

Social Media Intelligence: Protecting the Citizen - Enforcing the Law

M. Kandias, V. Stavrou, L. Mitrou

January 2015

Social Media Intelligence (SOCMINT):

Protecting the Citizen – Enforcing the Law

Miltos Kandias, Vasilis Stavrou, Lilian Mitrou

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



Outline

- Web 2.0 and Online Social Networks (OSN)
- Open Source (& Social Media) Intelligence (OSINT/SOCMINT)
- A selection of capabilities
- The **NEREUS** Framework
- OSINT and behavior prediction capabilities
 - Case 1:** Predisposition towards law enforcement
 - Case 2:** Detecting stress levels
 - Case 3:** Threat detection and narcissism
- Conclusions

Web 2.0 and Online Social Networks

- OSN and Web 2.0 enable users to add online content.
- Content can be
 - personalized
 - personalized
 - user/user
- Users are the sharers of the info
- Can content
 - User behavior
 - User profiles
 - Proactive minor



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



- Open Source Intelligence is produced from publicly available information, which is:
 - collected, exploited and disseminated in a **timely** manner,
 - offered to an **appropriate** audience, and
 - 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 and reporting (e.g. amateur spotters, radio monitors)
- OSINT defined by US DoD (Public Law 109-163, Sec. 931, "National Defense Authorization Act for Fiscal Year 2006").
- SOCMINT is produced from Online Social Networks and the Web 2.0

A selection of capabilities

- ✓ Identify and predict delinquent behavior
- ✓ Identify and predict deviant and delinquent behavior of minors
- ✓ Detect and prevent abuse of minors

Personal factors indicating delinquent behavior

- Predict delinquent behavior via psychosocial factors

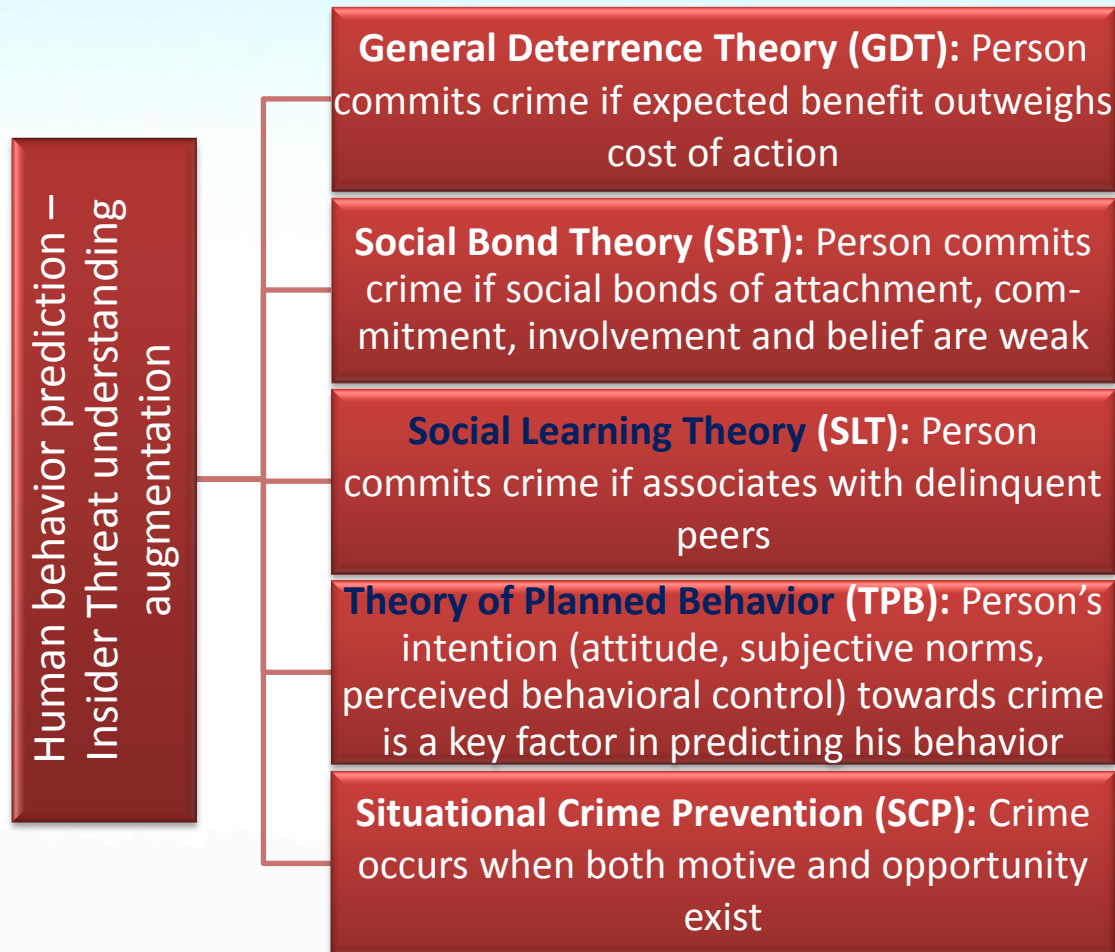
Shaw's factors

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

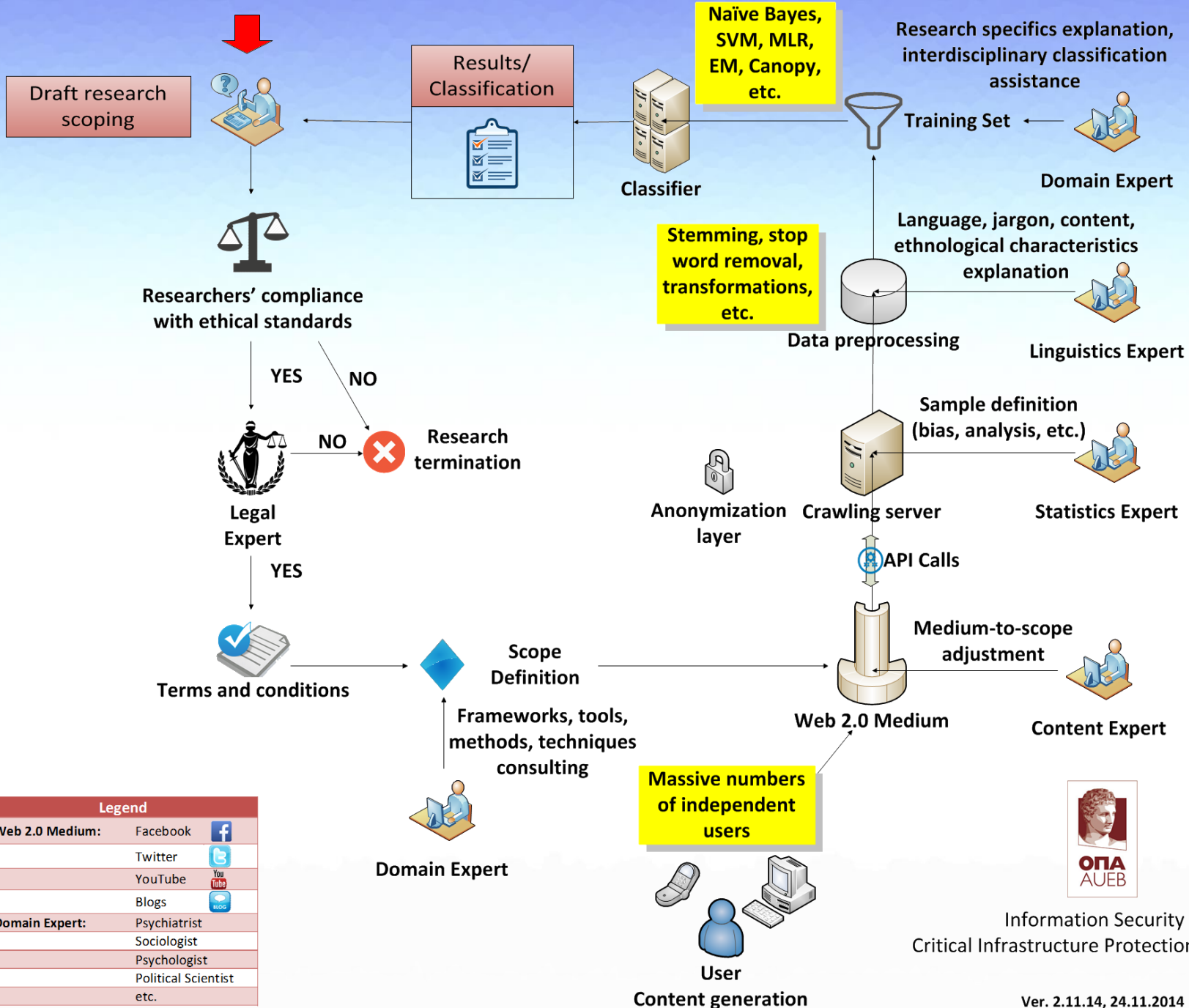
FBI's factors

- 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

Human behavior prediction: An indicative set of theories & frameworks



The NEREUS[©] Framework
Social Media Intelligence (COSMINT)




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



Case 1:

Revealing negative attitude towards law enforcement

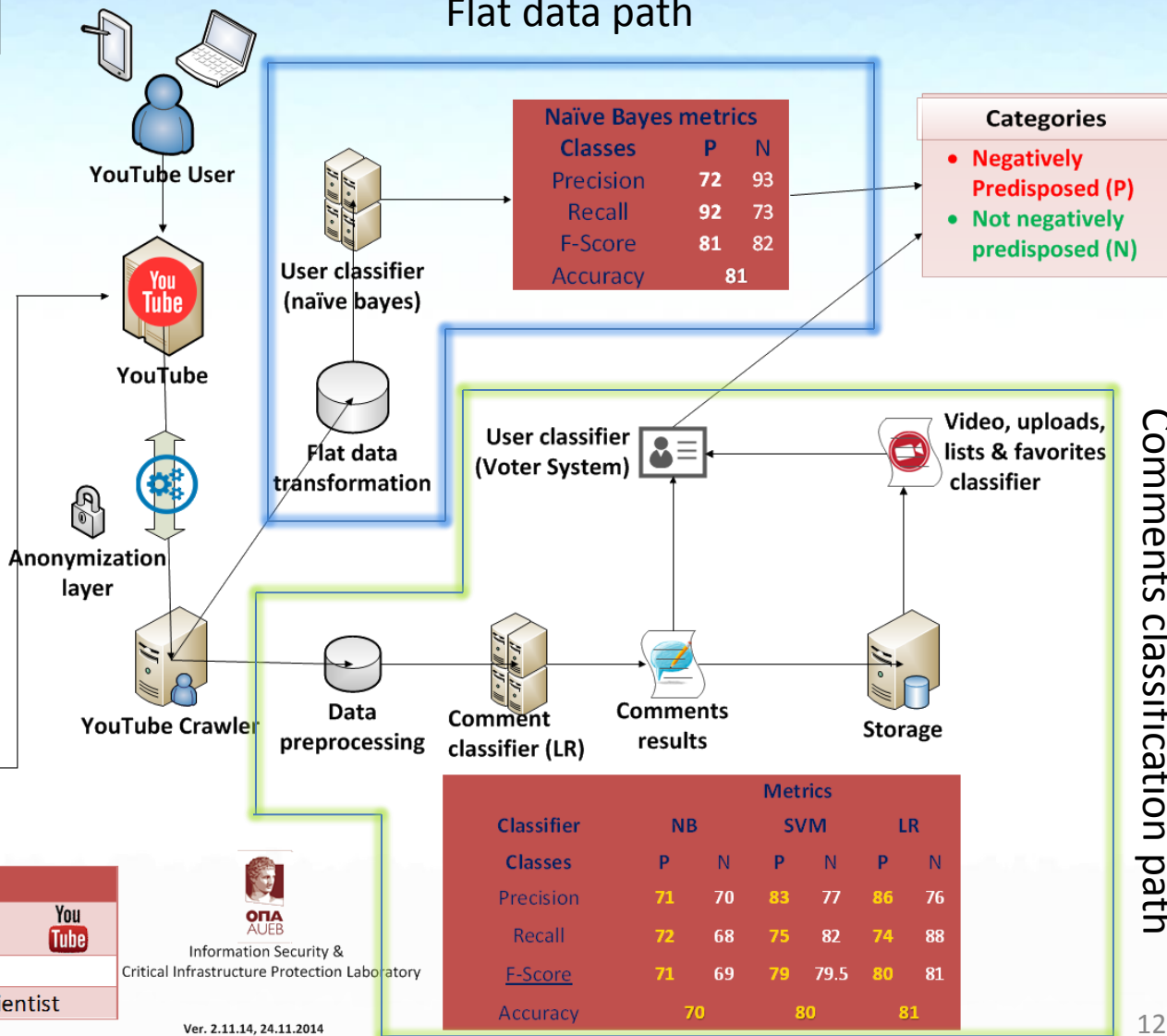
OSINT		OSN: YouTube	
Tools used for the analysis			
Science	Theory		
Computing	Machine Learning		
	Data Mining		
Sociology	Social Learning Theory		
Application: Detection of threats and delinquent behaviors, capabilities for detection of deviant behavior of minors			

In a nutshell

Detecting negative predisposition towards law enforcement

Flat data path

Comments classification 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)

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	

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

Information Security & Critical Infrastructure Protection Laboratory

Case 1: Revealing negative attitude towards law enforcement



Law enforcement predisposition

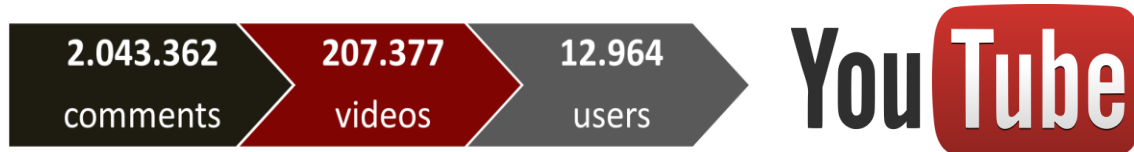
Study: Motive, anger, frustrations, predisposition towards law enforcement

Means: Machine Learning, comment classification, flat data classification.

- Individuals tend to transfer offline behavior online
- Extract results about users' negative **attitude towards law enforcement and authorities** (government, army, police)
- Trait of negative attitude towards law enforcement is connected to **delinquent behavior** via:
 - Sense of entitlement
 - Lack of empathy
 - **Anger and revenge syndrome** and
 - Inflated self-image

Dataset description

- Crawled YouTube and created dataset consists solely of **Greek** users.
- Utilized YouTube **REST-based API** (developers.google.com/youtube/):
 - Only publicly available data collected
 - Quote limitations (posed by YouTube) were respected
- Collected data were classified into three categories:
 - User-related information (profile, uploaded videos, subscriptions, favorite videos, playlists)
 - Video-related information (license, # of likes, # of dislikes, category, tags)
 - Comment-related information (comment content, # of likes, # of dislikes)

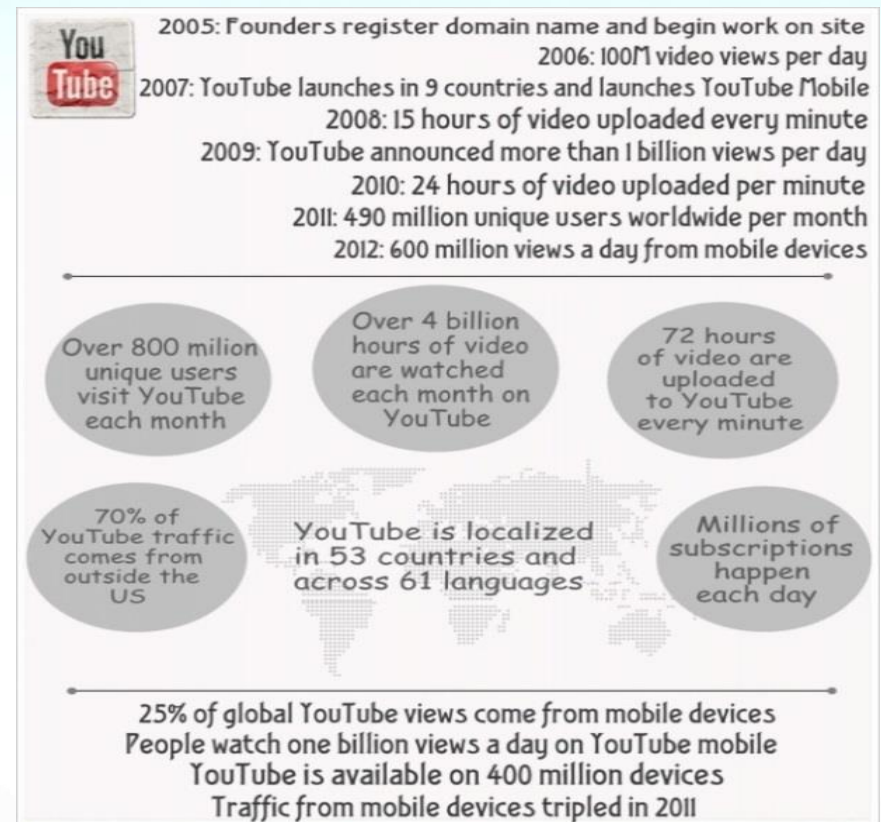


- Time span of collected data covered 7 years (Nov 2005 - Oct 2012).
- A basic anonymisation layer added to the collected data:
 - MD5 hashes instead of usernames

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, alleged anonymity, users interact with each other, contains political content



Machine Learning (1/3)

- Comment classified into categories of interest:
 - Process performed as **text classification**
 - Machine trained with **text examples** and the **category** each one belongs to
 - Excessive support by **field expert** (Sociologist)
- Test set used to evaluate efficiency 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 comments exist
- Training sets (Greeklish, Greek) were merged - One classifier 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 (2/3)

- **Comment** classification using:
 - Naïve Bayes (NB)
 - Support Vector Machines (SVM)
 - Logistic Regression (LR)
- Classifiers **efficiency** comparison:
 - Metrics (on % basis): Precision, Recall, F-Score, Accuracy
- **Logistic Regression** algorithm:
 - LR classifies a comment with **81% accuracy**

Precision: Measures the classifier exactness. Higher and lower precision means less and more false positive classifications, respectively.

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

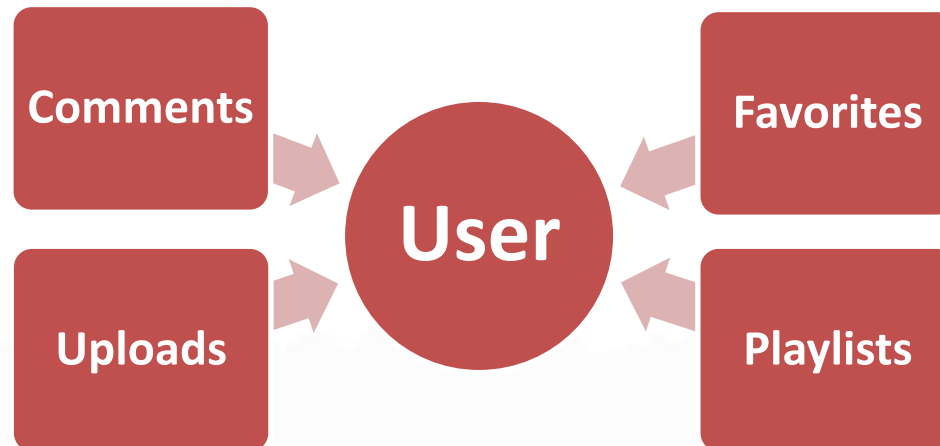
F-Score: Weighted harmonic mean of both metrics.

Accuracy: No. of correct classifications performed by the classifier. Equals to the quotient of good classifications by all data.

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

Machine Learning (3/3)

- **Video** classification:
 - Examination of a video on the basis of its comments
 - Voter process to determine category classification
- **(Video) Lists** classification:
 - Voter process to determine category classification (same threshold)
- Conclusions about **user behavior**:
 - If there is at least one category P attribute then the user is classified into category P



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**) machine trained by a set of users of categories P and N
 - Red:** Users of category P classified on the basis of their comments-only (**Machine Learning**)

- Data transformation:

- User repr comment



o ID the views).
eld expert).

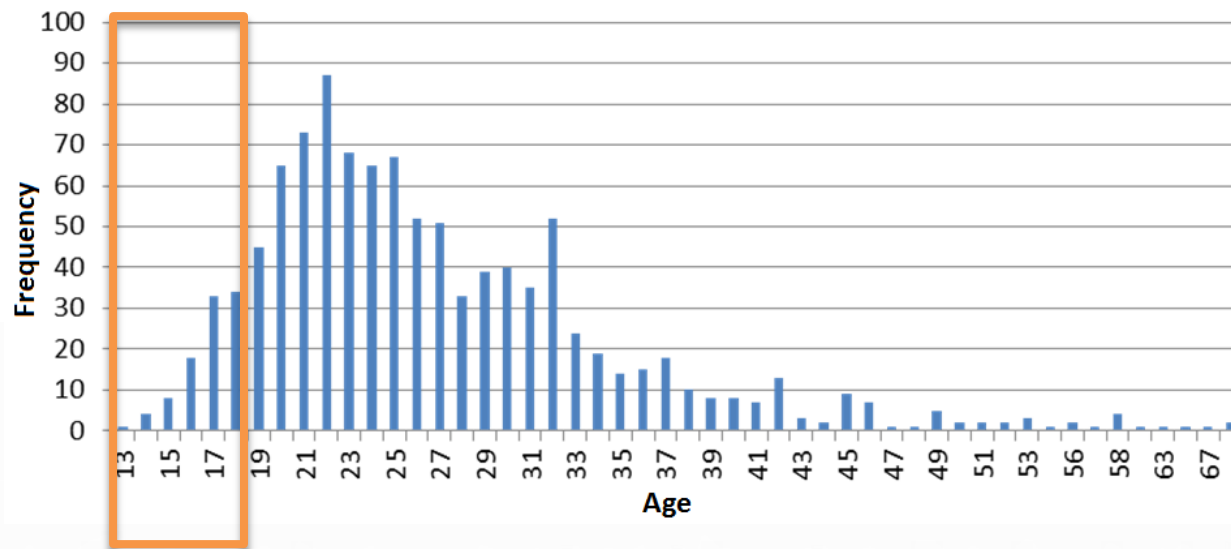
- Machine

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%	

Basic observations

- 6% of **comments** (of the 2.000.000 collected) include **negative attitude** towards law enforcement
- 3.5% (of the 200.000 collected) of **videos** classified into the specific category of interest
- 14% (of the 13.000 collected) of **users** express **negative attitude** towards law enforcement
- Ability to **detect and predict delinquent behaviour of minors**
 - Violent behaviour
 - Cyber bullying
 - Sexual or emotional harassment



Case 2:

Detecting stress level use patterns

OSINT

OSN: Facebook 

Tools used for the analysis

Science

Tool

Computing

Machine Learning

Data Mining

Psychiatry
Psychology

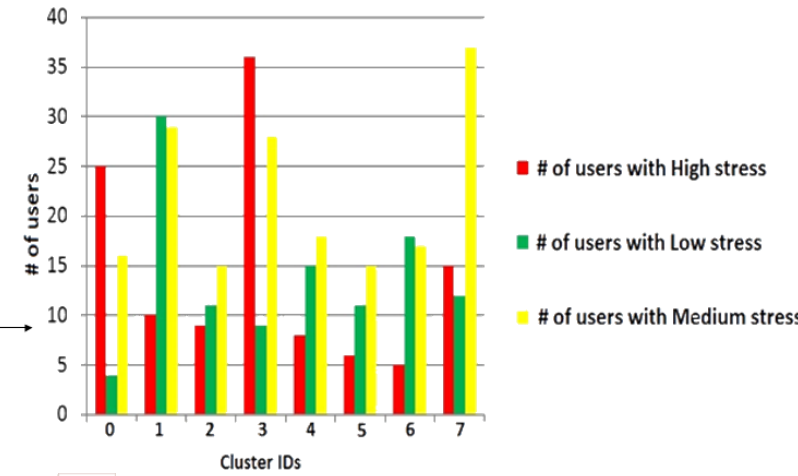
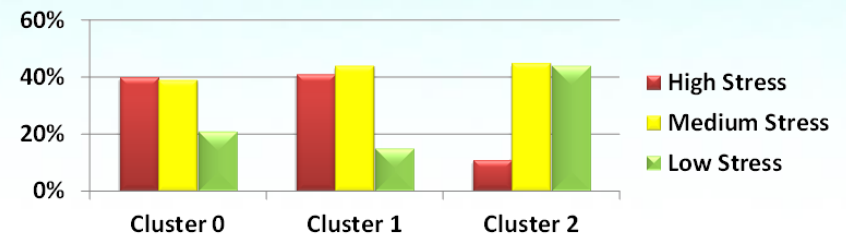
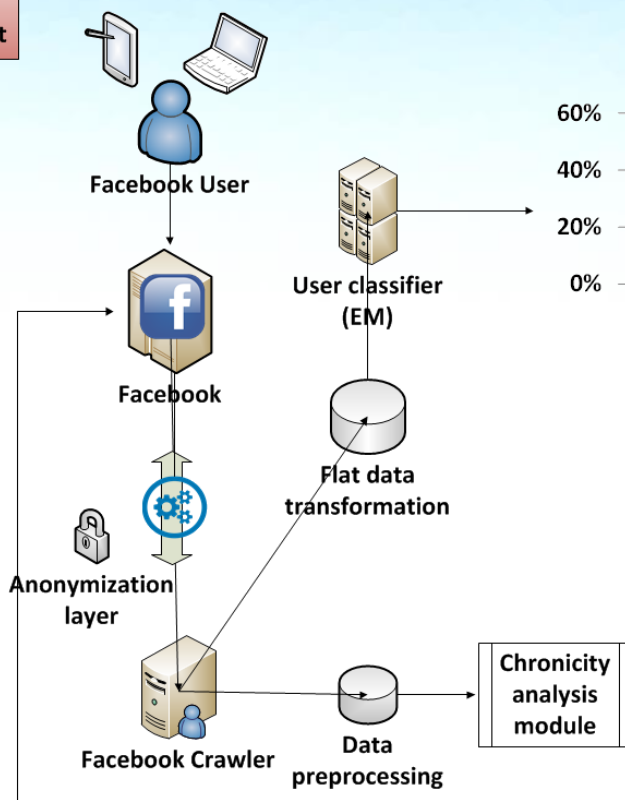
BAI stress test

Application: Detection/prediction of vulnerable individuals and potential threats, detection and mitigation of child abuse

In a nutshell



Detect individuals vulnerable to blackmail and moral inhibitions shift

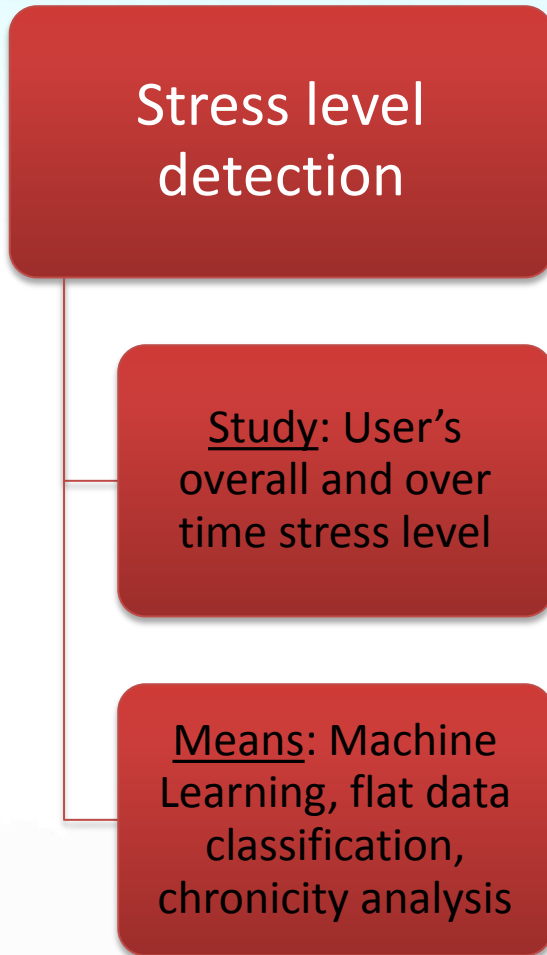


Legend	
Web 2.0 Medium:	Facebook
Domain Expert:	Psychiatrist Psychologist


 Information Security &
 Critical Infrastructure Protection Laboratory

Case 2:

Detecting stress level usage pattern

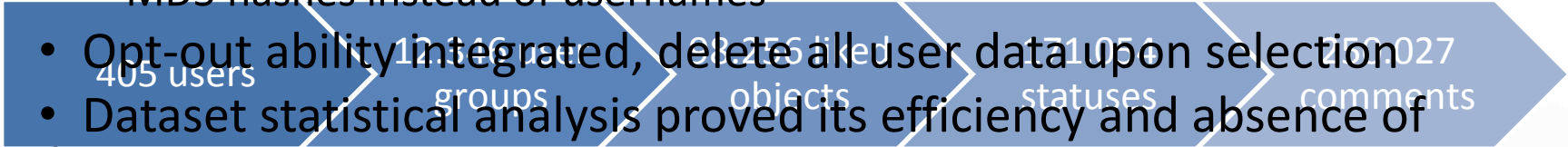
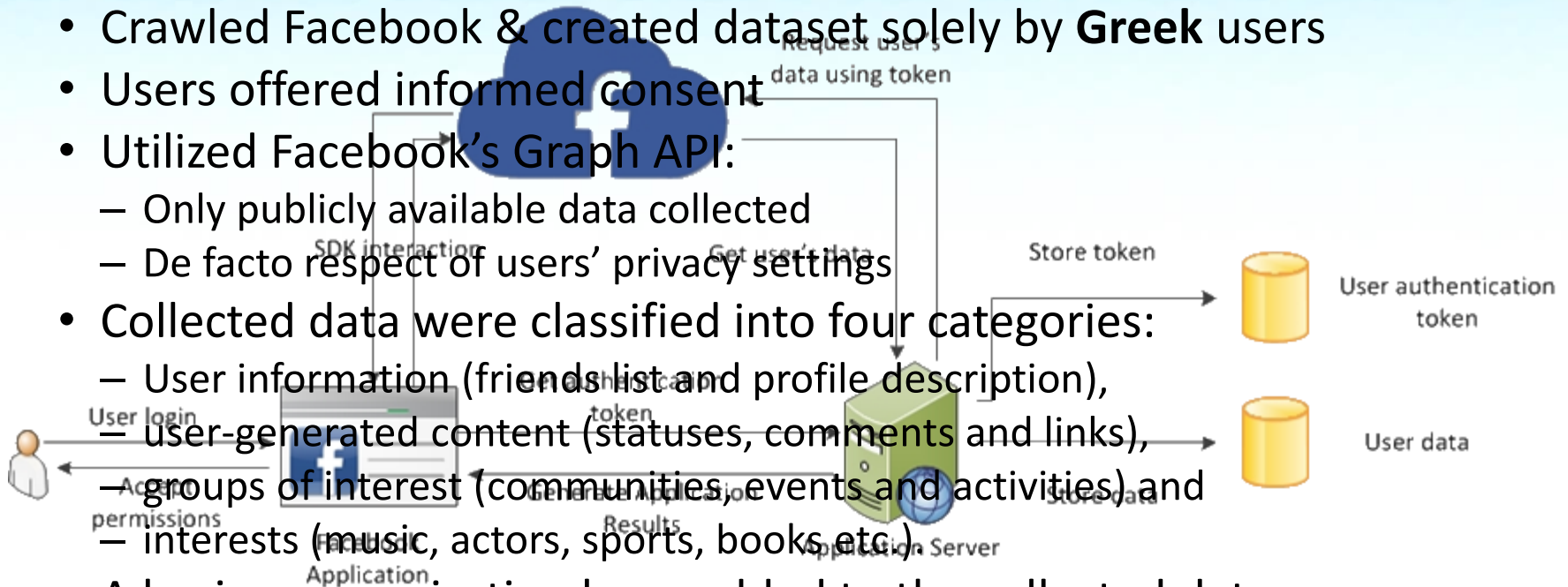


- Individuals tend to **transfer offline behavior online**
- Extract results about **usage pattern** depicted **stress level**
- Analyze each user under the prism of stress level both **overall** and **over time** (chronicity analysis)
- **High stress** has been found to:
 - Make individuals vulnerable to **fall prey** to third parties
 - Overcome **moral inhibitions**
- Analysis is based on **Social Learning Theory** and stress correlations are based on **Beck's Anxiety Inventory (BAI)** stress test



Dataset description

- Crawled Facebook & created dataset solely by **Greek** 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

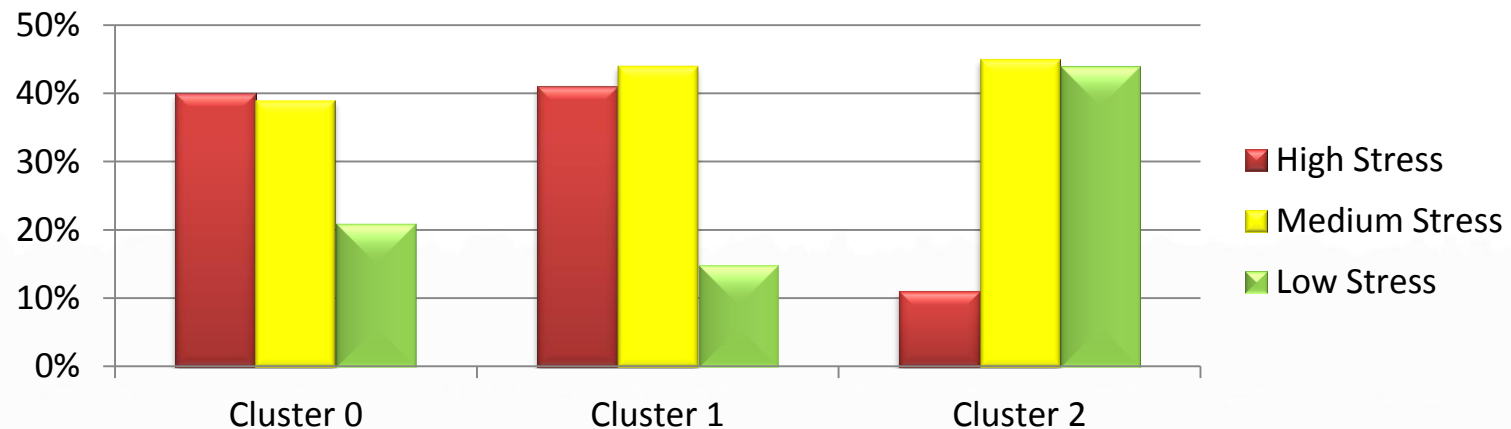


- Opt-out ability integrated, delete all user data upon selection
- Dataset statistical analysis proved its efficiency and absence of bias

Flat classification (overall indicators)



- Goal: Extract correlations between **usage patterns** and **users who share same stress valuation** (according to **BAI** test)
- Transformed relational database into a **single tuple record** containing solely users' **comments** and **status**
- Flat data tuple subjected to **stemming** process
- EM (Expectation Maximization) algorithm produced three clusters:
 - Cluster **0** has too few users (~20 users)
 - Cluster **1** includes users with high and medium-to-high stress score (~200 users)
 - Cluster **2** includes users with low and medium-to-low stress score (~200 users)



Chronicity analysis (indicators over time)



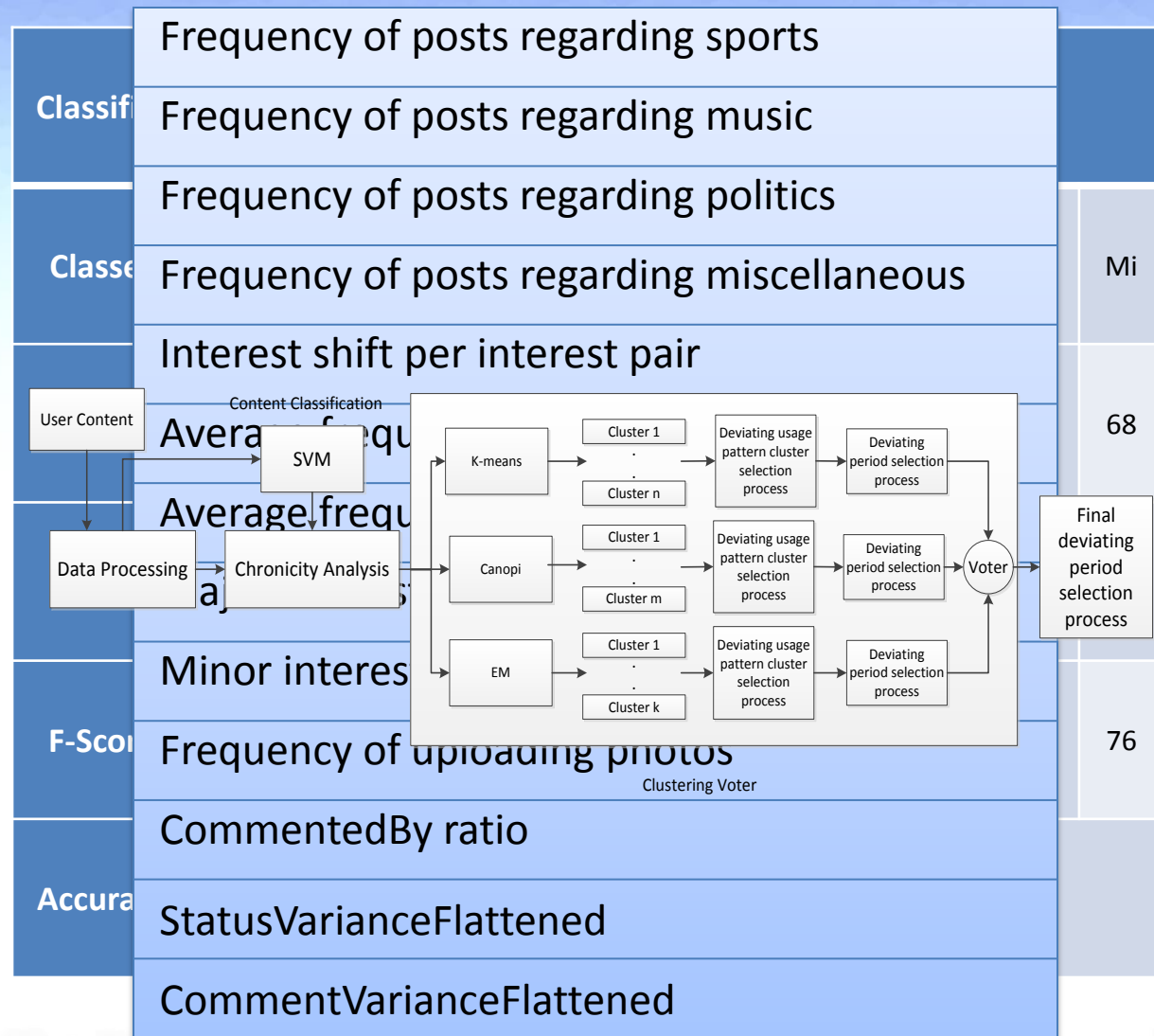
- Goal: **Detect differentiations** of OSN use patterns over **time related**, so as to depicted stress level
- Split users' use patterns into time periods (from 1 day to 1 month)
 - Time period of one week produced best results
- Chronicity analysis system consists of two modules:
 - Preprocessing data 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).

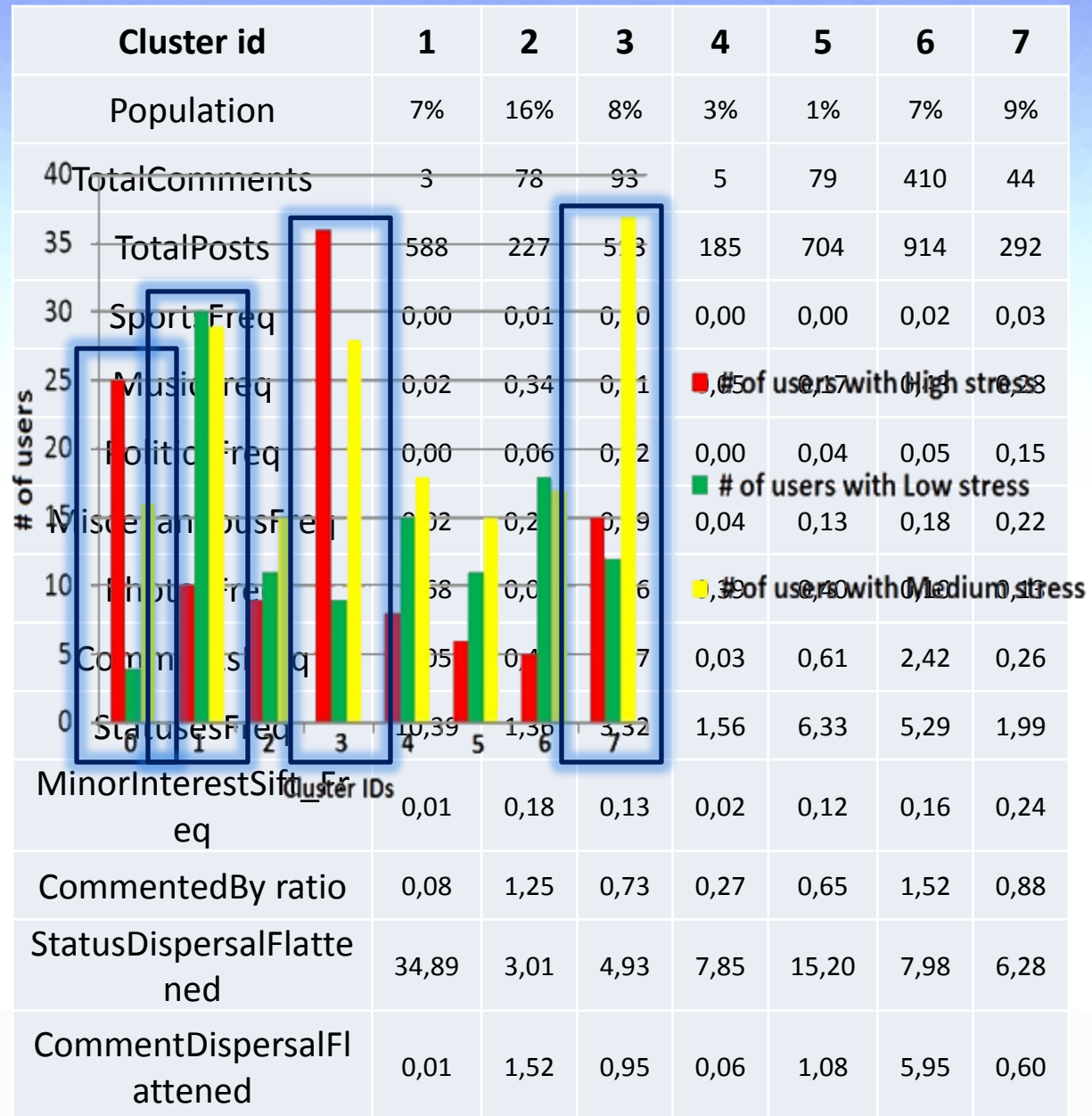
Step 2: Calculate following metrics for each user and time period (metrics developed on an ad-hoc basis according to our observations).

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 referring 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 miscellaneous content and also contain limited content referring to sports.



Case 3

Threat detection based on Narcissism

OSINT

OSN: Twitter



Tools used for the analysis

Science

Theory

Computing

Graph Theory

Sociology
Psychology

Theory of Planned Behavior

Social Learning Theory

Application: Emotional harassment and cyber bullying detection/prediction.

In a nutshell



Predicting & identifying potential insiders



Researchers' compliance with ethical standards

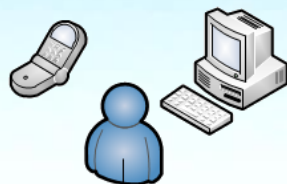
YES



Legal Expert

YES

Critical infrastructures
National security
Public interest



Twitter Users

Content generation



Twitter

Crawling & storing



Our crawling server



Klout score server

Klout score queries



Klout score api collector



Content Aggregator

Usage intensity valuation



Indegree/outdegree aggregator

Influence valuation



User classification according to categories

Legend

Web 2.0

Medium:

Domain Expert: Psychologist

Twitter



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

Case 3: Threat detection based on Narcissism



Narcissistic behavior detection

Study: Motive, ego/self-image,
entitlement

Means: Usage Intensity,
Influence valuation, Klout score

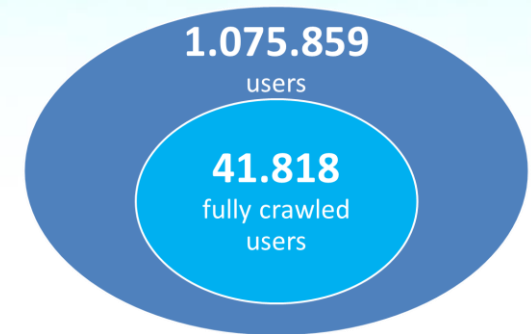
- Individuals tend to transfer offline behavior online.
- Trait of narcissism directly relates to **insider threats**, **OSN popularity** and **influence**.
- Utilize graph theoretic tools to perform analysis.
- Valuation of social media **popularity** and **usage intensity**.
- Twitter data to become open.
- Trait of narcissism relates to delinquent behavior via :
 - sense of entitlement
 - lack of empathy
 - anger and “revenge” syndrome
 - inflated self-image

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

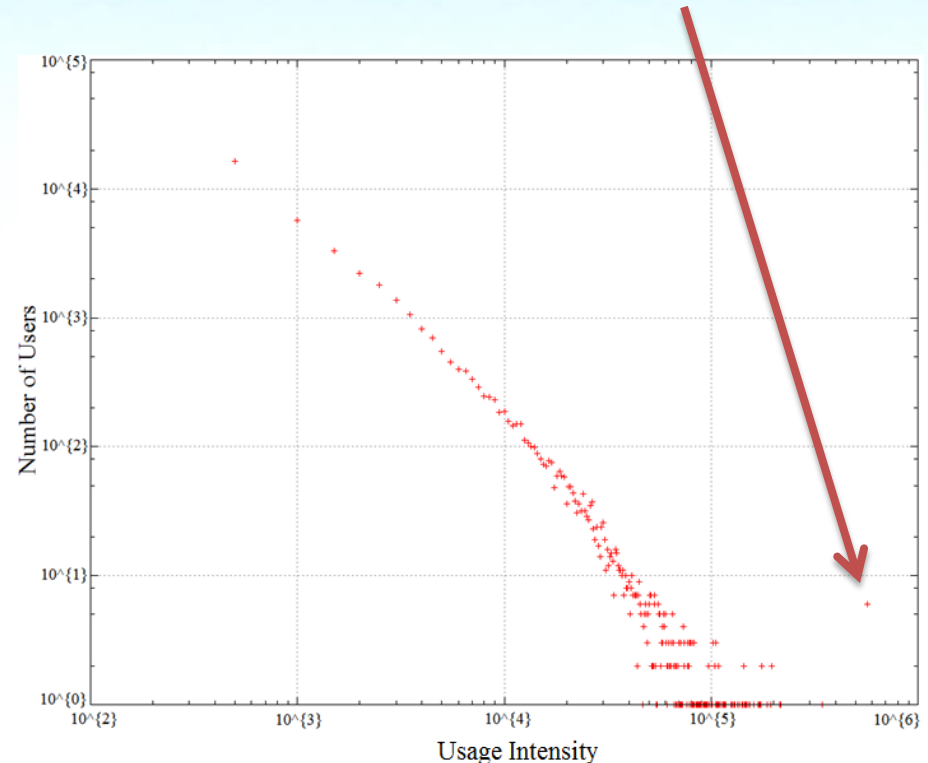


Graph Theoretical approach

- **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

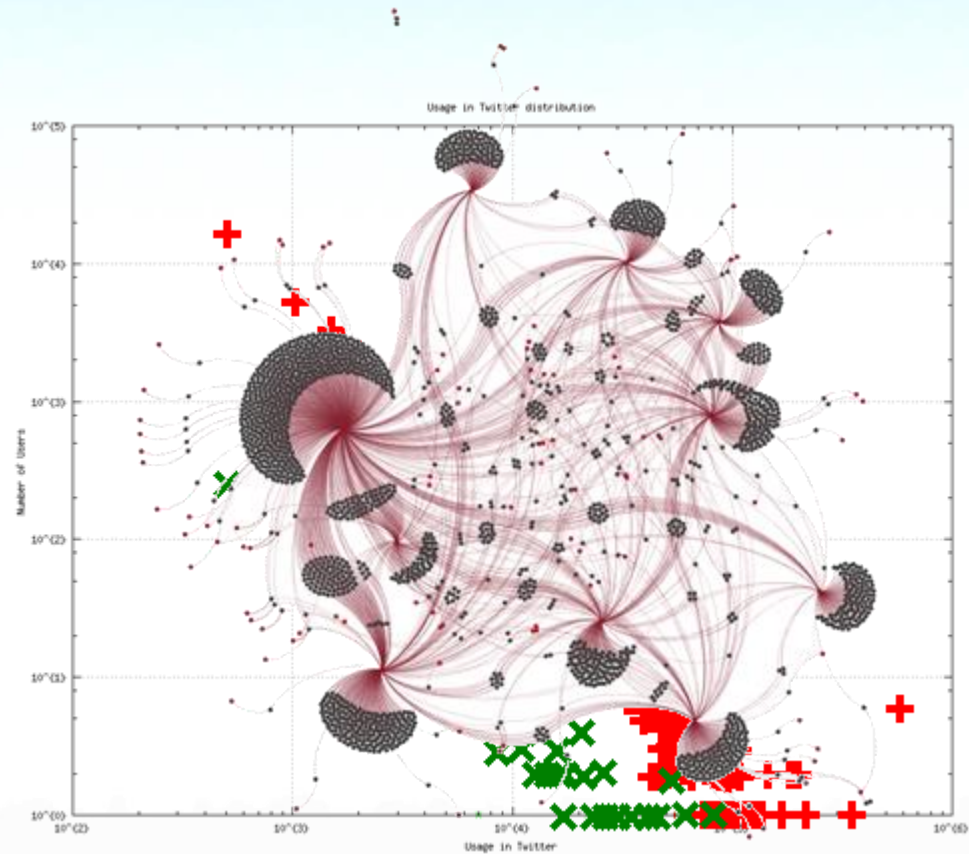
The excessive use of Twitter by persons who are not mass-media or personas could connect to narcissism and identify narcissists, i.e. persons who - inter alia - tend to turn delinquent

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**



Web 2.0 data exploitation options

- **Threat prediction:**
 - Applying Shaw and FBI psychosocial indicators (narcissism, anger syndrome, revenge syndrome, 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 (examination of delinquent behavior, etc.).

Some conclusions

- ✓ Web 2.0 produces vast amounts of **crawable** information and SOCMINT can transform this information into **intelligence**
- ✓ SOCMINT can assist in detecting **narcissistic behavior**, **predisposition towards law enforcement**, etc.
- ✓ SOCMINT can **predict threats**, **predict delinquent behavior**, **support law enforcement** and **enhance minor protection**
- ✓ SOCMINT intrusive nature dictates **specific** usage for clearly **legitimate** 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, Privacy, 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. Pipyros K., Mitrou L., Gritzalis D., Apostolopoulos T., "A Cyber Attack Evaluation Methodology", in *Proc. of the 13th European Conference on Cyber Warfare and Security (ECCWS-2014)*, Greece, 2014.
13. Stavrou V., Kandias M., Karoulas G., Gritzalis D., "Business Process Modeling for Insider threat monitoring and handling", in *Proc. of the 11th International Conference*