How Much do You Tell? Information Disclosure Behaviour in Different Types of Online Communities

Johann Schrammel* Christina Köffel* Manfred Tscheligi*‡

*CURE - Center for Usability Research and Engineering
Hauffgasse 3-5
1110 Vienna, Austria

‡ICT&S Center, University of Salzburg
Sigmund-Haffner-Gasse 18
5020 Salzburg, Austria

{schrammel, koeffel, tscheligi}@cure.at

ABSTRACT
Online communities of different types have become an important part of the internet life of many people within the last couple of years. Both research and business have shown interest in studying the possibilities and risks of these relatively new phenomena. Very controversial aspects of these communities are their implications and effects on privacy issues, as research has shown that users generally provide information rather freely on such communities. However, no systematic comparison of differences in information disclosure behavior considering different types of communities is available. Furthermore only few is known about the information disclosure behavior related to demographic variables, usage contexts and usage patterns. To better understand these aspects of online communities we conducted an online survey that questioned users of various popular online communities about their information disclosure behavior and usage patterns of these sites. More than 850 users responded to our questionnaire. In this paper we present the main results of the analysis and provide linear regression models that allow understanding the involved factors in detail.

Categories and Subject Descriptors
H.1.2 [**User/Machine Systems**]: Human Factors

General Terms
Design, Security, Human Factors

Keywords
Social Networking Sites, Information Disclosure, Privacy, Online Communities.

C&T'09, June 25–27, 2009, University Park, Pennsylvania, USA.
ACM 978-1-60558-601-4/09/06.

1. INTRODUCTION
Online communities of different types have become an important part in the internet life of many people within the last couple of years. Research has shown that already 59 percent of young adults [27] and 87 percent of students [15] are using social networking sites regularly. In terms of unique views MySpace[1] (114.5 million unique views in May 2008) and Facebook[2] (123.9 million unique views in May 2008) are the most popular [7].

At the same time business networking sites have become very popular. According to the business network LinkedIn[3] more than 25 million users are currently registered [28]. Xing[4], the German-founded competitor of LinkedIn claims to have more than 6 million registered and over 500,000 paying premium users [41].

Similar developments of increasing popularity can be observed in the areas of gaming communities, social news and bookmarking sites and community oriented content and media sharing.

To participate in these communities and to actualize their potential benefits it is required to provide these sites with information about oneself. This can be very limited (e.g. only username and password) or very voluminous (e.g. extensive collections of private materials). Even though the user typically has the possibility to participate with disclosing only very limited information, (social) mechanisms reward users that provide more information.

Even though most users declare that they are concerned about their privacy, e.g. because they are afraid of their boss or other people getting unwanted access to their data [5], they often act contradictory [9]. Several studies have revealed that most users are completely unaware of possible risks and consequences of online profiles without access restriction [1],[21]. This indicates that users have difficulties in estimating short- and long-term risks in connection to the disclosure of private information [2].

[1] http://www.myspace.com/
[2] http://facebook.com/
[3] http://www.linkedin.com/
[4] http://www.xing.com/

It has been already shown that there are real risks in disclosing sensitive information. There exist reports of people being deprived of employment opportunities due to employer reviews of social networking profiles [11]. Also social engineering [17] and social phishing [18] are main risks associated with freehanded information disclosure. A comprehensive overview of major risks for users of social networks can be found in [16].

2. RELATED WORK

2.1 Information disclosure

In general the potential risk for a user increases the more information s/he provides. Research has shown that people are providing possibly sensitive data rather freely. The extent of information disclosure has been researched in several studies, especially in the context of social networking sites (e.g. [13],[21], [22],[24],[38]). Table 1 provides an overview of the amount of data disclosed as found in the two most comprehensive studies available for the online social network Facebook (Gross et al. 2005 [13], Lampe et al. 2007 [24]).

Table 1: Information Disclosure on Facebook

	Gross et al. 2005	Lampe et al. 2007
Sample size	4,540	38,407
Restricted access	-	19 percent
Realistic name	89 percent	-
Picture	90.8 percent	-
Birthdate/birthday	87.8 percent	83.8 percent
Hometown	-	83.3 percent
Current residence	50.8 percent	45.1 percent
Phone number	39.9 percent	-
E-Mail address	-	92.3 percent
Gender	99.6 percent	93.8 percent
Relationship status	62.9 percent	78.5 percent

As can be seen in Table 1 a high percentage of users provide a realistic name and picture. Gross et al. 2005 [13] found that 61 percent of the profile-pictures were suitable for identification, 15 to 21 percent of the investigated user profiles are susceptible to real-world stalking and 77.7 percent could be victims of online stalking. Real-world stalking is made possible through the disclosure of address data such as hometown, current residence or phone number. The posting of the e-mail address or other online information such as instant messaging contacts on networks without any restriction facilitates online stalking, i.e. people seeking online contacts over different channels.

Another study conducted by Stutzmann [38] underlines the above indicated findings. According to Stutzmann a holistic and more subjective disclosure of identity information is enabled by online networks. Furthermore Stutzmann distinguishes a difference in the amount of information requested by different networks (i.e. Facebook, MySpace and Friendster[5]).

Facebook for example offers the possibility to disclose the most information, but only requires the posting of name, e-mail address, academic classification and school information. Friendster and MySpace on the other hand offer less optional information to be posted, but require more by default for registering.

In a similar study, Krishnamurthy and Wills [21] have compared eleven social networks and found that 55 to 90 percent of users allow their profile information to be viewable by the entire network, a number that even increases to around 80 percent in smaller networks such as university networks.

In contrast to the availability of research on information disclosure behaviour in social networking sites analysis of such data in other types of online communities such as collaborative gaming plattforms or content sharing sites is not available or focuses on different aspects, e.g. the disclosure of contents produced by the users and not of the information they explicetely provide.

Ahern et al [3] for example investigated the management of privacy settings in the online photo-sharing network Flickr[6] focusing on location disclosure. Another study by Van House [39] also deals with privacy awareness on Flickr. However, none of these studies provides information on the level of information disclosure in the profile.

2.2 Privacy Settings

An important point of discussion in online data disclosure are default privacy settings offered by online networks. Although default privacy settings are perceived to be too restrictive for contact-seeking users [20], recent studies have indicated that a high percentage of users (i.e. 79 percent of 3,851 users) did not change their default settings on MySpace and 99 percent of Twitter users preferred to keep the initial settings [22]. For Facebook Gross et al. 2005 [13] found that only 1.2 percent of users changed the default privacy settings. In order to inform users of potential risks of data disclosure, especially considering privacy settings, the Fraunhofer-SIT institute has conducted a survey investigating the privacy-friendliness of commonly used social and business networks [10]. The results show that none of the tested platforms is entirely safe to use. Although having many weaknesses, Facebook is ranked as the best social networking platform. Concerning business networks, LinkedIn allows for more anonymity and easier removal of accounts than Xing.

Besides the users' decline to change default privacy settings, people still do not entirely trust new technologies in combination with online networks. According to Andrews [4] people tend to trust persons they already know or have encountered offline, which implies that trust in the virtual vis-à-vis is as important as trust in the system to fulfill privacy requirements [6]. Different considerations have already been made in order to encourage sceptical users to trust the network [4],[19] or to recover trust after possible breakdowns [35],[40]. Nevertheless Preibush et al.[33] claim the area of privacy in social networks to be severely under researched.

[5] http://www.friendster.com/

[6] http://flickr.com/

3. MOTIVATION AND RESEARCH QUESTIONS

The review of related work above shows that valuable research and data is available on users' online information disclosure behaviour in different contexts. However, several aspects have not been addressed yet sufficiently.

Firstly, most studies deal either with only one specific community (most frequently Facebook) or at best with one type of online network. Various sets of items and measurements are used in the different studies, therefore the comparison of results is only possible in a very limited way. Consequently it remains complicated to analyse and understand the differences in information disclosure behaviour between different types of online networks.

Secondly, whereas first results of research investigating the influence of demographic variables concerning the participation in online communities become available [15], to the best of our knowledge no such data is available in relation to information disclosure behaviour.

Our main research questions therefore were whether there are systematic differences in the amount of information disclosure in different online communities and whether there is an important influences of the demographic background on the information disclosure behavior.

4. METHODOLOGICAL CONSIDERATIONS

To answer our research questions in a most valid way several methodological aspects were considered carefully. In the next sections we explicate the background and reasoning for the methodological decisions we made and briefly discuss limitations of our work.

4.1 Classification of Community Types

To be able to analyse and compare information disclosure on different networking sites we needed a classification scheme to start with. Numerous classifications for online communities have been proposed and used for conducting studies. Most of these classifications agree that the main purpose of the community plays a crucial role for classification [31],[32]. The purpose is typically further divided into several subcategories, commonly with the distinctions between work-related purposes, interests and games. Additionally the inclusion of commercial business models [25],[31], origin of content [30], or structural elements such as size and duration of existence [32] have been used.

Although all of these classification approaches enlighten specific aspects of online communities, none of the approaches was suited for our work without modification. Consequently we developed our own set of community categories (naturally based on above mentioned work). Concretely communities were differentiated into the following four categories:

Business Networking Sites. These are networking sites that are mainly used to maintain and administer existing and new business contacts. Typical examples for such networks are LinkedIn or Xing.

Social Networks. This term comprehends sites that are mainly used for maintaining private relationships and contacts. The most prominent example for such a site is Facebook, but there are also several other very popular communities such as Orkut[7] or StudiVZ[8].

Content and Media Sharing Networks. On these sites the major focus is on sharing content with others and not on maintaining relationships (even though sharing contents is an important aspect of maintaining relationships). Typically networks are specialised on different types of media, e.g. Flickr is focusing on pictures and YouTube's[9] main medium is video.

Social News and Bookmarking Sites. These sites are used to share and discover interesting links to news and contents in the web. Sites can be more focused on the collaborative bookmarking aspect (e.g. del.icio.us[10]) or the social news aspect (e.g. Digg[11]).

4.2 Common Elements

Another problem when comparing information disclosure on different networking sites is that each site uses its specific set of data. In order to be able to identify similarities and differences in information disclosure behaviour of different types of sites a set of common elements that the sites typically share must be identified. Stutzman [38] provides a helpful comparison of such elements for different social networking sites. Gross et al. [13] used a more differentiated approach and categorized profile names and images according to their consequences for identifiability. Lederer et al. [22] used an ordinal precision scale of personal information dimensions, with the levels precise, approximate, vague and undisclosed.

For our study we needed to define our own set of minimum comparable items, as in contrast to the work described above we were interested in users of a wide variety of online communities. We therefore performed a substantial analysis of different popular networks in all the targeted contexts and the data structures they use. We identified the following minimum list of common elements the users can communicate at almost all of the platforms:

Real name	Email-address
Nickname	Physical address
Picture of user	Phone number
Date of birth	Instant messaging contact
Network of friends	Link to Website of user

4.3 Automatic crawling of profiles versus self-reporting questionnaires

In general there are two major possibilities to study the information disclosure of people: by analysing actual profiles of users, typically collected by automatic crawling of profiles (used e.g. by [22],[26],[12],[30]) or by self-report data, typically collected by online questionnaires (used e.g. by [20],[27]) and paper-based questionnaires (e.g.[15]).

[7] http://www.orkut.com
[8] http://www.studivz.net/
[9] http://www.youtube.com/
[10] http://delicious.com/
[11] http://digg.com/

Analysing the actual data users provide on online networks allows collecting more data. The collected information is more precise, and researchers can use this data for further analysis. Unfortunately there are some serious limitations with this approach: Except in special cases, researchers do not have access to these data as this information is managed by the networking providers. Therefore automatic crawlers have frequently been used to collect the data. This has several important limitations: information not being available to the public can not be accessed in this way. Information only available for friends within the network can be accessed to a certain limit by requesting to become a friend, which implies a serious bias as only networks that are connected and users who easily accept friendships are researched. Additionally one has to consider that - ironically because network providers want to safeguard the privacy of their users - measures against crawling are implemented.

The other possibility to collect data is to rely on the user reporting its information disclosure behaviour by him/herself. This approach overcomes several of the above mentioned problems, but it also introduces its own weaknesses. Self-report data is not as precise as crawled data due to involuntary errors as well as deliberate misinformation due to social expectancy. Online questionnaires have an expected increased amount of self-selection i.e. users that decide not to participate in an online questionnaire might be different from users that agree on joining a questionnaire.

Due to our focus on numerous communities (which would make crawling very inefficient as the mechanism would have to be adapted to all communities) and our interest in publicly available data and information only for 'friends' we decided to follow a questionnaire-based approach. Besides we want to mention that we explicitly encourage researchers to complement our work by using crawling-based data.

4.4 Limitations of the study

It should be noted that the nature of the sampling method (see below) and the self-selection of respondents may have influenced the pattern of responses and overall levels of activity. Future research may wish to study a wider group of participants, or attempt to identify patterns of usage amongst non-respondents compared to respondents.

Another limitation of our study lies in a strong European and German-speaking dominated group of participants. As privacy-related behavior may be influenced by cultural factors, application of theses results to other populations should be done carefully and further research should indicate which role cultural aspects play with regard to information disclosure in online communities.

Due to our focus on comparability between communities we use a minimum set of common profile elements. However, user typically provide much information that is related to the purpose of the community and that was not covered by our questionnaire. Interpretations of our results should consider this fact.

5. PROCEDURE

We developed an online questionnaire to accomplish data about information disclosure behaviour, experiences and usage patterns concerning different online community types (see section 4.1).

The questionnaire was created using Sawtooth software and uploaded to our website. Participants were recruited by different channels. At first, all participants of our test user database, where people interested in participating in usability tests can sign up (which specified they use the internet intensely), were invited by mail. Several universities were contacted and asked for permission to distribute the questionnaire within their institutions. Therefore a larger audience of students could be reached. Universities from three countries (Austria, Denmark and Australia) offered their cooperation.

Additionally online forums, where we expected users of online communities to be present (e.g. gaming forums), were used to distribute the questionnaire. Furthermore several social and business networks (e.g. Xing, Facebook, etc.) were employed to post links to the questionnaire.

As an incentive users were given the possibility to participate in a raffle for prices. Amazon.com gift vouchers were chosen as the main incentive, altogether 50 gift vouchers with a value of 10 Euros, 5 vouchers with a value of 30 Euro and one voucher with a value for 50 Euro were raffled.

The questionnaire was available between Thursday, 29th May 2008 and Wednesday, 11th June 2008.

6. PARTICIPANTS

In total we received 856 completed questionnaires. The questionnaire's respondents came from different countries, with a vast majority of participants coming from German speaking countries (Austria: 632, Germany: 79), but also a notable amount of users from other countries in Europe (e.g. Belgium:48) and even the United States (28) participated in the questionnaire.

The distribution of participants with regard to their sex was almost equal with a slight majority of male users (44.6percent female versus 55.4percent male).

The age distribution of the participants can be found in Table 2 below. A predominant majority of users was in the age group of 21 to 30 years. This probably is caused by three effects: Firstly, this age group is most active in online communities. The popularity of Facebook and similar networks in universities is a good indicator for this. Secondly, students were directly recruited for the questionnaire through contacts to different universities. Thirdly, young people in education are typically easier to be motivated by a small amount of money than older people with an income.

Table 2: Age Distribution of Study Participants

Age group	<15	15-20	21-30	31-40	41-50	51-60	60+
Frequency	6	68	558	130	52	28	14
Percent	0.7	7.9	65.2	15.2	6.1	3.3	1.6

In compliance with the results regarding age described above, the dominating professions of the questionnaire participants were student or pupil (52.8 percent, 452 persons). Nevertheless 44.4 percent of the respondents indicated that they are employed or self-employed.

More than half of the questionnaire participants have a university or college degree (53.2 percent), and 40.2 percent finished grammar school or similar. Only few participants with compulsory or apprentice level education participated (6.7 percent).

Altogether the majority of participants considered themselves to be rather advanced computer users. Out of 856 respondents 470 (54.9 percent) indicated that they rarely need help when working with computers and 219 (25.6 percent) stated that they never need help. Only 2.8 percent stated that they often need help and 16.7 percent said that they sometimes need support.

Concerning the time spent online, the majority of participants (31.8 percent) actively use the internet between 16 and 30 hours a week. Over 80 percent of users spend between 5 and 50 hours a week online. Only 4.8 percent use the internet less than one hour, but 13.6 percent are online even more than 50 hours a week.

7. RESULTS

The following sections summarize the major results of our work. At first we focus on the usage of online communities, analyzing which participants use which communities, whether there are differences related to demographic indicators and if our findings are in line with existing research.

Next we analyze whether there are differences in the information disclosure with regard to different communities and demographics in detail.

Finally we present linear regression models that analyses the relationships between different influencing variables and allows predicting the information disclosure for the different networks.

7.1 Usage of Communities

Our research indicates that the most popular type of online networks are social networking sites. 67.52 percent of participants did actively use a social network. Also professional networks that allow people to establish work-related contacts are very popular: 41.59 percent of participants own and use a business network profile. Content sharing networks such as Flickr or YouTube are next in popularity with 37.38 percent of participants having an active account. Please note that this includes only active users, i.e. the sole usage of e.g. YouTube for consuming contents was not considered as participation in these communities. Social news communities are far less popular with 15 percent.

Another way to analyse the data is to consider the levels of network popularity by type of user attribute. Table 3 presents usage statistics broken down by gender, age group and profession. This breakdown is presented separately for the different types of networks.

The table shows the fraction of participants in percent that are active users of the different types of communities. Statistical analysis using Chi-Square-Tests that calculate whether derivations of the observed distributions from the expected distribution are statistically significant shows significant differences for all demographic variableas in all community types except for social news and bookmarking, where no distribution is significatly deviating.

As can be expected due to typical social roles males are more active in business networks and females in social networks. However, even though there are significant differences in the percentage of males and females using a certain type of network the participation rates are in similar ranges (44.7 vs. 37.7 percent respectively 62.0 vs. 74.4 percent). Bigger differences related to gender can be found in the usage of Content and Media Sharing Networks, where males are more active.

With regard to age the data shows a decline in the percentage of users for social networks, content and media sharing sites. Only for business networks the maximum of users is not below 30 but in the age group of 30 to 40 years.

Not surprisingly, interest in networks is matched to the participants' personal circumstances. Whereas for students and pupils the most used networks are social, employed persons prefer professional networks more than social networks.

Comparing these results to available work that mainly researched northern American populations (e.g. [13][21][24]) the general trend seems to be that the results are rather similar. However, to answer the question whether European and American populations differentiate or not research explicit targeted at this aspect is needed.

Many people use more than one type of online community. In our survey 28.0 percent only used one of the specified networking types, 28.2 percent used two, 19.3 percent used three and 8.5 percent used four or more types of networking sites.

Table 3: Usage of the different types of communities

	Businness Networks	Social Networks	Content & Media Sharing	Social News & Bookmarking
N	356	578	320	136
Gender				
Male	44.73	62.03	44.09	15.61
Female	37.70	74.35	29.06	16.23
Age Group				
15-20	7.35	91.18	42.65	13.24
20-30	43.73	79.39	42.11	17.20
30-40	54.62	38.46	28.46	16.92
40-50	44.23	25.00	19.23	11.54
50+	28.57	16.67	14.29	7.14
Profession				
Student/Pupil	27.56	84.00	42.67	15.11
(Self)Employed	59.95	50.13	31.30	17.24
Other	21.74	34.78	30.43	13.04

Table 4: Information disclosure to friends and unknown persons

	Business Networks		Social Networks		Content & Media Sharing		Social News & Bookmarking	
	Unknown Persons	Friends	Unknown Persons	Friends	Unknown Persons	Friends	Unknown Persons	Friends
Real name	68.5	96.6	55.0	88.2	10.9	31.8	30.1	47.7
Nickname	53.7	59.3	65.1	73.8	75.3	84.7	67.6	78.5
Picture of user	62.9	85.4	65.7	91.1	20.3	35.6	33.8	42.6
Date of birth	25.3	79.0	42.6	82.2	12.5	28.4	16.2	30.2
Network of friends	28.1	86.2	39.8	88.2	20.6	39.7	16.2	35.3
Email-address	8.7	62.6	12.5	64.7	8.1	35.9	16.9	38.2
Physical address	2.2	42.1	2.8	29.6	1.3	11.9	2.2	10.3
Phone number	3.1	46.1	2.1	29.1	1.3	12.6	1.5	8.1
Instant messaging cont.	15.7	63.2	17.6	60.0	9.1	26.9	11.0	21.3
Website	42.7	69.1	27.2	50.4	24.1	34.1	30.1	41.9

Analysis of which types of communities are used together by the same users show, that two combinations are especially frequent: social and professional networking is a common combination, and social networking and content and media sharing sites are used frequently together. In contrast, the combination of business networking and content and media sharing is much less frequent.

7.2 Disclosure of information

Table 4 provides an overview on how many users reveal which information to the entire network or only to friends.

A closer look on the percentages indicates that much more information is provided in networking sites with a social and business context than in other types of communities. As the possibility to correctly identify others and link their profile to a real person is essential for networking, it is easily understandable that users disclose this information to the entire network because otherwise these sites would not be able to provide their value for the user.

People also seem to be much more careful about disclosing their contact information (physical address and phone number) to the entire network than with other items such as name or date-of-birth.

Next, using logistic regression analyses, we analyzed the relationship of several factors and information disclosure on the different communities. To be able to do so we used a simple measure for characterizing information disclosure to friends and unknown persons based on the above described common elements of communities. Information disclosure to unknown persons is calculated as the number of disclosed items divided by the number of available items. Information disclosure to friends is calculated by the number of items disclosed to friends and unknown persons divided by the number of available items. According to these calculation methods, resulting values can vary between 0 and 1, with 0 meaning the person does not disclose any information at all and 1 meaning all information is made available.

The findings presented in Table 5 shows several interesting aspects. Students and pupils seem to be more freehanded in providing information than employed or self-employed persons. Please note that this is not an effect of age, as this factor was included in the model and did not show any significant influence.

There is also a marginally significant effect of gender on information disclosure to friends: Women seem to be more cautious than their male counterparts in providing information about themselves to friends.

Table 5: Results of logistic regression analyses for information disclosure to unknown persons and to friends

	Unknown Persons $F_{9,855}$=3.3, p<.000 R^2=.035			Friends $F_{9,855}$=9.1, p<.000 R^2=.088		
	β	t	Sig	β	t	Sig
Gender	-.015	-.048	.683	-.067	-1.913	.056
Age	.010	.278	.781	-.023	-.669	.503
Student/Employed	.117	2.947	**.003**	.087	2.245	**.025**
Computer Know.	.036	1.028	.304	-.025	-.735	.463
Online Time	-.026	-.705	.481	.016	.435	.663
Business Networks	-.019	-.515	.607	.111	3.029	**.003**
Social Networks	.087	2.260	**.024**	.112	2.993	**.003**
Content Sharing	-.095	-2.512	**.012**	-.239	-6.490	**.000**
Social News and Bookmarking	-.010	-.268	.789	.116	-3.284	**.001**

Table 6: Regression models for information disclosure to unknown persons

	Business Networks $F_{9,355}$= 7.1, p<.000, R^2=.16			Social Networks $F_{9,577}$= 16.9, p<.000, R^2=.21			Content & Media Sharing $F_{9,319}$= 6.8, p<.000, R^2=.17			Social News & Bookmarking $F_{9,135}$= 4.8, p<.000, R^2=.26		
	β	t	Sig	β	t	Sig	β	t	Sig	β	t	Sig
Gender	-,060	-1,136	,257	-,026	-,650	,516	,021	,386	,699	,046	,553	,581
Age	,044	,844	,399	-,033	-,843	,400	-,016	-,290	,772	,071	,892	,374
Student/Employed	,142	2,771	**,006**	,180	4,570	**,000**	,035	,639	,523	,192	2,365	**,020**
Computer KnowHow	,072	1,417	,157	,107	2,791	**,005**	,063	1,195	,233	,069	,863	,390
Online time	,082	1,584	,114	,042	1,055	,292	,079	1,444	,150	-,075	-,877	,382
Time on network	,098	1,954	,051	,029	,706	,481	,041	,752	,453	,147	1,817	,072
N Friends	,027	,535	,593	,042	1,063	,288	,194	3,513	**,001**	-,044	-,532	,596
Trust in network	-,259	-5,108	**,000**	-,197	-4,982	**,000**	-,269	-4,967	**,000**	-,439	-5,564	**,000**
Change Settings	,146	2,861	**,004**	,296	7,794	**,000**	,072	1,375	,170	,089	1,117	,266

Regarding the different types of networks the results show that there is a clear influence of the network type on the amount of disclosed information to friends. All four network variables in the model show significant relationships to the information disclosure to friends. In the model for disclosure to unknown persons however only social networks (increasing) and content sharing networks (decreasing) show such relations.

To better understand the varying information disclosure behavior in the next step we analyzed which variables are good predictors for the information disclosure behavior on the different types of networks.

7.3 Information disclosure to unknown persons

In addition to the demographic and usage data in the general model in the network specific models we included variables specific to these networks in the analysis. In detail the self reported average time on the network, the number of friends, whether the user trust the network or not and whether they have changed the default privacy settings were added.

Table 6 shows the results of the linear regression models for all four network types. As can be seen in the second line of the table the calculations resulted in a significant overall model for all networks.

In accordance with the results of the previous section the model shows that gender and age do not seem to be an important influencing factor for the information disclosure behavior of people towards unknown persons. Also average online time and time spend on the regarding network are not related to the information disclosure.

Students tend to provide more information than professionals in all networks except content and media sharing. This probably can be explained by the fact that content and media sharing networks frequently are used to share also contents with questionable copyright status and therefore users might be especially cautious of providing information that allows identifying them.

Another interesting aspect of content and media sharing sites is that these are the only networks - contrary to our expectations – were the analysis shows a significant relation between the number of friends and the disclosed information. We think this effect can be related to different usage purposes of sharing networks. Users might want to share contents with everybody, or they may use such sharing sites to maintain existing relationships. In the latter case the user will have more friends in the network, and he will probably also provide more information, so other users can find and identify him.

Not surprisingly trust in the network is related to the information disclosure behavior in all networks. The more the user trusts the provider of the site that he is handling the data with care the more information he provides.

In case of social and business networking sites also whether the user has changed the privacy settings or not is a predictor for their handling of data. In case the user had changed the privacy settings it can be expected that they provide less information to unknown persons.

Interestingly Computer know-how has a significant relation to the information disclosure on social networks. The more experienced and skilled a user is the more information he does provide to unknown persons.

The analysis shows also two marginally significant relationships. Time spend on the network seems to be related to the information disclosure to unknown persons in business networks and on social news and bookmarking sites.

Table 7: Regression models for information disclosure to friends

	Business Networks $F_{9,355}$= 7.6, p<.000, R^2=.17			Social Networks $F_{9,577}$= 10.7, p<.000, R^2=.16			Content & Media Sharing $F_{9,319}$= 6.6, p<.000, R^2=.16			Social News & Bookmarking $F_{9,135}$= 4.2, p<.000, R^2=.23		
	β	t	Sig	β	t	Sig	β	t	Sig	β	t	Sig
Gender	-,079	-1,522	,129	-,180	-4,352	**,000**	-,088	-1,616	,107	,142	1,660	,099
Age	,026	,503	,615	-,067	-1,643	,101	-,010	-,192	,848	,124	1,529	,129
Student/Employed	,165	3,228	**,001**	,141	3,444	**,001**	,106	1,934	,054	,181	2,196	**,030**
Computer KnowHow	,002	,039	,969	-,033	-,820	,412	-,094	-1,761	,079	-,070	-,866	,388
Online time	,061	1,170	,243	,081	1,939	,053	-,027	-,483	,629	-,046	-,530	,597
Time on network	,081	1,626	,105	,032	,747	,456	,051	,933	,352	,088	1,070	,287
N Friends	,026	,517	,605	,031	,761	,447	,222	4,012	**,000**	,026	,308	,759
Trust in network	-,322	-6,394	**,000**	-,258	-6,240	**,000**	-,248	-4,560	**,000**	-,388	-4,839	**,000**
Change Settings	-,175	-3,464	**,001**	-,076	-1,926	,055	-,071	-1,355	,176	-,048	-,597	,551

7.4 Information disclosure to friends

In general the results of this analysis show the same trends as the analysis in the previous steps. Profession of the users i.e. employed or student (only marginally significant for content and media sharing) and trust in the network are predictors for the information disclosure for all networks, the number of friends is especially important in content and media sharing networks and the change of privacy settings is relevant in business and social networks.

There is however a highly significant new aspect, that was not present before. Gender seems to play an important role for social networks. Here males to provide more information than females. We think this result is related to the in western societies traditionally more extrovert role of males in social relationships. An alternative explanation might relate this result to differences in the management of risks of males and females, with males typically being more willing to take risks.

8. DISCUSSION & CONCLUSION

In this paper we provided an overview of the information disclosure behavior on different communities and the influence of demographic, context and usage variables on the willingness to disclose personal information.

In accordance with existing research our study found that users of different online sites disclose sensitive information rather freely. However, extending existing research we could analyze the differences in information disclosure on different networks and identify several variables that are of main importance in this context.

There are significant differences in behavior and needs of users depending of the type of community they are in. There is an important differentiation between sites intended for networking (social or professional) and sites where networking and the provision of data serves a secondary purpose. This suggests the interpretation that users of networks typically only provide the information that is required to achieve the maximum gains of the membership.

According to our results the demographic variables gender and age are not relevant for the information disclosure behavior in different community types whereas the distinction between pupils/students and (self-)employed persons makes a relevant difference.

9. FUTURE WORK

In future we plan to expand our research with two important aspects. First we also want to include the effects of personality traits into our work. Especially we want to study whether extroverts and introverts follow different patterns in their online behavior with regard to information disclosure and privacy. Second, we want to expand our work and include network-specific content elements in the overall analysis of information disclosure.

Another area we think deserves further research is the development of information disclosure behavior over time i.e. are there typical patterns for beginners versus experienced users, does the behavior evolve and change together with the membership time or is it a rather fixed behavior that is not influenced by time.

Last but not least we think future research is needed on what happens when these networks become mobile.

10. AKNOWLEDGEMENTS

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2011) within the PICOS project (grant agreement n°215056).

11. REFERENCES

[1] Acquisti, A., Gross, R. Imagined Communities: Awareness, Information Sharing, and Privacy on the Facebook, in Post-Proceedings of the 6th International Workshop on Privacy Enhancing Technologies (PET 2006), Cambridge, UK, June 28-30, 2006, Revised Selected Papers, Springer LNCS, Volume 4258/2006, 36-58.

[2] Acquisti, A., Grossklags, J. Privacy Attitudes and Privacy Behaviour, in Camp, J. and Lewis, R. (Eds.) Economics of Information Security, 2004, Springer, Vol. 12, New York, NY, 165-178, Ch. 13.

[3] Ahern, S., Eckles, D., Good, N. S., King, S., Naaman, M.,and Nair, R. 2007. Over-exposed?: privacy patterns and considerations in online and mobile photo sharing. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (San Jose, California, USA, April 28 - May 03, 2007). CHI '07. ACM, New York, NY, 357-366.

[4] Andrews, D.C. Audience-specific online community design. Commun. ACM 45, 4 (Apr. 2002), 64-68.

[5] boyd, d.m. Friendster and publicly articulated social networking. In CHI '04 Extended Abstracts on Human Factors in Computing Systems (Vienna, Austria, April 24 - 29, 2004). CHI '04. ACM, New York, NY, 1279-1282. 2004.

[6] boyd, d.m. Reflections on Friendster, Trust and Intimacy. Ubiquitous Computing (Ubicomp 2003), Workshop application for the Intimate Ubiquitous Computing Workshop. Seattle, WA, October 12-15, 2003.

[7] Cnet. ComScore: Facebook is beating MySpace worldwide, http://news.cnet.com/8301-13577_3-9973826-36.html. Last access: 2008-12-15

[8] Ellison, N., Steinfield, C., Lampe, C. Spatially Bounded Online Social Networks and Social Capital: The Role of Facebook. paper to be presented at the Annual Conference of the International Communication Association (ICA), June 19-23, 2006 in Dresden, Germany

[9] Flinn, S., Lumsden, J. "User Perceptions of Privacy and Security on the Web," The Third Annual Conference on Privacy, Security and Trust (PST 2005). St. Andrews, New Brunswick, Canada. October 12-14, 2005.

[10] FraunhoferInstitut für Sichere Informationstechnologie SIT, 2008, Privatsphärenschutz in Soziale-Netzwerke-Plattformen. http://www.sit.fraunhofer.de/fhg/Images/SocNetStudie_Deu_Final_tcm105-132111.pdf. Last access: 2008-12-15

[11] Fuller, A. Employers snoop on Facebook, The Stanford Daily, 20 January 2006 http://daily.stanford.edu/article/2006/1/20/employersSnoopOnFacebook. Last access: 2008-12-15

[12] Gjoka, M., Sirivianos, M., Markopoulou, A., and Yang, X. 2008. Poking facebook: characterization of osn applications. In Proceedings of the First Workshop on online Social Networks (Seattle, WA, USA, August 18 - 18, 2008). WOSP '08. ACM, New York, NY, 31-36.

[13] Gross, R., Acquisti, A., and Heinz, H.J. Information revelation and privacy in online social networks. In Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society (Alexandria, VA, USA, November 07 - 07, 2005). WPES '05. ACM, New York, NY, 71-80.

[14] Hagel III, J., Armstrong, A.G. Net Gain: Expanding Markets through Virtual Communities. Harvard Business School Press.

[15] Hargittai, E. Whose Space? Differences Among Users and Non-Users of Social Network Sites. Journal of Computer-Mediated Communication, 13(1). 2007

[16] Hogben, G.H. (Editor). Security Issues and Recommendations for Online Social Networks. ENISA Position Paper No. 1. October 2007

[17] Hui, K., Tan, B. C., and Goh, C. 2006. Online information disclosure: Motivators and measurements. ACM Trans. Interet Technol. 6, 4 (Nov. 2006), 415-441.

[18] Jagatic, T. N., Johnson, N. A., Jakobsson, M., and Menczer, F. 2007. Social phishing. Commun. ACM 50, 10 (Oct. 2007), 94-100.

[19] Jensen, C., Davis, J., and Farnham, S. Finding others online: Reputation systems for social online spaces. Proc. SIGCHI Conference on Human Factors in Computing Systems, pp.447–454, 2002.

[20] Joinson, A.N. 'Looking at', 'Looking up' or 'Keeping up with' People? Motives and Uses of Facebook. In Proceedings of CHI 2008, April 5-10, 2008. Florence, Italy

[21] Krishnamurthy, B. and Wills, C. E. 2008. Characterizing privacy in online social networks. In Proceedings of the First Workshop on online Social Networks (Seattle, WA, USA, August 18 - 18, 2008). WOSP '08. ACM, New York, NY, 37-42.

[22] Krishnamurthy, B., Gill, P., and Arlitt, M. 2008. A few chirps about twitter. In Proceedings of the First Workshop on online Social Networks (Seattle, WA, USA, August 18 - 18, 2008). WOSP '08. ACM, New York, NY, 19-24.

[23] Lampe, C., Ellison, N., and Steinfield, C. A face(book) in the crowd: social Searching vs. social browsing. In Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work (Banff, Alberta, Canada, November 04 - 08, 2006). CSCW '06. ACM, New York, NY, 167-170.

[24] Lampe, C.A., Ellison, N., and Steinfield, C. A familiar face(book): profile elements as signals in an online social network. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (San Jose, California, USA, April 28 - May 03, 2007). CHI '07. ACM, New York, NY, 435-444. 2007

[25] Lechner, U., Hummel, J. Business Models and System Architectures of Virtual Communities: From a Sociological Phenomenon to Peer-to-Peer Architectures. International Journal of Electronic Commerce. Volume 6. Number 3. 2002

[26] Lerman, K. and Galstyan, A. 2008. Analysis of social voting patterns on digg. In Proceedings of the First Workshop on online Social Networks (Seattle, WA, USA, August 18 - 18, 2008). WOSP '08. ACM, New York, NY, 7-12.

[27] Li, C. How Consumers Use Social Networks. Forrester Research. June 21, 2007

[28] LinkedIn. About LinkedIn. http://www.linkedin.com/static?key=company_info. Last access: 2008-12-15

[29] Mislove, A., Koppula, H. S., Gummadi, K. P., Druschel, P., and Bhattacharjee, B. 2008. Growth of the flickr social network. In Proceedings of the First Workshop on online Social Networks (Seattle, WA, USA, August 18 - 18, 2008). WOSP '08. ACM, New York, NY, 25-30.

[30] Olsson, T., Toivola, H., Väänänen-Vainio-Mattila, K. Exploring Characteristics of Collective Content — A Field Study with Four User Communities. In Proceedings of CHI 2008, April 5-10, 2008. Florence, Italy

[31] Porter, C. E. A Typology of Virtual Communities: A Multi-Disciplinary Foundation for Future Research. JCMC 10 (1), Article 3, November 2004.

[32] Preece, J., Maloney-Krichmar, D., Abras, C. History of Emergence of Online Communities. In B. Wellman (Ed.), Encyclopedia of Community. Berkshire Publishing Group, Sage. 2003

[33] Preibusch, S., Hoser, B., Gürses, S., and Berendt, B. Ubiquitous social networks- opportunities and challenges for privacy-aware user modelling. In Proceedings of the Data Mining for User Modelling Workshop (DM.UM'07) at UM 2007, Corfu, June 2007

[34] Renaud, J.-F. Integrate Social Media into the Web Strategy: an overview. http://www.adviso.ca/en/integrate-social-media-into-the.html. Last access: 2008-12-15

[35] Riegelsberger, J., and Vasalou, A. Trust 2.1: advancing the trust debate. In CHI '07 Extended Abstracts on Human Factors in Computing Systems (San Jose, CA, USA, April 28 - May 03, 2007). CHI '07. ACM, New York, NY, 2137-2140. 2007

[36] Seay, A. F., Jerome, W. J., Lee, K. S., and Kraut, R. E. 2004. Project massive: a study of online gaming communities. In CHI '04 Extended Abstracts on Human Factors in Computing Systems (Vienna, Austria, April 24 - 29, 2004). CHI '04. ACM, New York, NY, 1421-1424.

[37] Stanoevska-Slabeva, K. and Schmid, B.F. A typology of online communities and community supporting platforms. In System Sciences, 2001. Proceedings of the 34th Annual Hawaii International Conference on Media and Communication, 3-6 Jan. 2001

[38] Stutzman, F. An Evaluation of Identity-Sharing Behavior in Social Network Communities. International Digital and Media Arts Journal, 3(1). 2006

[39] Van House, N. A. 2007. Flickr and public image-sharing: distant closeness and photo exhibition. In CHI '07 Extended Abstracts on Human Factors in Computing Systems (San Jose, CA, USA, April 28 - May 03, 2007). CHI '07. ACM, New York, NY, 2717-2722.

[40] Vasalou, A., and Riegelsberger, J. Recovering Trust and Avoiding Escalation: An overlooked design goal of social systems. In Proceedings of CHI 2008, April 5 – April 10, 2008, Florence, Italy.

[41] Xing. Dritte international Studie "Kommunikation & Networking im Internet", www.xing.com, 2007