



Information Revelation and Privacy in Online Social Networks (The Facebook case)

Vie privée en danger dans FaceBook

Ralph Gross and Alessandro Acquisti

Présenté par Guy-Michel FOPA

Décembre 2008



Intérêts

- Quels sont les informations fournies ?
- Quel est le degré d'exactitude des informations ?
- Pourquoi ces infos sont t'elles fournies ?
- Quels sont les techniques mis en œuvre pour l'exploitation des infos fournies ?
- Les utilisateurs sont-ils conscients ?



Qu'est qu'un site de socialisation ?

- Site qui facilite les interactions entre membres grâce aux profils
 - Profil : Moyen de se faire connaître , de se présenter aux autres membres
- On peut classifier ces sites selon les buts et méthodes d'utilisations



Historique des réseaux de socialisation

Phénomène ancien : 1960 mais boom avec l'arrivée d'Internet

- 1997: SixDegrees.com
- Après 2002: Explosion commerciale
- Friendster, Orkut, LinkedIn, ...,
 - croissance de plus de 20% par mois
- 7 millions Friendster ; 2 millions MySpace; (Leonard 2004)
- Revenues: publicité, data trading, subscriptions
- Attention des Médias: Salon, NYT, Wired, ...



Site de socialisation en ligne / Site de socialisation Hors Ligne

- Distinguer lien fort et lien faible entre membres
- Les lien faibles sont plus nombreux
 - Communauté imaginaire
- Confiance et intimité
- Echange des information privées avec un grand nombre d'amis (virtuel) et étrangers
- □ Plus de possibilité d'accéder aux information d'autrui
- □ Plus de possibilité d'interagir avec les autres membres

Autant d'amis

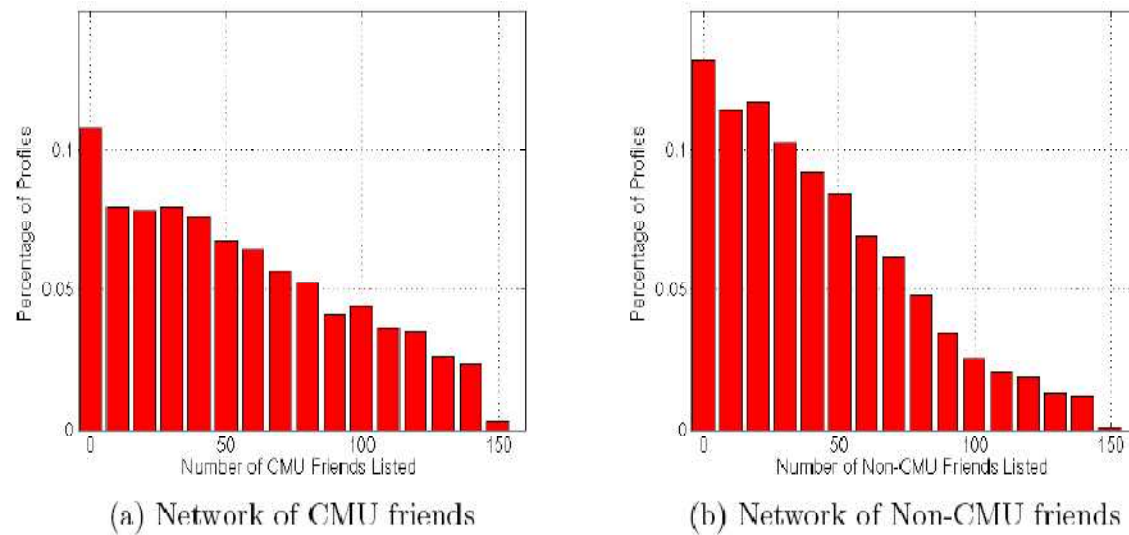


Figure 3: Histogram of the size of networks for both CMU friends (a) and non-CMU friends (b). Users maintain large networks of friends with the average user having 78.2 friends at CMU and 54.9 friends elsewhere.



Site de socialisation et sphère privée

- Quel information ?
 - Identification Nouvelle
 - *connu -> infos supplémentaires*
 - *Inconnu-> connu*
- • Qui peut accéder aux informations ?
 - Le site lui même
 - Parties Tierce (hackers, gouvernements, cyber-criminels, escrocs)
 - Utilisateurs (peu de contrôle sur le réseau et son expansion)
- • Comment les données privées peut être exploitée
 - Usurpation d'identité au harcèlement (espionnage) physique en ligne ou hors ligne
 - Spamming, campagne marketing, sondages, arnaque



Facebook

- www.facebook.com
- Début Février 2004
- Site de socialisation en milieu universitaire
- Etendue à 2000 écoles et 6.4 millions d'utilisateurs
- En moyenne, attire 80 pourcent des étudiants « Undergraduate » des universités américaines
- Aussi ,Anciens Elèves , corps professoral, le personnel académique
- Suscite de plus en plus l'attention des média

Quick Search

My Profile [edit](#)
 My Friends
 My Photos
 My Groups
 My Events
 My Messages
 My Account
 My Privacy

Stand Out

CALL FOR ENTRIES



[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, Guereros, End of Evangelion, anything with multiple atrocities
 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."

About Me:

--Hubertus Bigend, in "Pattern Recognition" by William Gibson
 This used to be angry.

Photos

William has 1 Album.

Eat babies
Created January



Communauté universitaire dans Facebook

- Compte e-mail d'une université requis pour l'enregistrement.
- Regroupement des étudiants d'une même université



Politique de protection de la vie privée

- *L'utilisateur n'est pas forcément au courant :*
 - *“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.”*



Information permettant l'identification

- Piste pour l'identification

- Adresse du domicile, domicile Actuel (Adresse IP)
- Réseau d'amis

- Identification unique

- Email de l'université
- Information de contact (Tel, nom, prénom , date de naissance)



Etude

- En Juin 2005, ils ont crée des profiles Facebook selon différentes caractéristiques tel que :
 - – Ex , degré de « connectedness », situation géographique, ...
- Ils ont recherchés des profiles sur Facebook
 - 4540 profiles ont été téléchargés
 - Ils ont déduit des informations supplémentaires qui ne sont pas immédiatement visibles des profils.
- Les servers de Facebook ont réagi (automatiquement ?) en désactivant les profils créés

Distribution des ages

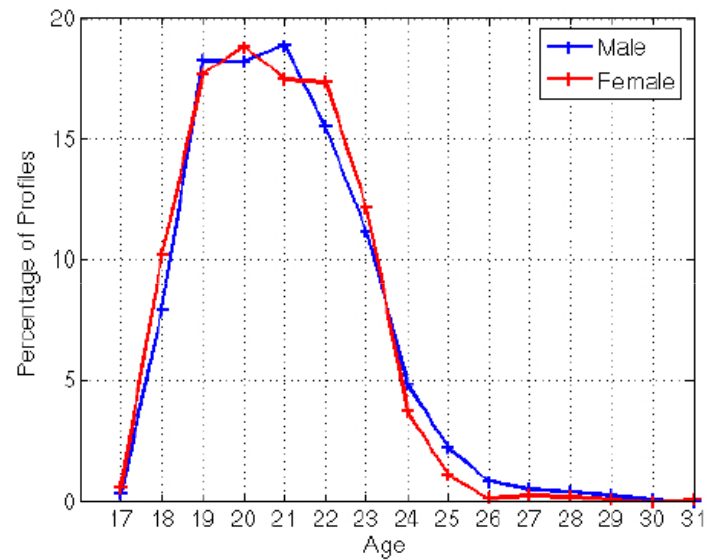


Figure 1: Age distribution of Facebook profiles at CMU. The majority of users (95.6%) falls into the 18-24 age bracket.



Démographie sur Facebook

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

Informations révélées

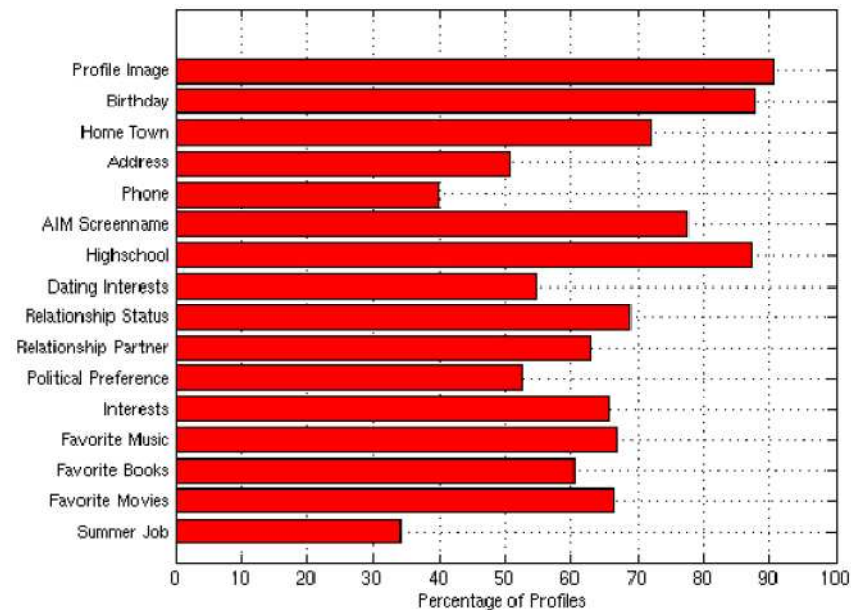


Figure 2: Percentages of CMU profiles revealing various types of personal information.



Informations Révélées

- Les hommes donnent plus d'infos que les femmes
 - 63 % des hommes donnent le numéro de téléphone sur une moyenne de 40%
- Les hommes célibataires révèlent plus facilement leurs numéros de téléphone



Exactitudes des informations

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%



Exactitude des données

Identification Nouvelle de la photo

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%



Exploitation des données fournies

- Espionnage dans la vie réel

- la vie à l'université est centré sur l'assistance aux cours
- les utilisateurs fournissent l'adresse de domicile et la liste de classe
- Le train quotidien est connu en grande partie

- Espionnage en ligne

- les profiles fournissent le login pour la messagerie AIM
- AIM permet d'ajouter des "buddies" sans préavis
- si la politique d' AIM n'a pas changé , un tiers peut tracker quant est qu'un utilisateur est en ligne



Exploitation des données fournies

- identification des membres

 - « 87% of US population is uniquely identified by *{gender, ZIP, date of birth}* »(Sweeney, 2001)

 - Les utilisateurs qui fournissent ces informations peuvent subir une usurpation d'identité



Exploitation des données fournies


- Nouvelle identification basé sur la photo
 - la plupart des photos sont de qualité
 - Les photos peuvent être utilisées pour repérer les utilisateurs dans des sites tierces
 - Match.com et Friendster.com utilisent les mécanismes de reconnaissance facial



Exploitation des données fournies

- Social Security Number

- Format du numéro : xxx yy zzzz
- Basé sur la ville de domiciliation et la date de naissance ,les paramètres xxx and yy deviennent déductible



Statistique sur les risques dues à l'exploitation des données

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

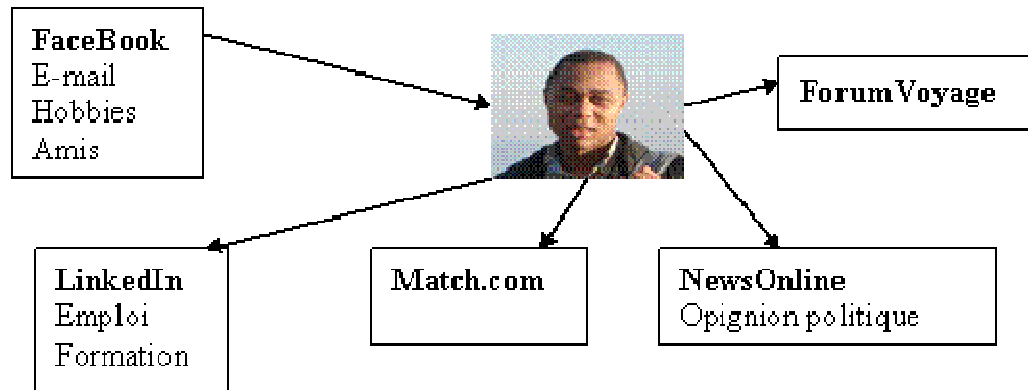
Risk	# CMU Facebook Profiles	% CMU Facebook Profiles
Real-World Stalking	280 (Female)	15.7 (Female)
	580 (Male)	21.2 (Male)
Online Stalking	3528	77.7
Demographics Re-Identification	1676	44.3
Face Re-Identification	2515 (estimated)	55.4



Le Dossier Digital

- Les utilisateurs révèlent les information sensibles (e.g. les partenaires actuels, les opinion politiques)
- Un simple programme de script permet aux adversaire d'extraire et de sauvegarder tous les infos du profil
- Les disques durs de plus en plus bon marché permettent un stockage infini

Enrichir le dossier digital : Puiser dans d'autres sources





Utilisation du dossier digital

- Les responsable universitaires utilisent les infos de Facebook pour rechercher :
 - Les images montrant les étudiants en plein ivresse lors des fêtes
 - Les images montrant des comportement illicites
- Les employeurs exploitent le dossier digital comme source d'information
- Les cyber-criminels et espions peuvent mettre sur pied des dossier digitaux



Test de mise en place d'un dossier digital à partir de Facebook

- Téléchargement du réseau Facebook une fois tous les jours
- A Présent:
 - 69 jours de données
 - 11.1 GB d'informations
 - les changements sur les profiles sont dynamiquement traquées

Configuration de la sécurité des données privées

Basic Settings **Advanced Settings**

Searches

Who can see you in searches? (Your friends at all schools can always see you in searches)

All of Facebook (Recommended)

Restricted (Some people will not be able to add you as a friend)

- Only people at your school
- Only your friends of friends at your school
- Only your friends

Profile

Who can see your profile? (Your friends can always see your profile)

Everyone at your school

Your friends of friends at your school

Only your friends

Among the group selected above, who is allowed to see your profile?

- Undergrads
- Grad Students
- Summer Students
- Alumni
- Faculty
- Staff

Profile Details

What do you want to show in your profile?

- Your Friends (Your friends can always see your friend list)
- Your Last Login Details
- Your Courses
- Your Wall
- Your Groupie Status



Accessibilité des données

- Profile Searchability (On peut rechercher)
 - *1.2% des utilisateurs (18 femmes et 45 hommes utilisent ce paramètre).*
- Profile Visibility :visibilité par les autres membres
 - Evaluation du nombre d'utilisateurs qui utilisent ce paramètre pour rendre leur profil visible qu'à un nombre restreint de personnes
 - *Seulement 0.06 % (3 pers) ont changé les valeurs par défaut*



Conscience des utilisateurs

- 21% croient que seuls les étudiants du CMU peuvent accéder à leurs profils et lire le contenu
- 40% ne réalisent pas que n'importe quel utilisateur peut rechercher leur profil
- 31% ne savent pas que les étudiants du CMU peuvent lire leur profil
- De l'autre côté, 23% croient faussement que tous les utilisateurs de Facebook peuvent lire leur profil



Conscience des utilisateurs

- *85 % ne savent pas que Facebook collecte les informations sur eux à partir d'autres sources .*
- *87% ne savent pas que les informations provenant d'autre source sont utilisées pour enrichir leur profil.*
- 60% ne savent pas que Facebook fournis des informations leur concernant à des organismes tierces et ce à des fins marketing



Conclusion

- Communauté imaginaire ?
- Les profils Facebook sont en réalité des informations publiques
 - Faible protection de la vie privée
 - Réseau d'amis incontrôlés et incontrôlable
 - Confiance aveugle
 - ...
- Il existe différents moyens d'exploiter les informations fournies à des fins divers