@denzil_correa, @ashish_sureka and @raghavsethi
Research Motivation

‣ Meteoric Rise of Social Media
  ✓ Facebook - 500 Million “active” users
  ✓ Twitter - 140 Million “active” users

‣ Existence of Niche Social Networks
  ✓ Foursquare - Location
  ✓ LastFM - Music
  ✓ YouTube - Video
  ✓ Flickr - Photo
Research Motivation

- A user has profiles on multiple social networks
- Eco-system of cross-syndication
- Connect profiles across social networks
YouTube

Sharing and Connected Accounts

You can choose to automatically share YouTube activity on YouTube and your favorite sites. Don't worry, we'll never share information about activity on private videos.

Connect your accounts
- Orkut
- Twitter
- Facebook

Share your activity
On connected accounts
- No accounts connected
- On YouTube
  - Upload a video*
  - Add video to playlist**
  - Like a video
  - Comment on a video
  - Favorite a video
  - Subscribe to a channel

* Public video uploads are always shared on YouTube
** Add To Playlist events are managed by the privacy settings of individual playlists
FourSquare

Settings

Sharing with other networks

If you want to be able to share with other networks, just link your accounts here.

Facebook posting isn't working for your check-ins. Reconnect your account by tapping the button below to fix this!

RECONNECT FACEBOOK

Your foursquare URL will include your Twitter username

ADD TWITTER
Cross Syndication

Tim Cook: "You can converge a toaster and fridge, but those things are probably not gonna be pleasing to the user." http://t.co/HzRB9mO4
@joshuatopolsky on Twitter · via Twitter

Like · Comment

35 people like this.
Cross Syndication

Lauren C.
@lilloco

Just ordered my drink correctly for the first time ever. Grande iced no classic iced coffee. (@Starbucks w/ 4 others) 4sq.com/I1iGFc

7:02 PM - 25 Apr 12 via foursquare · Embed this Tweet

Reply Retweet Favorite

from Starbucks
40 W Lake Street
Chicago, IL
(312) 553-0244

Map data ©2012 Google, Sanborn - Terms of Use
Research Motivation

- A user has profiles on multiple social networks
- Eco-system of cross-syndication & data flow
- Connect profiles across social networks
- ‘Sharing’ is Caring
  - 700 YouTube videos per minute on Twitter
  - 500 Years of YouTube video watched on Facebook
Application Scenarios

- User Data Privacy
- Digital Media Marketing
User Data Privacy

- Link profiles to create a social footprint
- Unintentional Personally Identifiable Information (PII) leakage
Example

<table>
<thead>
<tr>
<th>Username</th>
<th>Name</th>
<th>Gender</th>
<th>Description</th>
<th>Relationship Status</th>
<th>Location</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>dom.cobb</td>
<td>--</td>
<td>Male</td>
<td>--</td>
<td>--</td>
<td>Paris, France</td>
<td>--</td>
</tr>
<tr>
<td>domc</td>
<td>Dom Cobb</td>
<td>--</td>
<td>an engineer...</td>
<td>married</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>dco</td>
<td>--</td>
<td>Male</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>32</td>
</tr>
<tr>
<td>dco</td>
<td>Dom K C</td>
<td>--</td>
<td>loves cricket...</td>
<td>--</td>
<td>France</td>
<td>--</td>
</tr>
<tr>
<td>dom86</td>
<td>Dom Cobb</td>
<td>Male</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>32</td>
</tr>
</tbody>
</table>
Digital Marketing

- Linking profiles not necessarily bad!
- Acquire & harness customer data
- Understand customer engagement
- Target & contextual advertising
Digital Marketing

Lauren C. (@lilloco)

Just ordered my drink correctly for the first time ever. Grande iced no classic iced coffee. (@ Starbucks w/ 4 others) 4sq.com/I1iGFc

7:02 PM - 25 Apr 12 via foursquare · Embed this Tweet

Map data ©2012 Google, Sanborn - Terms of Use
Example

- She likes “Grande Iced coffee”
- ... but not “Classic Iced coffee”
- Business - Starbucks
- Location - Chicago, Illinois
Literature Survey

- Profile Information
- Social Network Information
- Folksonomy Information
Profile Information

- Limitations
  - ✓ Dependent on subset of PII attributes
  - ✓ Sensitive to PII conflicts
  - ✓ Involves guess work
Network Information

- Limitations
  - Friend networks not always publicly available
  - Non-overlapping friend networks
Folksonomy Information

- Limitations

✓ Only on networks which have tags

✓ Twitter, Facebook, Foursquare don’t have tags
<table>
<thead>
<tr>
<th></th>
<th>Bottom Line</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>Guess Work</td>
</tr>
<tr>
<td>X</td>
<td>Expensive</td>
</tr>
<tr>
<td>X</td>
<td>Infeasible to deploy</td>
</tr>
</tbody>
</table>
Intuition

Lauren C. @lilloco

Just ordered my drink correctly for the first time ever. Grande iced no classic iced coffee. (@ Starbucks w/ 4 others) 4sq.com/IiGFc

7:02 PM - 25 Apr 12 via foursquare · Embed this Tweet

Reply Retweet Favorite
Intuition

Lauren C.
@liloco

Just ordered my drink correctly for the first time ever. Grande iced no classic iced coffee. (@ Starbucks w/ 4 others)

4sq.com/IlGfc

Lauren checked in to Starbucks

"Just ordered my drink correctly for the first time ever. Grande iced no classic iced coffee.
20 hours ago via foursquare for iPhone

Points for this check-in:
Every check-in counts!
Solution Approach
Solution Approach

- Filter
- Extract
- Connect
Lauren C.  
@lilloco

Just ordered my drink correctly for the first time ever. Grande iced no classic iced coffee. (@ Starbucks w/ 4 others) 4sq.com/I1iGFC

7:02 PM - 25 Apr 12 via foursquare · Embed this Tweet

Reply Retweet Favorite
Flickr

Whoever thinks #Israel is a democracy: photos last night as police attacked activists exiting event for #nakba flic.kr/p/bBKz49

Activestills.org

14 RETWEETS 3 FAVORITES

11:36 AM - 26 Apr 12 via Flickr - Embed this Tweet
I uploaded a @YouTube video youtu.be/a4KFRoZZDo Tech Daily - Zeppelin Jailbreak Tweak
<table>
<thead>
<tr>
<th>Platform</th>
<th>Filter Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flickr</td>
<td>flic.kr -RT -@</td>
</tr>
<tr>
<td>Foursquare</td>
<td>4sq.com -RT -@</td>
</tr>
<tr>
<td>Last.fm</td>
<td>last.fm/user -RT #nowplaying</td>
</tr>
<tr>
<td>YouTube</td>
<td>I uploaded a youtu.be -RT @youtube</td>
</tr>
</tbody>
</table>
Twitter Search API

<table>
<thead>
<tr>
<th>Platform</th>
<th>Search Query</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>flickr</strong></td>
<td>flic.kr -RT -@</td>
</tr>
<tr>
<td><strong>foursquare</strong></td>
<td>4sq.com -RT -@</td>
</tr>
<tr>
<td><strong>last.fm</strong></td>
<td>last.fm/user -RT #nowplaying</td>
</tr>
<tr>
<td><strong>YouTube</strong></td>
<td>youtu.be -RT @youtube</td>
</tr>
</tbody>
</table>

Also, we put up a Christmas tree.  
[link](flic.kr/p/aPV3NK)

I'm at Starbucks (2245 Fenton Parkway).  
[link](4sq.com/syRQkJ)

#NowPlaying Vanessa Carlton – London  
[link](goo.gl/fb/sgXyY)

I uploaded a @YouTube video  
[link](youtu.be/-8TujhAlMuM) I can feed myself!!
Extract

- URL Extract
- URL Expand
Twitter Search API

<table>
<thead>
<tr>
<th>Service</th>
<th>Filter Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>flickr</td>
<td>flic.kr -RT -@</td>
</tr>
<tr>
<td>foursquare</td>
<td>4sq.com -RT -@</td>
</tr>
<tr>
<td>last.fm</td>
<td>last.fm/user -RT #nowplaying</td>
</tr>
<tr>
<td>YouTube</td>
<td>youtu.be -RT @youtube</td>
</tr>
</tbody>
</table>

Extract

URL Extraction

URL Expansion

1. Also, we put up a Christmas tree. flic.kr/p/aPV3NK
2. I'm at Starbucks (2245 Fenton Parkway) 4sq.com/syRQkJ
3. #NowPlaying Vanessa Carlton – London goo.gl/fb/sgXyY
4. I uploaded a @YouTube video youtu.be/-8TujhAlMuM? a I can feed myself!!
Connect

- Username Extraction

- Get data from the respective API
**Example**

<table>
<thead>
<tr>
<th>Username</th>
<th>Name</th>
<th>Gender</th>
<th>Description</th>
<th>Relationship Status</th>
<th>Location</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>dom.cobb</td>
<td>--</td>
<td>Male</td>
<td>--</td>
<td>--</td>
<td>Paris, France</td>
<td>--</td>
</tr>
<tr>
<td>domc</td>
<td>Dom Cobb</td>
<td>--</td>
<td>an engineer...</td>
<td>married</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>dcobb</td>
<td>--</td>
<td>Male</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>32</td>
</tr>
<tr>
<td>doco</td>
<td>Dom K C</td>
<td>--</td>
<td>loves cricket...</td>
<td>--</td>
<td>France</td>
<td>--</td>
</tr>
<tr>
<td>dom86</td>
<td>Dom Cobb</td>
<td>Male</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>32</td>
</tr>
</tbody>
</table>
Research Aim

- Investigation of activity feed sharing patterns for social profile identity mapping
Experimental Setup

- Twitter Search REST API
- Publicly Available
- 1st December - 31st December 2011
- Random Sample, No bias!
## Filter

<table>
<thead>
<tr>
<th>Social Network</th>
<th>Number of Tweets</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>flickr</strong></td>
<td>43438</td>
</tr>
<tr>
<td><strong>foursquare</strong></td>
<td>43245</td>
</tr>
<tr>
<td><strong>last.fm</strong></td>
<td>43319</td>
</tr>
<tr>
<td><strong>YouTube</strong></td>
<td>13037</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>143039</td>
</tr>
</tbody>
</table>
Just ordered my drink correctly for the first time ever. Grande iced no classic iced coffee. (@Starbucks w/ 4 others) 4sq.com/I1iGFc
Dataset

Activity-Feeds, 87.97%
Desktop-Client, 2.83%
Miscellaneous, 0.07%
Social-Plugin, 1.31%
Mobile-Client, 3.25%
Web, 1.17%
WebApp, 3.38%
# Extract

<table>
<thead>
<tr>
<th>Social Network</th>
<th>User Information in URL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>flickr</strong></td>
<td>username</td>
</tr>
<tr>
<td><strong>foursquare</strong></td>
<td>username</td>
</tr>
<tr>
<td><strong>YouTube</strong></td>
<td>video id</td>
</tr>
<tr>
<td><strong>last.fm</strong></td>
<td>username</td>
</tr>
</tbody>
</table>
# Connect

<table>
<thead>
<tr>
<th>Social Network</th>
<th>Unique Users Mapped</th>
</tr>
</thead>
<tbody>
<tr>
<td>flickr</td>
<td>14102</td>
</tr>
<tr>
<td>foursquare</td>
<td>32646</td>
</tr>
<tr>
<td>YouTube</td>
<td>22762</td>
</tr>
<tr>
<td>last.fm</td>
<td>76</td>
</tr>
<tr>
<td>facebook</td>
<td>16934</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>69496</strong></td>
</tr>
</tbody>
</table>
# Connect

<table>
<thead>
<tr>
<th>Number of Social Networks</th>
<th>Number of Users Mapped</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>86430</td>
</tr>
<tr>
<td>3</td>
<td>17216</td>
</tr>
<tr>
<td>4</td>
<td>97</td>
</tr>
</tbody>
</table>
Results
Normalized Attribute Leakage

- Normalized Attribute Leakage for a particular attribute = Number of users who have the attribute / Total Number of users
Unintended PII
## Username Uniqueness

<table>
<thead>
<tr>
<th>Social Network</th>
<th>Matching usernames</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>flickr</strong></td>
<td>3085 (21.88%)</td>
</tr>
<tr>
<td><strong>foursquare</strong></td>
<td>31610 (96.33%)</td>
</tr>
<tr>
<td><strong>YouTube</strong></td>
<td>6683 (30.36%)</td>
</tr>
<tr>
<td><strong>last.fm</strong></td>
<td>31 (40.79%)</td>
</tr>
<tr>
<td><strong>facebook</strong></td>
<td>4702 (27.77%)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>46311/69496 = 66.63%</td>
</tr>
</tbody>
</table>
Melina B.  @melimuses
So to clarify, I have started a second twitter account to support work-related PAC team efforts. Follow me @melinabaetti for PC me! haha
12:11 AM - 25 Apr 12 via web · Details

Melina B.
@melimuses
Coca-Cola Communications girl, UGA Grady grad, 'sun worshiper', lover of anything French, sports, traveling & Romeo. Recently back in ATL. Opinions are my own!
Atlanta, Georgia ·

Melina Baetti
@melinabaetti
Coca-Cola Public Affairs & Communications, University of Georgia Grady Graduate, Citizen of the world
Atlanta, Georgia ·
## Duplicate Profiles

<table>
<thead>
<tr>
<th>Social Network</th>
<th>Number of Duplicate Profiles</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>flickr</strong></td>
<td>255</td>
</tr>
<tr>
<td><strong>foursquare</strong></td>
<td>30</td>
</tr>
<tr>
<td><strong>YouTube</strong></td>
<td>85</td>
</tr>
<tr>
<td><strong>last.fm</strong></td>
<td>5</td>
</tr>
<tr>
<td><strong>facebook</strong></td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>375/69496 = 0.54%</td>
</tr>
</tbody>
</table>
# Evaluation

<table>
<thead>
<tr>
<th>Social Network</th>
<th>Users Mapped</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>whacky</strong></td>
</tr>
<tr>
<td><strong>flickr</strong></td>
<td>14064</td>
</tr>
<tr>
<td><strong>foursquare</strong></td>
<td>32646</td>
</tr>
<tr>
<td><strong>YouTube</strong></td>
<td>22672</td>
</tr>
<tr>
<td><strong>last.fm</strong></td>
<td>76</td>
</tr>
<tr>
<td><strong>facebook</strong></td>
<td>16934</td>
</tr>
</tbody>
</table>
Web Application

- Location, Gender, Relationship, Photo-count
- Hosted on Google App Engine
- Twitter Bootstrap UI
Demo
Whacky

What anybody could know about you

log in with Twitter!

FAQs

What does Whacky do?
Whacky scans your Twitter profile and gets publicly available information from your linked accounts. You can use Whacky to know which social network is leaking.

<table>
<thead>
<tr>
<th>attribute</th>
<th>information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>throwaway1408 or Raghu Sethi</td>
</tr>
<tr>
<td>Location</td>
<td>New Delhi, India</td>
</tr>
<tr>
<td>Gender</td>
<td>male</td>
</tr>
<tr>
<td>Relationship</td>
<td>single</td>
</tr>
<tr>
<td>Photocount</td>
<td>2</td>
</tr>
</tbody>
</table>

Authorize Whacky to use your account?

This application will be able to:
- Read Tweets from your timeline.
- See who you follow.

Authorize app  No, thanks

This application will not be able to:
- Follow new people.
- Update your profile.

<table>
<thead>
<tr>
<th>service</th>
<th>userid</th>
<th>information</th>
</tr>
</thead>
<tbody>
<tr>
<td>flickr</td>
<td>raghav_sethi</td>
<td>name: Raghu Sethi location: New Delhi, India gender: male</td>
</tr>
<tr>
<td>foursquare</td>
<td>21489712</td>
<td>name: Raghu Sethi location: New Delhi, India gender: male</td>
</tr>
<tr>
<td>facebook</td>
<td>526847183</td>
<td>name: Raghu Sethi location: New Delhi, India gender: male</td>
</tr>
<tr>
<td>youtube</td>
<td>throwawaysirs</td>
<td>name: throwawaysirs relationship: single</td>
</tr>
<tr>
<td>twitter</td>
<td>throwaway1408</td>
<td>name: throwaway1408</td>
</tr>
</tbody>
</table>
Contributions

‣ Investigation of activity feed sharing patterns for social profile identity mapping

‣ First focused study on social profile identity mapping on Twitter
In a nutshell ...

| ✓ | No Guess Work |
| ✓ | Explosive |
| ✓ | Infeasible to deploy |
Discussion
Foursquare API

- Foursquare API
  - Twitter username, Facebook username, e-mail, phone number
- Foursquare user ids - ‘n’ digit serial numbers
- Start from 1 to 15 Million
- Privacy concerns
Pros and Cons

Advantages
✓ Cheap & Elegant
✓ No Manual Evaluation, 100% Accuracy
✓ Scalable
✓ Not specific to Twitter

Limitations
✓ Only for networks with activity feeds
Future Directions

- Complement this approach to existing approaches
- Look for other activity signals
- Public Identity Mapping API
Conclusion

‣ Activity Feed Sharing Patterns to link social network profiles

‣ Cheap and Elegant

‣ Conservative but 100% Accuracy

‣ Scalable
Thank You!
References


Simplicity is the ultimate sophistication.