@I seek ‘fb.me’:
Identifying Users across Multiple Online Social Networks

Workshop on Web of Linked Entities (WoLE)

Paridhi Jain¶, Ponnurangam Kumaraguru¶, Anupam Joshi*

¶Indraprastha Institute of Information Technology (IIIT-Delhi)

*University of Maryland, Baltimore County (UMBC)
Motivation

Multiple OSNs

Multiple Identities

Difficult to manage? **Difficult to find?**
Motivation

Multiple OSNs

Motivation

Multiple Identities

Social Aggregation site

Difficult to manage? Difficult to find?
Motivation

Multiple OSNs

Social Aggregation site

Difficult to manage? Difficult to find?

Friend Finder?
Malicious user?
Influential user?
User of interest?
Motivation

Multiple OSNs

Multiple Identities

Social Aggregation site

Friend Finder? Malicious user? Influential user? User of interest?

Difficult to manage? Difficult to find?

Identity Resolution Problem
Identity Resolution

• For a user I, given a user identity I_A on a social network A, find user identity I_B on social network B.

{I_A}  
Alice

{I_B}  
??
Identity Resolution = Identity Search + Identity Matching

- Identity Search

For a user I, given her identity $I_A$ on a social network A, and a search parameter $S$, find the set of identities $I_{Bj}$ on social network B such that $S(I_A) = S(I_B)$.

$$\{I_A, S\} \rightarrow \{I_{B1}, ... I_{Bj}, ... , I_{BN}\} = Q$$

- Identity Matching

Given a user identity $I_A$ on a social network A, a set of candidate identities $Q$ on social network B, and a match function $M$, locate an identity pair $(I_A, I_{Bj})$ such that $M(I_A, I_{Bj}) = \max\{M(I_A, I_{B1}), M(I_A, I_{BN})\}$

$$\{I_A, Q, M\} \rightarrow \{I_A, I_{Bj}\} \rightarrow \{I_B\}$$
Research Gaps?

- Till now, focus on better identity matching algorithms
- Only profile attributes (private and public) for Identity Search
- Limitations of Profile Search -
  - Restrictive search, owing to non-availability of common attributes across networks. [Gender on Facebook, but not on Twitter]
  - Search with Limited attributes $\rightarrow$ Large candidate set size $\rightarrow$ Intensive Identity Matching computation
  - Users may choose different profile attributes $\rightarrow$ Miss out correct identity in the candidate set
- Little research on using content and network attributes to search for candidate identities
- Extensive use of both private and public attributes. Need user authorization for identity search
Research Gaps?

- Till now, focus on better identity matching algorithms
- Only profile attributes (private and public) for Identity Search
- Limitations of Profile Search -
  - Restrictive search, owing to non-availability of common attributes across networks. [Gender on Facebook, but not on Twitter]
  - Search with Limited attributes $\rightarrow$ Large candidate set size $\rightarrow$ Intensive Identity Matching computation
  - Users may choose different profile attributes $\rightarrow$ Miss out correct identity in the candidate set
- Little research on using content and network attributes to search for candidate identities
- Extensive use of both private and public attributes. Need user authorization for identity search
Research Gaps?

- Till now, focus on better identity matching algorithms
- Only profile attributes (private and public) for Identity Search
- Limitations of Profile Search -
  - Restrictive search, owing to non-availability of common attributes across networks. [Gender on Facebook, but not on Twitter]
  - Search with Limited attributes → Large candidate set size → Intensive Identity Matching computation
  - Users may choose different profile attributes → Miss out correct identity in the candidate set
- Little research on using content and network attributes to search for candidate identities
- Extensive use of both private and public attributes. Need user authorization for identity search
Proposal

• Include content and network attributes as search parameters
• Access **only publicly** accessible attributes
• Focus on two popular social networks - Twitter and Facebook
Contribution

- Proposed novel identity search methods on social networks
- Our identity resolution methods return correct Facebook identity for 39% Twitter users within top-2 ranks
- We observe an increase in accuracy of identity resolution by 11.6% owing to inclusion of content and network identity search, along with improvised profile search
Methodology

Candidate Identities

Search

Match

If self-identified returned by more than one search method

Yes

No

Syntactic and Image

Manual Verification

13/05/13

@I seek 'fb.me': Identifying Users across Multiple Online Social Networks
Identity Matching

• Syntactic Matching
  • Jaro Distance comparison between username and name
  • Example: {alice123, jane_alice}, {Alice Naura, Alice N. Janice}

• Image Matching

\[ IM_s(I_A, I_{Bj}) = \sqrt{\frac{(h_{IA} - h_{IBj})^2}{N_s}} \]

where \( h_{IA} \) and \( h_{IBj} \) are the RGB histograms of the profile image and \( N_s \) represent histogram size of \( I_A \)
Profile Search

Self - Identification

Non-Ranked Candidate Set

Username + Name + Location

Facebook Graph API

Username, Username Profile Image, URL

If URL has Facebook ID?

Yes

Non-Ranked Candidate Set

Facebook Graph API

Username + Name + Location

If URL has Facebook ID?

No

Profile Search
Profile Search

Twitter Search API

Name, Username Profile Image, URL

If URL has Facebook ID?

Yes

Self - Identification

No

Facebook Graph API

Username + Name + Location

Non-Ranked

Twitter Identity Details

Name: Kristina L
Username: KristinaLerman
URL: None
Description: full time working parent and social web researcher

Candidate Facebook Identities

Name: Kristina Lerman
Username: kristina.lerman
URL: www.facebook.com/kristina.lerman
Gender: female
Content Search

Hindupremi @hindupremi
Akhilesh Yadav’s cops unleash brutality on women protesting rape, murder of child bit.ly/15oGlhr
Expand

Hindu Adhiveshan
5 hours ago

Akhilesh Yadav’s cops unleash brutality on women protesting rape, murder of child

In a horrific display of police brutality, police in SP-ruled Uttar Pradesh unleashed violence on women and men who had come out into the streets of Aligarh to protest the rape of a six-year-old girl who went missing on Thursday. The girl’s body was found lying in a garbage pile this morning.
Content Search

@I seek ‘fb.me’: Identifying Users across Multiple Online Social Networks
Self-mention Search

Diagram:
- Twitter Search API
- Extract Tweets (Content)
- Extract URLs in tweets
- Social Network API (Extract user handle and name)
- Non-Ranked Candidate Set
- If any URL points to other social network?
  - Yes
  - No
- User can't be located
Self-mention Search

![Diagram showing the process of self-mention search involving Twitter and Facebook APIs to identify users across multiple social networks.]

Showcased @precog_iitd work to the Chief Minister of Delhi today! Some pics from the discussion and demo:
facebook.com/media/set/?set...
Network Search

Queried user - “alice123”
Name - Alice

Followees / Followers

bob123
Twitter URL - http://www.facebook.com/xyz

bob

Subset of Friends

Candidate identity
Name - Alice

Self-identification

IA connection network

IA Candidate friend-neighborhood

IA friend’s friend-neighborhood
Instance,

Name: Paridhi Jain
Username: pari_lakshya
URL: None
Description:

Name: Paridhi Jain
Username: paridhi.jain.399
Gender: female
Instance,

Name: Paridhi Jain
Username: pari_lakshya
URL: None
Description:

Public Friend List of a user extracted from public feeds
Integrated System – Finding Nemo

Search

Profile
Content
Self-mention
Network

Candidate Identities

If self-identified / returned by more than one search method

Yes

Syntactic and Image
Manual Verification

Match

No

@I seek ‘fb.me’: Identifying Users across Multiple Online Social Networks
### Evaluation

Accuracy = \( \frac{U_{\text{correct}}}{U_{\text{total}}} \)

<table>
<thead>
<tr>
<th>Method (543 users)</th>
<th># of users</th>
<th>% Accurate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Profile (P)</td>
<td>205</td>
<td>37.7</td>
</tr>
<tr>
<td>Content (C + SM)</td>
<td>34</td>
<td>6.3</td>
</tr>
<tr>
<td>Network (N)</td>
<td>1</td>
<td>0.2</td>
</tr>
<tr>
<td>Finding Nemo</td>
<td>212</td>
<td>39</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dataset</th>
<th># of users</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social Graph API</td>
<td>543</td>
</tr>
</tbody>
</table>
### Evaluation

*Accuracy* = \( \frac{U_{\text{correct}}}{U_{\text{total}}} \)

<table>
<thead>
<tr>
<th>Dataset</th>
<th># of users</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social Graph API</td>
<td>543</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method (543 users)</th>
<th># of users</th>
<th>% Accurate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Profile (P)</td>
<td>205</td>
<td>37.7</td>
</tr>
<tr>
<td>Content (C + SM)</td>
<td>34</td>
<td>6.3</td>
</tr>
<tr>
<td>Network (N)</td>
<td>1</td>
<td>0.2</td>
</tr>
<tr>
<td>Finding Nemo</td>
<td>212</td>
<td>39</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Search Algorithm</th>
<th># of users identified</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>P (without URL)</td>
<td>149</td>
<td>27.4%</td>
</tr>
<tr>
<td>P (with URL) + C + N + SM</td>
<td>149+56+6+1 = 149+71</td>
<td>27.4% + 11.6%</td>
</tr>
</tbody>
</table>
Mean Average Precision

\[
MAP = \frac{1}{U_{correct}} \sum_{j=1}^{U_{correct}} \frac{1}{R_j} \sum_{k=1}^{R_j} P(cand_k) \cdot rel(cand_k)
\]

\[
\downarrow
\]

\[
MAP = \frac{1}{U} \sum_{j=1}^{U} P(cand_k) \cdot rel(cand_k)
\]

<table>
<thead>
<tr>
<th>Matching algorithm</th>
<th>MAP Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image (profile image)</td>
<td>0.83</td>
</tr>
<tr>
<td>Syntactic (username)</td>
<td>0.76</td>
</tr>
<tr>
<td>Syntactic (name)</td>
<td>0.80</td>
</tr>
</tbody>
</table>
Demo
Inclusion of content and network attributes for identity search not only improves identity resolution accuracy but returns correct Facebook identity within top-2 ranks for majority of the Twitter users.
Current and Future Work

- Extend the social networks to search for a given identity. Example, Google+, Foursquare, etc.
- Extend the search methods to include social-network specific features
- Find multiple (fake) identities of users within social networks
Questions?

paridhij@iiitd.ac.in, pk@iiitd.ac.in, joshi@cs.umbc.edu
precog.iiitd.edu.in

Paper: http://precog.iiitd.edu.in/publications.html
THANK YOU!

October 7 – 8, 2013
Boston, USA
http://cosn.acm.org/
Paper submission due: June 28, 2013