

# Here comes the Brave New World of Social Media

Miltiadis Kandias

Athens University of Economics & Business

# Outline



- Social Media crawlable data (OSINT)
- OSINT exploitation
- A story of joy (?) and a horror story (?)
  - **Twitter:** Predicting malevolent insiders
  - **YouTube:** Revealing sensitive, personal data
- Emerging threats
- Legal (or maybe not?) applications



# Social Media Crawlable Data



- Social Media structure facilitates personalized usage
- Reveal psychosocial personality traits
- Multiple usage motives emerge:
  - Professional
  - Entertainment
  - Communication
  - Personal
- Users transfer their offline behavior, online<sup>1,2</sup>
- Concentrate enormous datasets of Open Source INTelligence (OSINT)
  - Information gathered from publicly available sources, utilized when promptly transmitted in order to address specific information needs (DoD, USA).

[1]. Ross C., Orr E, Sisic M., Arseneault J., Simmering M., Orr R., “Personality and motivations associated with Facebook use”, *Computers in Human Behavior*, Vol. 25, pp. 578-586, 2009.

[2]. Amichai-Hamburger Y., Vinitzky G., “Social network use and personality”, *Computers in Human Behavior*, Vol. 26, pp. 1289-1295, 2010.

# Open Source Intelligence (OSINT) Exploitation



- Usage pattern extraction
- User pattern extraction
- Capability of automated psychometric evaluations assessment
- Disclosure of sensitive, personal data<sup>3</sup>
  - Political beliefs
  - Religious beliefs
  - Sexual orientation

[3]. Kosinski M., Stillwell D., Graepel T., “Private traits and attributes are predictable from digital records of human behavior”, in *Proc. of the National Academy of Sciences*, 2013.



## Story of joy (?)

Predicting Malevolent Insiders via  
**Twitter**

# Detecting malevolent insiders via ... narcissism and ... Twitter!



- **Social Medium:**  
Twitter
- **Dataset:**  
1.075.859 users, 7.125.561 connections among them
- **Content:**  
41.818 fully crawled users
- **Graph-theoretic analysis<sup>4</sup>:**
  - Strongly Connected Components (*153.121 users form a graph, where there is a path from every user to another*)
  - Node Loneliness (*1.075.815 users are connected to  $\geq 1$  user*)
  - Small World Phenomenon (*every user is 6 hops away from everyone else*)
  - Indegree Distribution (*distribution of incoming edges at each node*)
  - Outdegree Distribution (*distribution of outgoing edges at each node*)

[4]. Kandias M., Galbogini L., Mitrou L., Gritzalis D., "Insiders trapped in the mirror reveal themselves in social media", in *Proc. of the 7<sup>th</sup> International Conference on Network and System Security*, Spain, June 2013.

# Detecting malevolent insiders via ... narcissism and ... Twitter!



## Statistical distribution & user analysis

### **Social Medium Usage Intensity**

Evaluated via the aggregation of the content that the user produced

### **Social Medium Influence Valuation**

Evaluated via the number of users who are exposed at the user's tweets (followers, retweets, mentions)

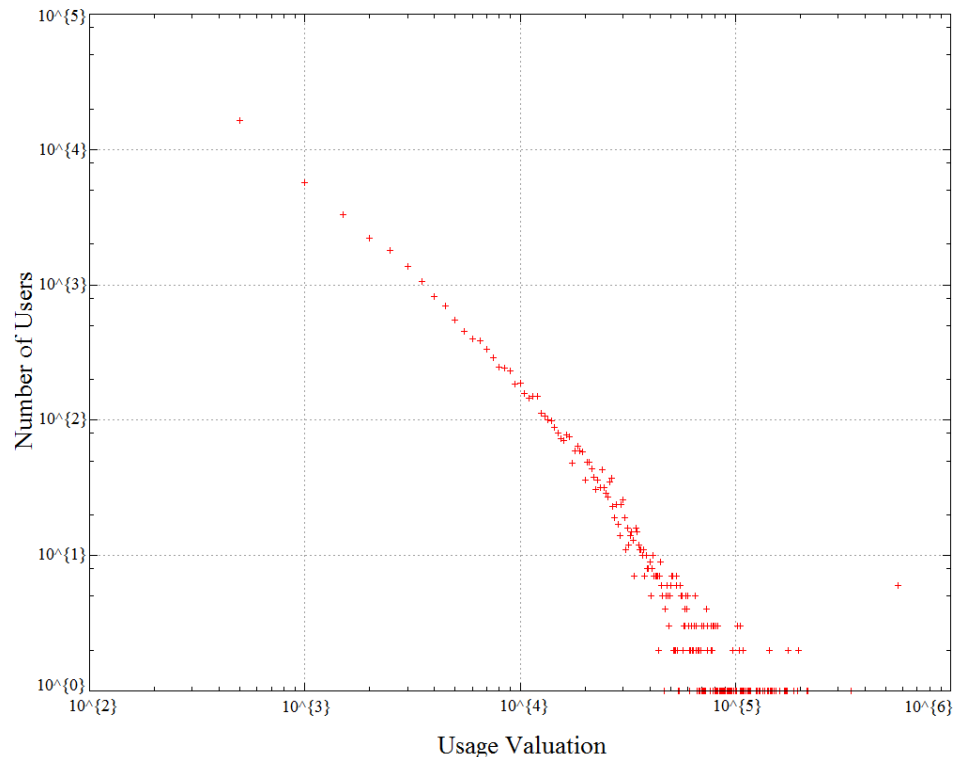
### **Klout score**

Assessment of the user's involvement within the Social Medium

# Detecting malevolent insiders via ... narcissism and ... Twitter!



- **Small World Phenomenon**
  - 99% of the users is  $\leq 6$  hops away from everyone else in the graph.
- **Indegree Distribution**
  - Distribution of incoming edges at each node. 13.2 followers/user on average.
- **Outdegree Distribution**
  - Distribution of outgoing edges at each node. 11 followings/ user on average.
- **Usage Intensity Distribution**
  - Distribution of the evaluation of usage intensity per user.



# Detecting malevolent insiders via ... narcissism and ... Twitter!



Category	Influence valuation	Klout score	Usage valuation
Loners	0 - 90	3.6 - 11.1	0 - 500
Individuals	90 - 283	11.1 - 26.0	500 - 4500
Known users	283 - 1011	26.0 - 50.0	4500 - 21000
News Media & Personas	1011 - 3604	50.0 - 81.99	21000 - 569000

# Conclusions



- ✓ Vast majority of users belongs to the first 3 categories.
- ✓ Individuals categorized to the 4<sup>th</sup> category are considered to manifest narcissistic behavior.
- ✓ Narcissistic behavior is closely connected to malevolent behavior (prediction indicator).
- ✓ A set of users may be studied under the prism of group homogeneity.
  - ✓ Homogeneity evaluation of an existing group.
  - ✓ How well could a newcomer fit into an existing group?
- ✓ Assess users under other prisms or groupings (profession, geolocation etc).



## Horror Story(?)

*Panopticon*: Revealing sensitive personal data via **YouTube**

*Panopticon*: An “ideal prison” designed by J. Bentham. A building designed in order to facilitate constant surveillance of the inmates from a central watch point.

# Revealing Political Beliefs via ... data mining and ... YouTube!



- **Social Medium:**

YouTube

- **Dataset** (November 2005 – December 2012):

12.964 users, 207.377 videos, 2.034.362 comments

Anonymization layer over the data, avoid correlation of usernames and political beliefs (MD5 hashes instead of usernames), according to Greek Law 2472/1997

- **Appropriate content within the medium:**

Political content, audio-visual stimuli, emotional arousal, broad user involvement

- **Methods of analysis:**

Graph-theoretic analysis (Small World Phenomenon, Indegree/Outdegree Distribution, Node Loneliness)

Content analysis (conclusion extraction via comment classification, opinion mining, machine learning)

Tag cloud analysis (concentration of videos' tags into one Tag Cloud in order to visualize and analyze word trends)



# Revealing Political Beliefs via ... data mining and ... YouTube!



- **Small World Phenomenon**
  - 99% of the users is  $\leq 6$  hops away from everyone else in the graph
- **Indegree Distribution**
  - Distribution of incoming edges at each node
- **Outdegree Distribution**
  - Distribution of outgoing edges at each node
- **Tag Cloud**
  - Concentration, visualization and analysis of video tags into one tag cloud



# Revealing Political Beliefs via ... data mining and ... YouTube!



## User Categories (indicative):

Centre & Center-left wing, Neutral, Centre and Centre-right wing

## Field expert contribution (pattern detection):

Sociologist, Political Scientist

### Algorithm: Multinomial Logistic Regression (MLR)

Categories Metrics	Centre & Centre- left	Neutral	Centre & Centre- right
Precision	83%	91%	77%
Recall	77%	93%	78%
F-Score	80%	92%	77%
Accuracy	87%		

**Precision:** Measures the classifier's exactness. Higher and lower precision means less and more false positive classifications.

**Recall:** Measures the classifier's completeness. Higher and lower recall means less and more false negative classifications

**F-Score:** The weighted harmonic mean of **Precision** και **Recall**.

**Accuracy:** Measures the number of correct classifications performed by the classifier.

# Conclusions



- ✓ Ability to automatically classify users into predefined categories of political beliefs
- ✓ Ability to reveal sensitive, personal data in an automated manner
- ✓ Ability to confirm pollster and demographic statistical data
- ✓ Raise and highlight severe ethical and legal issues

# Overall Emerging Threats



- ✓ Widening of social inequalities
- ✓ Deepening of social prejudices
- ✓ Creation of negative work/life environment
- ✓ Violation of human rights
- ✓ Exploitation of minorities
- ✓ Social and labor discriminations
- ✓ Widening of digital divide

# Legal Exploitation of Findings



- ✓ Assessment of predisposition towards delinquent behavior<sup>5</sup> (namely insider threats within critical infrastructures)
- ✓ User/usage profiling for personalized advertising (namely consumer profiling)
- ✓ Ability to predict some self-harming behaviors (namely teenage suicides)

[5]. U.S. Dept. of Justice, “The insider threat: An introduction to detecting and deterring insider spy”, FBI, USA, 2012, <http://www.fbi.gov/about-us/investigate/counterintelligence/the-insider-threat>



## References

Kandias, M., Stavrou, V., Bosovic, N., Gritzalis, D., "Proactive insider threat detection through social media: The YouTube case ", *Proc. of the 12<sup>th</sup> Workshop on Privacy in the Electronic Society*, ACM, Berlin, November 2013.

Kandias, M., Mitrou, L., Stavrou, V., Gritzalis, D., "Which side are you on? A new Panopticon vs. privacy", *Proc. of the 10<sup>th</sup> International Conference on Security and Cryptography*, pp. 98-110, SciTek Press, Iceland, July 2013.

Kandias, M., Galbogini, K., Mitrou, L., Gritzalis, D., "Insiders trapped in the mirror reveal themselves in social media", *Proc. of the 7<sup>th</sup> International Conference on Network and System Security*, pp. 220-235, Springer (LNCS 7873), Spain, June 2013.

Kandias, M., Virvilis, N., Gritzalis, D., "The Insider Threat in Cloud Computing", *Proc. of the 6<sup>th</sup> International Workshop on Critical Infrastructure Security*, pp. 93-103, Springer (LNCS 6983), Switzerland, September 2011.

Kandias, M., Mylonas, A., Virvilis, N., Theoharidou, M., Gritzalis, D., "An Insider Threat Prediction Model", *Proc. of the 7<sup>th</sup> International Conference on Trust, Privacy, and Security in Digital Business*, pp. 26-37, Springer (LNCS 6264), Spain, August 2010.