increase the number of user accounts instead of eliminating fake profiles. However, in July 2012, Facebook started an initiative asking users to identify fake names among their friends. We suppose that Facebook is trying to enhance the Facebook Immune System and strengthening their methods of detecting and eliminating fake profiles. Fake profiles can be used for data harvesting or other attacks on user privacy. We found that this was a rather easy way to gather private user information. Therefore, we conclude that the Facebook Immune System proposed by Stein et al. (2011) does also not sufficiently detect malicious behavior or abusive social network users. Moreover, we found that social attract-tiveness is an important factor for successful network infiltration. We determined human factors in effective social engineering attacks. From analyzing factors such as profile information on educational institutions, workplace, current location and interests, we determined that location is a strong factor, as is attending the same school or working at the same company. Furthermore, we concluded that having common interests is not a dominant factor. We have found that in general, female fake profiles are more successful in obtaining friends than male ones. However, the gender distribution shows that female profiles are not necessarily only attractive to male profiles.

4. FOCUS GROUPS

4.1 Framework

We have shown that a social engineering attack can be conducted successfully in a social network such as Facebook if the fake profiles are socially attractive. We managed to integrate some of our profiles into existing friendship networks, as shown in section 3.4.1. From the beginning of our research, we focused especially on reaching young people with our social engineering experiment. For young people, social networks such as Facebook are part of their daily routine. The generation of today’s teenagers is growing up with social media and Internet as a part of their everyday life. Especially with our teenage fake profiles Lena and Melissa we achieved a deep integration into existing friendship networks that were associated with the same schools as listed in the profiles and a high number of mutual friends. As the listed schools were located in Austria, we conducted focus groups as proposed by Bortz et al. (2006) and Thomas et al. (2001) to understand the perception of fake profiles in this context and strengthen the results obtained in the social engineering experiment. Furthermore, we wanted to learn about privacy protection strategies in order to estimate the overall data and privacy protecting behaviors of the participants.

4.2 Methodology

The focus groups were conducted in April and May 2012. First, we conducted a session with Austrian university students of computer science who participated in a specialized social media course. Two of the authors of this paper were present during the session. Afterwards, we discussed the outcome of the session with the students to get feedback and comments in order to further improve the method before performing the focus groups with high school students. In general, the session held with the university students helped us estimate whether our concept was feasible and possible to conduct within the timeframe of a double lesson. Therefore, the results from the focus group pilot session with the university students cannot be fully compared with the results from the actual focus groups and the following analysis does not contain results from the testing session at the university. We focused on teenagers, ast hey constituted the main part of the Facebook user data collected during the social engineering experiment. Furthermore, we assume that as Facebook is the most frequently used social media platform, it has a formative impact on the development of teenagers' social media behavior. As all of the participating high school students were underage, we sent consent forms to their parents about 2~3 weeks before the sessions. In the consent forms, we described the overall procedure in detail, including information on which kind of data was to be collected and how it would be used. All parents allowed their children to participate. We conducted three sessions with Austrian high school students, two in Vienna and one in Salzburg. Overall, 46 teenagers, from 14 to 17 years old, participated. The sessions were held during the computer science class at the participating schools. The supervising teachers and one of the authors were present. All sessions were recorded with a voice recorder and a memory protocol was made immediately following each session. We analyzed the material using a grounded theory as proposed by Adolph et al. (2011), Bortz et al. (2006) and Matavire et al. (2008).

4.3 Privacy Awareness

During the focus group sessions, we determined the respondents' general awareness for privacy. We found that there was definitely a lot of awareness, but that the participants' self-assessment concerning disclosure practices did not depict reality. This underlines the findings of Liu et al. (2011) and Lipford et al. (2008). To start the discussion on general privacy concerns we asked the participants "What can be found about you on the Internet, and particularly on Facebook?." At first, the most common answer was "Nothing." In all three sessions, however, a small amount of critical participants responded to their colleagues with remarks such as "There is even your name, school year and a picture of you on the school's homepage!" and "Well, you have a Facebook profile!." Furthermore, some of them reported that they had googled themselves before and mentioned what they had found. However, most of them dismissed this and said: "There is nothing special on Facebook, just useless stuff such as pictures and hobbies and so on." They were then instructed to examine their own profiles by reviewing them with their classmates if they wanted to. All participants who had a Facebook profile did so and were impressed by the amount of information they had provided on Facebook. Some of the participants asked their classmates to look at their profiles from their point of view to see what their disclosing habits were. Nevertheless, we observed that except for only four out of all participants, all had restricted their profiles by configuring their privacy settings.

4.4 Privacy Regulations-Privacy Enhancing Strategies

As stated in the previous section, almost all respondents had configured their privacy settings on Facebook, which means that they had spent some time in trying to understand the information-sharing model of the platform. We wanted to determine the reasons for them dealing with this topic. Most of them reported that specific incidents had led to this behavior. As shown by Strater and Lipford (2008), most users of a social network do not reflect their disclosure habits and privacy settings until an unpleasant incident occurs. Therefore, we assumed that the respondents had to have had a bad experience before reflecting on their privacy on Facebook. Fortunately, the only negative incident that was reported was that their parents commented on everything they posted on Facebook and the respondents found that embarrassing. Others reported that they had heard of risks such as stalking, mobbing or identity theft and furthermore simply thought that their disclosures were “not the business of a general audience.” Nevertheless, only two participants, both female, were able to inform us about how their privacy settings were configured without checking on Facebook. Some participants assumed that they had made some information public, but in general they were not sure. Many of the high school students immediately started checking their privacy settings. Only a handful did not even know where they could configure their settings. They started discussing this with their classmates and the more experienced ones among them explained how to configure them. They argued on the best boundary management principles. Most of them reported that having everything “friends-only” was the best privacy enhancement. One respondent reported that she liked the new feature that Facebook had introduced recently that allowed users to see their profile from a different audience’s perspective. Only a handful of participants reported ever having deleted a post. More precisely, they had deleted posts when switching from the old Facebook profile to Timeline (2012), as this view displayed older posts that they found were inappropriate or embarrassing. However, most of the participating respondents never reviewed or deleted posts. To estimate how much they had disclosed on Facebook, we requested the participants in our study to visualize their assumeptions. We handed them a pack of plain white A4 copy paper and asked them to imagine how high the stack of paper would be if they printed out all the information they had made available on Facebook. In the first round, the stacks were about 10 sheets of paper high. After discussing the recent disclosures and looking at their profiles on Facebook, most respondents changed the height of the stack by adding more paper. Many of them did not know in how many photos they were tagged or how many pictures they had uploaded themselves. Our results suggest that this task increased the awareness for responsible revealing of information to a wide audience.

4.5 Interactions with Others

We also wanted to know how the participants of the study interacted with others on Facebook. We determined that none of the respondents had ever considered that they might have fake Facebook profiles on their friend lists. We asked them about their experiences when getting incoming friend requests from strangers. Most of them reported that they usually browsed the request sender's profile to determine whether the person seemed friendly and interesting. Moreover, they said that if a person was of the same age or lived in the same city and attended the same school, they usually accepted them as friends in Facebook, even if they had never seen them before. Still, the main criterion for accepting friend requests was having mutual friends. Almost all of the respondents reported this as a major criterion for exclusion. Many high school students mentioned that they were friends with their parents on Facebook. Those who claimed that they would never list their parents as friends on Facebook said that this was because they were afraid of embarrassing situations and that their parents would spy on them over Facebook. In these discussions, some respondents started searching for their parents and teachers on Facebook. During one session, they found their favorite teacher's Facebook profile and discussed his pictures. We asked them whether they would add him as a friend and whether they would add teachers as friends in general. They immediately replied that they would definitely add this specific teacher and that he would surely accept the request. However, they said that they would not want to be friends with other teachers on Facebook. A participant in another session mentioned that the school forbid students and teachers to be friends on Facebook until after graduation.

4.6 Discussing the Social Engineering Facebook Experiment

In this work, we introduced the social engineering Facebook experiment as a research method. We had performed this experiment before conducting the focus groups, so we asked the participants in the groups whether they had ever seen one of our profiles and showed them the profile pictures of our fake Facebook users. We asked them whether they had ever seen those people before, either in reality or on Facebook. In the first group that we examined, almost all of the Facebook users were friends with at least one of our fake profiles, which was Melissa, who claimed to attend the same school. The second focus group, which was conducted in Salzburg, attended the same school as Lena. About half of them were friends with her, but some respondents claimed that they had rejected her friend request because they did not know her. The majority of our third focus group, again conducted in the same Viennese school as the first one, was not a friend of any of our fake profiles. Some, however, were friends with Melissa. Most of the participants in the study had many mutual friends with her. A female respondent reported that she had rejected the friendship request but had 89 mutual friends with Melissa. She started laughing when she observed in the focus group session that actually many of her friends and even her brother were friends with our fake profile Melissa. She was proud of not accepting the friendship request and promised to spread the word and explain the situation to her friends. In general, we observed that all of our participants who held a Facebook profile had mutual friends with either Melissa or Lena. Moreover, many participants had received friendship requests from either of them and some were even friends with them on Facebook.

4.7 Awareness Training

At the end of each focus group session, we performed a profiling task as privacy awareness training. We manually selected two public profiles on Facebook, belonging to people who were not familiar to the participants. The participants were given a short period of time to profile them using all the information they could find on Facebook or other online sites. Afterwards the results were collected on a whiteboard and discussed in detail. The respondents were amazed by the amount and precision of the gathered data. After discussing the results of the profiling task, we invited the participants to reflect on how much information they themselves disclosed and how it could be misused. Furthermore, we discussed the validity of the information disclosed in social networks. The respondents learned that the validity of information is generally hard to measure, but the validity can be checked by considering data from other user profiles and other Internet services, such as phone books or home pages of schools and companies. The respondents also reported that the profiling task had increased their awareness for data connections and privacy.

5. IMPLICATIONS

5.1 Discussion

Our results suggest that Facebook users are aware of several privacy related issues that are associated with publishing sensitive information. However, their behavior mostly does not map the level of awareness as people tend to behave differently than they perceive they do. Our research has shown that people are skeptical towards social media site operators, as most of the participants in our studies reported that they did not trust the platform in terms of data security and privacy, but they trust their interaction partners. Even though they are aware of the fact that some profiles contain fake information, they believe that their interaction partners indicate personal information correctly on their profiles. This shows that common Facebook users with even a lot of experience in online communication can have difficulties identifying fake profiles. Our research has furthermore shown that Facebook does not provide enough methods to reliably detect and eliminate fake profiles. As users either provide fake information themselves or interact with fake profiles without even noticing, they are not actively supporting the platform's need for reliable and correct user data. Even though Facebook clearly states in their legal terms that correct personal data must be provided, they are not able to automatically check the correctness or provide substantial countermeasures against entering fake data. Boshmaf et al. (2011) and Bilge et al. (2009) have shown that social engineering is very effective in social media. Based on their findings, we determined reasons for this effectiveness. To do so, we conducted the social engineering experiment on Facebook and monitored activities and observations. Furthermore, we analyzed the data that we gathered from the participants of the experiment. We have shown that the success of a social engineering attack is determined strongly by human factors and also influenced by the social attractiveness of a profile, as already shown by Boshmaf et al. (2011), Bilge et al. (2009) and Barracuda Labs (2011). We observed that many Facebook users do not doubt that there is a real person behind a profile. Especially if the profile indicates a location near them or claims to have attended the same school or university, Facebook users tend to believe that they know the person in real life. We also determined that having the same interests is not an important factor if they are not specific to a certain peer group. Thus, we determined the most important factor for successfully performing a social engineering attack in social media is having mutual friends with another Facebook user. We have shown that after a certain amount of mutual friends, a fake profile can easily integrate into a friendship network. We demonstrated that empty profiles (profiles that do not display social activities and a high number of friends) are more likely to be perceived as fake. After a certain number of interactions with other users, the social attractiveness of a profile increases and is more likely to be perceived as that of a real human being. We used several types of characters of different age groups and different levels of education. The results of our research suggest that the profile picture is only a secondary factor in determining a person behind a profile as even pictures with a higher number of artifacts were recognized as people from a user's real-life environment. The influence of the quality of a profile picture was not analyzed in this work but should be investigated in future research. From our research we determined that many Facebook users have a theoretical understanding of privacy concepts and the dangers of malicious attacks using fake profiles in the context of social media. However, we also showed that even though they have a theoretical understanding, many users are not able to apply their knowledge in practice when actively using social media and being confronted with fake profiles used for malicious attacks. Furthermore, they sometimes even have a wrong perception of their own actions, as determined by the focus groups. However, many people have learned from their previous mistakes and gathered a detailed knowledge about Facebook privacy settings and the use of third-party applications. We have shown that the core problem is the perception of audiences. We have shown that most people protect their private data by adjusting their privacy settings on Facebook, while others provide fake information to cover their private information. When finding unwanted content within the network, they use the reporting tools provided by Facebook or simply untag themselves in case of an unwanted post or picture. Facebook users are aware of these tools and do not hesitate to use them. This enables the Facebook platform to control the information within the network with the user's assistance. Overall, we learned that Facebook users do reflect on the behavior of themselves and other users. Especially when conducting the focus groups, we observed that people requested support and education concerning safe Facebook usage and privacy protection. Many of them criticized that Facebook had their data. Our results suggest that the problem of fake information in user data does not the Facebook business model. To date, to the best of our knowledge, Facebook is not able to detect and eliminate fake information, neither with their legal terms nor with automated procedures. They definitely rely on the users' awareness and assistance in overcoming this issue. The dilemma is that Facebook has to protect user privacy to satisfy the user on the one hand, but eliminate unwanted behaviors on the other hand, thereby possibly displeasing the users with stricter regulations.

5.2 Awareness Training as a Countermeasure

Fake profiles have a high impact on the sustainability of the business model as well as on advertising companies that rely on the accuracy of the user data. However, they also have a high impact on user privacy. As our results suggest, the users themselves are mostly unaware of the occurrence of fake profiles and their consequences. In this section we discuss countermeasures to increase user awareness. The results of our social engineering experiment may also be used to generate implications for business participants in social networks or the platform providers themselves. However, creating countermeasures for these parties is not within the scope of this work. We hypothesize that there is a lot of interest from all involved parties (students, parents and teachers) to discuss and learn about privacy-related issues in social media. The participants in the focus groups were motivated and contributed actively. During the sessions, they asked specific privacy and social media-related questions and whether we could show them how to handle configurations on Facebook. We also showed them how they could download all their data from Facebook. Many respondents were so interested that they even took notes without being prompted to. Furthermore, we showed them privacy awareness tools such as openbook.org and pleaserobme.com (2012). Some were scared and concerned when discussing the results of the profiling task and privacy risks and dangers. This also shows that this topic concerns them and that they want to learn more about it. In general, we recommend that teachers and parents discuss social media at home and in class to ensure responsible communication and disclosure in social networks. The core concept of our proposed awareness training is that the information-sharing model of social networks needs to be defined and explained in detail. According to our research results, a deep understanding of the information-sharing model supports the perception of audience boundary management. Furthermore, it is important to discuss risks that are associated with privacy leaks. Social media users need to learn the effects of the core functionalities provided by the platform. An example that is important concerning social engineering attacks with fake profiles is that establishing a friendship connection with someone usually provides them with more private information, as most users restrict their audiences and provide more detailed private information to people who they are friends with on Facebook. Facebook has already taken the first steps towards an increased awareness among its users by starting a campaign to involve them in reporting users who provide fake names. However, the question arises whether this procedure is not another violation of the right to privacy because of its surveillance-like approach.

6. CONCLUSIONS

We conducted a social engineering experiment on Facebook and determined factors that contribute to the successful integration of a fake profile into an existing friendship network. Furthermore, we have human behavior and interaction between common user profiles and our fake profiles and described the patterns we found. We also demonstrated that profiles that do not display social activities and a high number of friends are more likely to be perceived as fake than profiles that display social activities and interactions with others. Moreover, we examined Facebook users' privacy considerations and gained a deeper understanding of the connection between fake users and common users. In the end, we discussed the correlation between the number of registered profiles and the correctness of the user data and its impacts on the Facebook business model. In this work, we also proposed a training method to increase user awareness as a valid countermeasure.

ACKNOWLEDGEMENT

This work has been supported by Fem Tech under grant no. 835278 (FFG) and by COMET K1, FFG-Austrian Research Promotion Agency.

REFERENCES

Adolph S, Hall W., & Kruchten P (2006) Using grounded theory to study the experience of software development. In: Empirical Software Engineering 16(4):487-513. DOI 10. 1007/s10664-010-9152-6.

Agichtein E, Castillo C, Donato D, Gionis A, & Mishne G (2008) Finding high-quality content in social media. In: Proceedings of the international conference on Web search and web data mining, WSDM '08, ACM, New York, NY, USA:183-194. DOI 10.1145/1341531.1341557.

von Ahn L, Maurer B, McMillen C, Abraham D, & Blum M (2008) recaptcha: Human-based character recognition via web security measures. In: Science 321(5895):1465–1468.

Altman I (1975) The Environment and Social Behavior: Privacy, Personal Space, Territory, Crowding. In: Brooks/-Cole Pub. Co, Monterey, California.

Asuncion AU, Goodrich MT (2010) Turning privacy leaks into floods: Surreptitious discovery of social network friendships and other sensitive binary attribute vectors. In: Proceedings of the 9th annual ACM workshop on Privacy in the electronic society, WPES '10. ACM, New York, NY, USA:21-30. DOI 10.1145/1866919.1866923.

Baden R, Bender A, Spring N, Bhattacharjee B, & Starin, D (2009) Persona: An online social network with user-defined privacy. In: Proceedings of the ACM SIGCOMM '09 conference on Data communication, SIGCOMM '09. ACM, New York, NY, USA: 135-146. DOI 10.1145/1592568.1592585.

Barracuda Labs (2011) Barracuda labs social networking analysis. http://barracudalabs.com/fbinfographic/. Accessed 2012-5-5.

Besmer A, Richter LH (2010) Moving beyond untagging: Photo privacy in a tagged world. In: Proceedings of the 28th international conference on Human factors in computing systems, CHI '10. ACM, New York, NY, USA:1563-1572. DOI 10.1145/1753326.1753560.

Bilge L, Strufe T, Balzarotti D, & Kirda E (2009) All your contacts are belong to us: Automated identity theft attacks on social networks. In: Proceedings of the 18th international conference on World wide web, WWW '09. ACM, New York, NY, USA: 551-560. DOI 10.1145/1526709.1526784.

Bortz J, Döring N (2006) Forschungsmethoden und Evaluation: für Human- und Sozialwissenschaftler, überarb. edn. Springer, Heidelberg 4.

Boshmaf Y, Muslukhov I, Beznosov K, & Ripeanu M (2011) The socialbot network: When bots socialize for fame and money. In: Proceedings of the 27th Annual Computer Security Applications Conference, ACSAC '11. ACM, New York, NY, USA: 93-102. DOI 10.1145/2076732.2076746.

Boyd D, Ellison N (2007) Social network sites: Definition, history, and scholarship. Journal of Computer-Mediated Communication 13(1). URL http://jcmc.indiana.edu/vol13/issue1/boyd.ellison.html.

CAPTCHA (2012) http://www.captcha.net/. Accessed 5-15-2012.

Dorfman R (1943) The detection of defective members of large populations. In: Annals of Mathematical Statistics 14:436-440, Lee J, Lee J, & Feick L (2001) The impact of switching costs on the customer satisfaction-loyalty link: Mobile phone service in France. Journal of Services Marketing 15(1):35-48.

Facebook (2012) https://www.facebook.com/legal/terms. Accessed 7-19-2012.

Facebook Developers (2012) https://developers.facebook.com/docs/reference/api/. Accessed 6-25-2012.

Facebook Press (2012) http://www.facebook.com/press. Accessed 4-15-2012.

Facebook Timeline (2012) https://www.facebook.com/about/timeline. Accessed 6-6-2012.

Fruchterman TMJ, Reingold EM (1991) Graph drawing by force-directed placement. In: Software: Practice and Experience 21(11):1129-1164. DOI 10.1002/spe.4380211102.

Gao H, Hu J, Wilson C, Li Z, Chen Y, & Zhao B Y (2010) Detecting and characterizing social spam campaigns. In: Proceedings of the 17th ACMConference on Computer and Communications Security, CCS ’10. ACM, New York, NY, USA:681-683. DOI 10.1145/1866307.1866396.

Gephi: (2012) http://gephi.org/. Accessed 10-6-2012.

Gross R, Acquisti A (2005) Information revelation and privacy in online social networks. In: Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society, WPES ’05. ACM, New York, NY, USA:71-80. DOI 10.1145/1102199.1102214.

Huber M, Mulazzani M, Leithner M, Schrittwieser S, Wondracek G, & Weippl E (2011) Social snapshots: Digital forensics for online social networks. In: Proceedings of 27th Annual Computer Security Applications Conference (ACSAC):113-122.

JSON.org: (2012) http://json.org. Accessed 26-6-2012.

King J, Lampinen A, & Smolen A (2011) Privacy: Is there an app for that? In: Proceedings of the Seventh Symposium on Usable Privacy and Security, SOUPS ’11. ACM, New York, NY, USA:12.1-12.20. DOI 10.1145/2078827.2078843.

Krasnova H, Günther OSS, & Koroleva K (2009) Privacy concerns and identity in online social networks. In: Identity in the Information Society 2(1):39-63.

Lipford H R, Besmer A, Watson J (2008) Understanding privacy settings in facebook with an audience view. In: Proceedings of the 1st Conference on Usability, Psychology, and Security, UPSEC '08 . USENIX Association, Berkeley, CA, USA:2.1-2.8.

Liu Y, Gummadi K, Krishnamurthy B, & 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. ACM, New York, NY, USA:61-70. DOI 10.1145/2068816.2068823.

Mao H, Shuai X, & Kapadia A (2011) Loose tweets: an analysis of privacy leaks on twitter. In: Proceedings of the 10th annual ACM Workshop on Privacy in the Electronic Society, WPES '11. ACM, New York, NY, USA:1-12. DOI 10.1145/2046556.2046558.

Matavire R, Brown I (2008) Investigating the use of 'grounded theory' in information systems research. In: Proceedings of the 2008 annual research conference of the South African Institute of Computer Scientists and Information Technologists on IT Research in Developing Countries: Riding the Wave of Technology, SAICSIT '08. ACM, New York, NY, USA:139-147. DOI 10.1145/1456659.1456676.

Narayanan A, Shmatikov V (2009) De-anonymizing social networks. In: Proceedings of the 2009 30th IEEE Symposium on Security and Privacy, SP '09. IEEE Computer Society, Washington, DC, USA:173-187. DOI 10.1109/SP.2009.22.

NetworkX (2012) http://networkx.lanl.gov/. Accessed 10-6-2012.

pleaserobme.com (2012) http://pleaserobme.com/, Accessed 10-5-2012.

Puttaswamy KPN, Sala A, & Zhao BY (2009) Starclique: Guaranteeing user privacy in social networks against intersection attacks. In: Proceedings of the 5th International Conference on Emerging Networking Experiments and Technologies, CoNEXT '09. ACM, New York, NY, USA:157-168. DOI 10.1145/1658939.1658958.

Python (2012) http://www.python.org/. Accessed 10-6-2012.

Smith A, O'Hara K, & Lewis P (2011) Visualising the past: Annotating a life with linked open data. In: Web Science Conference '11. URL http://eprints.ecs.soton.ac.uk/22324/.

SnapshotSurvey (2012) http://is.gd/snapshotsurvey. Accessed 22-5-2012.

Stanton JM (2003) Socio-technical and human cognition elements of information systems. In: S Clarke, E Coakes, GM Hunter, A Wenn (eds.) Information Technology and Privacy, chap. Information technology and privacy: a boundary management perspective. IGI Publishing, Hershey, PA, USA:79-103.

Stein T, Chen E, & Mangla K (2011) Facebook immune system. In: Proceedings of the 4th Workshop on Social Network Systems, SNS '11. ACM, New York, NY, USA:8.1-8.8. DOI 10.1145/1989656.1989664.

Strater K, Lipford HR (2008) Strategies and struggles with privacy in an online social networking community. In: Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction, BCS-HCI '08. British Computer Society, Swinton, UK, UK 1:111-119.

Stutzman F, Capra R, & Thompson J (2011) Factors mediating disclosure in social network sites. Computers in Human Behavior 27(1):590-598. DOI 10.1016/j.chb.2010.10.017.

Stutzman F, Kramer-Duffield J (2010) Friends only: Examining a privacy-enhancing behavior in Facebook. In: Proceedings of the 28th International Conference on Human Factors in Computing Systems, CHI '10, ACM, New York, NY, USA:1553-1562. DOI 10.1145/ 1753326.1753559.

Sun X, Sun L, & Wang H (2011) Extended k-anonymity models against sensitive attribute disclosure. In: Computer Communications 34(4):526-535. DOI 10.1016/j.comcom.2010. 03.020.

TechCrunch (2012) http://techcrunch.com/2012/02/15/facebook-verifiedaccounts-alternate-names/. Accessed 7-30-2012.

TechCrunch (2012) http://techcrunch.com/2012/07/30/startup-claims-80-of-itsfacebook-ad-clicks-are-coming-from-bots/. Accessed 7-30-2012.

Thomas JC (2001) Qualitative vs. quantitative: Myths of the culture and practical experience. In: Proceedings of the 34th Annual Hawaii International Conference on System Sciences:10.

Wang N, Xu H, & Grossklags J (2011) Third-party apps on Facebook: Privacy and the illusion of control. In: Proceedings of the 5th ACM Symposium on Computer Human Interaction for Management of Information Technology, CHIMIT '11. ACM, New

York, NY, USA:4.1-4.10. DOI 10.1145/2076444.2076448.

Wang Y, Norcie G, Komanduri S, Acquisti A, Leon PG, & Cranor L (2011) I regretted the minute I pressed share: a qualitative study of regrets on Facebook. In: Proceedings of the Seventh Symposium on Usable Privacy and Security, SOUPS ’11. ACM, New York, NY, USA:10.1-10.16. DOI 10.1145/2078827.2078841.

Zheleva E, Getoor L (2009) 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, WWW ’09. ACM, New York, NY, USA:531-540. DOI 10.1145/1526709.1526781.

AUTHOR BIOGRAPHIES

Katharina Krombholz is researcher and Ph.D. student at SBA Research in Vienna, Austria. She received a master degree in Media Informatics from the Vienna University of Technology. Her research interests include security, privacy, social networks, human-computer interaction and interaction design.

Dieter Merkl is Associate Professor of Applied Computer Science at the Institute of Software Technology and Interactive Systems at the Vienna University of Technology (Austria). His main research interests are in the areas of information retrieval, data mining, and interaction design. He has published more than 140 scientific papers in these areas.

Edgar Weippl is research director of SBA Research and Associate Professor at the Institute of Software Technology and Interactive Systems at the Vienna University of Technology (Austria). His research focuses on applied concepts of IT-security and e-learning.