Private Attributes

1

Study of Privacy Revelation

Ralph Gross and Alessandro Acquisti, *Information revelation and privacy in online social networks* WPES '05 Proceedings of the 2005 ACM workshop on Privacy in the Electronic Society ACM

2

Online networking sites share a core of features:
an individual offers a "profile" - a representation of their selves (and, often, of their own social networks) - to others to peruse

WHY?

- contacting or being contacted by others,
- to meet new friends or dates (Friendster, Orkut),
- find new jobs (LinkedIn),
- receive or provide recommendations (Tribe), and
- much more.

Growth (in 2005)

- "well over a million self-descriptive personal profiles are available across different web-based social networks" in the United States [18]

- "seven million people have accounts on Friendster. [...] Two million are registered to MySpace. A whopping 16 million are supposed to have registered on Tickle for a chance to take a personality test." [16]

3

The paper focus on patterns of personal information revelation and privacy implications associated with online networking.

Τι είδους προσωπική πληροφορία αποκαλύπτουν οι χρήστες και τι αυτό σημαίνει για την ιδιωτικότητα

How?

Using actual field data about the usage and the inferred privacy preferences of more than 4,000 users of Facebook

4

- Participation rates to online social networking and amount and type of information participants freely reveal among certain demographics

- (private attributed) category-based representations of a person's broad interests
E.g., person's literary or entertainment interests, as well as political and sexual ones.
- (ids) personally identified or identifiable data (as well as contact information).

Apparent openness to reveal personal information to vast networks of loosely defined acquaintances and complete strangers

5

The most common model

- the presentation of the participant's profile and
- the visualization of her network of relations to others

In *matchmaking sites*, like Match.com or Nerve and Salon Personals, the profile is critical and the network of relations is absent.

In *diary/online journal sites* like LiveJournal, profiles become secondary, networks may or may not be visible, while participants' online journal entries take a central role.

Patterns of personal information revelation are quite variable

6

First, the pretense of **identifiability** changes across different types of sites.

- **use of real names** to (re)present an account profile to the rest of the online community may be encouraged
(through technical specifications, registration requirements, or social norms) in college websites like the Facebook, that aspire to connect participants' profiles to their public identities.

- use of real names may be tolerated but filtered
in dating/connecting sites like Friendster, that create a thin shield of weak pseudonymity between the public identity of a person and her online persona by making *only the first name* of a participant visible to others, and not her last name.

- use of real names and personal contact information could be openly discouraged,
as in pseudonymous-based dating websites like Match.com, that attempt to protect the public identity of a person by making its linkage to the online persona more difficult.

However, most sites encourage the publication of identifiable **personal photos** (such as clear shots of a person's face).

7

Second, the **type of information revealed** or elicited

These include:

- often around hobbies and interests, but also

- semi-public information such as current and previous schools and employers (as in Friendster);

- private information such as drinking and drug habits and sexual preferences and orientation (as in Nerve Personals); and

- open-ended entries (as in LiveJournal).

8

Third, visibility of information is highly variable.

- In certain sites (especially the ostensibly pseudonymous ones) *any member* may view any other member's profile.

- On weaker pseudonym sites, access to personal information may be limited to participants that are *part of the direct or extended network* of the profile owner. Such visibility tuning controls become *even more refined* on sites which make no pretense of pseudonymity, like the Facebook.

And yet, across different sites, anecdotal evidence suggests that participants are happy to disclose as much information as possible to as many people as possible.

It is not unusual to find profiles on sites like Friendster or Salon Personals that list their owners' personal email addresses (or link to their personal websites), in violation of the recommendation or requirements of the hosting service itself.

9

Social Network Theory and Privacy

Πως σχετίζεται το κοινωνικό δίκτυο και οι απαιτήσεις μας για ιδιωτικότητα

The relation between privacy and a person's social network is multi-faceted.

- In certain occasions we want information about ourselves to be known only by a small circle of close friends, and not by strangers.
- In other instances, we are willing to reveal personal information to anonymous strangers, but not to those who know us better.

Social network theorists have discussed the relevance of
- relations of different depth and strength in a person's social network
- the importance of so-called weak ties in the flow of information across different nodes in a network.

Also, network theory has been used to explore how distant nodes can get interconnected through relatively few random ties

10

Strahilevitz has proposed applying formal social network theory as a tool for aiding interpretation of privacy in legal cases.

basing conclusions regarding privacy "on what the parties should have expected to follow the initial disclosure of information by someone other than the defendant" (op cit, p. 57).

how information is expected to flow from node to node in somebody's social network should also inform that person's expectations for privacy of information revealed in the network.

11

significant differences between the offline and the online scenarios.

- Offline social networks are made of ties that can only be loosely categorized as weak or strong ties, but in reality are extremely diverse in terms of how close and intimate a subject perceives a relation to be.

- Online social networks, on the other side, often reduce these nuanced connections to simplistic binary relations: "Friend or not".

Danah Boyd notes that "there is no way to determine what metric was used or what the role or weight of the relationship is. While some people are willing to indicate anyone as Friends, and others stick to a conservative definition, most users tend to list anyone who they know and do not actively dislike. This often means that people are indicated as Friends even though the user does not particularly know or trust the person"

12

- the number of strong ties that a person may maintain on a social networking site may not be significantly increased by online networking technology,

Donath and Boyd note that "the number of weak ties one can form and maintain may be able to increase substantially, because the type of communication that can be done more cheaply and easily with new technology is well suited for these ties"

- an offline social network may include up to a dozen of intimate or significant ties and 1000 to 1700 "acquaintances" or "interactions",
- an online social networks can list hundreds of direct "friends" and include hundreds of thousands of additional friends within just three degrees of separation from a subject.

13

Thus,

Online social networks are both vaster and have more weaker ties, on average, than offline social networks.

Thousands of users may be classified as friends of friends of an individual and become able to access her personal information, while, at the same time, the threshold to qualify as friend on somebody's network is low.

This may make the online social network only an imaginary (or, to borrow Anderson's terminology, an imagined) community.

Hence, trust in and within online social networks may be assigned differently and have a different meaning than in their offline counterparts.

14

Online social networks are also more leveled:
the same information is provided to larger amounts of friends connected to the subject through ties of different strength.

While privacy may be considered conducive to and necessary for intimacy,
intimacy resides in selectively revealing private information to certain individuals, but not to others, trust may decrease within an online social network.

At the same time, a new form of intimacy becomes widespread: the sharing of personal information with large and potential unknown numbers of friends and strangers altogether.

The ability to meaningfully interact with others is mildly augmented, while the ability of others to access the person is significantly enlarged.

15

It remains to be investigated how similar or different are the mental models people apply to personal information revelation within a traditional network of friends compared to those that are applied in an online network.

16

Privacy Implications

depend on the level of identifiability of the information provided, its possible recipients, and its possible uses.

"Quasi-attributes"

Even social networking websites that do not openly expose their users' identities may provide enough information to identify the profile's owner. This may happen, for example, through face re-identification.

a 15% overlap in 2 of the major social networking sites they studied [18].
Since users often re-use the same or similar photos across different sites, an identified face can be used to identify a pseudonym profile with the same or similar face on another site.

Similar re-identifications are possible through demographic data, but also through category-based representations of interests that reveal unique or rare overlaps of hobbies or tastes.

17

"Sensitive Information

Information revelation can work in two ways:

- by allowing another party to identify a pseudonymous profile through previous knowledge of a subject's characteristics or traits; or

- by allowing another party to infer previously unknown characteristics or traits about a subject identified on a certain site.

18

To whom may identifiable information be made available?

the hosting site, that may use and extend the information (both knowingly and unknowingly revealed by the participant) in different ways).

within the network itself, whose extension in time (that is, data durability) and space (that is, membership extension) may not be fully known or knowable by the participant.

the easiness of joining and extending one's network, and the lack of basic security measures (such as SSL logins) at most networking sites make it easy for third parties (from hackers to government agencies) to access participants data without the site's direct collaboration (already in 2003, LiveJournal used to receive at least five reports of ID hijacking per day)

19

How can that information be used?

It depends on the information actually provided

Risks range from identity theft to online and physical stalking; from embarrassment to price discrimination and blackmailing.

Yet, Tribe.net CEO Mark Pincus noted that "social networking has the potential to create an intelligent order in the current chaos by letting you manage how public you make yourself and why and who can contact you."

20

While privacy may be at risk in social networking sites, information is willingly provided.

Different factors:

- Signalling, because the perceived benefit of selectively revealing data to strangers may appear larger than the perceived costs of possible privacy invasions;
- peer pressure and herding behavior;
- relaxed attitudes towards (or lack of interest in) personal privacy;
- incomplete information (about the possible privacy implications of information revelation);
- faith in the networking service or trust in its members;
- myopic evaluation of privacy risks; or also
- The service's own user interface, that may drive the unchallenged acceptance of permeable default privacy settings.

21

Facebook

the Facebook has spread "to 573 campuses and 2.4 million users. [...]

attracts 80% of a school's undergraduate population

the Pentagon manages a database of 16-to-25-year-old US youth data, containing around 30 million records, and continuously merged with other data for focused marketing [6].

22

analyze data gathered from the network of Carnegie Mellon University (CMU) students enlisted on Facebook

validates CMU-specific network accounts by requiring the use of CMU email addresses for registration and login.

- Its interface grants participants very granular control on the searchability and visibility Visibility of their personal information (by friend or location, by type of user, and by type of data).
- The default settings:

the participants profile searchable by anybody else in any school in the Facebook network, the actual content visible to any other user at the same college or at another college in the same physical location

Privacy policy [30]: the site will collect additional information about its users (for instance, from instant messaging), not originated from the use of the service itself. The policy also reports that participants' information may include information that the participant has not knowingly provided (for example, her IP address), and that personal data may be shared with third parties.

23

In June 2005, separately searched for all "female" and all "male" profiles for CMU Facebook members using the website's advanced search feature and extracted their profile IDs.

Using these IDs, downloaded a total of 4540 profiles - virtually the entire CMU Facebook population at the time of the study.

24

Demographics I

The majority of users of the Facebook at CMU are undergraduate students This corresponds to 62.1% of the total undergraduate population at CMU

The majority of users is male (60.4% vs. 39.2%).

	# 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

Table 2: Gender distribution for different user categories.

		# Profiles	% of Category	% of CMU Population
Overall	Male	2742	60.4	-
	Female	1781	39.2	-
Undergraduate Students	Male	2025	60.5	62.0
	Female	1320	39.5	62.3
Alumni	Male	484	56.7	-
	Female	369	43.3	-
Graduate Students	Male	191	70.7	6.3
	Female	79	29.3	6.3
Staff	Male	23	65.7	-
	Female	12	34.3	-
Faculty	Male	17	100	3.4
	Female	0	0.0	0.0

25

Demographics II

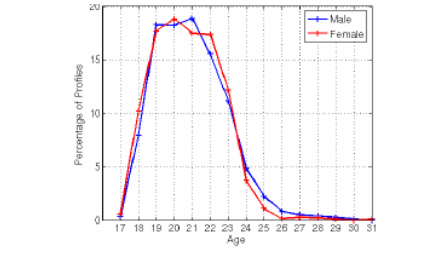


Figure 1: Age distribution of Facebook profiles at CMU. The majority of users (95.6%) falls into the 18-24 age bracket.

The strong dominance of undergraduate users is also reflected in the user age distribution
The vast majority of users (95.6%) falls in the 18-24 age bracket.
Overall the average age is 21.04 years.

26

Types and Amount of Information Disclosed

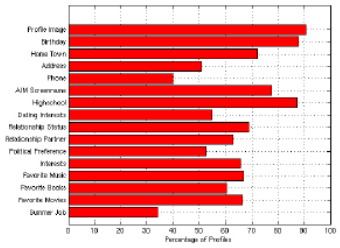


Figure 2: Percentages of CMU profiles revealing various types of personal information.

90.8% of profiles contain an image,
87.8% of users reveal their birth date,
39.9% list a phone number (including 28.8% of profiles that contain a cellphone number), and
50.8% list their current residence
The majority
their dating preferences (male or female), current relationship status (single, married,or in a relationship), political views (from "very liberal" to "very conservative"), and various interests (including music, books, and movies).
A large percentage of users (62.9%) that list a relationship status other than single even identify their partner by name and/or link to their Facebook profile.

27

Facebook profiles tend to be fully identified with each participant's real first and last names, both of which are used as the profile's name.

Across most categories, the amount of information revealed by female and male users is very similar.

A notable exception is the phone number, disclosed by substantially more male than female users (47.1% vs. 28.9%). Single male users tend to report their phone numbers in even higher frequencies, thereby possibly signalling their elevated interest in making a maximum amount of contact information easily available..

28

Data Validity and Data Identifiability

encourage users to only publish profiles that directly relate to them and not to other entities, people or fictional characters

to sign up with the Facebook a valid email address of one of the more than 500 academic institutions that the site covers has to be provided

Tested
- how valid the published data appears to be.
- how identifiable or granular the provided data is

In general, determining the accuracy of the information provided by users on the Facebook (or any other social networking website) is nontrivial for all but selected individual cases.

Restrict our validity evaluation of manually determined perceived accuracy of information on a randomly selected subset of 100 profiles.

29

1. Profile Names

Manually categorized the names given on Facebook profiles as :

1. Real Name: Name appears to be real.
2. Partial Name: Only a first name is given.
3. Fake Name: Obviously fake name.

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%

89% of all names to be realistic and likely the true names (for example, can be matched to the visible CMU email address provided as login), with only 8% of names obviously fake.

The percentage of people that choose to only disclose their first name was very small: 3%.

In other words, the vast majority of Facebook users seem to provide their fully identifiable names, although they are not forced to do so by the site itself.

30

As comparison, 98.5% of the profiles that include a birthday actually report the fully identified birth date (day, month, and year),

although, again, users are not forced to provide the complete information (the remaining 1.5% of users reported only the month or the month and day but not the year of birth).

Assessing the validity of birth dates is not trivial.

However, in certain instances we observed friends posting birthday wishes in the comments section of the profile of a user on the day that had been reported by the user as her birthday. In addition, the incentives to provide a fake birth date (rather than not providing one at all, which is permitted by the system) would be unclear.

31

2. Images

The vast majority of profiles contain an image (90.8%)

While there is no explicit requirement to provide a facial image, the majority of users do so.

In order to assess the quality of the images provided we manually labelled them into one of four categories:

1. Identifiable Image quality is good enough to enable person recognition.
2. Semi-Identifiable The profile image shows a person, but due to the image composition or face pose the person is not directly recognizable. Other aspects however (e.g. hair color, body shape, etc.) are visible.
3. Group Image The image contains more than one face and no other profile information (e.g. gender) can be used to identify the user in the image.
4. Joke Image Images clearly not related to a person (e.g. cartoon or celebrity image).

32

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%

In the majority of profiles the images are suitable for direct identification (61%).

Overall, 80% of images contain at least some information useful for identification.

Only a small subset of 12% of all images are clearly not related to the profile user.

We repeated the same evaluation using 100 randomly chosen images from Friendster, where the profile name is only the first name of the member (which makes Friendster profiles not as identifiable as Facebook ones).
Here the percentage of "joke images" is much higher (23%) and the percentage of images suitable for direct identification lower (55%).

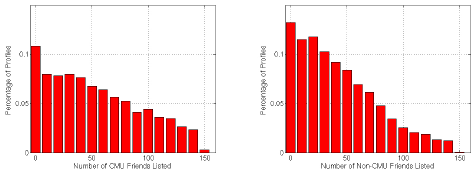
33

3. Friends Networks

the network of friends may function as profile fact checker, potentially triggering questions about obviously erroneous information.

Facebook users typically maintain a very large network of friends.
On average, CMU Facebook users list 78.2 friends at CMU and 54.9 friends at other schools. 76.6% of users have 25 or more CMU friends, whereas 68.6% of profiles show 25 or more non-CMU friends.

Histogram plots of the distribution of sizes of the networks for friends at CMU and elsewhere. This represents some effort, since adding a friend requires explicit confirmation.



(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.

34

PRIVACY IMPLICATIONS

35

Stalking

can determine the likely physical location of the user for large portions of the day.

Facebook profiles include information about residence location, class schedule, and location of last login.

In the CMU population 860 profiles (280 female, 580 male: disclose both their current residence and at least 2 classes they are attending.

study outside of the semester, we speculate this number to be even higher

36

cyber-stalking using the AOL instant messenger (AIM):

AIM allows users to add “buddies” to their list without knowledge of or confirmation from the buddy being added.

Once on the buddy list the adversary can track when the user is online. In the CMU population 77.7% of all profiles list an AIM screen name for a total of more than 3400 users.

37

Reidentification

the linkage of datasets without explicit identifiers such as name and address to datasets with explicit identifiers through common attributes [

Demographics reidentification

It has been shown that a large portion of the US population can be re-identified using a combination of 5-digit ZIP code, gender, and date of birth

The vast majority of CMU users disclose both their full birthdate (day and year) and gender on their profiles (88.8%).

For 44.3% of users (total of 1676) the combination of birthdate and gender is unique within CMU.

In addition, 50.8% list their current residence, for which ZIP codes can be easily obtained.

Overall, 45.8% of users list birthday, gender, and current residence.

38

Face ReIdentification

able to correctly link facial images from Friendster profiles without explicit identifiers with images obtained from fully identified CMU web pages using a commercial face recognizer

39

Social Security Numbers and Identity Theft

additional re-identification risk in making birthdate, hometown, current residence, and current phone number publicly available at the same time.
can be used to estimate a person's social security number (SSN) and expose to identity theft.

The first three digits of a SSN reveal where that number was created (specifically, the digits are determined by the ZIP code of the mailing address shown on the application for an SSN).
The next two digits are group identifiers, which are assigned according to a peculiar but predictable temporal order.
The last four digits are progressive serial numbers

- When a person's hometown is known, the window of the first three digits of her SNN can be identified with probability decreasing with the home state's populousness.
- When that person's birthday is also known, and an attacker has access to SSNs of other people with the same birthdate in the same state as the target (for example obtained from the SSN death index or from stolen SSNs), it is possible to pin down a window of values in which the two middle digits are likely to fall.
- The last four digits (often used in unprotected logins and as passwords) can be retrieved through social engineering.

40

Social Security Numbers and Identity Theft

Vast majority of the Facebook profiles not only include birthday and hometown information,

but also current phone number and residence (often used for verification purposes by financial institutions and other credit agencies),

users are exposing themselves to substantial risks of identity theft.

41

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

42

Building a Digital Dossier

Possible to continuously monitor the evolution of the network and its users' profiles, thereby building a digital dossier for its participants.

College students, even if currently not concerned about the visibility of their personal information, may become so as they enter sensitive and delicate jobs a few years from now - when the data currently mined could still be available.

43

FRAGILE PRIVACY PROTECTION

44

Fake Email Address

Facebook verifies users as legitimate members of a campus community by sending a confirmation email
containing a link with a seemingly randomly generated nine digit code to the (campus) email address provided during registration.

An adversary simply needs to gain access to the campus network for a very short period of time,
e.g. by attempting to remotely access a hacked or virus-infected machine on the network or physically accessing a networked machine in e.g. the library, etc.

45

Manipulating Users

obtain confidential information by manipulating legitimate users

Implementation of this practice on the Facebook is very simple: just ask to be added as someone's friend.

Demonstrated by a Facebook user who, using an automatic script, contacted 250,000 users of the Facebook across the country and asked to be added as their friend.

75,000 users accepted

46

Advanced Search Features

While not directly linked to from the site, the Facebook makes the advanced search page of any college available to anyone in the network.

Using this page various profile information can be searched for, e.g. relationship status, phone number, sexual preferences, political views and (college) residence.

By keeping track of the profile IDs returned in the different searches a significant portion of the previously inaccessible information can be reconstructed.

47

Fragile Privacy Protection

thus, personal information even on sites with access control and managed search capabilities effectively becomes public data.

48

Protecting Private Attributes in Social Networks

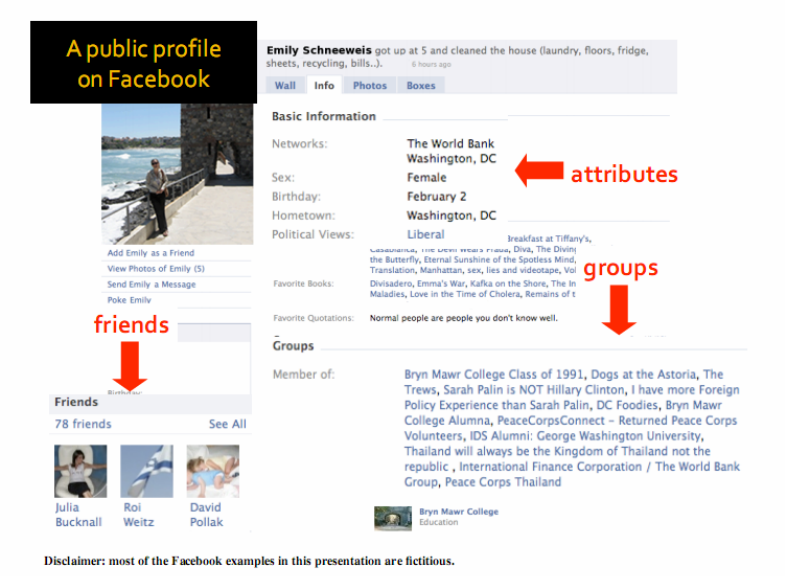
49

Privacy of Private Profiles

E. Zheleva and L. Getoor, *To Join or not to Join: The illusion of privacy in social networks with mixed private and public profiles.* WWW 2009

50

Motivating Example



Disclaimer: most of the Facebook examples in this presentation are fictitious.

51

Motivating Example

52

Problem addressed:

sensitive attribute inference in social networks

Inferring the private information of users given a social network in which some profiles and all links and group memberships are public

53

- Assumptions of this work:
  - an online social network
  - public AND private profiles
  - friendship links and group affiliations are public
- Question: can we predict private attributes based on public information?
  - links
  - groups
  - public profiles

54

Privacy Problem Addressed

- Identity disclosure
  - E.g., Elise Labbort refers to the CNN reporter Elise Labbort --
- Attribute disclosure
  - E.g., Elise Labbort is 30
- Link/relationship disclosure
  - E.g., Elise Labbort is friends with Emily
- Group membership disclosure
  - E.g., Elise Labbort is a member of the group "Sarah Palin is NOT Hillary Clinton"

55

Privacy Problem Addressed: Private Attribute Disclosure

- If an adversary is able to determine the value of a user attribute that the user intended to stay private
  - E.g. Is Paul liberal? Is Elise liberal?

56

- Mixing private and public profiles in a social network
  - For example, in Facebook many users choose to set their profiles to private, yet fewer people hide their friendship links and even if they do, their friendship links can be found through the backlinks from their public-profile friends.
- **Group participation** information
  - even if a user makes her profile private, her participation in a public group is shown on the group's membership list. Currently, neither Facebook nor Flickr allow users to hide their group memberships from public groups.

57

Why privacy?

Both commercial and governmental entities may employ privacy attacks e.g., for targeted marketing, health care screening or political monitoring

Different Problem than Anonymized Publishing:

Goal is not to release anonymized data but to illustrate how social network data can be exploited to predict hidden information

58

Model I

A social network as a graph G = (V, E, H),
V is a set of n nodes of the same type,
E is a set of (directed) edges (the friendship links), and
H is a set of groups
A group as a *hyper-edge* h ∈ H among all the nodes who belong to that group

- h.U set of users who are connected through hyper-edge h and
- v.H the groups that node v belongs to
- v.F is the set of nodes that v has connected to (friends)
- A group can have a set of properties h.T.

59

Model II

Each node v has a **sensitive attribute** v.a that can take on one of a set of possible values $\{a_1...a_m\}$.

A user profile has a unique id

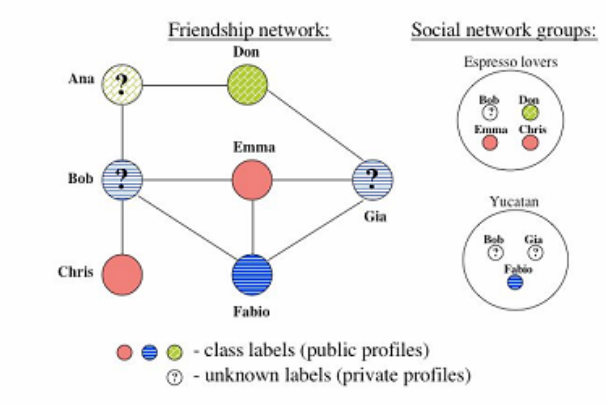
Each profile is associated with a sensitive attribute, either *observed* or *hidden*. *Private* profiles (the sensitive attribute value is unknown) and *Public* profiles *Sensitive* set of nodes Vs and *Observed* set Vo.

The adversary's goal is to predict Vs.A, the sensitive attributes of the private profiles.

study the case where nodes have no other attributes beyond the sensitive attribute

60

Example



61

Approach

The sensitive attribute value of an individual is modeled as a random variable.

The distribution of this random variable can depend
- on the overall network's attribute distribution,
- the friendship network's attribute distribution and/or **LINK**
- the attribute distribution of each group the user joins. **GROUP**

The problem of sensitive attribute inference is to infer the hidden sensitive values, Vs.A, conditioned on the observed sensitive values, links and group membership in graph G.

62

- Assume adversary can apply a probabilistic model M to predict it

$$v_s.\hat{a}_M = \arg\max_{a_i} P_M(v_s.a = a_i; G)$$

P_M is the probability that the sensitive attribute value of node vs $\in$ Vs is ai according to model M and the observed part of graph G.

- Overall distribution is either known or can be found (using the public profiles) -> Baseline attack
- Succesful attack: if significantly higher accuracy than the baseline, given extra knowledge (e.g., friendship links, group affiliations)
- extra knowledge compromises the privacy of the users

63

Attack Types (aka Attribute Inference Models)

Based on overall network distribution

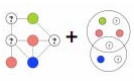
BASIC

Based on friendship links

AGG, CC, BLOCK, LINK

Based on social groups

CLIQUE, GROUP, GROUP*



Based on both links and groups

LINK-GROUP

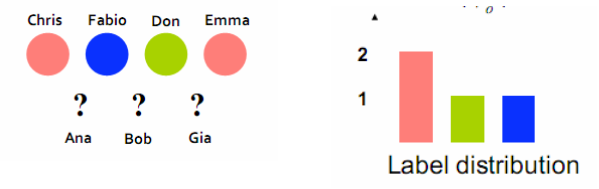
64

Baseline attack

Model based on overall network distribution

In the absense of links and groups, BASIC assigns majority label

$$P_{BASIC}(v_s.a = a_i; G) = P(v_s.a = a_i \mid V_o.A) = \frac{|V_o.a_i|}{|V_o|}$$

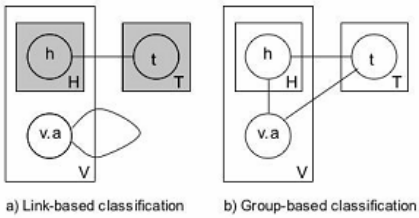


- at least as good as a random guess

Attacks using the Friendship Links

Take advantage of autocorrelation, the property that the attribute values of linked objects are correlated

There is a random variable associated with each sensitive attribute v.a, and the sensitive attributes of linked nodes are correlated.



a) Link-based classification b) Group-based classification

Figure 2: Graphical representation of the models. Grayed areas correspond to variables that are ignored in the model.

Attacks using the Friendship Links

Friends-Aggregate Model (AGG): looks at the attribute's distribution among the friends

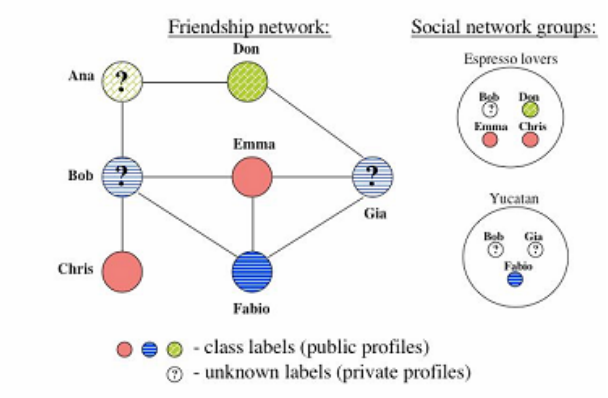
$$P_{AGG}(v_s.a = a_i;G) = P(v_s.a = a_i \mid V_o.A,E) = \frac{|V_o'.a_i|}{|V_o'|}$$

The adversary picks the most probable attribute value (i.e., the mode of the friends' attribute distribution).

67

Attacks using the Friendship Links

Example



Bob, same as Emma and Chris, Ana? Gia?

68

Attacks using the Friendship Links

Collective Classification Model (CC): learning and inferring class labels of linked objects together

Instead of each instance being classified independently, use also the inferred attributes for connected private profiles

Various Implementations -
ICA (Iterative Classification)
first assigns a label to each private profile based on the labels of the friends with public profiles, then
it iteratively re-assigns labels considering the labels of both public and private-profile friends.

The assignment based on a local classifier which takes the friends' class labels as features.
For example, a simple classifier could assign a label based on the majority of the friends labels. A more sophisticated classifier can be trained using the counts of friends' labels.

69

Attacks using the Friendship Links

Flat-link model (LINK): "Flatten" the data by considering the adjacency matrix of the graph.

1. Each row (user) a list of binary features of the size of the network: value 1 if the user is friends with the person who corresponds to this feature, and 0 otherwise.
2. The user instance also has a class label known if the user's profile is public
3. The instances with public profiles are the training data which can be fed to any traditional classifier, such as Naivee Bayes, logistic regression or SVM.
4. The learned model can then be applied to predict the private profile labels.

70

Attacks using the Friendship Links

Blockmodeling attack (BLOCK)

Basic idea: users form *natural clusters or blocks*, and their interactions can be explained by the blocks they belong to.

The link probability between two users is the same as the link probability between their corresponding blocks.

If sensitive attribute values separate users into blocks, then based on the observed interactions of a private-profile user with public-profile users, one can predict the most likely block the user belongs to and thus discover the attribute value.

71

Attacks using the Friendship Links

Block Bi the set of public profiles with value ai

λi,j the probability that a link exists between users in block Bi and users in block Bj

λi is the vector of all link probabilities between block Bi and each block B1, ...,Bm.

λ(v)j the probability of a link between a single user v and a block Bj

λ(v) the vector of link probabilities between v and each block.

To find the probability that a private-profile user v belongs to a particular block, look at the maximum similarity between the interaction patterns (link probability to each block) of v and the overall interactions between blocks.

After finding the most likely block, the sensitive attribute value is predicted.

$$P_{BLOCK}(v_s.a_i;G) = P(v_s.a_i \mid V_o.A, E, \lambda) = \frac{1}{Z} sim(\lambda_i, \lambda(v_s))$$

sim (any similarity, minimum L2 norm

Z a normalization factor

72

Attacks using the Friendship Links: Summary

- Link-based models
  - AGG: aggregate over public friends' labels (majority)

$$P_{AGG}(v_s.a = a_i;G) = P(v_s.a = a_i \mid V_o.A, E) = \frac{|V_o'.a_i|}{|V_o'|}$$

  - CC: collective classification
    - uses approximate inference and local classifiers
  - LINK: use friends as classification features
    - uses a global classifier, e.g. SVM, Naïve Bayes, LR
  - BLOCK: statistical blockmodeling
    - assumes nodes form blocks according to labels
    - finds most likely block using similarity of linking

$$P_{BLOCK}(v_s.a_i;G) = P(v_s.a_i \mid V_o.A, E, \lambda) = \frac{1}{Z} sim(\lambda_i, \lambda(v_s))$$

73

Attacks using Groups

Groupmate-link model (CLIQUE)

groupmates as friends to whom users are implicitly linked.

Each group is a clique of friends: a friendship link between users who belong to at least one group together.

apply any of the link-based models

(+) simplifies the problem to a link-based classification problem
(-) doesn't account for the strength of the relationship between two people, e.g. number of common groups

74

Attacks using Groups

Group-based Classification (GROUP): each group as a feature in a classifier

Three steps

Step 1: identify which groups are likely to be predictive -- apply feature selection

Step 2: learn a global function f, (e.g., train a classifier, that takes the relevant groups of a node as features and returns the sensitive attribute value).

Uses only the nodes from the observed set whose sensitive attributes are known.

Each node v ras a binary vector where each dimension corresponds to a unique group: {groupId : isMember}, v.a. Only memberships to relevant groups and v.a is the class coming from a multinomial distribution which denotes the sensitive attribute value.

Step 3: return the predicted sensitive attribute for each private profile.

Algorithm 1 Group-based classification model

1: Set of relevant groups $H_{relevant} = \emptyset$
2: **for** each group $h \in H$ **do**
3: **if** $isRelevant(h)$ **then**
4: $H_{relevant} = H_{relevant} \cup \{h\}$
5: **end if**
6: **end for**
7: $trainClassifier(f, V_o, H_{relevant})$
8: **for** each sensitive node $v \in V_s$ **do**
9: $v.\hat{a} = f(v.H_{relevant})$
10: **end for**

75

Attacks using Groups

Group Selection

members of a very large number of groups, identify which groups are likely to be predictive

apply standard feature selection criteria: If there are N groups, the number of candidate group subsets is 2^N, and finding an optimal feature subset is intractable.

Prune groups based on their properties (e.g., density, size and homogeneity).

Size and Homogeneity

Smaller groups may be more predictive than large groups, and groups with high homogeneity may be more predictive of the class value.

One way to measure group homogeneity is by computing the entropy of the group and the confidence in the computed group entropy

One way to measure confidence is through the percent of public profiles in the group.

76

Attacks using Groups: Summary

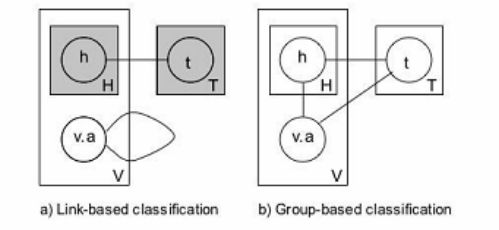


Figure 2: Graphical representation of the models. Grayed areas correspond to variables that are ignored in the model.

77

Attacks using Groups: Summary

- Group-based models
  - CLIQUE: assumes links between groupmates
    - applies a link-based model (e.g., CLIQUE-LINK)
  - GROUP: uses groups as classification features
    - uses a global classifier, e.g. SVM, Naïve Bayes, LR
  - GROUP*: chooses informative groups as features
    - chosen based on group properties (size, homogeneity, etc.)
    - expect higher accuracy than GROUP
    - lower node coverage (fewer nodes participate)

78

Attacks using both Links and Groups

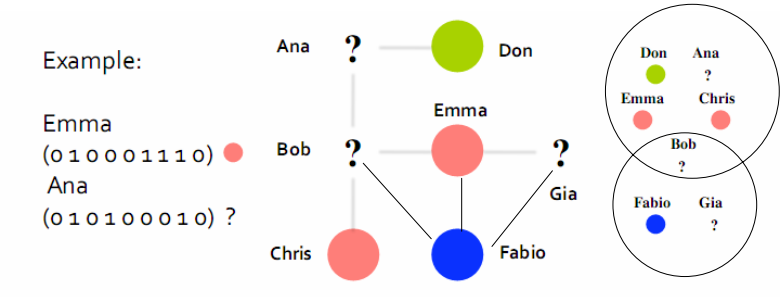
a method which uses both links and groups to predict the sensitive attributes of users

a simple method which combines the flat-link and the group-based classification models into one:
LINK-GROUP: uses all links and groups as features,

Like LINK and GROUP, LINK-GROUP can use any traditional classifier.

79

Both groups and links



80

Evaluation: Datasets

- Flickr: snowball sample
  - ~9,000 profiles, 1 million links, 50,000 groups
  - sensitive: location (55 values)
- Facebook: all freshmen (Harvard)
  - ~1,600 profiles, 86,000 links, 3,000 groups
  - sensitive: gender (2) and political views (6)
- Dogster: random sample
  - ~2,600 profiles, 4,500 links, 1,000 groups
  - sensitive: breed category (7)
- BibSonomy: ECML 2008 dataset
  - ~30,000 profiles, 130,000 groups
  - sensitive: whether spammer (2)

81

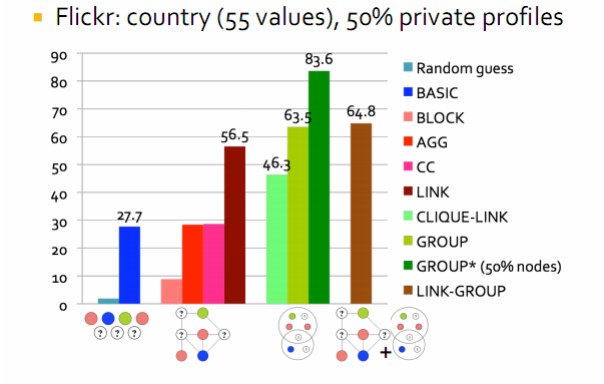
Evaluation: Experimental Setup

- Assign each profile to be public with prob. n%
  - public profiles = training data
  - private profiles = test data
- Classifier: SVM^{multiclass}
- Output – avg. over 5 trials
  - accuracy
  - node and group coverage
- GROUP*: informative groups chosen based on:
  - size, entropy and % public profiles in group

succesful attack?

82

- Flickr: country (55 values), 50% private profiles



83

Baseline achieved a relatively low accuracy (27.7%)

Link-based attacks

AGG's accuracy was 28.4%, predicting that most users were from the United States. The iterative collective classification attack, CC, performed slightly, but not significantly, better (28.6%).

- Flickr users do not form friendships based on their country of origin and country attribute in Flickr is not autocorrelated (only 23% of the links are between users from the same country), or
- the class had a very skewed distribution which persisted in friendship circles.

The blockmodeling attack, BLOCK, performed worse, with only 8.8% accuracy, showing that users from a particular country did not form a natural block to explain their linking patterns.

The "flattened" link model, LINK, with simple binary features: 56.5% accuarcy.

Results were slightly better using undirected links (which are those reported)

84

Group-based attacks.

For the CLIQUE model:
groupmate relationships converted into friendship relationships.
extremely high densification of the network.: From an average of 142 friends per user became 7,239 (out of maximum possible 9, 178).

CLIQUE-LINK model 46.3% accuary
+ due to the lack of sparsity, its training took much longer time than any of the other approaches.

85

Group-based attacks.

GROUP on all group memberships: prediction accuracy was 63.5%

Size
- If larger groups are excluded, the accuracy improves even further (72.1%).
- Medium to small-sized groups are more informative.

Entropy
- Choosing based solely on their entropy shows even better results
- Using the groups with entropy lower than 0.5 resulted in the best accuracy.

On varying percentages of public profiles per group
- Raised the accuracy even further

86

Group-based attacks.

Choosing relevant groups
+ reduced the group space by 71.2% ; SVM training time was much shorter.
(-) some of the users do not belong to any of the chosen groups, thus the node coverage decreases

51% of the private profile attributes were predicted with 83.6% accuracy.
Groups can help an adversary predict the sensitive attribute for half of the users with private profiles with a high accuracy.

The more the private profiles in the network, the worse the accuracy.
Even in the case of mostly private profiles, the GROUP attack is still successful (63.4%).

Results for the case when the minimum portion of public profiles per group is equal to the portion in the overall network and the cutoff for the maximum group entropy is at 0.5.

87

Group-based attacks.

The most heterogeneous group found ”worldwidewondering - a travel atlas.”
Some of the larger homogeneous groups include ”Beautiful NC,””Disegni e scritte sui muri”and ”*Nederland belicht*”.
One”::PONX::” which turned out to be the title of a Mexican magazine. For one user looked at, this group helped determine that although he claims to be from all over the world, he is most likely from Mexico.

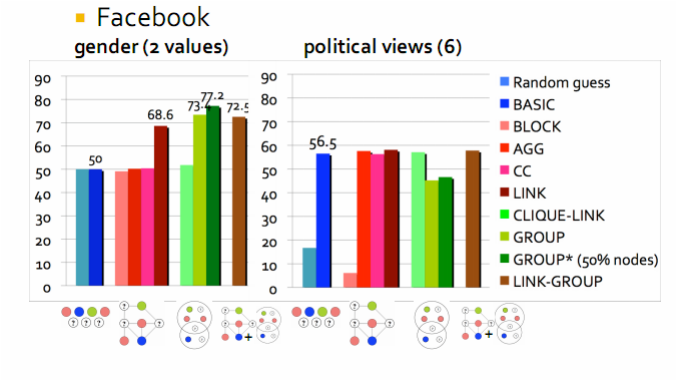
Mixed model.

LINK-GROUP did not perform statistically diferent from the GROUP model (64.8%).

Conclusion

Not participating in low-entropy groups helps people preserve their privacy better. (If users with private profiles do not join low-entropy groups, then GROUP is no longer successful.)

88



89

Link-based attacks.

In **predicting gender**, AGG, CC and BLOCK performed similarly to the baseline, LINK's accuracy varied between 65.3% and 73.5%.

In **predicting the political views**, the link-based methods performed similarly to the baseline
Binary classification to predict whether someone is liberal or not and the results were similar.
The best-performing method was LINK with 61.8% accuracy.

While it is easy to predict gender, it is hard to predict the political views of Facebook users based on their friendships.

90

Group-based attacks.

The GROUP attack was successful in predicting gender (73.4%) when using all groups.

Selecting groups that have at least 50% public profiles per group raised the accuracy by 4% but dropped the node coverage by a half.

Predicting political views with GROUP was not successful (45.2%);

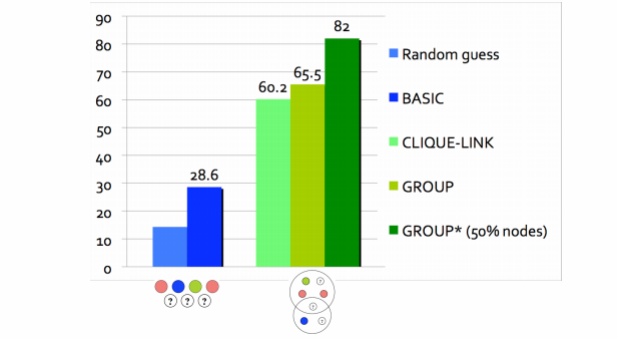
some possible explanations are
- the groups considered are not real social groups and that books, movies and music taste of first-year college students may not be related to their political views
- relatively low number of groups

Mixed model.

Again, LINK-GROUP did not perform statistically different from the other best-performing models (72.5% for gender, 57.8% for political views)

91

- Dogster: breed category (7 possible values)



92

Link-based attacks.

Due to the fact that this was a random rather than a snowball sample, there were only 432 nodes with links, do not report results

Group-based attacks.

Baseline accuracy: 28.6%.
CLIQUE-LINK's accuracy significantly higher (60.2%),

GROUP's accuracy (65.5%) when there were 50% public profiles.

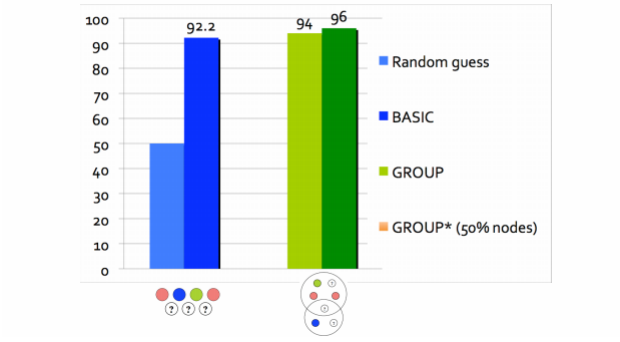
Pruning groups based on entropy led to even higher accuracy (88.9%) but had lower node coverage (14.9%) -- results for 0.5.

The accuracy increased significantly as the number of public profiles in the network increased with one exception: the accuracies for 70% and 90% public profiles did not have a statistically significant difference.

A group named "All Fur Fun"was the least homogeneous of all groups, i.e., had the highest group entropy of 2.7 (a group that invites all dogs to party together)

93

- BibSonomy: spammer or not (2 possible values)



94

Group-based attacks.

Large class skew in the data: most of the labeled user profiles are spammer profiles and the baseline accuracy is 92.2%.

Using all groups when 50% of the profiles are public leads to a statistically significant improvement in the accuracy (94%) and very good node coverage (98.5%);
this covers almost all users with tags that at least one other user uses (98.7%).

minimum entropy 0, i.e., only completely homogeneous groups were chosen.
The coverage gets lower when the most homogeneous groups are chosen (which in the spam case is actually undesirable).

Precision was 99.9-100% in all group-based classification cases, meaning that virtually all predicted spammers were such, whereas in the baseline case, it is 92.2%.

The results also suggest that if more profiles were labeled, then more covered spammers would be caught. Some of the homogeneous tags with many taggers include ”mortgage” and ”refinance.”

95

Privacy Scores

K. Liu and E. Terzi, *A Framework for Computing the Privacy Scores of Users in Online Social Networks.* ICDM 2009

96

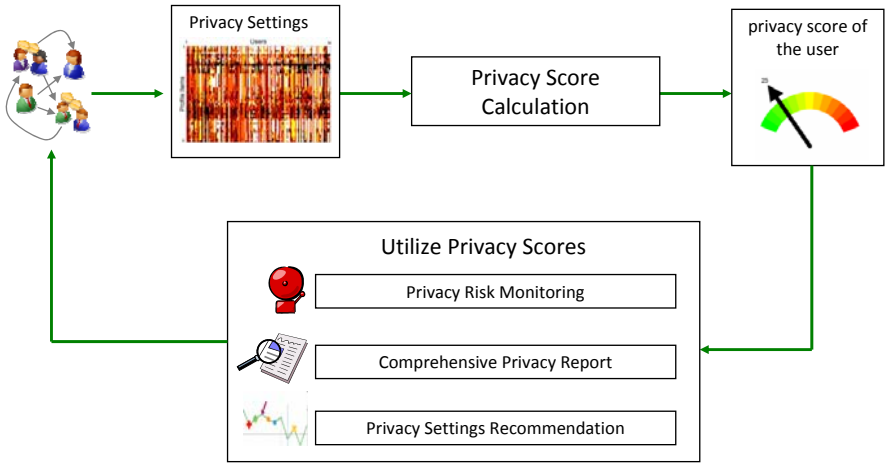
What is privacy risk score and why is it useful?

- What?
  - It is a credit-score-like indicator to measure the potential privacy risks of online social-networking users.
- Why?
  - It aims to boost public awareness of privacy, and to reduce the cognitive burden on end-users in managing their privacy settings.
    - privacy risk monitoring & early alarm
    - comparison with the rest of population
    - help sociologists to study online behaviors, information propagation

97

Privacy Score Overview

Privacy Score measures the potential privacy risks of online social-networking users.



IBM Almaden Research Center -- http://www.almaden.ibm.com/cs/projects/iis/ppn/

98

How is Privacy Score Calculated? – Basic Premises

- **Sensitivity**: The more sensitive the information revealed by a user, the higher his privacy risk.
- **Visibility**: The wider the information about a user spreads, the higher his privacy risk.

Mathematical models to estimate both sensitivity and visibility

Model

A social-network G of *N nodes*, every node associated with a user.

Every user has a profile of *n items*.

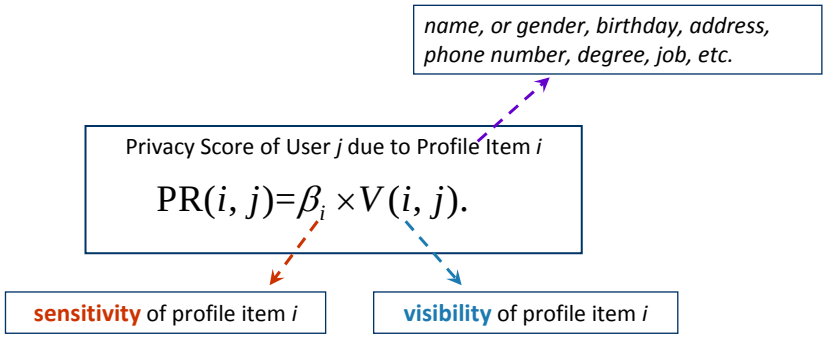
For each profile item, users set a privacy level

n* × *N* response matrix *R stores the privacy levels of all N users for all n profile items; R(i, j) the privacy setting of user j for item i

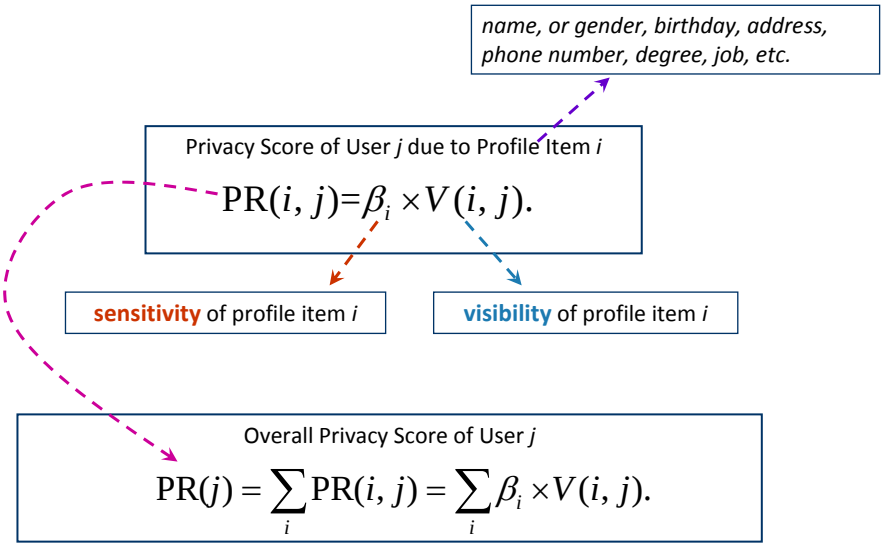
- Dichotomous, R(i, j)=0 means private, R(i, j)=1 means publicly available.
- Polytomous, R(i, j)=0 means private; R(i, j)= k with k ≥ 1 means that j discloses information regarding item i to users that are at most k-links away

In general, R(i, j) ≥ R(i', j) means that j more conservative privacy settings for i than for i'

Privacy Score Calculation



Privacy Score Calculation



As random variables described by a probability distribution.

Observed response matrix just a sample of responses that follow this probability distribution.

For dichotomous response matrices, we use Pij to denote the probability that user i selects R(i, j) = 1.

$$P_{ij} = \text{Prob}\{\mathbf{R}(i,j) = 1\}$$

For polytomous,

$$P_{ijk} = \text{Prob}\{\mathbf{R}(i,j) = k\}$$

103

For dichotomous response matrices, observed visbility

$$\mathbf{V}(i,j) = \mathbf{I}_{(\mathbf{R}(i,j)=1)}$$

True visibility,

$$\mathbf{V}(i,j) = P_{ij} \times 1 + (1 - P_{ij}) \times 0 = P_{ij}$$

$$P_{ij} = \text{Prob}\{\mathbf{R}(i,j) = 1\}$$

104

Polytomous Setting

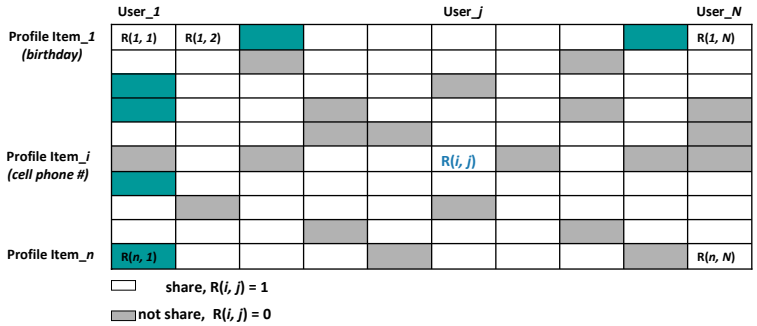
Definition 1. *The sensitivity of item* $i \in \{1,\ldots,n\}$ *with respect to privacy level* $k \in \{0,\ldots,\ell\}$, *is denoted by* β_{ik}. *Function* β_{ik} *is monotonically increasing with respect to* k; *the larger the privacy level* k *picked for item* i *the higher its sensitivity.*

Definition 2. *The visibility of item* i *that belongs to user* j *at level* k *is denoted by* $\mathrm{V}(i,j,k)$. *The* observed visibility *is computed as* $\mathrm{V}(i,j,k) = \mathrm{I}_{(\mathbf{R}(i,j)=k)} \times k$. *The* true visibility *is computed as* $\mathrm{V}(i,j,k) = P_{ijk} \times k$, *where* $P_{ijk} = Prob\{\mathbf{R}(i,j) = k\}$.

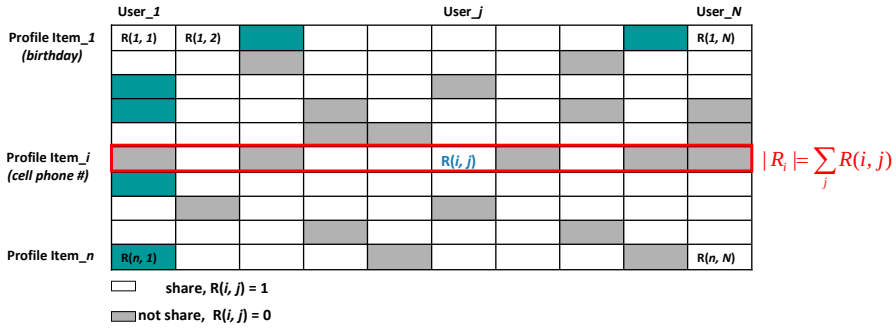
$$\mathrm{PR}(j) = \sum_{i=1}^{n} \sum_{k=0}^{\ell} \beta_{ik} \times \mathrm{V}(i,j,k)$$

105

The Naïve Approach



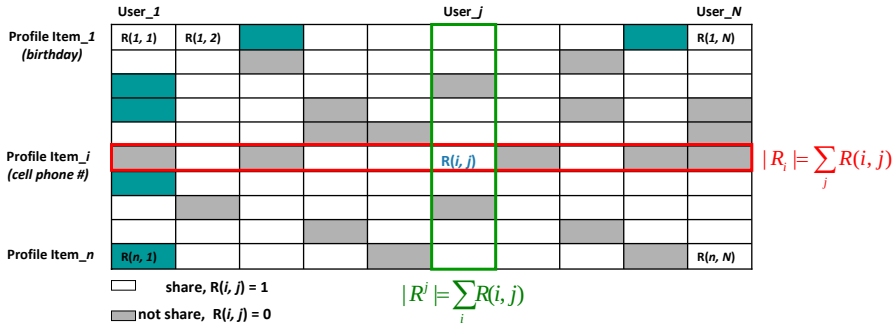
The Naïve Approach



$$|R_i| = \sum_j R(i,j)$$

Sensitivity: $\beta_i = \dfrac{N - |R_i|}{N}$

The Naïve Approach



$$|R_i| = \sum_j R(i,j)$$

$$|R^j| = \sum_i R(i,j)$$

Sensitivity: $\beta_i = \dfrac{N - |R_i|}{N}$

Visibility: $V(i,j) = \Pr\{R(i,j) = 1\}$

$$P_{ij} = \Pr\{R(i,j) = 1\} = \frac{|R_i|}{N} \times \frac{|R^j|}{n} = (1 - \beta_i) \times \frac{|R^j|}{n}$$

Advantages and Disadvantages of Naïve

- Computational Complexity O (*Nn*) – best one can hope
- Scores are sample dependent
  - Studies show that Facebook users reveal more identifying information than MySpace users
  - Sensitivity of the same information estimated from Facebook and from MySpace are different
- What properties do we really want?
  - Group Invariance: scores calculated from different social networks and/or user base are comparable.
  - Goodness-of-Fit: mathematical models fit the observed user data well.

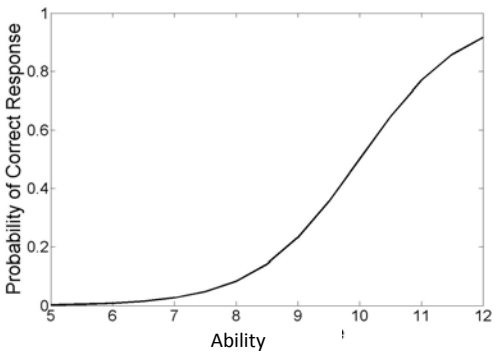
109

using true visibility

110

Item Response Theory (IRT)

- IRT (Lawley,1943 and Lord,1952) has its origin in psychometrics.
- It is used to analyze data from questionnaires and tests.
- It is the foundation of Computerized Adaptive Test like GRE, GMAT



Item Response Theory (IRT)

In psychometrics, to analyze data from questionnaires and tests.

Measure:
1. the abilities of the examinees,
2. the difficulty of the questions and
3. the probability of an examinee to correctly answer a given question.

Examinee j characterized by his **ability level** θ , $\theta \in (-\infty,\infty)$.

Question q_i characterized by a pair of parameters $\xi_i = (\alpha_i, \beta_i)$.
- Parameter β_i , $\beta_i \in (-\infty,\infty)$, represents the **difficulty** of q_i.
- Parameter α_i , $\alpha_i \in (-\infty,\infty)$, quantifies the **discrimination power** of q_i.

Basic random variable of the model: the **response** of examinee j to a particular q_i.

Item Response Theory (IRT)

- θ_j (ability): is an unobserved hypothetical variable such as intelligence, scholastic ability, cognitive capabilities, physical skills, etc.
- β_i (difficulty): is the location parameter, indicates the point on the ability scale at which the probability of correct response is .50
- α_i (discrimination): is the scale parameter that indexes the discriminating power of an item

Response of examinee j to a particular q_i. Dichotomous (correct or wrong) the probability that j answers q_i ***correctly*** is given by

$$\text{ICC: } P_{ij} = \Pr\{R(i,j)=1\} = \frac{1}{1+e^{-\alpha_i(\theta_j-\beta_i)}}$$

113

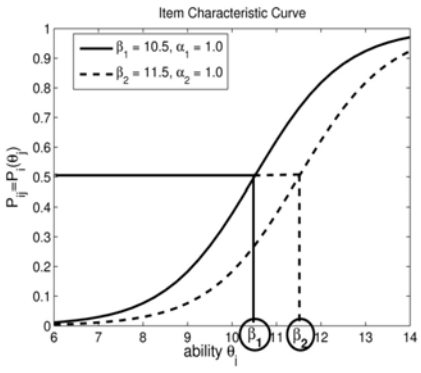
Item Characteristic Curve (ICC)

ability level $\theta \in (-\infty,\infty)$, **difficulty** of q_i, $\beta_i \in (-\infty,\infty)$, **discrimination power** of q_i $\alpha_i \in (-\infty,\infty)$

as a function of θ_j

ICCs for two questions q_1 and q_2

with $\alpha_1 = \alpha_2$ and $\beta_1 < \beta_2$



β: the point on the ability scale at which $P_{ij} = 0.5$

β and θ on the same scale

- If ability higher than the difficulty, better chance to get the answer right, and vice versa.

group invariance: the difficulty of an item is a property of the item itself, not of the people that responded to the item.

114

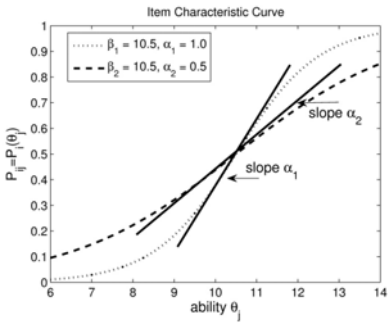
Item Characteristic Curve (ICC)

ability level $\theta \in (-\infty,\infty)$, **difficulty** of q_i, $\beta_i \in (-\infty,\infty)$, **discrimination power** of q_i $\alpha_i \in (-\infty,\infty)$

as a function of θ_j

ICCs for two questions q_1 and q_2

with $\alpha_1 > \alpha_2$ and $\beta_1 = \beta_2$



α: proportional to the slope of $P_{ij} = P_i(\theta_j)$ at the point $P_{ij} = 0.5$

- the steeper the slope, the higher the discriminatory power

this question can well differentiate among examinees whose abilities are below and above the difficulty of this question.

115

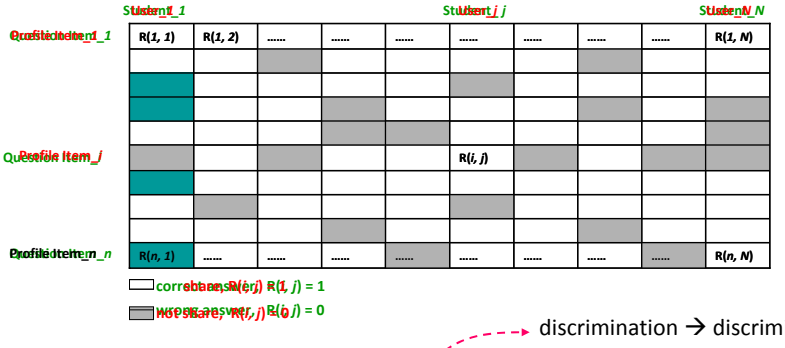
Item Response Theory for Sensitivity

Estimate the probability Prob(R(i, j)) = 1, using the IRT Equation

- examinee mapped to user
- question mapped to a profile item
- ability θ of an examinee corresponds to the attitude of user
  - attitude: how concerned j is about his privacy
- difficulty β to quantify the sensitivity of profile item i
  - to maintain the monotonicity of the privacy score need $\beta_i \geq 0$ for all $i \in \{1,...,n\}$
- α is ignored

116

Mapping from PRS to IRT



$$P_{ij} = \Pr\{R(i,j)=1\} = \frac{1}{1+e^{-\alpha_i(\theta_j-\beta_i)}}$$

discrimination → discrimination

ability → attitude/privacy concerns

difficulty → sensitivity

Prob of correct answer → Prob of share the profile

Item Response Theory (IRT)

To compute the privacy score, need to compute the sensitivity βi for all items i ∈{ 1, ..., n} and the probabilities, using the IRT Equation

For the probabilities, need to know all the parameters ξi = (αi ,βi) for all items i, 1 ≤ i ≤ n and θj for all users 1 ≤ j ≤ N.

estimate these parameters using as input the response matrix and Maximum Likelihood Estimation (MLE) techniques.

Three independence assumptions:
(i) independence between items;
(ii) independence between users; and
(iii) independence between users and items.

Experiments show that parameters learned based on these assumptions fit the real-world data

Computing PRS using IRT

Overall Privacy Risk Score of User *j*

$$PR(j) = \sum_i \beta_i \times V(i, j)$$

Sensitivity: β_i

Visibility: $V(i,j) = P_{ij} \times 1 + (1 - P_{ij}) \times 0 = P_{ij}$, where $P_{ij} = \Pr\{R(i,j) = 1\}$

$$P_{ij} = \Pr\{R(i,j) = 1\} = \frac{1}{1 + e^{-\alpha_i(\theta_j - \beta_i)}}$$

Byproduct: profile item's discrimination and user's attitude

119

Calculating Privacy Score using IRT

Overall Privacy Score of User *j*

$$PR(j) = \sum_i \beta_i \times V(i, j)$$

Sensitivity: β_i

Visibility: $V(i,j) = \Pr\{R(i,j) = 1\}$

$$P_{ij} = \Pr\{R(i,j) = 1\} = \frac{1}{1 + e^{-\alpha_i(\theta_j - \beta_i)}}$$

byproducts: profile item's discrimination and user's attitude

120

All the parameters can be estimated using Maximum Likelihood Estimation and EM.

Advantages of the IRT Model

- The mathematical model fits the observed data well
- The quantities IRT computes (*i.e.*, sensitivity, attitude and visibility) have intuitive interpretations
- Computation is parallelizable using e.g. MapReduce

121

Computation of β_i of a particular item i

- Since items are independent, computation done separately for every item
- Assuming that the attitudes of the N individuals $\theta = (\theta_1, .., \theta_N)$ are given

The likelihood function is maximized

$$\prod_{j=1}^{N} P_{ij}^{\mathbf{R}(i,j)} \left(1 - P_{ij}\right)^{1-\mathbf{R}(i,j)}$$

Users form K non-overlapping groups Fi, all users in a group Fg share the same attitude θg

$$\prod_{g=1}^{K} \binom{f_g}{r_{ig}} \left[P_i\left(\theta_g\right)\right]^{r_{ig}} \left[1 - P_i\left(\theta_g\right)\right]^{f_g - r_{ig}}$$

where rig number of users with Rij = 1

Estimating the Parameters of a Profile Item

Known Input:

Attitude level	Share	Not share	# of users at this attitude level
θ_1	r_i1	$f\_1 - r\_i1$	f_1
θ_2	r_i2	$f\_2 - r\_i2$	f_2
θ_3	r_i3	$f\_3 - r\_i3$	f_3
.....	.....	.....	.....
θ_g	r_ig	$f\_g - r\_ig$	f_g
.....	.....	.....	.....
.....	.....	.....	.....
θ_K	r_iK	$f\_g - r\_iK$	f_K

$$\sum_{g=1}^{K} f_g = N$$

Log likelihood: $L = \log\left[\prod_{g=1}^{K}\binom{f_g}{r_{ig}}\left(P_{ig}\right)^{r_{ig}}\left(1-P_{ig}\right)^{f_g-r_{ig}}\right]$, where $P_{ig} = \frac{1}{1+e^{-\alpha_i(\theta_g-\beta_i)}}$

Newton-Raphson: $\begin{bmatrix}\alpha_i \\ \beta_i\end{bmatrix}_{t+1} = \begin{bmatrix}\alpha_i \\ \beta_i\end{bmatrix}_t - \begin{bmatrix}L_{11} & L_{12} \\ L_{21} & L_{22}\end{bmatrix}_t^{-1} \times \begin{bmatrix}L_1 \\ L_2\end{bmatrix}_t$,

where $L_1 = \frac{\delta L}{\delta \alpha_i}$, $L_2 = \frac{\delta L}{\delta \beta_i}$,

$L_{11} = \frac{\partial^2 L}{\partial \alpha_i^2}$, $L_{22} = \frac{\partial^2 L}{\partial \beta_i^2}$, $L_{12} = L_{21} = \frac{\partial^2 L}{\partial \alpha_i \beta_i}$.

123

Here, the goal is again to find $\vec{\xi} = (\xi_1, \ldots, \xi_n)$ to maximize $\mathbf{P}\left(\mathbf{R} \mid \vec{\xi}\right)$. The only difference is that the elements of vector $\vec{\theta}$ are unknown. We tackle this problem using an *Expectation-Maximization* (EM) procedure.

Algorithm 1 The EM algorithm for estimating item parameters $\xi_i = (\alpha_i, \beta_i)$ for all items $i \in \{1, \ldots, n\}$.

Input: Response matrix $\mathbf{R}$ and the number K of user groups. Users in the same group have the same attitude.
Output: Item parameters $\vec{\alpha} = (\alpha_1, \ldots, \alpha_n)$, $\vec{\beta} = (\beta_1, \ldots, \beta_n)$.

```
1:  for i = 1 to n do
2:      α_i ← initial_values
3:      β_i ← initial_values
4:      ξ_i ← (α_i, β_i)
5:  ξ⃗ ← (ξ_1, ..., ξ_n)
6:  repeat
    // Expectation step
7:      for g = 1 to K do
8:          Sample θ_g on the ability scale
9:          Compute f⃗_g using Equation (6)
10:         for i = 1 to n do
11:             Compute r̄_ig using Equation (7).
    // Maximization step
12:     for i = 1 to n do
13:         (α_i, β_i) ← NR_Item_Estimation(R_i, {f⃗_g, r̄_ig, θ_g}_{g=1}^K)
14:         ξ_i ← (α_i, β_i)
15: until convergence
```

125

Expectation Step: In this step, we calculate the *expected grouping* of users using previously estimated $\vec{\xi}$. In other words, for $1 \leq i \leq n$ and $1 \leq g \leq K$, we compute $\mathbf{E}[f_g]$ and $\mathbf{E}[r_{ig}]$ as follows:

$$\mathbf{E}[f_g] = \overline{f}_g = \sum_{j=1}^{N} \mathbf{P}\left(\theta_g \mid \mathbf{R}^j, \vec{\xi}\right) \text{ and} \quad (6)$$

$$\mathbf{E}[r_{ig}] = \overline{r}_{ig} = \sum_{j=1}^{N} \mathbf{P}\left(\theta_g \mid \mathbf{R}^j, \vec{\xi}\right) \times \mathbf{R}(i,j). \quad (7)$$

The computation relies on the *posterior probability distribution* of a user's attitude $\mathbf{P}\left(\theta_g \mid \mathbf{R}^j, \vec{\xi}\right)$. Assume for now that we know how to compute these probabilities. It is easy to observe that the membership of a user in a group is probabilistic. That is, every individual belongs to every group with some probability; the sum of these membership probabilities is equal to 1.

126

Maximization Step: Knowing the values of $\overline{f}_g$ and $\overline{r}_{ig}$ for all groups and all items allows us to compute a new estimate of $\vec{\xi}$ by invoking the Newton-Raphson item-parameters estimation procedure (`NR_Item_Estimation`) described in Section V-B1.

127

The Posterior Probability of Attitudes: By the definition of probability, this posterior probability is:

$$\mathbf{P}\left(\theta_j \mid \mathbf{R}^j, \vec{\xi}\right) = \frac{\mathbf{P}\left(\mathbf{R}^j \mid \theta_j, \vec{\xi}\right) \mathbf{g}(\theta_j)}{\int \mathbf{P}\left(\mathbf{R}^j \mid \theta_j, \vec{\xi}\right) \mathbf{g}(\theta_j)\, d\theta_j}. \quad (8)$$

Function $\mathbf{g}(\theta_j)$ is the probability density function of attitudes in the population of users. It is used to model our prior knowledge about user attitudes and its called the *prior distribution* of users attitude. Following standard conventions [5], we assume that the prior distribution $\mathbf{g}(.)$ is Gaussian and is the same for all users. Our results indicate that this prior fits the data well.

128

Polytomous Setting

Transform a polytomous response matrix R into l+1 dichotomous response matrices R_0^*, R_1^*, .., R_l^*

129

Survey dataset

Ask users

5 privacy levels {0, 1, 2, 3, 4}

49 profile items

153 responses (users)

Construct

a polytomous response matrix R, and

4 dichotomous matrices

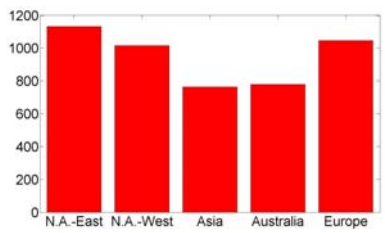
130

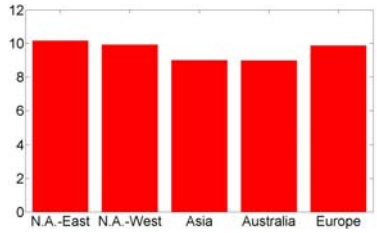
Survey dataset

IRT vs non-IRT

131

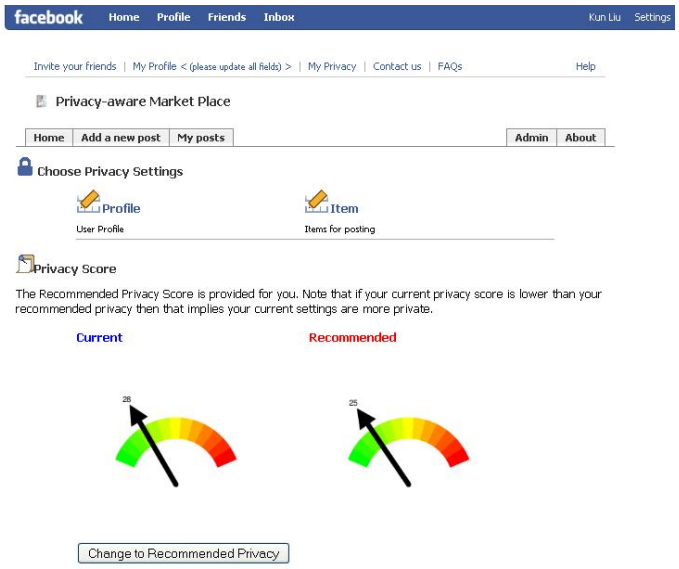
Experiments – Interesting Results





132

132


facebook
Home
Profile
Friends
Inbox
Kun Liu
Settings
Invite your friends | My Profile < (please update all fields) > | My Privacy | Contact us | FAQs
Help
Privacy-aware Market Place
Home
Add a new post
My posts
Admin
About
Choose Privacy Settings
Profile
User Profile
Item
Items for posting
Privacy Score
The Recommended Privacy Score is provided for you. Note that if your current privacy score is lower than your recommended privacy then that implies your current settings are more private.
Current
Recommended
26
25
Change to Recommended Privacy

133

133