

The **NEREUS** Platform: An Open
Source Intelligence software tool for
exploiting national defense
capabilities

D. Gritzalis, M. Kandias, V. Stavrou

January 2015

Invited Lecture,
Hellenic National Defense College
Athens, January 2015

The **NEREUS** Platform: An Open Source Intelligence software tool for exploiting national defense capabilities



Dimitris Gritzalis, Miltos Kandias, Vasilis Stavrou

Information Security & Critical Infrastructure Protection (INFOSEC) Lab
Dept. of Informatics, Athens University of Economics & Business

Presentation outline

Open Source (& Social Media) Intelligence

A selection of capabilities

The **NEREUS** Framework

Behavior prediction capabilities using OSINT/SOCMINT

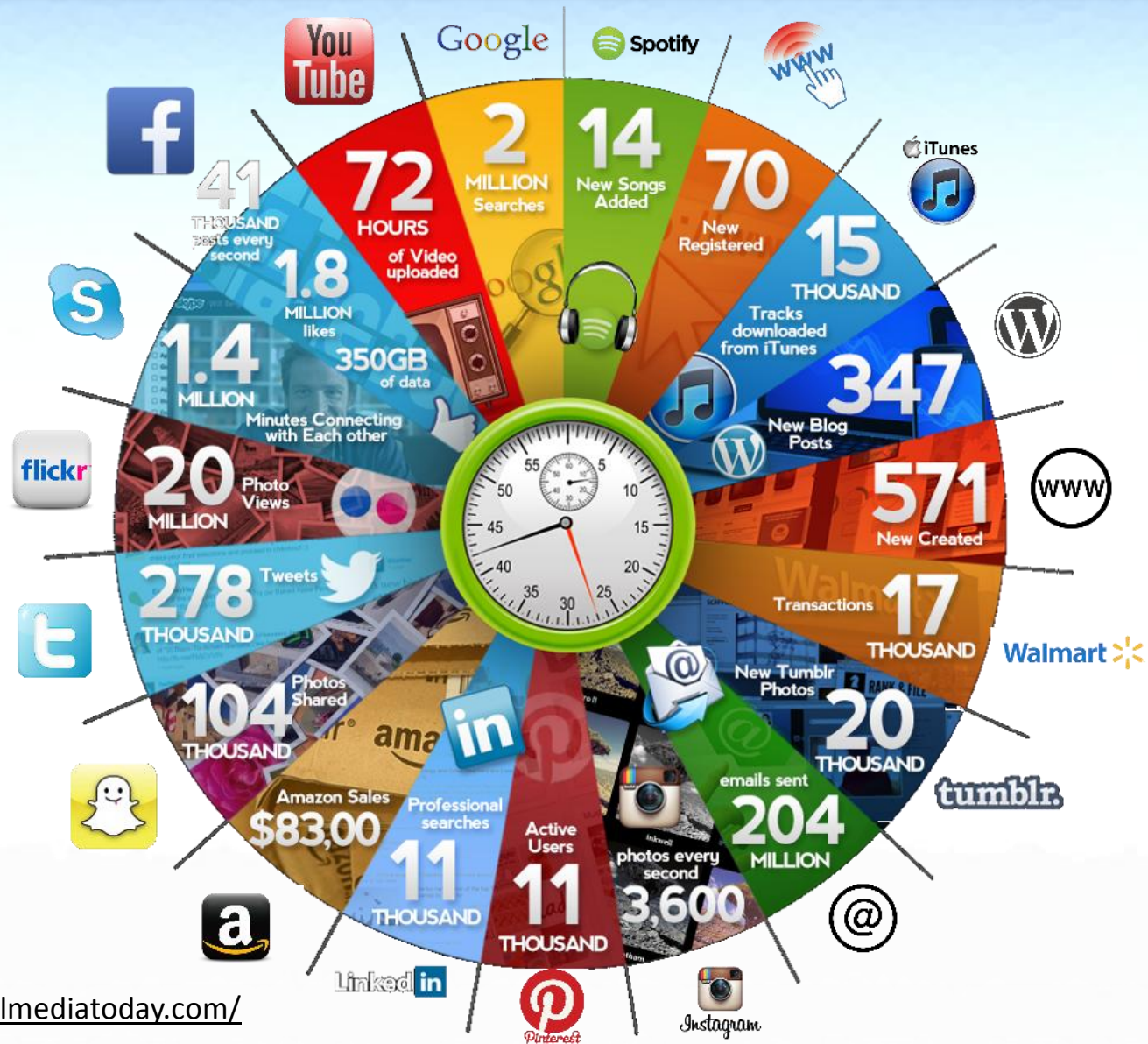
Case 1: Insider detection and narcissism (**Twitter**)

Case 2: Predisposition towards law enforcement (**YouTube**)

Case 3: Detecting stress levels (**Facebook**)

Conclusions

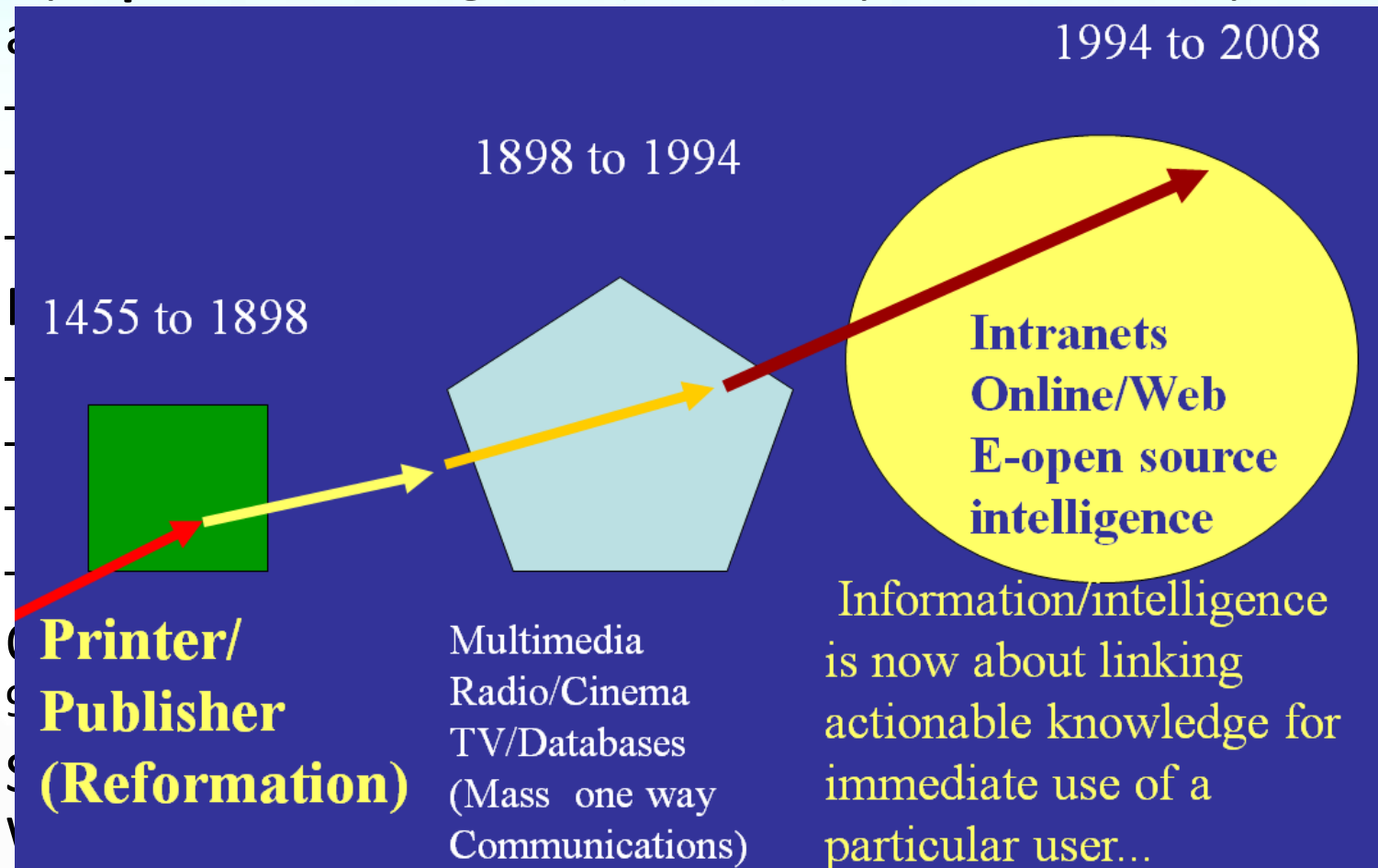
Web 2.0 and Online Social Networks



Open Source (& Social Media) Intelligence (OSINT/SOCMINT)



- Open Source Intelligence (OSINT) Transition 1455 to 2008



- Information/intelligence is now about linking actionable knowledge for immediate use of a particular user...

Who can exploit OSINT/SOCMINT?



Business



Law Enforcement



Government



Criminals



Special Agencies

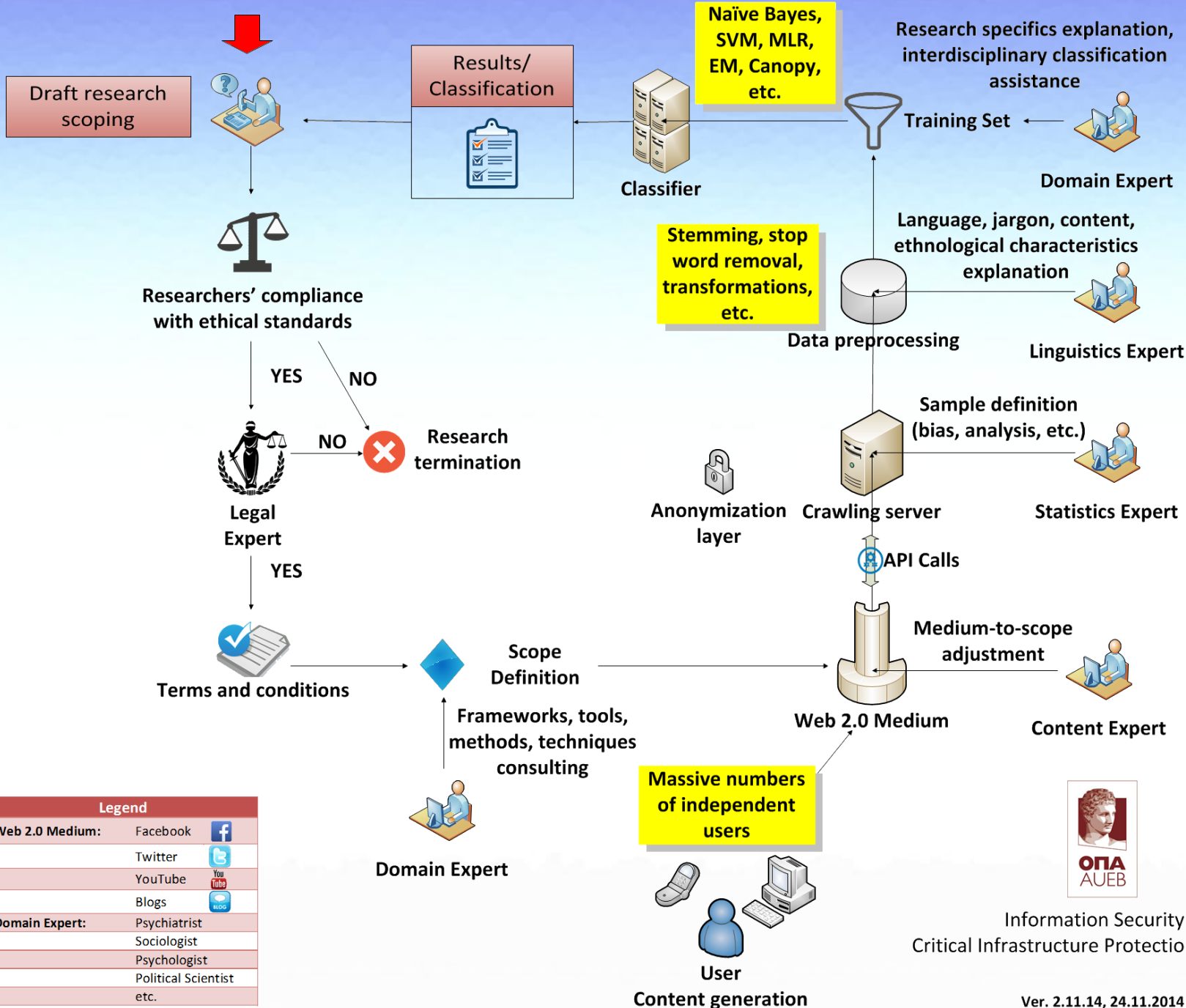


Military

A selection of OSINT/SOCMINT capabilities

- Identify and predict **insiders**
- Detect **sentiment of population** w.r.t. specific intelligence demands
- Identify and predict **public opinion** on specific intelligence demands
- Identify and predict **public opinion fluctuations**
- Detect **influential users** capable of supporting a cause
- Detect **appropriate means and content** of communication for achieving optimum results
- Facilitate achievement of **influence goals and success tactics**
- Optimize efficiency of communication and **influence tactics**

OSINT from Web 2.0 Media
The NEREUS[©] Framework




Legend		
Web 2.0 Medium:	Facebook	
	Twitter	
	YouTube	
	Blogs	
Domain Expert:	Psychiatrist	
	Sociologist	
	Psychologist	
	Political Scientist	
	etc.	



Information Security & Critical Infrastructure Protection Laboratory

Case 1

Insider threat prediction based on Narcissism

OSINT		OSN: Twitter 
Tools used for the analysis		
Science	Theory	
Computing	Graph Theory	
Sociology	Theory of Planned Behavior	
Psychology	Social Learning Theory	
Application: Insider threat detection/prediction, influential users detection, means of communication evaluation		

Personal factors indicating delinquent behavior

Prediction of delinquent behavior via psychosocial factors

Shaw's factors

- **Introversion**
- **Social and personal frustrations**
- Computer dependency
- Ethical "flexibility"
- **Reduced loyalty**
- **Entitlement-Narcissism**
- Lack of empathy
- **Predisposition towards law enforcement**

FBI's factors

- Greed/financial need
- **Anger/Revenge**
- **Problems at work**
- **Ideology/Identification**
- **Divided loyalty**
- Adventure/Thrill
- Vulnerability to blackmail
- **Ego/self-image (Narcissism)**
- Ingratiation
- Compulsive and destructive behavior
- Family problems

In a nutshell



Predicting & identifying potential insiders



Researchers' compliance with ethical standards

YES



Legal Expert

YES

Critical infrastructures
National security
Public interest



Twitter Users

Content generation



Twitter

Crawling & storing



Our crawling server



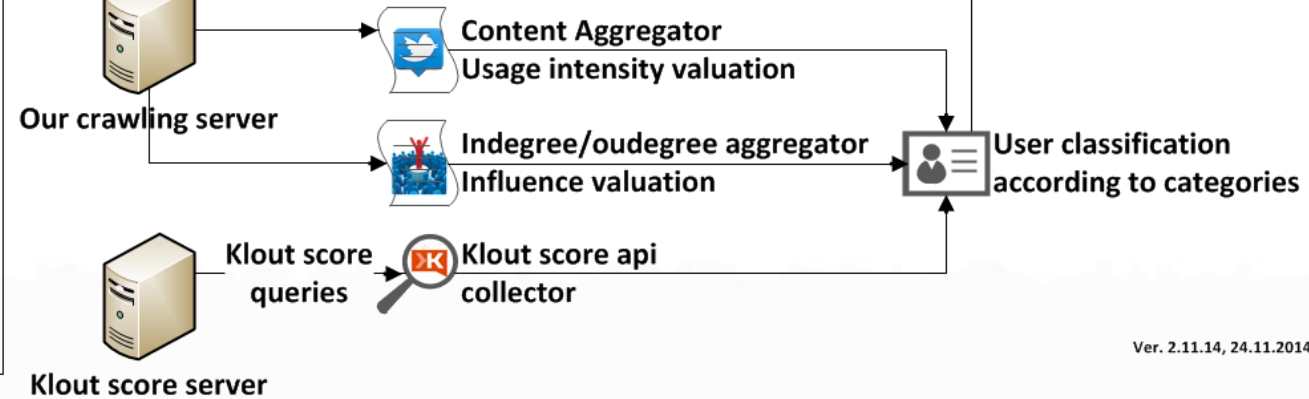
Klout score server

Legend		
Web 2.0 Medium:	Twitter	
Domain Expert:	Psychologist	



Information Security & Critical Infrastructure Protection Laboratory

Category	Influence valuation	Klout score	Usage valuation
Loners	0 - 90	3.55 - 11.07	0 - 500
Individuals	90 - 283	11.07 - 26.0	500 - 4.500
Known users	283 - 1.011	26.0 - 50.0	4.500 - 21.000
Mass Media & Personas	1.011 - 3.604	50.0 - 81.99	21.000 - 56.9000

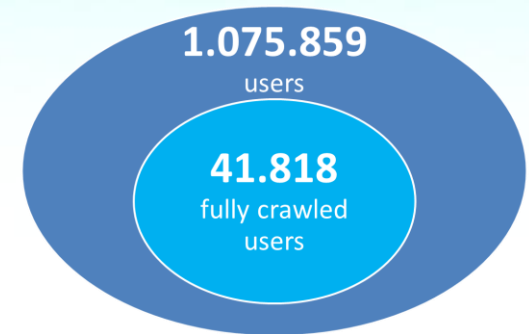


Dataset description



- 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
- Graph:
 - Each user is a node
 - Every interaction is a directed edge
- **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)



7.125.561 connections
among them

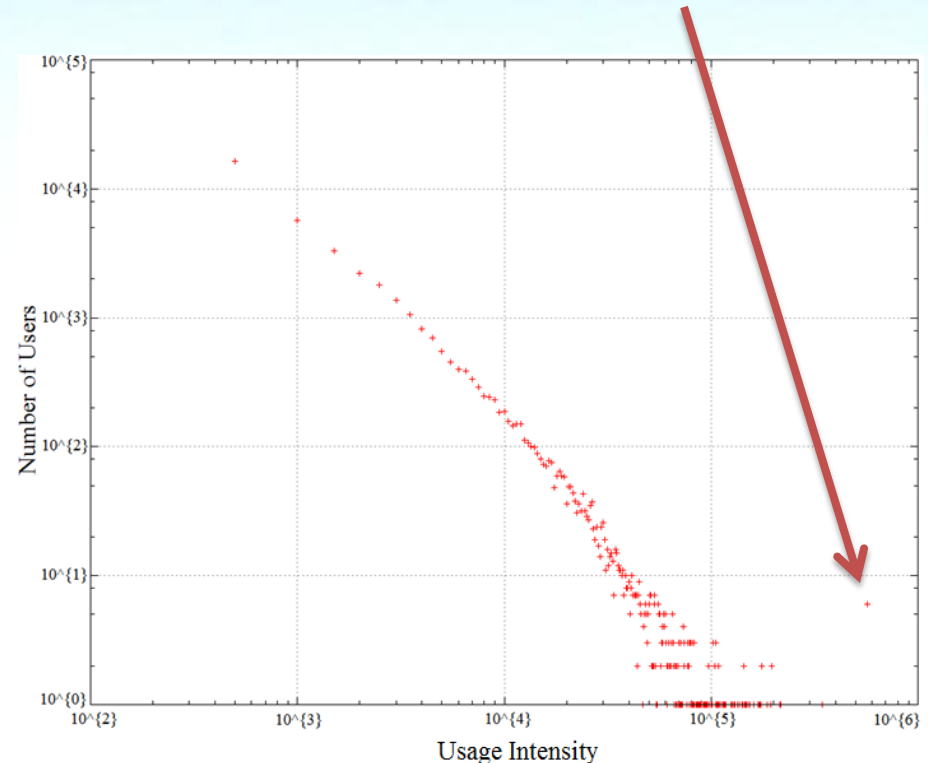


Graph Theoretical approach

- **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
- **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}

Important cluster of users





Narcissism detection

- Majority of users make limited use of Twitter
 - A lot of “normally” active users and very few “popular” users
 - Users classified into four categories, on the basis of specific metrics (influence valuation, Klout score, usage valuation)
- Above a threshold:
 - User becomes **quite influential/perform intense** medium use
 - User get a “**mass-media & persona**” status

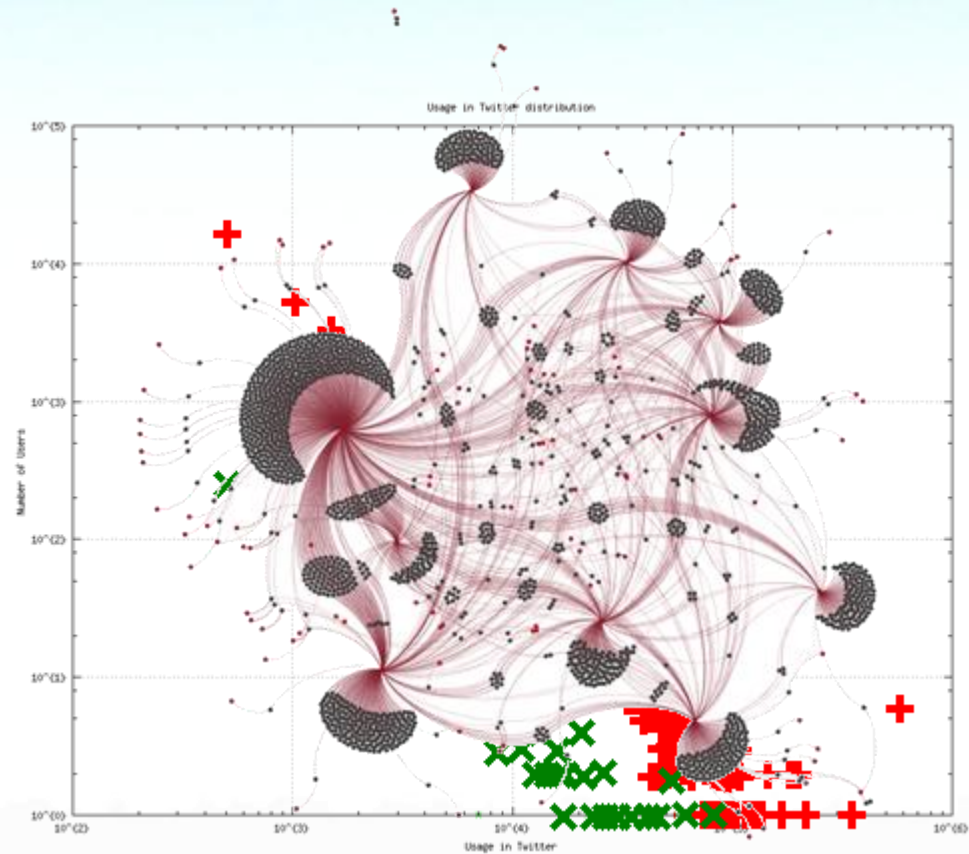
The excessive use of Twitter by persons who are not mass-media or personas could connect to narcissism and identify narcissists, i.e. persons who - inter alia - tend to turn insiders

Category	Influence valuation	Klout score	Usage valuation
Loners	0 - 90	3.55 - 11.07	0 - 500
Individuals	90 - 283	11.07 - 26.0	500 - 4.500
Known users	283 - 1.011	26.0 - 50.0	4.500 - 21.000
Mass Media & Personas	1.011 - 3.604	50.0- 81.99	21.000 - 56.9000

Group dynamics




- Create reliable graphs of interconnection, i.e. visualization of groups of people according to their **relationships** and **common interests**
- Compare deviating usage behavior according to a set of parameters, **maximize efficiency**



Case 2:

Revealing negative attitude towards law enforcement

OSINT		OSN: YouTube	
Tools used for the analysis			
Science	Theory		
Computing	Machine Learning		
	Data Mining		
Sociology	Social Learning Theory		
Application: Detection/prediction of threats, capabilities for influence and divided loyalty.			

In a nutshell

Detecting negative predisposition towards law enforcement



Researchers' compliance with ethical standards

YES



Legal Expert

YES

Critical infrastructures
National security
Public interest



YouTube User



YouTube

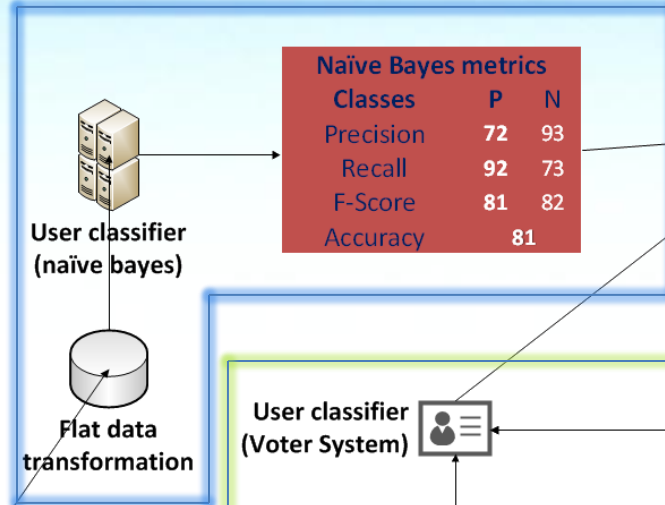


Anonymization layer



YouTube Crawler

Flat data path

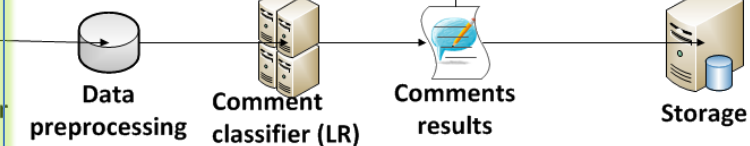


Naïve Bayes metrics		
Classes	P	N
Precision	72	93
Recall	92	73
F-Score	81	82
Accuracy	81	

Categories	
•	Negatively Predisposed (P)
•	Not negatively predisposed (N)

User classifier (Voter System)

Video, uploads, lists & favorites classifier



Classifier	Metrics					
	NB		SVM		LR	
Classes	P	N	P	N	P	N
Precision	71	70	83	77	86	76
Recall	72	68	75	82	74	88
F-Score	71	69	79	79.5	80	81
Accuracy	70		80		81	

Comments classification path

Legend		
Web 2.0 Medium:	YouTube	
Domain Expert:	Sociologist	
	Political Scientist	



Information Security & Critical Infrastructure Protection Laboratory

Dataset description



- Crawled YouTube and created dataset consists solely of **Gre-ek** users.
- Utilized YouTube **REST-based API** (developers.google.com/youtube/):
 - Only publicly available data collected
 - Qu-o-te li-mi-tations (posed by YouTube) were respected
- Collected data were classified into three cate--gories:
 - User-related information (pro-fi-le, uploaded videos, subscriptions, favorite vi-de-os, playlists)
 - Video-related in-for-ma-tion (license, # of likes, # of dislikes, category, tags)
 - Com-ment--re-la-ted information (com---ment content, # of likes, # of dislikes)

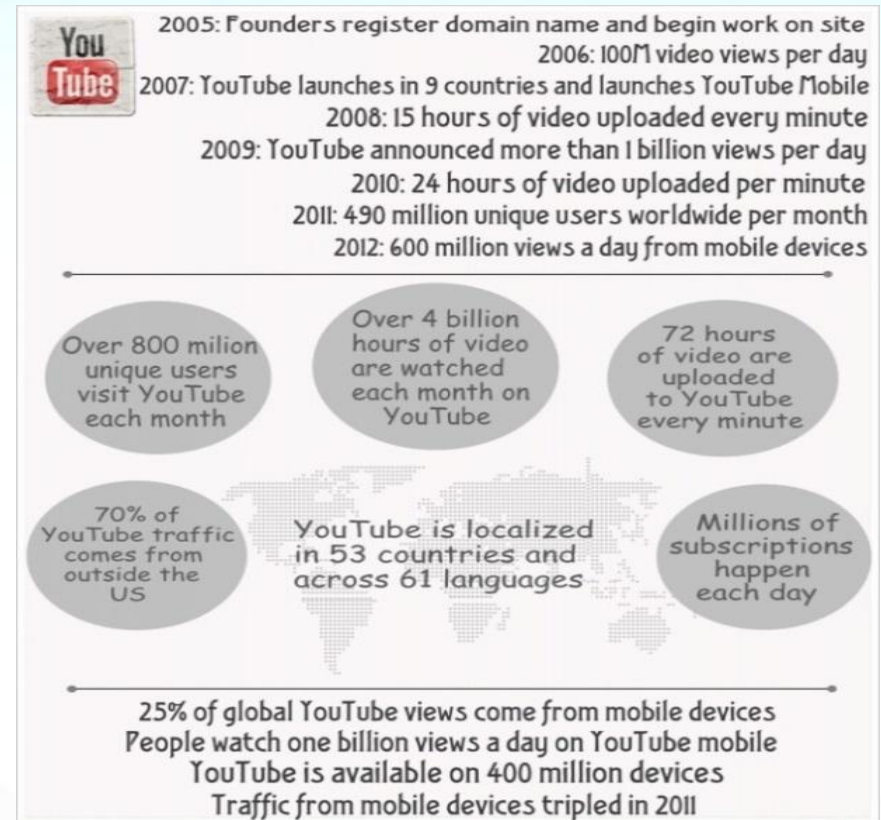


- Ti-me span of collected data covered 7 years (Nov 2005 - Oct 2012).
- A basic anonymization layer added to the col-lec-t-ed data:
 - MD5 hashes instead of usernames

Graph Theory and Content Analysis



- **Small World Phenomenon:**
 - Every user of the community is 6 hops away from everyone else
- **Indegree Distribution:**
 - Presentation of statistical distribution of incoming edges per node
- **Outdegree Distribution:**
 - Presentation of statistical distribution of outgoing edges per node
- **Tag Cloud:**
 - Axis of content of the collected data via tag cloud analysis
- **YouTube's nature:**
 - Popular social medium, emotional-driven responses, audio-visual stimuli, allegedly anonymous, users interact with each other, contains political content



Machine Learning (1/2)

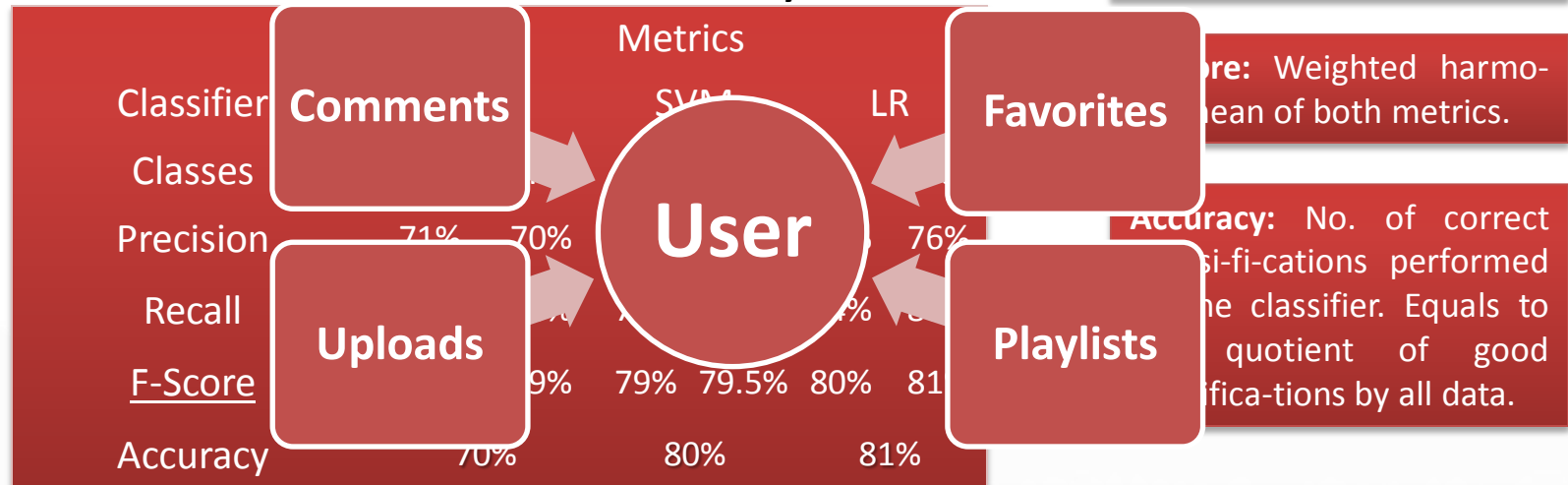
- Comment classified into categories of interest:
 - Process performed as **text clas-si-fi-ca-tion**
 - Machine trai-n-ed with **text examples** and the **cate-go-ry** each one belongs to
 - Excessive support by **field expert** (Sociologist)
- Tes-t set used to evaluate efficien-cy of resulting classifier:
 - Contains pre-labeled data fed to machine, labeled by field expert
 - Check if initial assigned label is equal to predicted one
 - Testing set labels assigned by field expert
- Most comments are written in Greek/Greeklish
- Training sets (Greeklish and Greek) were merged – One clas-si-fi-er was trained
- Two categories of content were defined:
 - Users with a **negative** attitude (**P**re-disposed negatively (P))
 - Users with a **not negative** attitude (**N**ot-pre-disposed negatively (N))

Machine Learning (2/2)

- **Video classification:**
 - **Comment classification using:**
 - Examination of a video on the basis of its comments
 - Naïve Bayes (NB)
 - Support Vector Machines (SVM)
 - **Video process to determine category classification**
- **(Video) Lists classification:**
- **Classifier efficiency comparison:**
- **Conclusion about user behavior:**
 - Metrics (on % basis) Precision, Recall, F-Score, Accuracy
 - If there is at least one category P attribute then the user is classified into category P
- **Logistic Regression algorithm:**
 - LR classifies a comment with **81% accuracy**

Precision: Me-a-su-res the classifier exactness. Higher and lower pre-cision means less and more false positive clas-si-fi-ca-tions, respectively.

Recall: Measures the clas-sifier completeness. Higher and lower recall means less and more false negative classifications, respectively.



Flat Data

- Addressing the problem from a different perspective:
 - Connection between users of category P and confidence of accuracy of comments belonging to category P.
 - assumption-free and easy-to-scale method
 - verify (or not) the results of the Machine Learning approach.

Blue: Users of category P classified on the basis of the comment-oriented tuple (**Flat Data**).
Red: Users of category P classified on the basis of their comments-only (**Machine Learning**).

- Data transformation:

- User repr...
 - comment



ID the views).
eld expert).

- Machine

1721 users are (almost certainly) negative predisposed toward law enforcement

Approach	Metrics			
	Machine Learning		Flat Data	
Classifier	Logistic Regression		Naïve Bayes	
Classes	P	N	P	N
Precision	86%	76%	72%	93%
Recall	74%	88%	92%	73%
<u>F-Score</u>	80%	81%	81%	82%
Accuracy	81%		81%	

data
ica-
hine
ents

Case 3:

Detecting stress level use patterns

OSINT

OSN: Facebook 

Tools used for the analysis

Science

Tool

Computing

Machine Learning

Data Mining

Psychiatry
Psychology

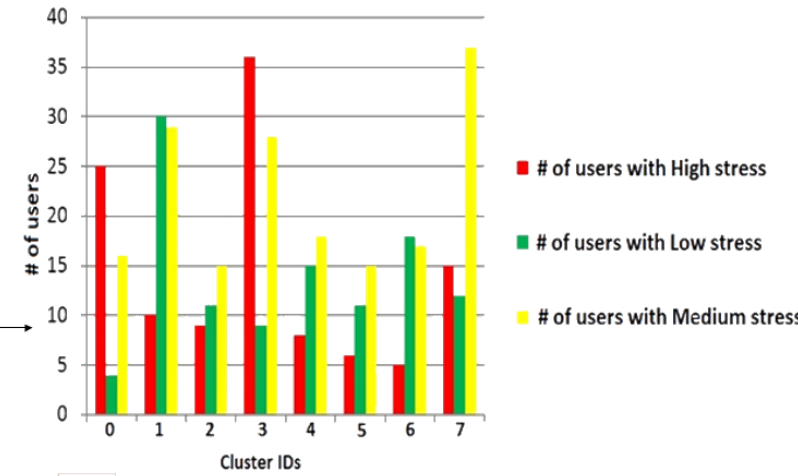
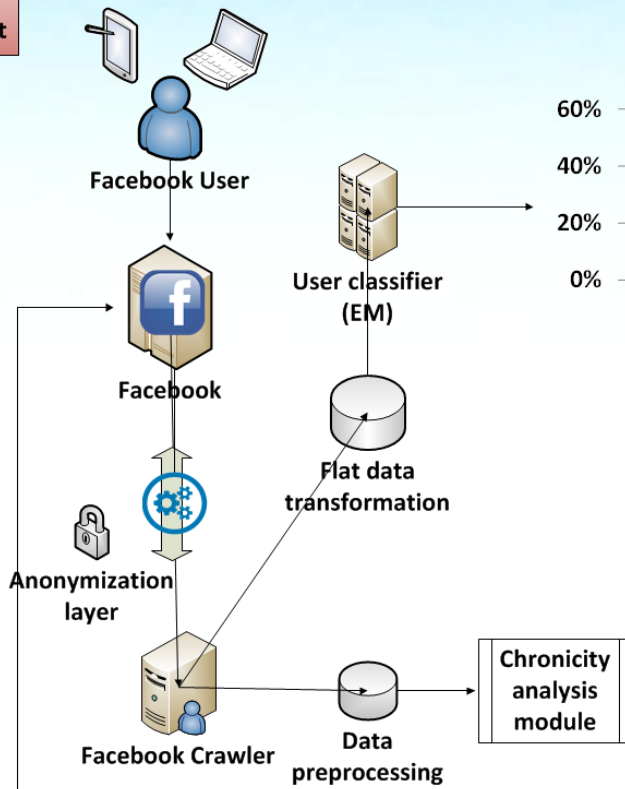
BAI Stress Test


Application: Detection/prediction of vulnerable individuals and potential threats, momentum of engagement, etc.

In a nutshell



Detect individuals vulnerable to blackmail and moral inhibitions shift



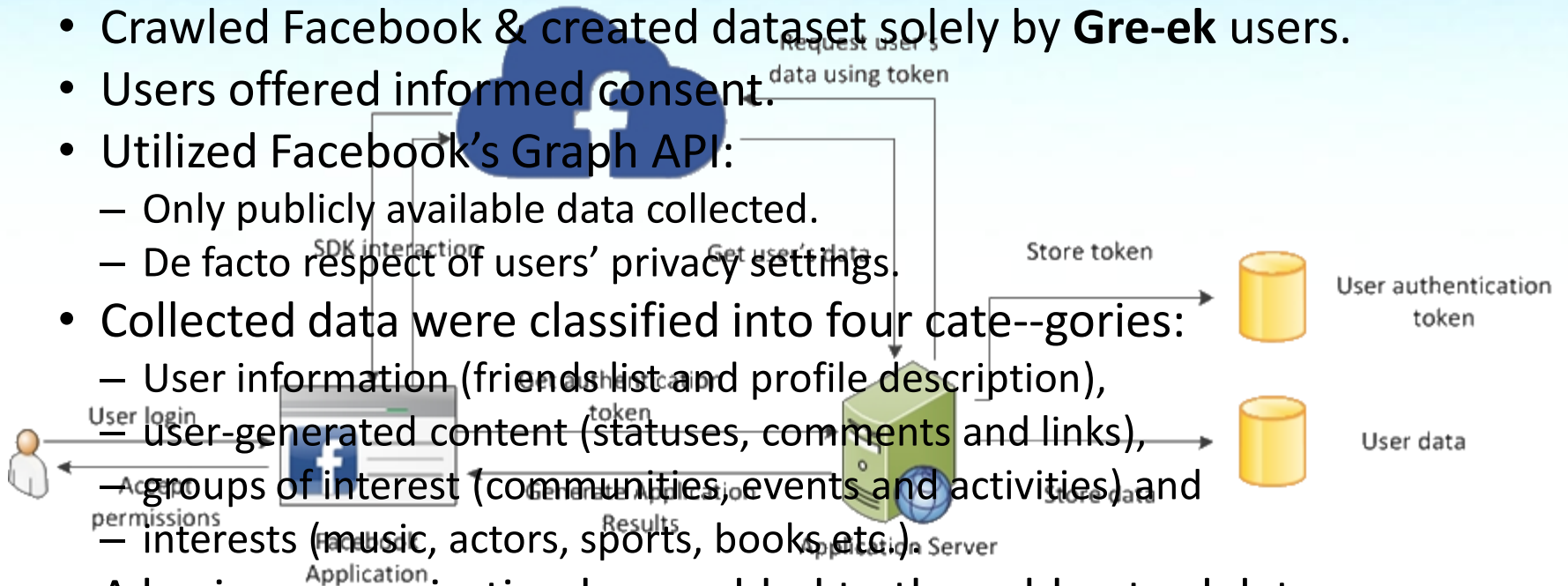
Legend	
Web 2.0 Medium:	Facebook 
Domain Expert:	Psychiatrist Psychologist


 Information Security &
 Critical Infrastructure Protection Laboratory



Dataset description

- Crawled Facebook & created dataset solely by **Gre-ek** users.
- Users offered informed consent.
- Utilized Facebook's Graph API:
 - Only publicly available data collected.
 - De facto respect of users' privacy settings.
- Collected data were classified into four categories:
 - User information (friends list and profile description),
 - user-generated content (statuses, comments and links),
 - groups of interest (communities, events and activities) and
 - interests (music, actors, sports, books etc.)
- A basic anonymization layer added to the collected data:
 - MD5 hashes instead of usernames.



- Opt-out ability integrated, delete all user data upon selection.
- Dataset statistical analysis proved its efficiency and absence of bias.

405 users

110 user groups

2000 objects

1000 statuses

1000 comments

Chronicity analysis (indicators over time)



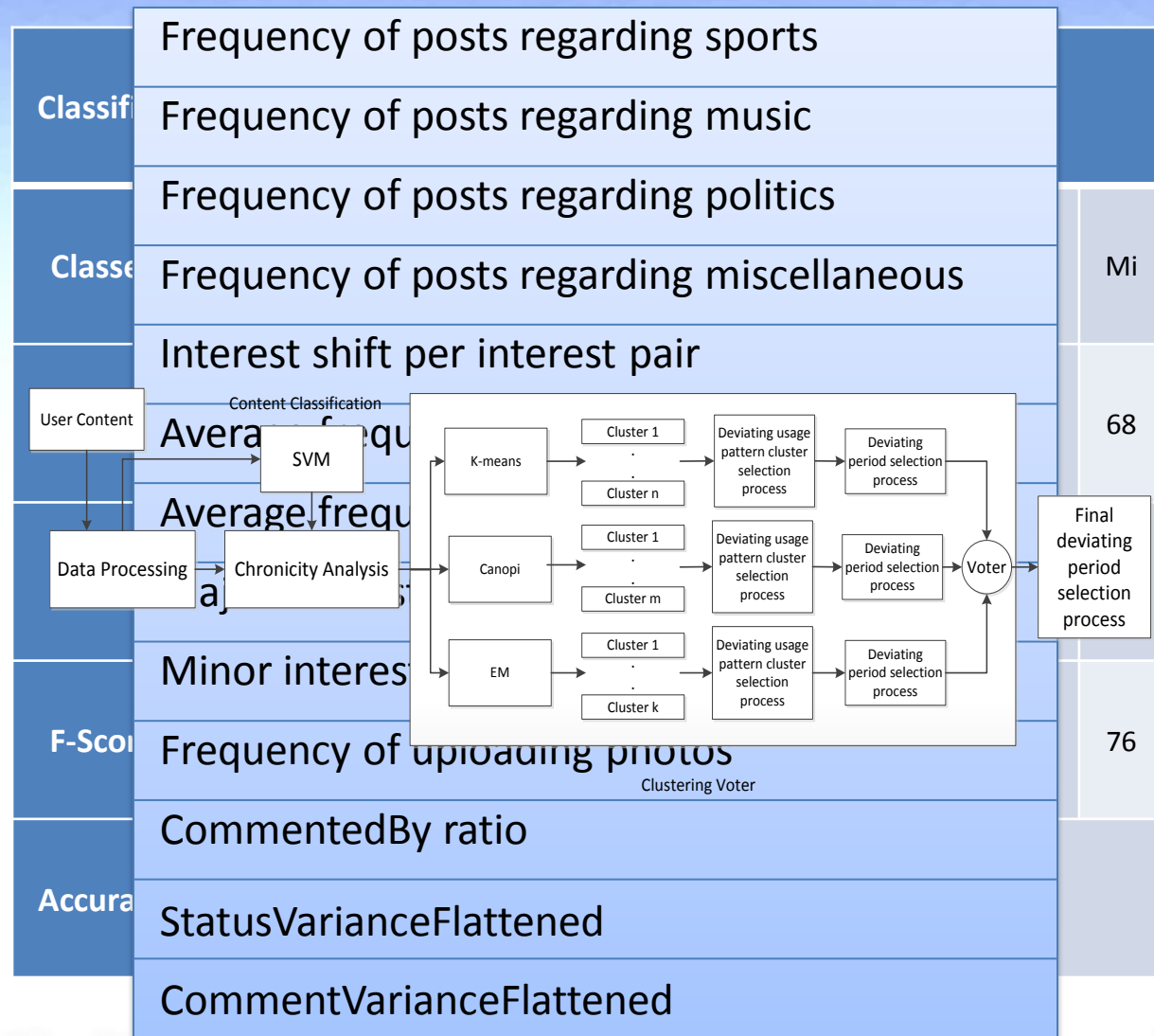
- Goal: **Detect differentiations** of OSN use patterns over **time related** so as to depicted stress level
- Split users' use patterns into time periods (from 1 day to 1 month)
 - Time period of one week produced best results
- Chronicity analysis system consists of two modules:
 - Preprocessing data module (responsible for the processing of input data)
 - Usage pattern analysis module (responsible for analyzing usage patterns based on a set of metrics)
- Use pattern fluctuations depict differentiated medium use

Chronicity analysis steps

Step 1: Classify user generated content into 4 predefined categories ('S' stands for sports, 'M' for music, 'P' for politics and 'Mi' for miscellaneous).

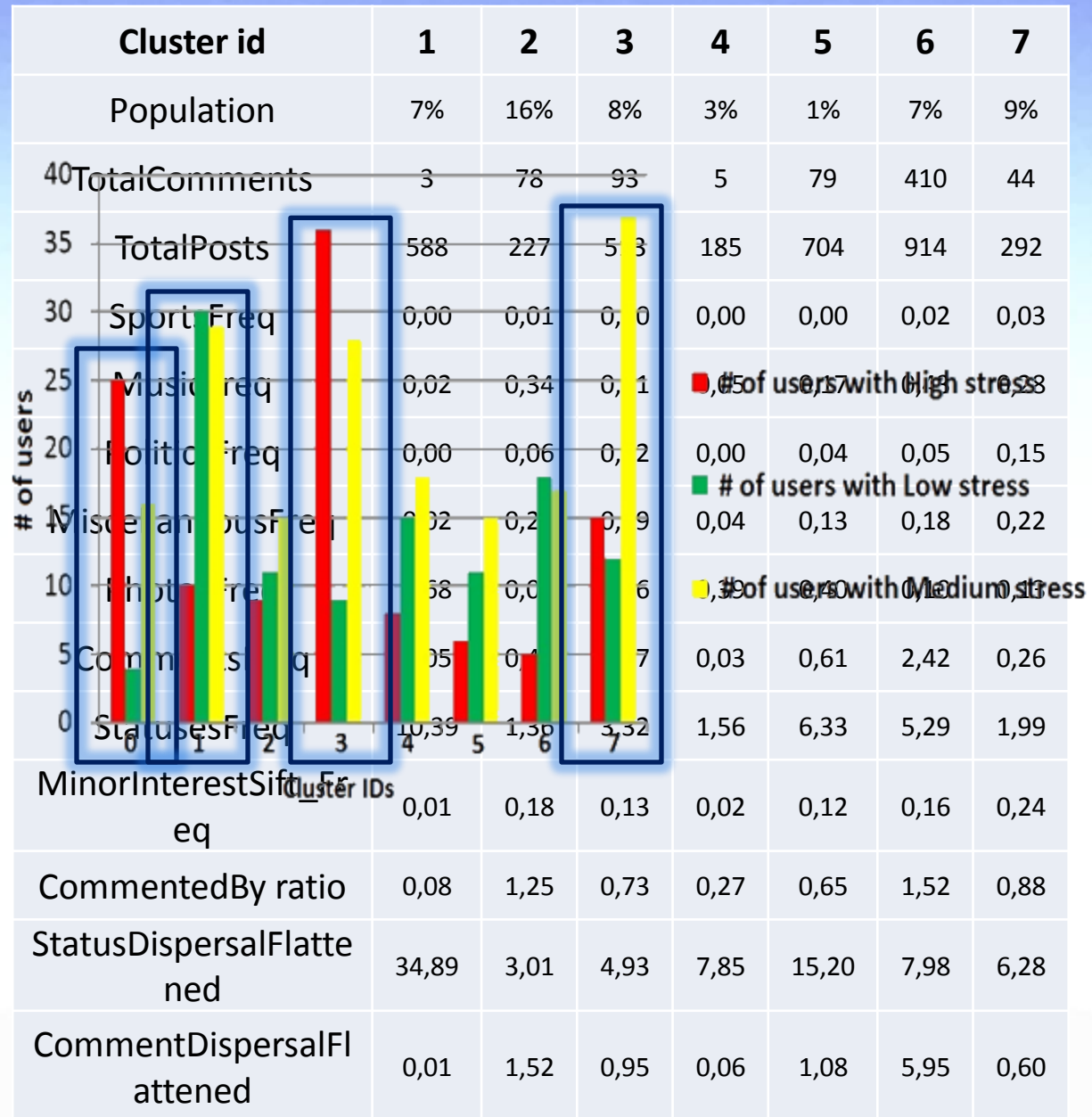
Step 2: Calculate following metrics for each user and time period (metrics developed on an ad-hoc basis according to our observations).

Step 3: Transform metrics results into arithmetic vectors and perform data mining on them using (a) *K-means*, (b) *EM*, and (c) *Canopy* algorithms. Utilize voter to decide fluctuations.



Chronicity analysis results

- Metrics results per detected cluster.
- Visual representation of users belonging to each cluster.
- **Clusters 0 and 3** contain mainly users classified in high stress category.
- In **cluster 0**, users post mainly photos.
- In **cluster 3** users post photos, discuss about music, whereas a small fraction of the content is re-fer-ring to miscellaneous information.
- **Clusters 1 and 7** contain many users classified in medium or low stress category.
- **Clusters 1 and 7** refer mainly to music and mis-cel--la-ne-ous content and also contain limited content referring to sports.



The **NEREUS** Framework:

Selected exploitation capabilities

- **(Insider) Threat prediction:**
 - Applying Shaw and FBI psychosocial indicators (narcissism, anger syndrome, revenge syndrome, etc.).
- **Influence opportunities exploitation:**
 - Analyzing communication graphs, correlating psycho-social characteristics, assessing engagement tactics, etc.
- **Delinquent behavior prediction:**
 - Analysis of psycho-social characteristics (narcissism, anger syndrome, revenge syndrome, etc.).
 - Predisposition analysis (Graph Theory and Content Analysis through Social Learning Theory, etc.).
- **Forensics analysis support:**
 - Suspect profiling and analysis (prediction of delinquent behavior, etc.).

Some conclusions

- ✓ Web 2.0 produces vast amounts of **crawlable** information and OSINT/SOCMINT can transform it into **intelligence**.
- ✓ OSINT/SOCMINT can assist in detecting **narcissistic behavior**, **predisposition towards law enforcement**, etc.
- ✓ OSINT/SOCMINT can help in **predicting insiders**, in **predicting delinquent behavior**, in **supporting law enforcement** and in **enhancing national defense**.
- ✓ OSINT/SOCMINT intrusive nature dictates **specific** uses for **legitimate** only purposes.

References

1. Gritzalis D., Stavrou V., Kandias M., Stergiopoulos G., "Insider Threat: Enhancing BPM through Social Media", in *Proc. of the 6th IFIP International Conference on New Technologies, Mobility and Security (NMITS-2014)*, Springer, UAE, 2014.
2. Gritzalis D., "Insider threat prevention through Open Source Intelligence based on Online Social Networks", Keynote address, *13th European Conference on Cyber Warfare and Security (ECCWS-2014)*, Greece, 2014.
3. Gritzalis D., Kandias M., Stavrou V., Mitrou L., "History of Information: The case of Privacy and Security in Social Media", in *Proc. of the History of Information Conference*, Law Library Publications, Athens, 2014.
4. Kandias M., Mitrou L., Stavrou V., Gritzalis D., "Which side are you on? A new Panopticon vs. privacy", in *Proc. of the 10th International Conference on Security and Cryptography (SECRYPT-2013)*, pp. 98-110, Iceland, 2013.
5. Kandias M., Galbogini K., Mitrou L., Gritzalis D., "Insiders trapped in the mirror reveal themselves in social media", in *Proc. of the 7th International Conference on Network and System Security (NSS-2013)*, pp. 220-235, Springer (LNCS 7873), Spain, June 2013.
6. Kandias M., Virvilis N., Gritzalis D., "The Insider Threat in Cloud Computing", in *Proc. of the 6th International Conference on Critical Infrastructure Security (CRITIS-2011)*, pp. 93-103, Springer (LNCS 6983), United Kingdom, 2013.
7. Kandias M., Stavrou V., Bozovic N., Mitrou L., Gritzalis D., "Can we trust this user? Predicting insider's attitude via YouTube usage profiling", in *Proc. of 10th IEEE International Conference on Autonomic and Trusted Computing (ATC-2013)*, pp. 347-354, IEEE Press, Italy, 2013.
8. Kandias M., Stavrou V., Bosovic N., Mitrou L., Gritzalis D., "Proactive insider threat detection through social media: The YouTube case", in *Proc. of the 12th Workshop on Privacy in the Electronic Society (WPES-2013)*, pp. 261-266, ACM Press, Germany, 2013.
9. Kandias M., Virvilis N., Gritzalis D., "The Insider Threat in Cloud Computing", in *Proc. of the 6th International Workshop on Critical Infrastructure Security (CRITIS-2011)*, Bologna S., et al (Eds.), pp. 93-103, Springer (LNCS 6983), Switzerland, 2011.
10. Kandias M., Mylonas A., Virvilis N., Theoharidou M., Gritzalis D., "An Insider Threat Prediction Model", in *Proc. of the 7th International Conference on Trust, Pri-vacy, and Security in Digital Business (TrustBus-2010)*, pp. 26-37, Springer (LNCS-6264), Spain, 2010.
11. Mitrou L., Kandias M., Stavrou V., Gritzalis D., "Social media profiling: A Panopticon or Omniopticon tool?", in *Proc. of the 6th Conference of the Surveillance Studies Network*, Spain, 2014.
12. Stavrou V., Kandias M., Karoulas G., Gritzalis D., "Business Process Modeling for Insider threat monitoring and handling", in *Proc. of the 11th International Conference on Trust, Privacy & Security in Digital Business (TRUSTBUS-2014)*, pp. 119-131, Springer (LNCS 8647), Germany, September 2014.
13. Virvilis N., Gritzalis D., "Trusted Computing vs. Advanced Persistent Threats: Can a defender win this game?", in *Proc. of 10th IEEE International Conference on Autonomic and Trusted Computing (ATC-2013)*, pp. 396-403, IEEE Press, Italy, 2013.