

Fred Stutzman
May 8, 2007

Abstract:

Social network websites (SNS), such as Myspace.com or Facebook.com, enable individuals to represent their social networks in a computer-mediated context. While the exact value and meaning of the social connections represented in online social networks is variable (Donath, 2004), there are distinct connections between the social networks represented in a computer-mediated context and the real world (Lampe, 2006). In the following analysis, types of information disclosed in social network websites are analyzed. Three indices of social network information – identity information, presence information and preference information – are developed, and the relationships between the disclosure of information type and information volume are explored. While all three indices have significant explanatory power, identity information has the strongest effect on information sharing volume, while preference information has the greatest explanatory power.

Introduction

Of the many narratives exploring use and outcomes of social network websites, perhaps the most common explores the public sharing of personal identity information. Over the past four years, social network websites (SNS) have achieved strong market penetration with a wide range of participants (Lenhart, 2007). While broad age ranges are represented in these sites, such as Myspace.com and Facebook.com, youth sharing behaviors remain one of the main points of concern as we examine and attempt to understand the popularity of social network websites.

Social network sites are inherently identity-centric. On a social network website, the individual serves as the core “social object,” around which the site orients. Common activities, such as the addition of friends, the posting of messages to the wall, or the publishing of photos can all be thought of as “identity transactions.” These identity transactions, generally carried out in public, shape the representation of the individual’s identity as presented in the particular social network website. In essence, the shaping of the individual identity is directly tied to outcomes from use of the site.

The presentation of identity in a social network site is an iterative process, in which individuals tune and “tweak” their profiles over time (boyd, 2006). For example, the disclosure of one’s political identity, or favorite books is an explicit representation of identity. However, the structural aspects of many social network sites enable identity to be represented in a more “conversational” nature. For example, the individual controls many profile elements, but they do not direct which photos they are “tagged” in, or what

individuals will write on their wall.¹ It is important to understand this distinction as it serves to remind us that identity is socially constructed in social network sites.

The construction of identity in a social network site is valuable for many reasons. Primarily, the construction of an identity renders one “findable” – their social network identity is discoverable by friends, both old and new. Furthermore, the creation of identity serves as a virtual representation of self in the social network context; in the real world, identity is a performance (Goffman, 1959), and there is no particular reason to think that construct does not apply to virtual worlds. Of course, with such wide use we must remain cautious and cognizant that many alternative motives exist for the creation of identity; brand and marketing avatars create representations for the explicit purpose of advertisement, for example.

In the following study, a sample of data collected from the University of North Carolina’s Facebook (<http://facebook.com>) will be examined in an attempt to understand the nature of identity disclosure in a social network context. This analysis will examine the types of identity information disclosed in a social network site, with a goal of exploring what factors of a social network profile are most distinctly tied to identity disclosing behavior in a social network profile.

¹ Individuals generally do have the authority to “delete” externally generated content, but they do not have editorial control over the content.

Literature Review

There are a number of factors that motivate use of a social network site. In research at Michigan State University, Ellison, Steinfeld and Lampe discovered the students used Facebook to pass time, to connect with others, and to manage their offline contacts (Ellison, 2006). Bumgarner's exploration of uses and gratifications of Facebook at the University of North Carolina further confirm it uses as social network manager of sorts; Bumgarner found that students primarily used Facebook to explore their friend networks, to update their personal information, and to extract practical information about the social world around them (Bumgarner, 2006).

These two analyses provide insight into the value of the publicly articulated identity in social network sites. While these sites afford entertainment and distraction (in the browsing of profiles, etc.), they also deliver a core utility in the sense they allow the exploration of pre-existing social networks. Additionally, they serve a very core purpose as a "social directory," enabling personal findability. At the center of all of these actions is the profile; whether it be browsing, exploring or refinding, the information provided in the profile serves a valuable purpose to users. In essence, we can see that the personal information shared in social network sites is the catalyst for interest in the sites.

As the shared information provides motivation for the use of social network sites, a number of studies have explored the nature of information sharing in these contexts. Stutzman explored the nature of identity sharing on a number of different social network sites, finding high disclosure of identity information generally perceived of as "personal"

in nature (Stutzman, 2006). Gross and Aquisti explored the nature of information disclosure in Facebook, uncovering similar findings about identity disclosure. Furthermore, Gross and Acquisti discovered that individual concern about privacy is a weak predictor of network participation (Gross, 2006). While Stutzman (Stutzman, 2007) has uncovered an ongoing trend towards increased privacy in Facebook, levels of identity disclosure in social network sites remain high.

In Lampe, Ellison and Steinfeld (Lampe, 2006), the directional nature of friendship in Facebook is explored. The authors found that relationships in Facebook primarily originated offline, migrating into the online context. Mazer, Murphy and Simonds (Mazer, 2007) further explored this situation, assessing the perceptions created by teacher use of Facebook in a higher education context. In each case, the amount and type of disclosure provided was a valuable factor in third-party evaluations of social network profiles.

We see two key trends emerge from our analysis of the literature. First, individuals create robust representation of identity in the social network context, at generally high levels of disclosure. Second, the type and levels of information disclosed in social network sites are tied to real-world outcomes, such as evaluations, friend judgments, and social awareness. Using data collected from Facebook, we will explore which types or elements of disclosure best predict levels of disclosure; that is, we will examine the relationship between certain types of information sharing and certain levels of information sharing. As much literature has focused on the amount of information

disclosed in the social network context, this particular study will isolate potential key factors inducing information sharing.

Data

The data used in the following study was sourced from the University of North Carolina's Facebook (<http://facebook.com>). It was collected over the course of two days, July 15 and 16, 2005. Data collection was completed via automated scripts, written in the Perl language and utilizing the LWP module. Once downloaded, the data was processed via a Perl program utilizing the HTML::TokeParser module; the data was eventually segmented into 77 variables, including a data-integrity variable, and analyzed with the Stata computer program. All data has been anonymized, and use of the data has been approved by the University of North Carolina IRB and Facebook.com.

The data set analyzed in this study contains 12,938 individuals, comprising the undergraduate population of Facebook users at the University of North Carolina. While this data appears to be population data, rather than a sample, it is important to remember that some individuals that enable privacy are not collected in the sample. While it is impossible to know exactly who was *not* collected, based on enrollment information and third party literature, a non-collection rate of 5% is to be assumed. As this analysis is primarily concerned with those who create public profiles, we do not seek to generalize outside the set of 12,938.

In the analysis, “level of disclosure” will be operationalized as the independent variable.

Level of disclosure corresponds to an awarded index based on how many profile elements an individual shares publicly. The levels of disclosure per element are not considered for this analysis, which is a limitation. Sharing is treated as a binary yes/no measure, and the respective shared profile elements comprising the index score.

Dependent measures will include indexes of the types of information shared in Facebook.

Three main categories emerge: Demographic/identity information, presence information and preference information. These groups will be explored in greater detail in the corresponding descriptive analysis.

This data was collected from one institution, the University of North Carolina at Chapel Hill. As such, it has limited power to speak as a representative set for all users of Facebook. However, in comparison to work done by Gross and Aquisti (Gross, 2006) and Lampe, Ellison and Steinfeld (Lampe 2006, Lampe 2007), similarities between the datasets collected at their disparate institutions and the data collected at UNC- Chapel Hill point to strong similarities between users at these institutions. Furthermore, in thesis work by Vanden-Boogart (Vanden-Boogart, 2006), strong similarities in Facebook use between a representative set of institutions were uncovered.

Due to the process of collecting a large sample from the internet, there may be corruption in the data. Attempts to uncover corrupted data did not result in significant findings, though it remains possible that a number of profiles may have been cut off prematurely in their download, etc.

Descriptive Analysis

A Facebook profile provides students a number of channels through which they can share identity information. The majority of information shared in a profile is self-disclosed, meaning the student has to fill in their own information. A minority of information shared in the service is created directly by Facebook; for example, the time of a profile's last update is a measure shared by Facebook without user intervention. Finally, third-party individuals are responsible for The Wall, in which a post written by someone else can show up on a student's profile.

A content analysis revealed three main types of information shared in a student Facebook profile. These three groups are identity information, presence information and preference information. For the descriptive analysis, we will only explore information that is self-disclosed; therefore, information created by Facebook or by other students will not be explored. In the inferential analysis, various aspects of user- and site-generation information will be introduced as potential control or explanatory variables.

Identity information is the first main content area to be explored in the descriptive analysis (Table 1). This category contains the following profile elements: Major, date of birth, relationship status, "looking for" status, "About Me", political status, hometown, high school and high school graduation year. Political status and "looking for" information ("Looking for" refers to dating preference) were included in identity information rather than preference information as we felt one's political status and dating status were identity claims.

Table 1. Student identity information sharing behavior

<i>Identity Elements</i>	<i>Total Sharing (n=12,938)</i>	<i>Percentage Sharing (n=12,938)</i>
Academic Major	10,349	.7998
Age	10,467	.8090
“Looking for” Status	5,153	.3968
Relationship Status	8,872	.6857
Political Status	8,603	.6649
Hometown	9,531	.7366
High School	10,685	.8258
High School Grad Year	9,557	.7386
About Me	6,419	.4961
<i>Shares at Least One</i>	11,967	.9249

The next content area explored was presence information (Table 2). This set of identity information was isolated as it provides external individuals with direct means to contact the sharing individual. This content area is comprised of the following elements: Instant messenger name, mobile phone number, home phone number, residence information, current address and class schedule. Each of these elements would provide an outsider information that would allow direct contact external of Facebook.

Table 2. Student presence information sharing behavior.

<i>Presence Elements</i>	<i>Total Sharing (n=12,938)</i>	<i>Percentage Sharing (n=12,938)</i>
Instant Messenger Name	9,782	.756
Mobile Phone Number	2,575	.1990
Home Phone Number	1,467	.1134
Current Address	608	.0469
Residence Information	6,794	.5251
Class Schedule	5,678	.4388
<i>Shares at Least One</i>	11,307	.8739

Finally, the sharing of preference information was examined (Table 3). This information included in this area is self-disclosed preference information. For example, students are asked to share their favorite books or movies, creating a profile that describes their preferences. The content areas included in this category are: Preferences, clubs/jobs, music, movies, books, favorite quotes and “about me”.

Table 3. Student preference information sharing behavior.

<i>Preference Elements</i>	<i>Total Sharing (n=12,938)</i>	<i>Percentage Sharing (n=12,938)</i>
Interests	9,301	.7188
Clubs and Jobs	8,030	.6206
Favorite Music	9,781	.7559
Favorite Movies	9,547	.7379
Favorite Books	8,803	.6803
Favorite Quotes	8,596	.6643
<i>Shares at Least One</i>	10,890	.8417

Each content area contains a variable, entitled “*Shares at Least One*,” which is set when an individual shares at least one element in the content area. In each content area, we see a large portion of our sample sharing information; the most popular content area was identity information, followed by presence information and then preference information. The high response in the identity content area affirms the idea that Facebook is a place to create identity, though the identity created is more like a “search identity,” in which students flesh out profile areas in an attempt to be findable to other students. This line of logic is continued as presence elements are the second more likely to be shared. Just as students want to be findable, they want to be contactable. We see Facebook as a social directory in this situation.

Multivariate Analysis

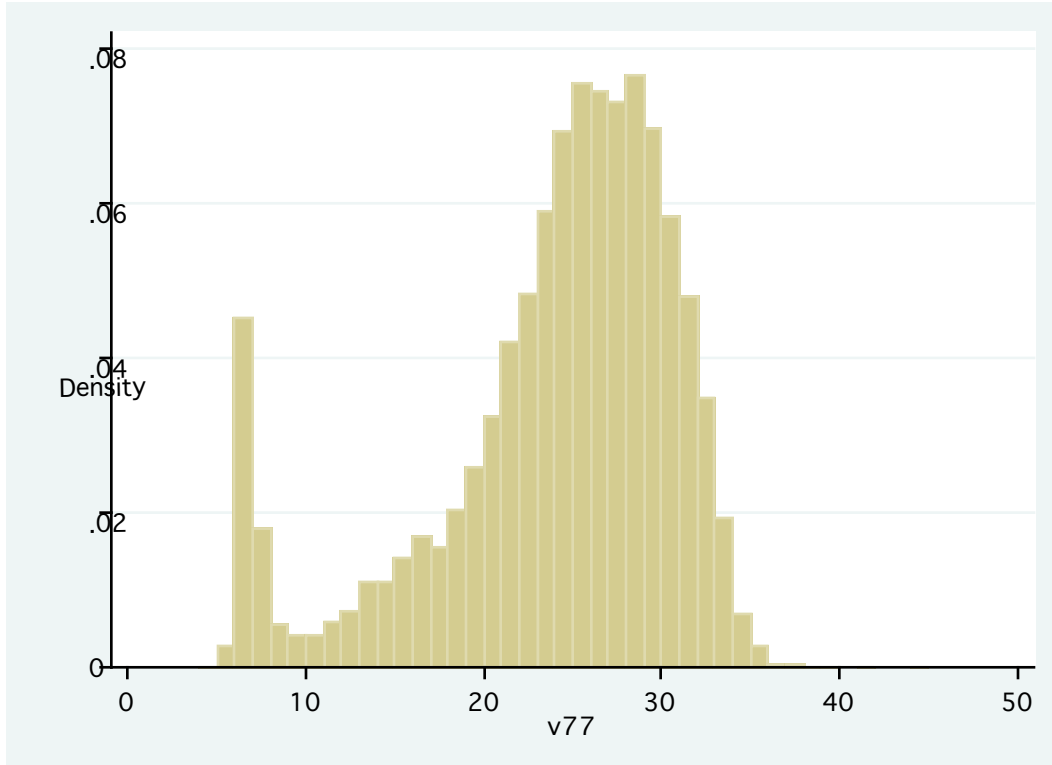
In the multivariate analysis, we will explore the effects of sharing certain types of information on the volume of information shared in a Facebook profile. This exploration will give us insight into how college students value information, and what types of information sharing most highly incentivize the disclosure of personal information in an online context. The independent variable (Table 4) we will explore is level of disclosure, the index of profile information explained previously.

Table 4. Index of shared profile elements

<i>Variable</i>	<i>Obs</i>	<i>Mean</i>	<i>Std. Dev.</i>	<i>Min</i>	<i>Max</i>
Disclosure Indes	12938	23.5473	6.875361	4	45

The distribution of the variable index is presented in Figure 1. The distribution is skewed negatively (-1.075) due to persons who set up and abandoned Facebook profiles (this particular cluster involves people who share 6 or 7 elements). However, as these dormant profiles are searchable and make identity statements about the individual who filled them out, they are important for the analysis. While the distribution is skewed, we have not translated the index into a log index, as a skewness 1 is manageable.

Figure 1. Distribution of identity index (v77)



Furthermore, three measures will be introduced as controls in our multivariate analysis. Those measures are gender, length of time used and number of friends. The first measure, gender, is a dichotomous categorical variable representing the subject's gender. Gender is included as a control, as studies (Lenhart 2007) have found significant differences between gender utilization of social networks. The second measure, length of time used, is an interval variable representing the number of days the subject had been "on" the Facebook at the time of data collection. This control is included to uncover potential recency effects. The final measure, number of friends, is an interval variable representing the number of friends the individual had added to their profile at the time of data collection. This control is included to further explore the hypothesis of Lampe,

Ellison and Steinfeld (2007), who found that certain types of disclosure predict friend accumulation behavior.

Table 5. Distribution of control variables

<i>Variable</i>	<i>Obs</i>	<i>Mean</i>	<i>Std. Dev.</i>	<i>Min</i>	<i>Max</i>
Gender (0=M/1=F)	11378	.5902619	.4918069	0	1
#Days on Facebook	12061	241.0668	137.5946	2	445
# Friends	12910	79.79574	83.983	0	3436

The multivariate analysis will explore the effects of disclosure types on levels of disclosure. We will look at whether sharing identity, presence and preference information affects the level of disclosure in a social network profile. To conduct this analysis, a dummy variable indicating whether or not an individual has shared a particular information type is created. We will explore the relationship between sharing an information type and level of disclosure.

Our first analysis examines the effect of identity information sharing on overall information sharing. In the first model (Table 6), we see that the sharing of identity information has a strong effect with a beta of 18.29 and a R-square of .49. In model two, gender is introduced, and is not significant. In models three and four, days active and number of friends are introduced to the model. While significant, the betas for these variables are very small. Additionally, the model's R-square falls substantially, indicating poorness of fit as an explanatory measure.

Table 6. Identity information sharing model

	(1)	(2)	(3)	(4)
	Disclosure index	Disclosure index	Disclosure index	Disclosure index
Identity Information	18.291	14.704	14.160	13.602
Gender==Male	(0.164)**	(1.262)**	(1.236)**	(1.203)**
Days Active		0.107 (0.093)	0.134 (0.091)	0.093 (0.089)
Number of Friends			0.006 (0.000)**	0.002 (0.000)**
Constant	6.629 (0.157)**	10.431 (1.262)**	9.449 (1.236)**	9.722 (1.203)**
Observations	12938	11378	11349	11323
R-squared	0.49	0.01	0.04	0.09
Standard errors in parentheses				
* significant at 5%; ** significant at 1%				

The second analysis examines the effect of presence information sharing on overall information sharing. In the first model (Table 7), we see that the sharing of identity information has a strong effect with a beta of 13.995 and a R-square of .46. In model two, gender is introduced, and is not significant. In models three and four, days active and number of friends are introduced to the model. While significant, the betas for these variables are very small. Additionally, the model's R-square falls substantially, indicating poorness of fit as an explanatory measure.

Table 7. Presence information sharing module

	(1)	(2)	(3)	(4)
	Disclosure index	Disclosure index	Disclosure index	Disclosure index
Presence Information	13.995	7.342	6.905	6.626
	(0.134)**	(0.195)**	(0.194)**	(0.190)**
Gender==Male		0.150 (0.088)	0.171 (0.087)*	0.129 (0.085)
Days Active			0.005 (0.000)**	0.002 (0.000)**
Number of Friends				0.014 (0.001)**
Constant	11.317 (0.126)**	18.134 (0.197)**	17.279 (0.203)**	17.270 (0.199)**
Observations	12938	11378	11349	11323
R-squared	0.46	0.11	0.13	0.17
Standard errors in parentheses				
* significant at 5%; ** significant at 1%				

The third analysis examines the effect of preference information sharing on overall information sharing. In the first model (Table 8), we see that the sharing of identity information has a strong effect with a beta of 14.559 and a R-square of .60. In model two, gender is introduced, and is not significant. In models three and four, days active and number of friends are introduced to the model. While significant, the betas for these variables are very small. Additionally, the model's R-square falls moderately, indicating poorness of fit as an explanatory measure.

Table 8. Preference information sharing model

	(1)	(2)	(3)	(4)
	Disclosure index	Disclosure index	Disclosure index	Disclosure index
Interest Information	14.559	10.274	9.973	9.670
	(0.105)**	(0.131)**	(0.132)**	(0.130)**
Gender==Male		0.038	0.052	0.020
		(0.076)	(0.075)	(0.073)
Days Active			0.004	0.001
			(0.000)**	(0.000)**
Number of Friends				0.011
				(0.001)**
Constant	11.293	15.784	15.111	15.170
	(0.096)**	(0.133)**	(0.141)**	(0.139)**
Observations	12938	11378	11349	11323
R-squared	0.60	0.35	0.36	0.38
Standard errors in parentheses				
* significant at 5%; ** significant at 1%				

Discussion

All three of the models (Table 9) presented had significant explanatory power in their basic form. As an index, the categories represented more than one variable; however, the largest category contained nine variables, out of 77 possible. The model with the greatest base power, preference information, had an index comprised of only six variables.

In the first analysis, identity information, we see a strong beta and r-square for the model.

When a person adds publicly shares identity information, they are incited to share information publicly. In fact, the individual who shares identity information shares an average of 18.29 items about his or her personal life. Introducing gender, days active and

number of friends into the model substantially reduced the explanatory power of the model (Table 10), indicating there may be interaction between identity sharing and the controls.

Table 9. Uncontrolled model beta and r-squares

<i>Model</i>	<i>Beta</i>	<i>R-Square</i>
Identity Information	18.29	.49
Presence Information	13.995	.46
Preference Information	14.559	.60

The second analysis, presence information, again produced a strong beta and R-square. As the controls were introduced, we saw the beta fall, again indicating interaction. The beta on the first model of this analysis indicates that the sharing of presence or location information incites users to share at higher levels, but not at levels higher than personal information.

Table 10. Controlled model beta and r-squares (with all controls)

<i>Model</i>	<i>Beta</i>	<i>R-Square</i>
Identity Information	13.602	.09
Presence Information	6.626	.17
Preference Information	9.670	.38

The final analysis, preference information, had the highest explanatory power, though its betas were lower than identity information. This indicates that preference information explains a good deal of variance in the model, but the sharing of preference information does not incite disclosure at levels higher than identity information.

These models are interesting as they shed light on the types of information that most strongly affect information sharing in a social network website. To this extent, the results are in line with our expectations. Social networks are inherently identity-centric, and they are used as social directories. Identity information, presence information and preference information all fill those needs, and the sharing of these information types closely correlates to information sharing outcomes.

References

- Bumgarner, B. (2006). You have been poked: Exploring the uses and gratifications of facebook among emerging adults. Senior Honors Thesis, University of North Carolina at Chapel Hill.
- Donath, J. and danah boyd (2004). Public displays of connection. *BT Technology Journal*, 22(4):71–82.
- boyd, d. (2006). Friends, friendsters and top 8: Writing community into being in social network sites. *First Monday*, 11(12).
- Ellison, N., Steinfeld, C., and Lampe, C. (2006). Spatially bounded online social networks and social capital: The role of facebook. In *Proceedings of ICA 2006*.
- Goffman, E. *The Presentation of Self in Everyday Life*. Doubleday, New York, NY, 1959.
- Gross, R. and Acquisti, A. (2005). Information revelation and privacy in online social networks. In *Proceedings of WPES 2005*.
- Lampe, C., Ellison, N., and Steinfeld, C. (2006). A face(book) in the crowd: Social searching vs. social browsing. In *Proceedings of CSCW 2006*, pages 167–170.
- Lampe, C., Ellison, N., and Steinfeld, C. (2007) A familiar Face(book): Profile elements as signals in an online social network. In *Proceedings of CHI 2007*.
- Lenhart, A. and Madden, M. (2007). *Teens, privacy & online social networks*. Research report, Pew Internet & American Life Project.
- Mazer, J., Murphy, R. and Simons, C. I'll see you on Faebook: The effects of computer-mediated teacher self-disclosure on student motivation, affective learning and classroom climate. *Communication Education*, 56(1):1-17.
- Stutzman, F. (2006). An evaluation of identity-sharing behavior in social network communities. *iDMA Journal*, 3(1).
- Stutzman, F. (2007). *Logitudinal analysis social network privacy settings*. White paper, University of North Carolina.
- Vanden-Boogart, M. (2006). *Uncovering the social impacts of facebook on a college campus*. Master's thesis, Kansas State University.