An Evaluation of Identity-Sharing Behavior in Social Network Communities

Frederic Stutzman

School of Information and Library Science
The University of North Carolina at Chapel Hill
213 Manning Hall CB #3456, Chapel Hill, NC 27599, USA
`fred@metalab.unc.edu`

Abstract. Social network communities facilitate the sharing of identity information in a directed network. Compared with traditional methods for identity information disclosure, such as a campus directory, the social network community fosters a more subjective and holistic disclosure of identity information. In the following paper, the results of a quantitative analysis of identity information disclosure in social network communities, as well as subject opinions regarding identity protection and information disclosure are presented. Through comparative analysis, the need for further analysis of the value and jeopardy of identity information sharing in social network communities is identified.

1 Introduction

The management and protection of student identity information is a high priority for academic institutions. Federal legislation, such as the Family Educational Rights and Privacy Act (FERPA), provide guidelines for academic institutions with regards to the disclosure of identity information. In compliance with legislation such as FERPA, and mindful of the prevalence of identity theft[2–4], academic institutions have taken the necessary, difficult steps to protect student identity information. As institutions work to protect student identity information, is it possible that students are actively undermining these protections through participation in social network communities?

In recent years, social network communities (SNC) such as Friendster and MySpace[1] have drawn significant press from the business and academic communities. boyd (sic), in [1] describes SNC's as technologies that enable the public articulation of social networks. Indeed, the inherent sociality of these communities have led to strong adoption trends[2], particularly among the college demographic.

SNC's allow more than the public articulation of social networks; in each service, a user creates a richly detailed personal profile. The data in an individual's profile ranges from the relatively innocuous (favorite book or movie) to the

---

[1] Friendster.com, MySpace.com

[2] As of July 2005, Friendster reports over 17 million users and MySpace reports over 20 million users

potentially invasive (sexual orientation, political views, photo). From a research and administrative standpoint, it seemed wise to evaluate the breadth of SNC penetration, and the scope of identity information shared in SNC profiles on a typical college campus.

Mindful of recent trends in identity theft [2–4], and particularly identity theft on the web [5], a pilot study was commissioned to extract quantitative metrics on student SNC participation and identity information disclosure. The data contained in SNC's, while differing in levels of accessibility [6], is generally trivial for an outsider to access. As we have seen in the work of Hogg et al [7] and Liu et al [8], the notion of outsiders harvesting data in SNC's for ancillary purposes is established. Just as SNC data can be harvested for recommender and reputation systems, third parties may mine a SNC for an individual's identity information. Indeed, SNC's are dramatically changing how identity information is shared online; through this primary analysis, we develop a measure of just how SNC's are redefining the identity sharing behavior of a campus population.

2 Study Perspective

The primary goal of the pilot study was to develop quantitative metrics on SNC participation on a college campus. The secondary goal of the pilot study was to investigate and comparatively analyze population attitudes about participation in SNC's, and population attitudes about online identity sharing in general. Understanding that outsiders (in this case, entities not linked to a social group) in SNC's generally have the lowest level of access to data [9], and that third party identity information harvesters will *at most* be outsiders, the analysis is conducted from the standpoint of the outsider. Additionally, this sets a reasonable baseline for future investigative research from different, more-connected standpoints.

3 Methodology

The pilot study was guided by a number of goals, included among them a viability test for conducting research in SNC's. Research goals were guided by the following questions:

- Which SNC's do students participate in?
- What identity information is disclosed in the SNC's? How does it compare to identity information previously disclosed by the university?
- How much identity information are students disclosing in SNC's?
- What are student opinions about identity information disclosure in SNC's?

3.1 Procedure

A random selection of students were asked to complete a survey about their use of SNC's, and their feelings about disclosure of identity information. The first part

of the survey was entirely quantitative; students indicated which, if any, SNC's they participated in. A list of common SNC's, as well as an option to share other SNC's was made available. In the second part of the survey, students were asked to respond to a number of statements about identity information disclosure, indicating their level of agreement with the statement. The statements dealt primarily with how students feel about their SNC profiles being accessed, and how students feel about sharing identity information in general.

SNC's that occurred more than once in student responses were profiled. Profiling involved the construction of an identity information matrix for each service; respondents who indicated participation were then discovered in the service, and their level of identity disclosure recorded in the disclosure matrix. Student SNC participation data, identity information disclosure matrices, and the opinion data were then analyzed.

3.2 Participants

Of our randomly selected participants (N=200), 19 percent (N=38) completed the survey. Of the respondents, 20 were undergraduates, and 18 were graduate/professional (G/P). We are able to accept the respondents for generalization about the undergraduate and G/P sub-populations ($x^2_{(1)} = .6306, p < .1$).

4 Findings

4.1 Social Network Community Breadth

71 percent of all respondents indicated participation in a SNC, with participation skewing heavily towards undergraduates (90 percent reporting participation) as compared to G/P students (44 percent reporting participation). The most popular SNC was TheFacebook[3], with 90 percent of undergraduates reporting use. Friendster and MySpace were the other common (used by more than two respondents) SNC's reported by respondents.

4.2 Social Network Community Identity Data Analysis

The three common SNC's, TheFacebook, Friendster, and MySpace were profiled for identity information disclosure, and a common element comparison is presented in Table 1. To retain perspective, publicly accessible campus directory information was included in the comparison.

TheFacebook and MySpace requested the disclosure of identity information beyond common elements, as described in Table 2. The non-common elements are presented here to display the notable level of identity information disclosure these communities request. It is important to note that terms have been pooled to handle space considerations, and only when the pooling clearly didn't change the intended meaning of the term.

---

[3] TheFacebook.com

Table 1. Common requested identity disclosure elements in three Social Network Communities, compared with identity information disclosed in a FERPA-compliant student directory service (UNC). Terms have been recoded to handle semantic difference between services. All disclosure elements are optional, except those marked by an asterisk

Common Identity Elements	UNC Dir.	TheFacebook	MySpace	Friendster
Name	Yes	Yes*	Yes*	Yes*
Email Address	Yes	Yes*	Yes*	Yes*
Physical Address	Yes	Yes	No	No
Phone Number	Yes	Yes	No	No
Academic Classification	Yes	Yes*	No	No
Major	Yes	Yes	No	No
Website/Rss	Yes	Yes	No	Yes
Academic Status	Yes	Yes	No	No
Gender	No	Yes	Yes*	Yes*
Hometown	No	Yes	Yes	Yes
Birthdate	No	Yes	Yes*	Yes*
Photo	No	Yes	Yes	Yes
Friend Network	No	Yes	Yes	Yes
Group Affiliation	No	Yes	Yes	Yes
Sexual Orientation	No	Yes	Yes	Yes
Relationship Status	No	Yes	Yes	Yes*
Interests	No	Yes	Yes	Yes
Job/Occupation	No	Yes	Yes	Yes
Favorite Music	No	Yes	Yes	Yes
Favorite Books	No	Yes	Yes	Yes
Favorite Movies	No	Yes	Yes	Yes
Personal Statement	No	Yes	Yes	Yes
Favorite TV Shows	No	No	Yes	Yes
School Information	Implied	Yes*	Yes	Yes
Zip Code	Implied	Yes	Yes*	Yes*
Country	Implied	No	Yes*	Yes*

Table 2. Additional, non-common identity elements requested by TheFacebook and MySpace. Terms have been pooled in cases marked by an asterisk

Service	Identity Element
TheFacebook	AIM Screenname, Favorite Quotes, Summer Plans*, School Course Schedule
MySpace	Heroes, Religion, Drinking Status, Smoking Status, Children, Income, Networking*, Ethnicity, Body Type, Height

4.3 Identity Information Disclosure in TheFacebook

The SNC with the highest level of campus participation was TheFacebook. A relatively new SNC, TheFacebook is heavily utilized by undergraduates (90 percent report use), and lightly utilized by G/P students (22 percent report use). As a result, TheFacebook was selected as the SNC that would be analyzed for student identity information disclosure.

The analysis process is described as follows: for each student that indicated use of TheFacebook, the student's profile is "discovered "in the service. Students are located in the service *only* with information publicly disclosed in the student directory, thereby ensuring that the investigator remains an outsider[4]. Student responses to information requested by TheFacebook are marked in the disclosure matrix as a positive response. Students that indicate use of TheFacebook but aren't found in the service receive are marked with a negative response in each field in the disclosure matrix. No attempts are made to verify the veracity of information disclosed[5].

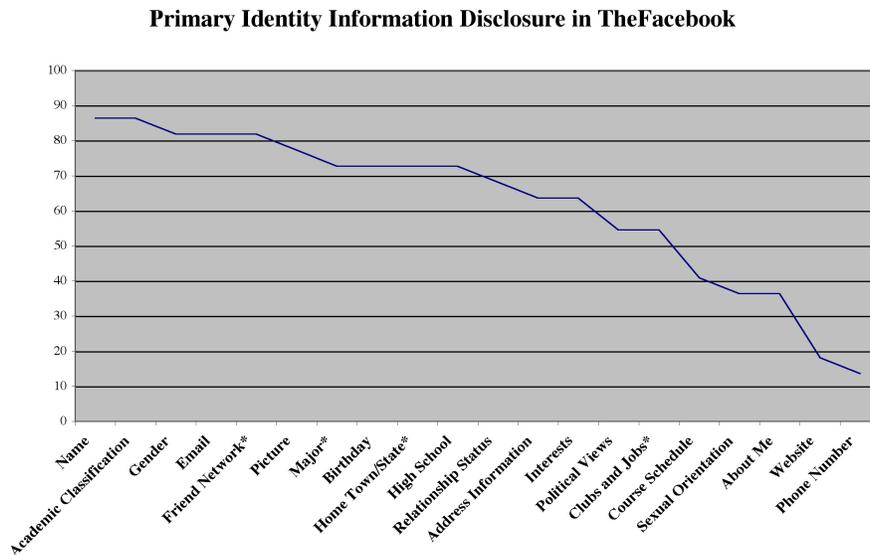**Primary Identity Information Disclosure in TheFacebook**

Fig. 1. This graph explores selected primary identity information disclosed in TheFacebook, indexed by percentage of campus users that disclose the particular element. Elements marked with an asterisk have been pooled.

---

[4] The outsider perspective assumes a crawl of student directory information.

[5] The challenge of verifying the veracity of information disclosed is beyond the scope of this study. Additionally, the veracity of identity information disclosed publicly may be irrelevant to outsiders, especially those who wish to use the information for disingenuous motives.

The results are presented in Figure 1, providing insight into the metrics of identity disclosure in a SNC. It is important to note that to gain access to TheFacebook, an individual must posses an email address that ends with the institution's domain name. This measure exists largely for quality control, rather than as an information security measure. Terms were pooled when appropriate in the analysis presented in Figure 1.

4.4   Student Opinions on Identity Information Disclosure

Students were asked to react to a number of statements regarding identity information disclosure in SNC's. On a scale of 1 to 5, with 1 reflecting a level of strong disagreement with the statement, and 5 reflecting a level of strong agreement with the statement, the responses are presented in Table 3.

Table 3. Average level of student agreement to selected statements about identity information disclosure. The scale ranges from 1, with 1 reflecting a level of strong disagreement with the statement, to 5, with 5 reflecting a level of strong agreement with the statement

Statement	Avg. Resp.
I am OK with friends accessing my SNC profile	4.55
I am OK with family accessing my SNC profile	3.78
I am OK with classmates accessing my SNC profile	3.76
I am OK with strangers accessing my SNC profile	3.15
It is important to me to protect my identity information	4.21
I am concerned with the consequences of sharing identity info.	3.29
I am likely to share my identity information online in the future	3.34
I believe my identity information is well-protected online	2.66

5   Limitations and Future Directions

There are a number of limitations of the pilot study, including a) sample size, b) characteristics of survey respondents, c) lexical differences between SNC's, and d) the effective, but ad-hoc nature of being an outside analyst of SNC's. Concerns a) and b) can be easily addressed in the revised methodology of the full study. The sample size can be increased, and surveying methods may move away from the on-line survey, which may account for a disproportionately tech-savvy sample. Concern c) can also be addressed in the revised methodology required for a full study. A more thorough lexical normaliztion/recoding will take place to ensure parity between meanings of identity elements in SNC's. While the status of an outsider investigator (concern d) will remain service-dependent, it may be worthwhile to complete documentation of the bounds of outsider investigator behavior, so that future studies can use and improve the methodology.

6 Discussion and Conclusions

A number of key findings have been presented as a result of the pilot study. First, a quantitative analysis of SNC's on a typical college campus revealed a number of interesting trends. As might be expected, undergraduates use SNC's more commonly than G/P students. Additionally, the percentage of undergraduates utilizing the particular SNC TheFacebook is significant.

In Tables 1 and 2, we explore the level of identity information disclosure requested by common SNC's. From an outsider's perspective, some of the information is very interesting; relationship status, location information, and political views are just a few of the many identity information elements that are disclosed in SNC's.

In Figure 1, we are presented with results of identity information discovery among survey respondents for TheFacebook. A large number of students share particularly personal information online. If we are to compare the trends we observe in Figure 1, with the opinions students present in Table 3, it strongly suggests that there is a disconnect between the value of traditional identity information (Name, SSN) and the new types identity information being disclosed (photo, political views, sexual orientation) in SNC's. This disconnect identifies the need for a new discussion of identity information protection on campus, one that is effectively holistic and SNC-aware.

References

1. boyd, d.: Friendster and Publicly Articulated Social Networking. Conference on Human Factors and Computing Systems. ACM Vienna (2004)
2. Federal Trade Commission: National and State Trends in Fraud & Identity Theft, January-December 2004. Federal Trade Commission, Washington DC (2004)
3. Newman, G.: Identity Theft. Problem-Oriented Guides for Police, Guide Number 25. U.S. Department of Justice, Washington DC (2004)
4. Goldman, E.: The Growing Thread of Identity Theft. Educase Review, June/July (2004) 66-67
5. Berghel, H.: Identity Theft, Social Security Numbers, and the Web. Commuunications of the ACM, Vol. 43 No. 2, February (2000) 17-21
6. O'Murchu, I., Breslin, J., and Decker, S.: Online Social and Business Networking Communities. DERI Technical Report 2004-08-11, Galway Ireland August (2004)
7. Hogg, T., and Adamic, L.: Enhancing Reputation Mechanisms via Online Social Networks. EC 2004, New York New York (2004)
8. Liu, H. and Maes, P.: InterestMap: Harvesting Social Network Profiles for Recommendations. IUI 2005, Workshop: Beyond Personalization, San Diego California (2005)
9. Watts, D., Dodds, P.S., and Newman, M.E.J.: Identity and Search in Social Networks. Science 296 (2002) 1302-1305