A Privacy Disaster Waiting To Happen?
-
The Facebook and Privacy on Social Networking Sites

Ralph Gross and Alessandro Acquisti

Quick Show of Hands

- Students with a Facebook profile?

- How many are concerned about privacy issues in connection with their Facebook profile?

- How many list their full actual birthday?

- How many list their current residence?

Motivation

- Dichotomy between stated privacy attitudes and behavior

  – People claim to care about the privacy of their data

  – People give their data away for very little in return

Agenda

- Privacy implications of OSN

  – Patterns of information revelation

  – Decision making process

- The Facebook.com

  – Usage data & users survey

  – Privacy risks & security flaws

What are Online Social Networks?

- Sites that facilitate interaction between members through self-published personal profiles

  - Core: representations of sel[ves] to others

- Progressive diversification and sophistication of purposes and usage patterns

  - *Classifieds <> OSN <> blogs*

A History of Online Social Networks

- ….
- 1997: SixDegrees.com
- After 2002: commercial explosion
  - Friendster, Orkut, LinkedIn, …,
  - Viral growth with participation expanding at rates topping 20% a month
  - 7 million Friendster users; 2 millions MySpace users; 16 million registered on Tickle to take personality test (Leonard 2004)
  - Revenues: advertising, data trading, subscriptions
 - Media attention: Salon, NYT, Wired, …

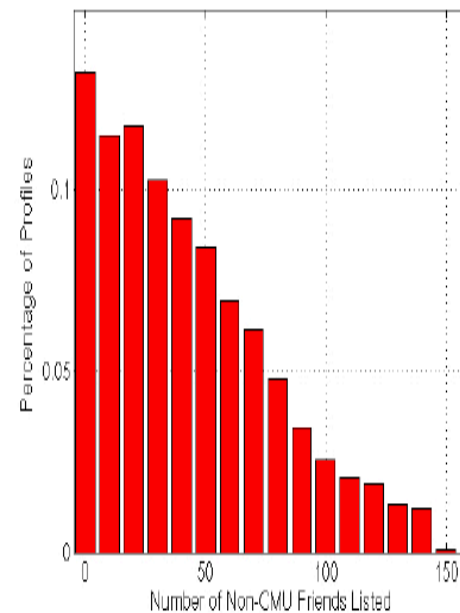# *Online* Social Networks

- Vaster, with more weaker ties than offline social networks
  - *An imagined community?*

- Intimacy and trust
  - Sharing *same* personal information with a large and potential unknown number of friends and strangers
  - *The "unknown buddy"?*

➢ Ability to meaningfully interact with others is mildly augmented

➢ Ability of others to access the person is significantly enlarged

So Many Friends...
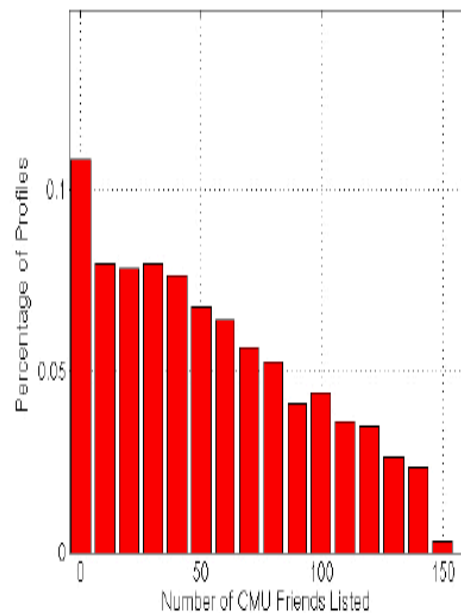
(a) Network of CMU friends   (b) Network of Non-CMU friends

Figure 3: Histogram of the size of networks for both CMU friends (a) and non-CMU friends (b). Users maintain large networks of friends with the average user having 78.2 friends at CMU and 54.9 friends elsewhere.

Online Social Networks and Privacy

- What information?
  - Re-identification
    - *known* $\longrightarrow$ *additional information*
    - *unknown* $\longrightarrow$ *known*

- Who can access the information?
  - Site (non obvious information)
  - Third-parties (hackers, government)
  - Users (little control on social network and its expansion)

- Risks
  - From identity theft to online and physical stalking; from embarrassment and blackmailing to spam and price discrimination

The Facebook

- www.facebook.com
- Started February 2004
  - Attracted Silicon Valley funding
- Has spread to 2000 schools and 4.2 million users 6.4 million
- On average, attracts 80 percent of a school's undergraduate population
  - Also gets graduate students, faculty members, staff, and alumni
- Now targeting *high schools*
- Growing media attention

William Gronim's Profile

Carnegie Mellon

Quick Search

My Profile   edit
My Friends
My Photos
My Groups
My Events
My Messages
My Account
My Privacy

View More Photos of William (4)
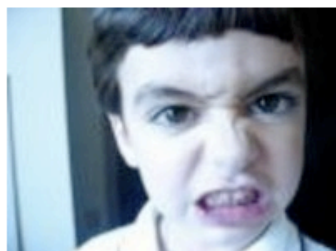Send William a Message
Poke Him!
Add William as a Friend
Report this Person

Friends at Carnegie Mellon

William has 7 Carnegie Mellon friends.

Jonah Sherman   Jonathan Mendelson   Waylon Lu

Sara Brooks   Pauline Law   Carolina Kowalczyk

[see all]

Other Schools

William has friends at...

Appalachian State (1)	Johns Hopkins (2)
Bard (1)	NYU (1)
Bryn Mawr (1)	Polytechnic (1)
Case Western (1)	RPI (1)
Claremont (1)	Stanford (1)
Columbia (1)	USC (1)
Cornell (2)	Wesleyan (1)
Duquesne (1)	

Information

Account Info
Name: William Gronim [add to friends]
Member Since: December 1, 2004
Last Update: January 23, 2006

Basic Info
Geography: Pittsburgh, PA
School: Carnegie Mellon '06
Status: **Staff**
Sex: Male
Concentration: Computer Science
Birthday: 01/01/1984
Home Town: IdentityTheft, PA 15210
High School: Stuyvesant High School '02

Contact Info
Contact Email: eat.a.baby@gronim.com
School Email: wwg@andrew.cmu.edu
AIM Screenname: nope sorry

Personal Info
Looking For: Friendship
Interested In: Women
Relationship Status: In a Relationship
Political Views: Other
Clubs and Jobs: I do software stuff for a living.
Favorite Music: jungle, biatch
Favorite Movies: Visitor Q, Guererros, End of Evangelion, anything with multiple attrocities
Favorite Books: Neuromancer
Favorite Quote: "We have no idea, now, of who or what the inhabitants of our future might be. In that sense, we have no future. Not in the sense that our grandparents had a future, or thought they did. Fully imagined cultural futures were the luxury of another day, one in which 'now' was of some greater duration. For us, of course, things can change so abruptly, so violently, so profoundly, that futures like our grandparents' have insufficient 'now' to stand on. We have no future because our present is too volatile. We have only risk management. The spinning of the given moment's scenarios. Pattern recognition."

--Hubertus Bigend, in "Pattern Recognition" by William Gibson

About Me: This used to be angry.

Photos

William has 1 Album.

Eat babies
Created January

Campus Community

- College/university email account required to sign up for Facebook

- Difference between users at the same school vs. elsewhere

Facebook's Privacy Policy

- …is lax, but straightforwardly so:

"Facebook also collects information about you from other sources, such as newspapers and instant messaging services. This information is gathered regardless of your use of the Web Site."

…

"We use the information about you that we have collected from other sources to supplement your profile unless you specify in your privacy settings that you do not want this to be done."

…

"In connection with these offerings and business operations, our service providers may have access to your personal information for use in connection with these business activities."

Facebook and Unique Privacy Issues

- Unique data
  - Includes home location, current location (from IP address), etc.
- Uniquely identified
  - College email account
  - Contact information
- Ostensibly bounded community
 - "Shared real space"

  ➢ *…or imagined community?*

Data Gathering

- In June 2005, we created Facebook profiles with different characteristics
  - E.g., degree of connectedness, geographical location, …
- We searched for CMU Facebook members' profiles
  - Downloaded 4540 profiles
  - Inferred additional information not immediately visible from profiles
- Facebook's servers reacted (probably automatically) by deactivating our accounts
  - Change in terms of use, negotiation with Counsel

Demographics
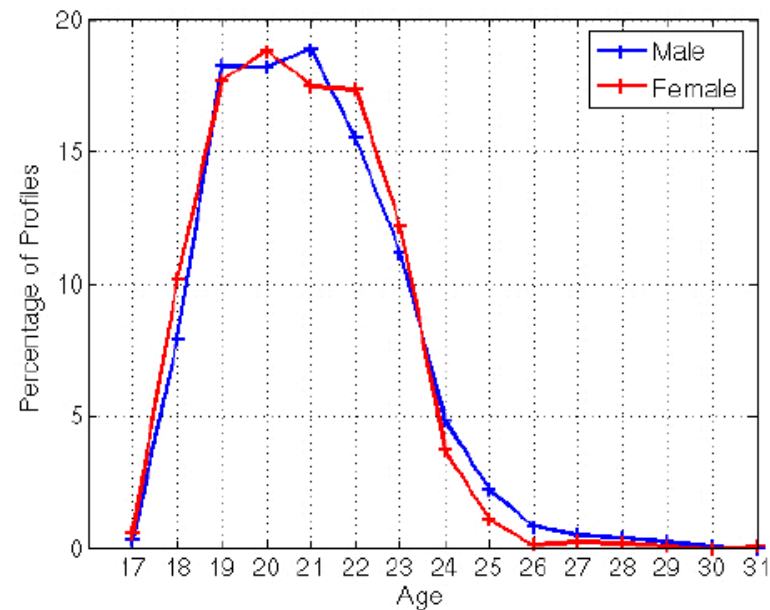
Figure 1: Age distribution of Facebook profiles at CMU. The majority of users (95.6%) falls into the 18-24 age bracket.

Demographics

Table 1: Distribution of CMU Facebook profiles for different user categories. The majority of users are undergraduate students. Where available the table lists the percentage of the CMU population (for that category) that are users of the Facebook.

	# Profiles	% of Facebook Profiles	% of CMU Population
Undergraduate Students	3345	74.6	62.1
Alumni	853	18.8	-
Graduate Students	270	5.9	6.3
Staff	35	0.8	1.3
Faculty	17	0.4	1.5

Information Revelation
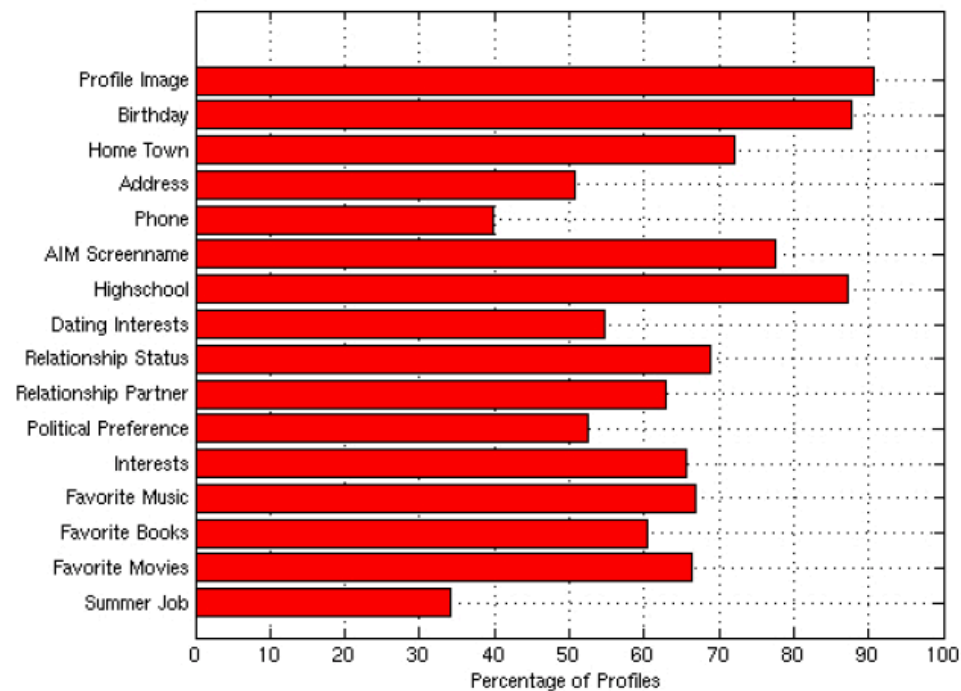
Figure 2: Percentages of CMU profiles revealing various types of personal information.

Information Revelation

- Male users 63% more likely to leave phone number than female users
  - *On average, 40% of users reveal phone number*

- *Single* male users tend to report their phone numbers in even higher frequencies

Data Verifiability - "Fakesters"?

Table 3: Categorization of name quality of a random subset of 100 profile names from the Facebook. The vast majority of names appear to be real names with only a very small percentage of partial or obviously fake names.

Category	Percentage Facebook Profiles
Real Name	89%
Partial Name	3%
Fake Name	8%

Data Verifiability - "Fakesters"?

Table 4: Categorization of user identifiability based on manual evaluation of a randomly selected subset of 100 images from both Facebook and Friendster profiles. Images provided on Facebook profiles are in the majority of cases suitable for direct identification (61%). The percentage of images obviously unrelated to a person ("joke image") is much lower for Facebook images in comparison to images on Friendster profiles (12% vs. 23%).

Category	Percentage Facebook Profiles	Percentage Friendster Profiles
Identifiable	61%	55%
Semi-Identifiable	19%	15%
Group Image	8%	6%
Joke Image	12%	23%

Privacy Risks: Stalking

- **Real-World Stalking**
  - College life centers around class attendance
  - Facebook users put home address and class list on their profiles; whereabouts are known for large portions of the day

- **Online stalking**
  - Facebook profiles list AIM screennames
  - AIM lets users add "buddies" without notification
  - Unless AIM privacy settings have been changed, adversary can track when user is online

Privacy Risks: Re-identification

- Demographics re-identification
  - 87% of US population is uniquely identified by *{gender, ZIP, date of birth}* (Sweeney, 2001)
  - Facebook users that put this information up on their profile could be linked up to outside, de-identified data sources

Privacy Risks: Re-Identification

- Face re-identification
    - Facebook profiles often show high quality facial images
    - Images can be linked to "de-identified" profiles on e.g. Match.com or Friendster.com using **face recognition**

Privacy Risks: Re-Identification

- Social Security Number re-identification
  - Anatomy of a social security number: xxx yy zzzz
  - Based on hometown and date of birth xxx and yy can be narrowed down substantially

Privacy Risks

Table 5: Overview of the privacy risks and number of CMU profiles susceptible to it.

Risk	# CMU Facebook Profiles	% CMU Facebook Profiles
Real-World Stalking	280 (Female) 580 (Male)	15.7 (Female) 21.2 (Male)
Online Stalking	3528	77.7
Demographics Re-Identification	1676	44.3
Face Re-Identification	2515 (estimated)	55.4

Privacy Risks: Digital Dossier

- Users reveal sensitive information (e.g. current partners, political views) in profiles

- Simple script programs allow adversaries to continuously retrieve and save all profile information

- Cheap hard drives enable essentially indefinite storage

Use of "Digital Dossiers" Today

- University officials browsing Facebook profiles to find:
  - "Party" images showing underage drinking
  - Pictures showing "inappropriate" behavior

- Employers using Facebook profiles as source of "information"

How We Are Assembling Dossiers

- Download **complete** CMU Facebook network once **every day**

- So far:
  - 69 days worth of data
  - 11.1 GB of information
  - Track dynamic changes of the profiles

Data Accessibility - Privacy Settings

Data Accessibility

- Profile Searchability
  - We measured the percentage of users that changed search default setting away from being searchable to everyone on the Facebook to only being searchable to CMU users
  - *1.2% of users (18 female, 45 male) made use of this privacy setting*

- Profile Visibility
  - We evaluated the number of CMU users that changed profile visibility by restricting access from unconnected users
  - *Only 3 profiles (0.06%) in total fall into this category*

Attacks

1. Fake email addresses
2. Manipulating users
3. Using other users
4. Advanced search feature ~~~~~ eographical location
   - Bypass loca~~~~~
   - Search for a~~~~~tes to reconstruct profiles

Bug (?!) Fixed

➢ *An imagined community?*
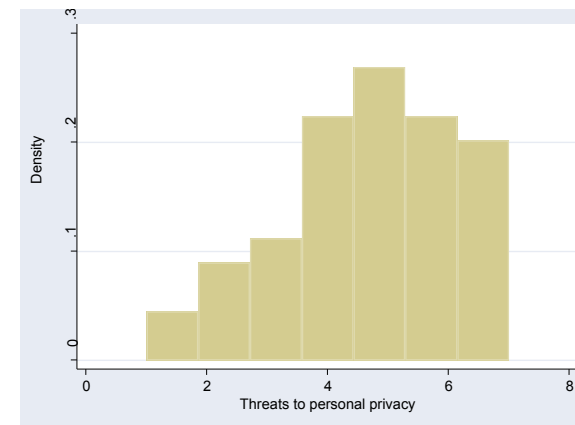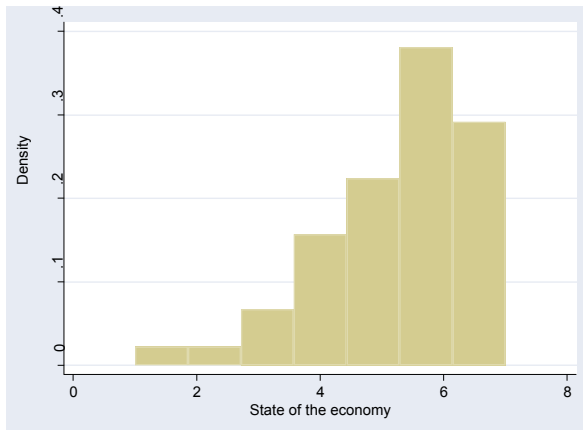➢ *Facebook profiles are, effectively, public data*

Why?

- Low privacy sensitivity
- Herding behavior
- Peer pressure
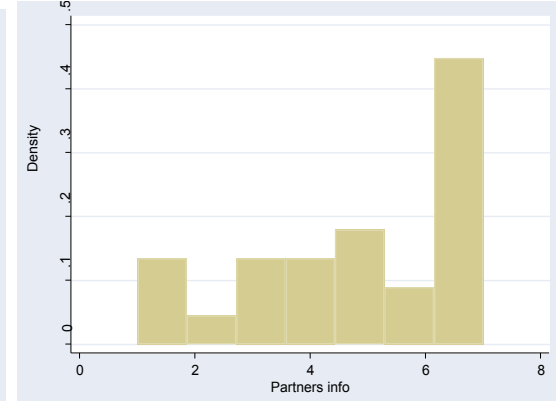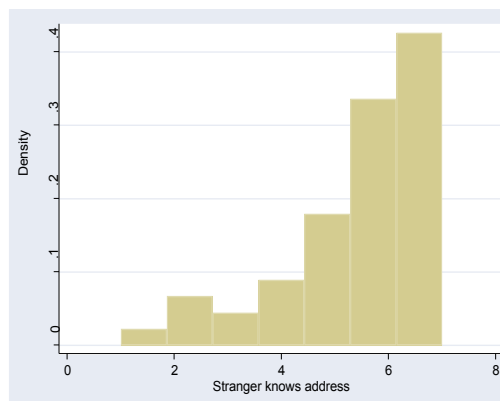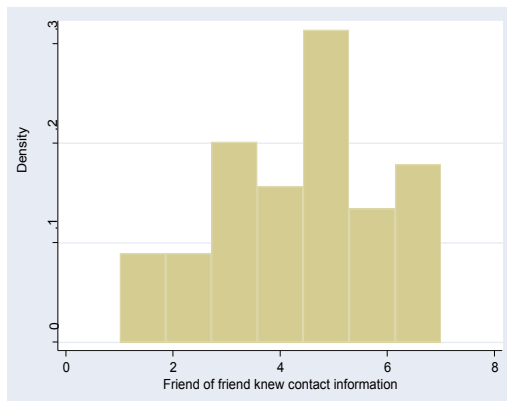- Myopic discounting
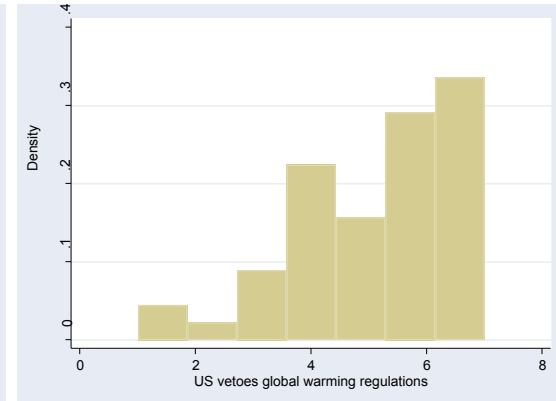- Incomplete information
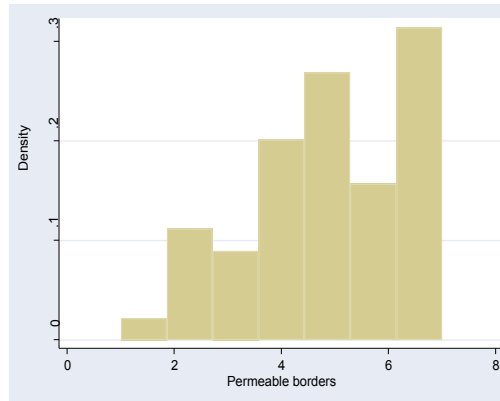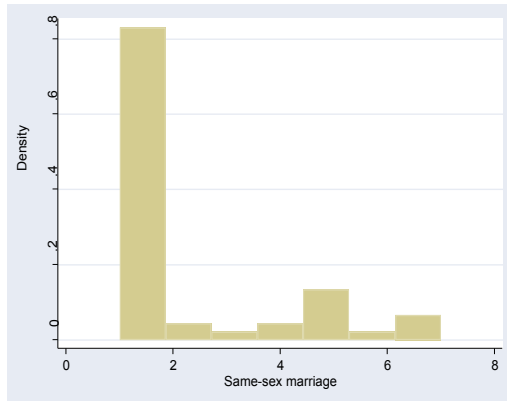- …

Survey

- Thirty-six online questions

- Anonymous, paid

- Pilot

Generic Concerns
(7-point Likert scale)
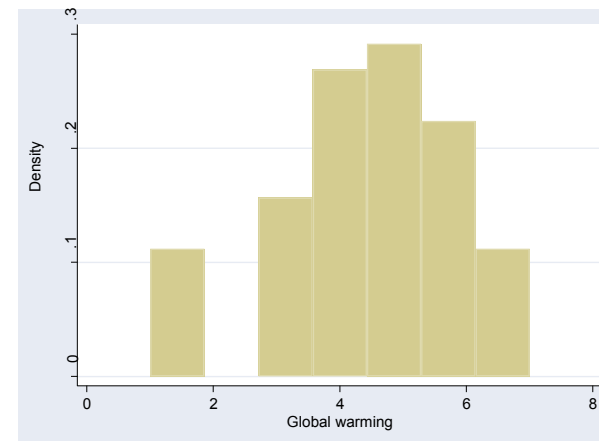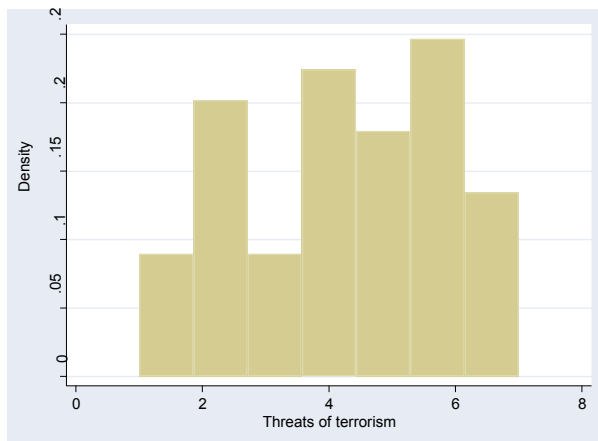
Specific Concerns
(7-point Likert scale)

Attitudes vs. Behavior

- Share of users with high sensitivity (Likert >5) to partner/sexual orientation information who provide it on Facebook: ~70%

- Share of users with high sensitivity (Likert >5) to home location and class schedule information who provide it on Facebook: ~32%

- Share of users with high sensitivity (Likert >5) to contact information who provide it on Facebook: ~42%

Awareness:
Visibility and Searchability

- 21% incorrectly believe only CMU users can search their profiles

- 71% do not realize that everybody at UPitt can search their profiles

- 40% do not realize that anybody on Facebook can search their profiles

- 31% do not realize that everybody at CMU can read their profiles

- On the other side, 23% incorrectly believe that everybody on Facebook can read their profiles

Facebook's Privacy Policy Revisited

"Facebook also collects information about you from other sources, such as newspapers and instant messaging services. This information is gathered regardless of your use of the Web Site."

- 85% believe that is *not* the case

"We use the information about you that we have collected from other sources to supplement your profile unless you specify in your privacy settings that you do not want this to be done."

- 87% believe that is *not* the case

"In connection with these offerings and business operations, our service providers may have access to your personal information for use in connection with these business activities."

- 60% believe that is *not* the case

Privacy Concerns

- 69% believe that the information *other* Facebook users reveal may create privacy risks for those users
- But:
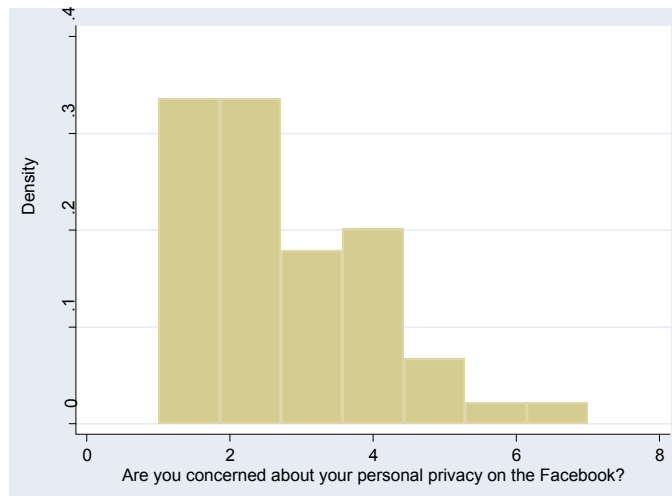
Information Revelation

- Reasons to provide more personal information (in order of importance):
    1. No factor in particular, it's just fun
    2. No factor in particular, but the amount of information I reveal is necessary to me and other users to benefit from the *FaceBook*
    3. No factor in particular, rather I am following the norms and habits common on the site
    4. Quite simply, expressing myself and defining my online persona
    5. Showing more information about me to "advertise" myself

    …..

    – Getting more potential dates

Other Privacy Concerns

- Reasons for low privacy concerns (in order of importance):
  1. Control on information
  2. Control on access
  3. CMU environment
  4. Student environment
  …

So, Why?

- Signaling? Perhaps
- Low privacy concerns? Not quite

- Peer pressure?
- Myopic discounting?

- Default settings (Mackay 1991)
  – Less than 2% make their profiles less searchable
  – Less than 1% make their profiles less visible
- Incomplete information
- Biased perspectives
  – An *imagined* community