Information Revelation and Computer-Mediated Communication in Online Social Networks

R.J.Davey and A.D.Phippen

Centre for Information Security and Network Research,
University of Plymouth, Plymouth, United Kingdom
e-mail: info@cscan.org

Abstract

There is much information being disseminated through user profiles and communication channels on social networking websites. This publication examines user demographics, the types, and volume of information exposed through social networking profiles found in the conceptual region of Plymouth.

Keywords

Information Revelation, Social Networking, Personal Privacy

1 Introduction

Social networking websites complement established real life social patterns and as a result, have rapidly grown in popularity throughout recent years. The volume and persistent nature of user visits to such websites demonstrate the level of integration and the grasp that social networking has on the lives of its users. However, there is growing concern over the numerous potential threats afforded through the revelation of information on social networking websites. These threats pose a genuine risk and are highly publicised through high profile media channels. It is unclear as to what information is being exposed on social networking websites and the communication channels that exist within them.

In this paper, patterns of detailed information revelation and the exposure of personal information through computer-mediated communication on social networking websites within the region of Plymouth are presented. The region of Plymouth was selected since it holds local interest, and taking a narrowed geographical picture of the SN state helped differentiate this study from other similar online studies. This paper also highlights the potential for harm, explains the superficial attitudes of social network users towards personal privacy, describes the ethical issues regarding online research, and presents an analysis of the findings.

2 The Evolution of Online Social Networking

Rheingold (1993) declares that the convergence of technologies with everyday life was foreseen in the late 1970's and was predicted to affect everyone, whether they knew or cared about the future direction of technology. This claim was made in the

time of simple textual based social networks, and these were still in early stages of development and had not yet finished expanding into the integrated communities, as they are known today. The first truly collaborative and user-driven websites were established around 2003 as a result of new 'Web 2.0' technologies. These websites gave users the opportunity to create their own content and were generally divided into one of two types; a site gathering information as part of a collaboration, and a site that hosts and allows interaction between a collection of personal profiles. This user generated content and interaction between profiles provided the foundation for the development of social networking websites.

The launch of MySpace in 2003 allowed bands to promote their music. The website provided a place where young people could post pictures of themselves, find friends, and let people listen to their music. The demographics of MySpace became evident around 2004 with the launch of Facebook. Facebook was the creation of a former Harvard student and membership was restricted to Harvard students. The website was later opened through invitation only to other educational institutions, and quickly became a cultural status among teenagers (Boyd, 2007). Social networking websites are rapidly changing and the latest addition is the developer-orientated architecture of Facebook, which allows the creation of 'applications' or embeddable chunks of code that allows for the incorporation of external content with interactive user participation. This has led to concerns in the ethical view held by the developers who have control over the personal data being passed between these applications.

3 Personal Privacy

The labels 'public' and 'private' are prevalent online as metaphors for Internet interaction, as they can be easily interpreted. Nevertheless, while a social network profile page may be publicly viewable and accessible, it does not ensure that the user recognises the extent of the exposure of the information and interaction given on that page, which may be deemed private by the individual. It has been noted that connecting to public forums from private homes and workplaces can give the impression of privacy (Rheingold, 1993). Social network users often consider online identities separate from those offline, and this also applies to the information they disclose online. This gives a possible explanation to why potentially sensitive information can sometimes be disclosed (Stern, 2004). However, the Bakhtinian theory contradicts the expected perceptions of social network users. When translated into a virtual context, this theory shows that people can participate in online conversations and other online social activities, only while understanding and respecting its privacy space (Bakhtin, 1984). There is no particular attitude shown towards privacy from teenagers and younger Facebook users. However, it was confirmed in a recent study that some of the younger Facebook users are aware of active threats involved with the exposure of personal information, but did not understand the potential impact of the risks and were still happy to disregard protective advice. This was generally due to hidden motives in the hegemonic demographic. On the other hand, the subaltern teenagers had a better understanding of the consequences of disregarding personal privacy (Boyd, 2007). Users attitude towards breached privacy differs but is generally found to be fairly weak. This could be down to the lack of awareness regarding support, or not feeling empowered

enough to take legal action on infringement of personal privacy due to social or financial status (Atkinson, 2007).

3.1 Privacy Implications and the Potential for Harm

Before the creation of social networking websites, the only accessible form of personal information on the Internet was through a personal homepage or group bulletin board. This data was selective with a strong sense of self-presentation, and was hard to process in large quantities since the information was not standardised (Bober, 2004). Facebook is growing rapidly and this has a direct effect on the amount of standardised information available. Prior to the development of social networking websites, there was much emphasis on keeping data private. However, many successful Internet start-ups such as Flickr, initially disregarded personal privacy. This led to the sharing of personal information and opened the doors to social networking, but is nevertheless now leading to issues in personal privacy (Torkington, 2005). The Internet allows data to be moved, transformed or manipulated, which raises the core issues of authorship and authenticity of material. This can be expanded to cover the topics of confidentiality, integrity, availability and accountability (Furnell, 2005). As a result, it is not the technology that should cause concern, but the possession of the information and how it could be used to cause harm. Many studies have been conducted to evaluate the extent of the potential for harm through probing social network accounts for exposed information. A survey of 800 parents and children has recently shown that 25 percent of the children when questioned had given out personal information. In contrast, only 13 percent of the children's parents were aware of the data being publicly posted (Vine, 2008). The target group in this survey was aimed particularly towards children, as younger age groups are presumed to be at the highest risk. The Sophos Facebook survey (2007) showed that this is not the case since 41 percent of randomly selected social network users are willing to share their personal information with potential identity thieves.

4 Information Revelation in Online Social Networks

It has been established that websites can be used as objects of analysis, due to their potential source for both qualitative and quantitative content. The type of information available in this environment is the same as traditionally available to the researcher, such as interview, observational, document, and audio-video materials (Creswell, 2007). In the case of this study, the natural setting spans several popular social network domains, linked through the conceptual physical location of the users' online identity. This can in few situations include other sites through an online 'mashup' of web applications, ties to a personal homepage, personal Blog, or re-identification across other social networks. However, to define clear boundaries for the study, the research was limited to the data provided directly on users' social network profile pages.

4.1 Selection of Population

Social networking websites have the potential to reach users internationally due to the size of the audience and the scope of the Internet. Facebook is becoming popular among teenagers, but more so for those from a wealthy demographic and those with

a higher emphasis on education. The main factor in the educational divide among the social networking websites could be a product of Facebook's origin, when in its infancy it was initially limited to university students and individuals with an email address from an academic institution. A strong demographic divide is apparent between social networking websites, and one community cannot be given as an accurate representation of the ethnicity, educational background or income of the population at large (Boyd, 2007). Nevertheless, some studies have shown no real difference in demographics, but this is most likely due to the fact that these studies were conducted on a limited self-selected group or community (Ellison *et al*., 2007).

In June 2008, a preliminary study was undertaken with a small sample of 50 profiles. Later in July 2008, the profiles of 384 users across Facebook, MySpace and Bebo were gathered in accordance with our initial findings. In cases such as this where a population is spread over multiple defined target groups, stratified or weighted sampling is usually employed. By reflecting the distribution of the population based on a criterion, this ensured an accurate representative sample of the population was held and an extra increment of precision was injected into the probability sampling process. The sample population was stratified using the national population statistics for active online social network use (Burmaster, 2007). The sampling fraction was calculated and indicated that the probability of inclusion in the sample was 1 in 190. Using the sampling fraction, it was expected to include 144 Facebook, 141 MySpace and 99 Bebo profiles in our final sample.

4.2 Demographics

It is important to note at this point the similarities between the preliminary findings and the main findings to establish reliability within our results. A line graph of the user age distribution across the pilot and final sample groups is given in Figure .

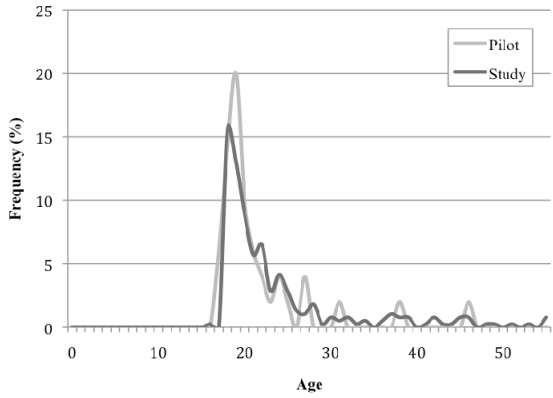

Figure 1: Line graph showing the user age distribution across the pilot and final sample groups

This figure shows a very similar pattern in the distribution of users appearing in the range between 16 and 28 years of age. The most frequent age is 18 years, with

slightly more outliers present in the tailing age groups, although this is expected in a larger sample group. The grouping of these age demographics backs the idea of a 'Generation Y' (Gribben, 2007). This is defined as an ambitious generation born between 1978 and 1998, who have grown up with the Internet and have become accustomed to the freedom and instant global connectivity found online.

Users up to the age of 14 are restricted from registering with any social networking website, due to the terms and conditions of these services. Some users from the age of 14 are allowed use of social networking services, but these profiles carry heavy restrictions preventing them to be searched or browsed, without first directly gaining their consent and confirming them as friends. This explanation justifies the drop off and lack of users bellow the age of 18 in the findings, thus preserving the line of reasoning behind the 'Generation Y' theory.

4.3 Types and Frequency of Information Exposure

These results have shown that information revelation is a genuine issue across the sample group. The frequency of every identifiable piece of information has been measured for each information category, and the percentage of the likelihood of revelation has been calculated. The resulting bar chart can be seen in Figure 2.

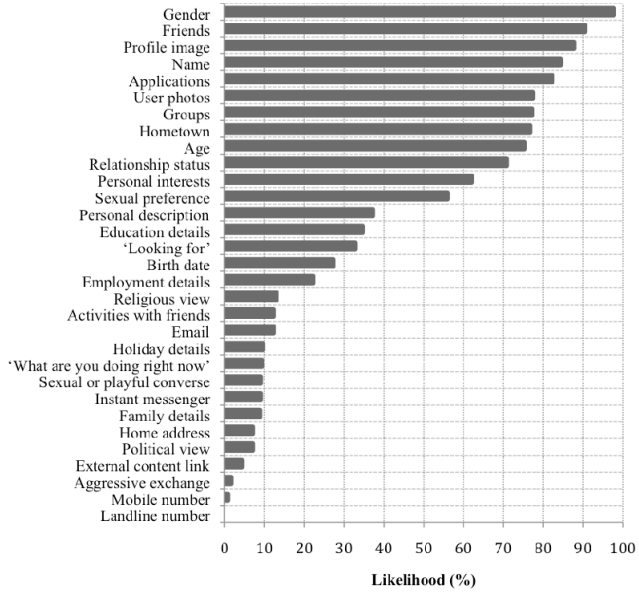

Figure 2: Bar chart showing the likelihood of personal information revelation in the sample group

It was noticed that the social networking website had an effect on the likelihood of the type of information that may be available on any given profile. For example, Facebook users were more likely to either have a political opinion or be willing to reveal it, whereas MySpace users were ready to present their email address. Across

all the target websites in the region, there is much information being disseminated, which could be used to build an accurate personal profile upon each user, and open them to evident risks.

4.4 The Perceived Level of Privacy

It was accepted in Section 3 that the social network users' attitude towards privacy is seen to be fairly weak. This proposition is confirmed by these findings; where users show signs of awareness of privacy, but are still willing to reveal sensitive information or not facilitate the full use of their privacy settings. It is suggested that this could be due to the act of signalling (Gross and Acquisti, 2005). This theory of signalling splits the target group by gender and assumes that males have a significantly higher chance of revealing information. This suggestion is explored in Figure 3 with an abstract view of the users' perceived privacy settings, grouped by gender.

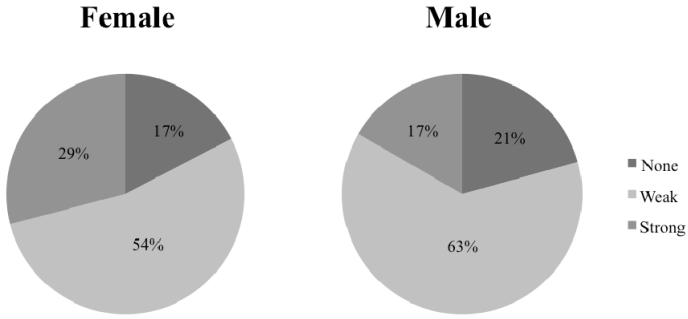

Figure 3: Pie charts showing the presence of user privacy settings grouped by gender in the sample group

As expected, the male group generally featured a lower level of privacy protection. In comparison the female group showed a surprisingly higher perceived level of control over their profiles, with nearly double the amount of profiles having a stronger level of privacy protection.

4.5 Data Validity

In several profiles analysed, it was made apparent through computer-mediated communication channels, that specific items in a profile were noticeably false. Users appeared to possess the attitude that this was intended and valued as a joke, such as a false sexual preference. This could have a negative effect in the reliability and validity of the information in a single instance, but should not negatively affect the study in regards to the risks posed though exposing this or other such information. When the study was in the process of being conducted and false information was identified, it was dismissed at the discretion of the researcher, based on other available information in context. Some of the information recorded in the profiles observed, were noted as containing a false positive. One Facebook profile seemed to possess extremely lax privacy settings, and had much personal information on

display. On deeper analysis of its content it appeared intentionally presented, with the apparent aim of advertising an advance powerboat tuition service that the user offered. This user lacked any form of privacy settings, nevertheless seemed to understand the risks as the information was purposely selected to advertise his business. This demonstrates a particular understanding of privacy and it's application to both personal and professional information.

5 Conclusion

Privacy is a real concept and a growing concern due to the wide scope of the Internet. Social networking websites give the users the right and freedom of control over the flow of their information, but inevitable hidden risks always pose a concern. It can be hard to identify risks to personal privacy online, and measure the resulting impact should a risk materialise, as many risks do not clearly expose themselves or give any details of origin. This study has identified several important characteristics of the risks posed through Plymouth's presence on social networking websites. Users in the region of Plymouth show a willingness to expose information, and this is exposing them to a high risk of online grooming, harassment, and identity theft. Although it is not possible to clearly define the users who may pose many of these threats, they are rarely caused by broken strong ties, but are more likely to be held by 'friends' who can be classed as weak ties. It is inevitable that these weak ties will exist, since it is part of the nature of social network users and is built into the culture surrounding such websites. Most users show awareness of privacy through some form of privacy setting, but they show willingness to share information through unprotected communication channels due to social necessity and trust. Evidence exists in the findings that suggests signalling may share a blame in the revelation of information, but only as an amplification factor in particular information types among the male group. The convergent validity of the data with existing theories supports the validity of the findings and the direction of causality of research is also self-evident. The recommendation is to minimise the escalation of the threats by removing any direct contact information from the view of these low intensity relationships. This is made increasingly possible due to the advancement in the customisation of privacy settings available to the individual. However, further research is needed into the state of the default privacy settings and the users attitude towards myopic discounting.

6 References

Atkinson, S. (2007), "Risk Reduction through Technological Control of Personal Information", Ph.D. Thesis, University of Plymouth.

Bakhtin, M. (1984), *Problems of Dostoevsky's Poetics*, Minneapolis: University of Minnesota Press, ISBN: 978-0816612284.

Bober, M. (2004), "Virtual Youth Research: An Exploration of Methodologies and Ethical Dilemmas from a British Perspective", in Buchanan, E. (Ed.) *Readings in Virtual Research Ethics: Issues and Controversies*, Hershey: Information Science Publishing, ISBN: 978-1591401520.

Boyd, D. (2007), "Viewing American class divisions through Facebook and MySpace", http://www.danah.org/papers/essays/ClassDivisions.html, (Accessed 18 January 2008).

Burmaster, A. (2007), "Facebook is Now UK's Most Popular Social Network", http://www.nielsen-netratings.com/pr/pr_070925_UK.pdf, (Accessed 2 February 2008).

Creswell, J. (2007), *Qualitative Inquiry and Research Design: Choosing Among Five Approaches* (2nd Edition Ed.), California: Sage Publications, ISBN: 978-1412916073.

Ellison, N., Steinfield, C. and Lampe, C. (2007), "The Benefits of Facebook Friends: Social Capital and College Students' Use of Online Social Network Sites", *Journal of Computer-Mediated Communication*, Vol. 12, No. 4, pp1143-1168.

Furnell, S. (2005), *Computer Insecurity: Risking the System*, London: Springer, ISBN: 978-1852339432.

Gribben, R. (2007), "Generation Y talking about a revolution", http://www.telegraph.co.uk/money/main.jhtml?xml=/money/2007/11/29/cmgen29.xml, (Accessed 20 May 2008).

Gross, R. and Acquisti, A. (2005), "Information Revelation and Privacy in Online Social Networks", http:// www.heinz.cmu.edu/~acquisti/papers/privacy-facebook-acquisti-slides.ppt, (Accessed 23 May 2008).

Rheingold, H. (1993), *The Virtual Community: Homesteading on the Electronic Frontier*, New York: Addison-Wesley, ISBN: 978-0262681216.

Sophos (2007), "Sophos Facebook ID probe shows 41% of users happy to reveal all to potential identity thieves", http://www.sophos.com/pressoffice/news/articles/2007/08/facebook.html, (Accessed 1 December 2007).

Stern, S. (2004), "Studying Adolescents Online: A Consideration of Ethical Issues", in Buchanan, E. (Ed.) *Readings in Virtual Research Ethics: Issues and Controversies*, Hershey: Information Science Publishing, ISBN: 978-1591401520.

Torkington, N. (2005), "A Web 2.0 Investment Thesis", http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html?page=4, (Accessed 17 January 2008).

Vine, J. (2008), "One Click from Danger", http://news.bbc.co.uk/1/hi/programmes/panorama/7180769.stm, (Accessed 14 January 2008).