Psychiatrists, Psychologists and Computer Scientists: Exploring new research collaboration opportunities

Miltiadis Kandias, Dimitris Gritzalis
September 2015
Psychiatrists, Psychologists & Computer Scientists: Exploring new research collaboration opportunities

Miltiadis Kandias & Dimitris Gritzalis
Information Security & Critical Infrastructure Protection Laboratory
Dept. of Informatics | Athens University of Economics & Business

ARIADNE Network
Thessaloniki, Greece
September 2015
Web 2.0 and Online Social Networks (OSN)
Open Source Intelligence (OSINT) is produced from publicly available information, which is:

- Collected, exploited and disseminated in a timely manner
- Offered to an appropriate audience
- Used for the purpose of addressing a specific intelligence requirement

Publicly available information refers to (not only):

- Traditional media (e.g. television, newspapers, radio, magazines)
- Web-based communities (e.g. social networking sites, blogs)
- Public data (e.g. government reports, official data, public hearings)
- Amateur observation/reporting (e.g. amateur spotters, radio monitors)

Social Media Intelligence (SOCMINT) is produced from Online Social Networks and the Web 2.0.
NEREUS: Architecture in a nutshell

Draft research scoping

Results/Classification

Naïve Bayes, SVM, MLR, EM, Canopy, etc.

Naïve Bayes, SVM, MLR, EM, Canopy, etc.

Research specifics explanation, interdisciplinary classification assistance

Training Set

Domain Expert

Language, jargon, content, ethnological characteristics explanation

Linguistics Expert

Sample definition (bias, analysis, etc.)

Statistics Expert

Data preprocessing

Crawling server

API Calls

Medium-to-scope adjustment

Content Expert

Web 2.0 Medium

Massive numbers of independent users

Domain Expert

User

Content generation

Legal Expert

Researchers’ compliance with ethical standards

YES

NO

Research termination

NO

TERMS AND CONDITIONS

Scope Definition

Frameworks, tools, methods, techniques consulting

Domain Expert

Anonymization layer

Table:

<table>
<thead>
<tr>
<th>Web 2.0 Medium</th>
<th>Psychiatrist</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facebook</td>
<td>Sociologist</td>
</tr>
<tr>
<td>Twitter</td>
<td>Psychologist</td>
</tr>
<tr>
<td>YouTube</td>
<td>Political Scientist</td>
</tr>
<tr>
<td>Blogs</td>
<td>etc.</td>
</tr>
</tbody>
</table>

Information Security & Critical Infrastructure Protection Laboratory

Ver. 2.11.14, 24.11.2014
Narcissistic behavior detection

**Study:** Motive, ego/self-image, entitlement

**Means:** Usage Intensity, Influence valuation, Klout score

- Individuals tend to transfer offline behavior online.
- Utilize *graph theoretic* tools to perform analysis.
- Valuation of social media *popularity* and *usage intensity*.
- Assist in detecting *delinquent behavior*
- Assist in predicting *deviant behavior of minors*
Dataset description

- Focus on a Greek Twitter community:
  - Context sensitive research
  - Utilize ethnological features rooted in locality
  - Extract and analyze results
- Analysis of content and measures of user influence and usage intensity
- User Categories: Follower, Following and Retweeter
- 41.818 fully crawled users (personal and statistical data)
  - Name, ID, personal description, URL, language, geolocation, profile state, lists, # of following/followers, tweets, # of favorites, # of mentions, # of retweets

Twitter (Greece, 2012-13)

- 1.075.859 users
- 41.818 fully crawled users
- 7.125.561 connections among them
**Strongly connected components:**
There exists 1 large component (153,121 nodes connected to each other) and several smaller ones

**Node Loneliness:**
99% of users connected to someone

**Small World Phenomenon:**
Every user lies <6 hops away from anyone else

**Indegree Distribution:**
# of users following each user
Average 13.2 followers/user

**Outdegree Distribution:**
# of users each user follows
Average 11 followers/user

**Usage Intensity Distribution:**
Weighted aggregation of {# of followers, # of followings, tweets, retweets, mentions, favorites, lists}
Narcissism detection

Majority of users make limited use of Twitter

- A lot of “normally” active users
- Very few “popular” users
- Users classified into 4 categories, on the basis of specific metrics (i.e., Influence valuation, Klout score, Usage valuation)

Above a threshold:

- User becomes **quite influential/perform intense** medium use
- User gets a “**mass-media & persona**” status

<table>
<thead>
<tr>
<th>Category</th>
<th>Influence valuation</th>
<th>Klout score</th>
<th>Usage valuation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loners</td>
<td>0 - 90</td>
<td>3.55 - 11.07</td>
<td>0 - 500</td>
</tr>
<tr>
<td>Individuals</td>
<td>90 - 283</td>
<td>11.07 - 26.0</td>
<td>500 - 4.500</td>
</tr>
<tr>
<td>Known users</td>
<td>283 - 1.011</td>
<td>26.0 - 50.0</td>
<td>4.500 - 21.000</td>
</tr>
<tr>
<td>Mass Media &amp; Personas</td>
<td>1.011 - 3.604</td>
<td>50.0 - 81.99</td>
<td>21.000 - 56.9000</td>
</tr>
</tbody>
</table>
Research collaboration opportunities

- Several psychosocial characteristics can be studied via SOCMINT methods.

- Contribution from psychiatrists and psychologists could facilitate further understanding of online human behavior.

- New opportunities for analysis of results and findings.

- Psychiatrists and psychologists could benefit from massive, automated data processing offered by data scientists.
General conclusions

✓ Major research opportunities do exist for collaboration between psychiatrists, psychologists and computer scientists.
✓ SOCMINT can transform into intelligence the vast amount of data produced by Web 2.0.
✓ SOCMINT is an intrusive technology and could put in danger civic rights.
✓ SOCMINT utilization is not - and should not be considered as - a solely technical issue.
✓ SOCMINT could assist in predicting attitude towards law infringement and narcissism.
✓ SOCMINT could assist in predicting delinquent behavior of minors.


