To Join or Not to Join
the illusion of privacy in social networks with mixed public and private user profiles

Elena Zheleva and Lise Getoor
WWW, April 23, 2009
Motivation

- Online social networks (OSN) are popular
  - allow complex interactions

- Bring many privacy concerns
  - preserving privacy is challenging
  - some OSN allow private profiles
A public profile on Facebook

Attributes

Groups

Friends

Disclaimer: most of the Facebook examples in this presentation are fictitious.
Emily’s friends and groups

- Friends
- Group affiliation
- Private profile
- Public profile

Group affiliations cannot be hidden!
Focus of this work

- The privacy of private profiles
- 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
Privacy breaches in networks

- **Identity disclosure**
  - E.g. “Elise Labott” refers to the CNN reporter Elise Labott

- **Attribute disclosure**
  - E.g. Elise Labott is 30

- **Link/relationship disclosure**
  - E.g. Elise Labott is friends with Emily

- **Group membership disclosure**
  - E.g. Elise is a member of “Sarah Palin is NOT Hillary Clinton”

---

1: Backstrom et al. WWW ‘07, Hay et al. VLDB ‘08, Liu & Terzi SIGMOD ’08, Narayanan SP ’08, ’09
2: He et al. ISI ‘06, Lindamood et al. WWW ‘09
3: Zheleva & Getoor PinKDD ’07, Korolova et al. CIKM ‘08
4: Choffnes et al. Northwestern TR ‘09
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?
Sensitive attribute inference

- Sensitive attribute is a random variable
- Depends on:
  - overall network’s attribute distribution
  - the friendship network’s attribute distribution
  - the attribute distribution of each user group
  - no other attributes are available for private profiles
- Assume adversary can apply a probabilistic model $M$ to predict it

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

Friendship network:

Social network groups:

- known sensitive attributes (public profiles)
- unknown attributes (private profiles)
Inference models
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
Model based on overall network distribution
- in the absence of links and groups
- BASIC: assigns majority label

\[ P_{\text{BASIC}}(v_s.a = a_i; G) = \frac{|V_o.a_i|}{|V_o|} \]

Label distribution:
- Chris: 2
- Fabio: 1
- Don: 0
- Emma: 0
- Ana: 0
- Bob: 1
- Gia: 0
Attribute inference models

- Link-based models
  - AGG: aggregate over public friends’ labels (majority)
    \[
P_{AGG}(v_s, a = a_i; G) = P(v_s, a = a_i | V_o\cdot A, E) = \frac{|V_o'.a_i|}{|V_o'|}
    \]
  - 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
      \[
P_{BLOCK}(v_s, a_i; G) = P(v_s, a_i | V_o\cdot A, E, \lambda) = \frac{1}{Z} \text{sim}(\lambda_i, \lambda(v_s))
      \]
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)
Attribute inference models

- Model using both links and groups
  - LINK-GROUP: friends and groups are features
    - uses a global classifier, e.g. SVM, Naïve Bayes, LR

Example:

Emma (0 1 0 0 0 1 1 1 0)
Ana (0 1 0 1 0 0 0 1 0)
Evaluation
Data description

- **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)
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
Flickr: country (55 values), 50% private profiles
Attribute inference results

- **Facebook**
  - Gender (2 values)
  - Political views (6)

Bar charts showing accuracy percentages for different methods:
- Random guess
- BASIC
- BLOCK
- AGG
- CC
- LINK
- CLIQUE-LINK
- GROUP
- GROUP* (50% nodes)
- LINK-GROUP
Attribute inference results

- Dogster: breed category (7 possible values)
Attribute inference results

- BibSonomy: spammer or not (2 possible values)
GROUP*: informative groups based on homogeneity
- measured by entropy

Accuracy with varying entropy on Flickr

<table>
<thead>
<tr>
<th>Maximum group entropy</th>
<th>Accuracy</th>
<th>Node coverage</th>
<th>Group coverage</th>
<th>Baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>73.3%</td>
<td>86.7%</td>
<td>28.8%</td>
<td>27.7%</td>
</tr>
<tr>
<td>1</td>
<td>75.3%</td>
<td>94.0%</td>
<td>47.4%</td>
<td>27.7%</td>
</tr>
<tr>
<td>2</td>
<td>73.9%</td>
<td>97.4%</td>
<td>65.7%</td>
<td>27.7%</td>
</tr>
<tr>
<td>3</td>
<td>70.7%</td>
<td>98.8%</td>
<td>81.8%</td>
<td>27.7%</td>
</tr>
<tr>
<td>4</td>
<td>67.7%</td>
<td>99.6%</td>
<td>91.4%</td>
<td>27.7%</td>
</tr>
<tr>
<td>5</td>
<td>64.8%</td>
<td>100.0%</td>
<td>93.3%</td>
<td>27.7%</td>
</tr>
</tbody>
</table>
- **GROUP**\(^*\) - informative groups chosen based on % public profiles per group

<table>
<thead>
<tr>
<th>Minimum percent public profiles per group</th>
<th>Accuracy</th>
<th>Node coverage</th>
<th>Baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td>75.0%</td>
<td>88.1%</td>
<td>27.7%</td>
</tr>
<tr>
<td>30%</td>
<td>77.1%</td>
<td>83.3%</td>
<td>27.7%</td>
</tr>
<tr>
<td>50%</td>
<td>83.0%</td>
<td>52.3%</td>
<td>27.7%</td>
</tr>
<tr>
<td>70%</td>
<td>83.4%</td>
<td>13.8%</td>
<td>27.7%</td>
</tr>
<tr>
<td>90%</td>
<td>100.0%</td>
<td>0.02%</td>
<td>27.7%</td>
</tr>
</tbody>
</table>
GROUP*: vary n, % public profiles

Accuracy with varying % public profiles

<table>
<thead>
<tr>
<th>Percent public profiles</th>
<th>Accuracy</th>
<th>Node coverage</th>
<th>Baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td>10%</td>
<td>63.4%</td>
<td>85.7%</td>
<td>27.7%</td>
</tr>
<tr>
<td>30%</td>
<td>71.6%</td>
<td>74.7%</td>
<td>27.7%</td>
</tr>
<tr>
<td>50%</td>
<td>83.6%</td>
<td>51.2%</td>
<td>27.7%</td>
</tr>
<tr>
<td>70%</td>
<td>86.2%</td>
<td>48.6%</td>
<td>27.7%</td>
</tr>
<tr>
<td>90%</td>
<td>87.6%</td>
<td>28.9%</td>
<td>27.7%</td>
</tr>
</tbody>
</table>
Joining more homogeneous groups poses higher threat to privacy

Privacy and policy
- allow private group affiliations
- not join certain groups at all

Would hiding affiliations to homogeneous groups help?
GROUP: joining only diverse groups

<table>
<thead>
<tr>
<th>Minimum entropy of groups that users join</th>
<th>Accuracy</th>
<th>Node coverage</th>
<th>Group coverage</th>
<th>Baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>65.2%</td>
<td>100.0%</td>
<td>92.9%</td>
<td>27.7%</td>
</tr>
<tr>
<td>1</td>
<td>57.7%</td>
<td>99.8%</td>
<td>56.6%</td>
<td>27.7%</td>
</tr>
<tr>
<td>2</td>
<td>50.4%</td>
<td>99.5%</td>
<td>29.6%</td>
<td>27.7%</td>
</tr>
<tr>
<td>3</td>
<td>40.0%</td>
<td>99.1%</td>
<td>11.9%</td>
<td>27.7%</td>
</tr>
<tr>
<td>4</td>
<td>20.3%</td>
<td>96.8%</td>
<td>1.6%</td>
<td>27.7%</td>
</tr>
<tr>
<td>5</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>27.7%</td>
</tr>
</tbody>
</table>

Accuracy for users who do not join low-entropy groups

Minimum entropy of groups that users join
Showed how easy it can be to predict private user attributes

Groups have a high potential for leaking private information
  - more informative than friendship links

Hope: motivate social media websites to enable greater control over release of information
THANK YOU!

To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles

Elena Zheleva and Lise Getoor
WWW, April 23, 2009