Security and Trust in social media networks

Prof. Touradj Ebrahimi
touradj.ebrahimi@epfl.ch

DMP meeting
San Jose, CA, USA
11 January 2014
Social media landscape

http://www.fredcavazza.net/2012/02/22/social-media-landscape-2012/
Popularity of social networks

Number of Users on Popular Social Networking Sites

May 2012

Monthly Visits on Top Social Networking Websites

March 2012

Popularity of social networks (cont.)

- Instagram: 1+ billion photos
- Picasa: 7+ billion photos
- Facebook: ~140 billion photos
- YouTube: 72 hours of video every minute

April 2012
Popularity of social networks (cont.)

Facebook:
- 510,000 Post comments
- 293,000 Status updates

Twitter:
- 120,000 Tweets

Foursquare:
- 2,083 Check-ins

February 2012
Popularity of social networks (cont.)

http://online.wsj.com/article/SB10001424052970204653604577249341403742390.html
Trust & security issues in social networks
• **Socialcam** application *automatically shares* with your Facebook friends videos you watch
• May be doing you more harm than good!
• **Spotify** and **Yahoo News Activity** automatically publish songs and news you have listened to or read on your profile page.
• “Responsible” sharing (e.g. a friend posted an embarrassing photo of a young political candidate Emma Kiernan enjoying a party)

Anybody can post any picture of you on Facebook at any time! Even if deleted, it is there forever, stored in the vast Internet memory bank to be found again and again!

• “Responsible” sharing (e.g. top model Rosanagh Wypych posted photos of her and friends drinking alcohol and consuming cannabis)

Trust & security issues in social networks (cont.)

• **Wrong tags in Flickr**

Only around 50% of tags are truly related to an image

Challenges:
- Identify the most appropriate tags
- Eliminate noisy or spam tags

[Kennedy et al., ACM MIR 2006]
• **Spam bookmarks** in Delicious

![Spam bookmarks in Delicious](image-url)
Trust & security solutions in social networks
• **Limit number of people** to share content and communications with – create private groups (e.g. Google+ Circles, Facebook Smart Lists)
• **Limit time** to watch shared content (e.g. SnapChat application – though they are **not able to guarantee** that your messaging data will be deleted in all instances)
Trust & security solutions in social networks (cont.)

- **Manage** your account (friends, photos, comments, posts)
- **Make sure** that what you are sharing or posting is not going to cause regret

---

http://www.huffingtonpost.com/2012/03/20/social-media-privacy-infographic_n_1367223.html
Trust & security before social media

- E-mail
- Web search
- Web videos
- Blogs
- Online shopping systems
- Peer-to-peer (P2P) networks
Anti-spam strategies in online systems

Detection-based
- Text classification
- Link analysis
- User behavior analysis
- Manual user identification
- Manual moderator identification

Demotion-based
- Spam-hardened queries

Prevention-based
- CAPTCHAs
  - Account fees
  - Proof of work
  - Pay per action
  - Community size imits
- Hidden or partitioned input interfaces

[Heymann, Koutrika, Garcia-Molina, IEEE IC 2007]
Trust and reputation online systems

Customer Reviews
Citizen Men's AT0200-05E Eco-Drive Chronograph Canvas Watch

190 Reviews
Average Customer Review: 4.5 stars (159 customer reviews)

See most helpful viewpoints

Nice Addition., August 25, 2010
By DaveH "Deal Hunter" (L, NY) - See all my reviews
Amazon Verified Purchase (What's this?)
This review is from: Citizen Men's AT0200-05E Eco-Drive Chronograph Canvas Watch (Watch)

Just received my Citizen AT0200-05E today and the watch looks better in person. Read lots of reviews about the stock band being short but it wasn't a concern for me as I ordered my watch with the Modena Peneral black band upgrade. The rubber band really dresses up the watch nicely and adds a diving watch quality to it. I've included some pictures with my review with the "wet" look and I think it's a great upgrade to an already awesome watch!

Help other customers find the most helpful
Was this review helpful to you? Yes | No

PageRank
Trust and reputation online systems (cont.)

Email spam filtering

Bayes rule

In this system, spam probability is calculated using Bayes' rule. The spam probability is 5.98%.

Sample database

Delete as spam

Emails are classified as spam or non-spam based on their content and past interactions.

Incoming e-mail

Bayes rule

\[ P_s(E) = \frac{P(H)P_s(E)}{P(H)P_s(E) + P(\neg H)P_\neg(E)} \]

Contact member
View items for sale
View more options

In-box

5.98% spam probability
Model of social tagging system

[Ivanov, Vajda, Lee, Ebrahimi, IEEE SPM 2012]
Categorization of trust models

- Trust Models
  - Content Trust
  - User Trust
    - Static Trust
    - Dynamic Trust

[Ivanov, Vajda, Lee, Ebrahimi, IEEE SPM 2012]
User versus content trust modeling

- **User trust** modeling is *more popular* than content trust modeling:
  - User trust models has a *less complexity* than content trust models
  - User trust models can *quickly adapt* to the constantly evolving and changing environment in social systems due to the type of features used for modeling, and thus be applicable longer than content trust models, without need for creation of new models

- **User trust modeling** has a disadvantage of “*broad brush*”, i.e. it may be excessively strict if a user happens to post one bit of questionable content on otherwise legitimate content

- **“Subjectivity”** in classifying spam and non-spam content/users remains as a fundamental issue, i.e. what is spam content/user to one person may be interesting to another one
## Summary of representative recent techniques

<table>
<thead>
<tr>
<th>Reference</th>
<th>Trust model</th>
<th>Media</th>
<th>Method</th>
<th>Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gyongyi <em>et al.</em> 2004</td>
<td>content</td>
<td>web pages</td>
<td>an iterative approach, called TrustRank, to propagate trust scores to all nodes in the graph by splitting the trust score of a node among its neighbors according to a weighting scheme</td>
<td>AltaVista, real</td>
</tr>
<tr>
<td>Koutrika <em>et al.</em> 2008</td>
<td>content</td>
<td>bookmarks</td>
<td>a coincidence-based model for query-by-tag search which estimates the level of agreement among different users in the system for a given tag</td>
<td>Delicious, real &amp; simulated</td>
</tr>
<tr>
<td>Wu <em>et al.</em> 2005</td>
<td>content</td>
<td>images</td>
<td>a computer vision technique based on low-level image features to detect embedded text and computer-generated graphics</td>
<td>SpamArchive &amp; Ling-Spam, real</td>
</tr>
<tr>
<td>Liu <em>et al.</em> 2009</td>
<td>content &amp; user</td>
<td>bookmarks</td>
<td>an iterative approach to identify spam content by its information value extracted from the collaborative knowledge</td>
<td>Delicious, real</td>
</tr>
<tr>
<td>Bogers and Van den Bosch 2008</td>
<td>content &amp; user</td>
<td>bookmarks</td>
<td>KL-divergence to measure the similarity between language models and new posts</td>
<td>BibSonomy &amp; CiteULike, real</td>
</tr>
<tr>
<td>Ivanov <em>et al.</em> 2012</td>
<td>user</td>
<td>images</td>
<td>an approach based on the feedback from other users who agree or disagree with a tag associated with an image</td>
<td>Panoramio, real</td>
</tr>
</tbody>
</table>

[Ivanov, Vajda, Lee, Ebrahimi, IEEE SPM 2012]
### Summary of representative recent techniques (cont.)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Trust model</th>
<th>Media</th>
<th>Method</th>
<th>Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xu <em>et al.</em> 2006</td>
<td>user</td>
<td>bookmarks</td>
<td>an iterative approach to compute the <strong>goodness</strong> of each tag with respect to a content and the <strong>authority scores</strong> of the users</td>
<td>MyWeb 2.0, real</td>
</tr>
<tr>
<td>Krestel and Chen 2008</td>
<td>user</td>
<td>bookmarks</td>
<td>a <strong>TrustRank-based</strong> approach using features which model tag co-occurrence, content co-occurrence and co-occurrence of tag-content</td>
<td>BibSonomy, real</td>
</tr>
<tr>
<td>Benevenuto <em>et al.</em> 2009</td>
<td>user</td>
<td>videos</td>
<td>a <strong>supervised learning</strong> approach applied on features that reflect users behavior through video responses</td>
<td>YouTube, real</td>
</tr>
<tr>
<td>Lee <em>et al.</em> 2010</td>
<td>user</td>
<td>tweets</td>
<td>a machine learning approach applied on <strong>social honeypots</strong> including users’ profile and tweets’ features</td>
<td>Twitter, real &amp; simulated</td>
</tr>
<tr>
<td>Krause <em>et al.</em> 2008</td>
<td>user</td>
<td>bookmarks</td>
<td>a machine learning approach applied on a user’s profile, bookmarking activity and context of tags</td>
<td>BibSonomy, real</td>
</tr>
<tr>
<td>Markines <em>et al.</em> 2009</td>
<td>user</td>
<td>bookmarks</td>
<td>a machine learning approach applied on tag-, content- and user-based features</td>
<td>BibSonomy, real</td>
</tr>
<tr>
<td>Caverlee <em>et al.</em> 2008</td>
<td>user</td>
<td>user profiles</td>
<td>an approach to compute a <strong>dynamic trust score</strong>, called SocialTrust, depending on the quality of the relationship and personalized feedback ratings</td>
<td>MySpace, real</td>
</tr>
</tbody>
</table>
Publicly available datasets:

- Publication of datasets from different social networks and even different datasets of one social network for evaluation of trust modeling approaches is rarely found, which makes it difficult to compare results and performance of different trust models.

- Most of the datasets provide data for evaluating only one aspect of trust modeling, either user or content trust modeling, while evaluation of the other aspect requires introducing simulated objects in the real-world social tagging datasets.
Dynamics of trust:

- User’s trust tends to vary over time according to the user’s experience and evolvement of social networks
- Only a few approaches deal with dynamics of trust by distinguishing between recent and old tags
Multilingualism:

- Most of the existing trust modeling approaches based on text information assume **monolingual environments**.
- Some text information may be regarded as wrong due to the **language difference** – people from various countries, so various languages simultaneously appear in tags and comments.
Open issues & challenges (cont.)

- **Interaction across social networks:**
  - How *trust models across domains* can be effectively connected and shared? E.g. users can use their Facebook accounts to log in some other social network services.
Open issues & challenges (cont.)

- **Multimodal analysis:**
  - Most of the current techniques for noise and spam reduction focus only on textual tag processing and user profile analysis, while audio and visual content features of multimedia content can also provide useful information about the relevance of the content and content-tag relationship.
References


Any question?
Prof. Touradj Ebrahimi
Touradj.Ebrahimi@epfl.ch

Thank you!