

Open Source Intelligence from Web 2.0 Media

Miltos Kandias, Vasilis Stavrou

November 2014

Technical Report AUEB/INFOSEC/Rev-1114/v.1.1
INFOSEC Laboratory, Dept. of Informatics
Athens University of Economics & Business
November 2014

Open Source Intelligence from Web 2.0 Media



Miltos Kandias, Vasilis Stavrou

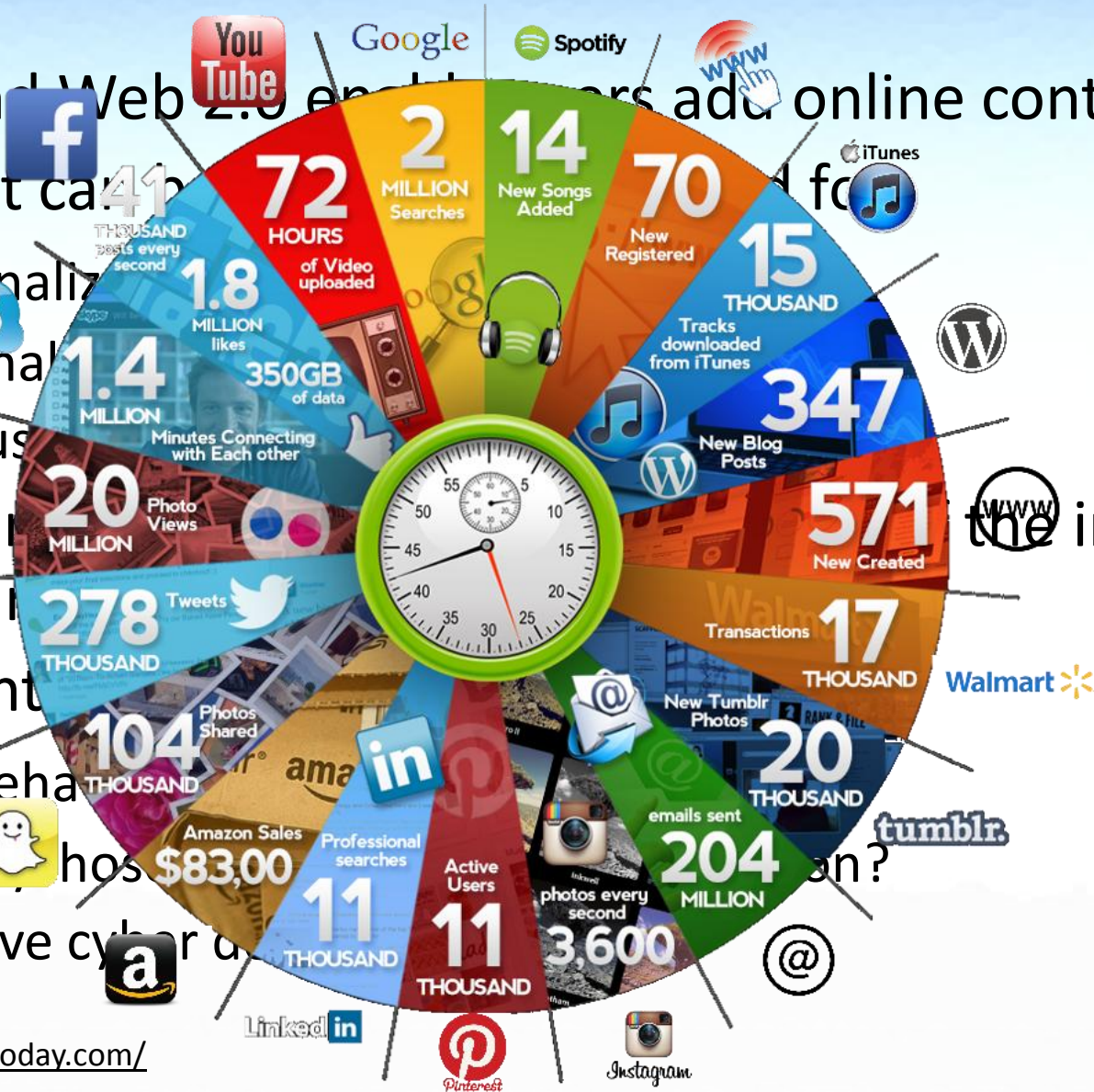
Information Security & Critical Infrastructure Protection Laboratory
Dept. of Informatics, Athens University of Economics & Business, Greece

Outline

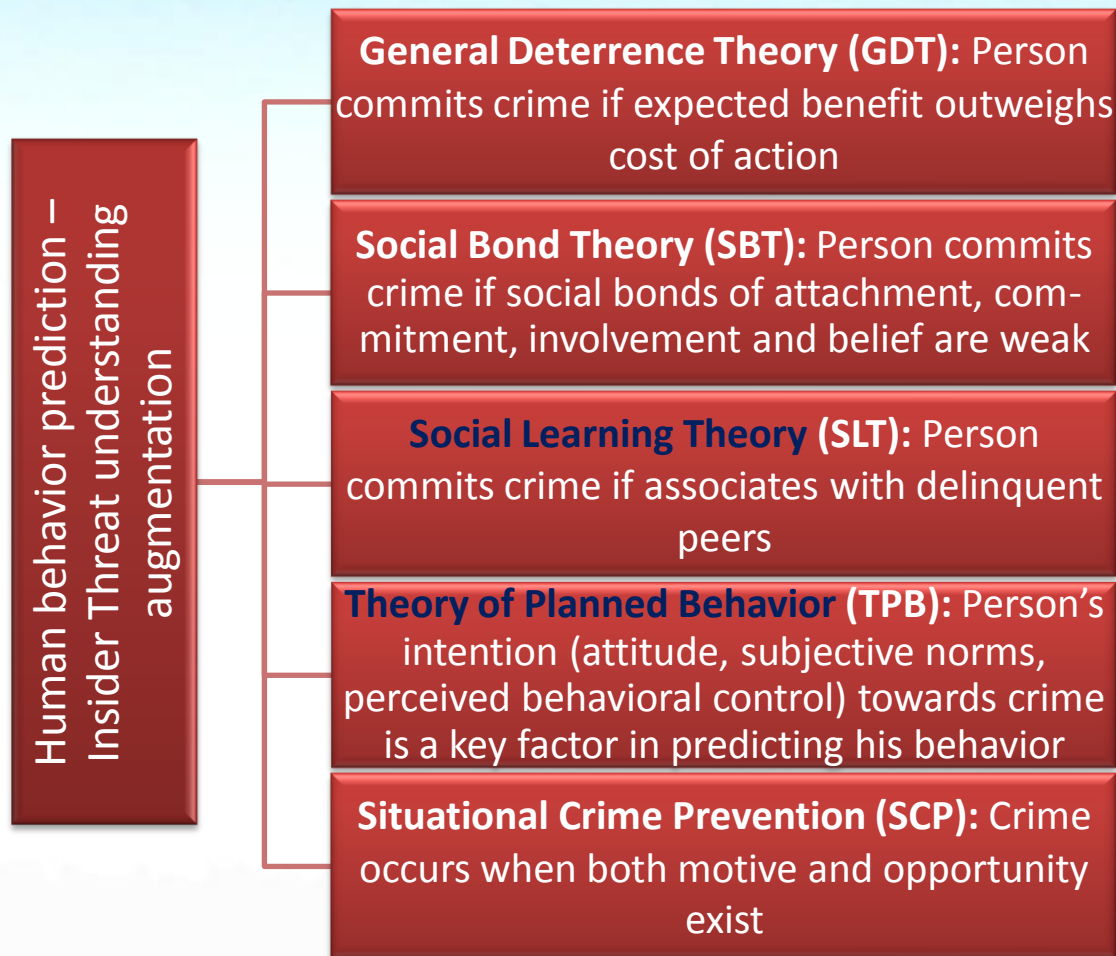
- Web 2.0 and Online Social Networks (OSN)
- Open Source Intelligence (OSINT)
- Threats and Opportunities
- Behavior prediction capabilities
 - Case 1:** Success story - Insider detection and narcissism
 - Case 2:** Success story - Predisposition towards law enforcement
 - Case 3:** Success story - Detecting stress levels
 - Case 4:** Horror story - Revealing political beliefs
- Conclusions

Web 2.0 & Online Social Networks

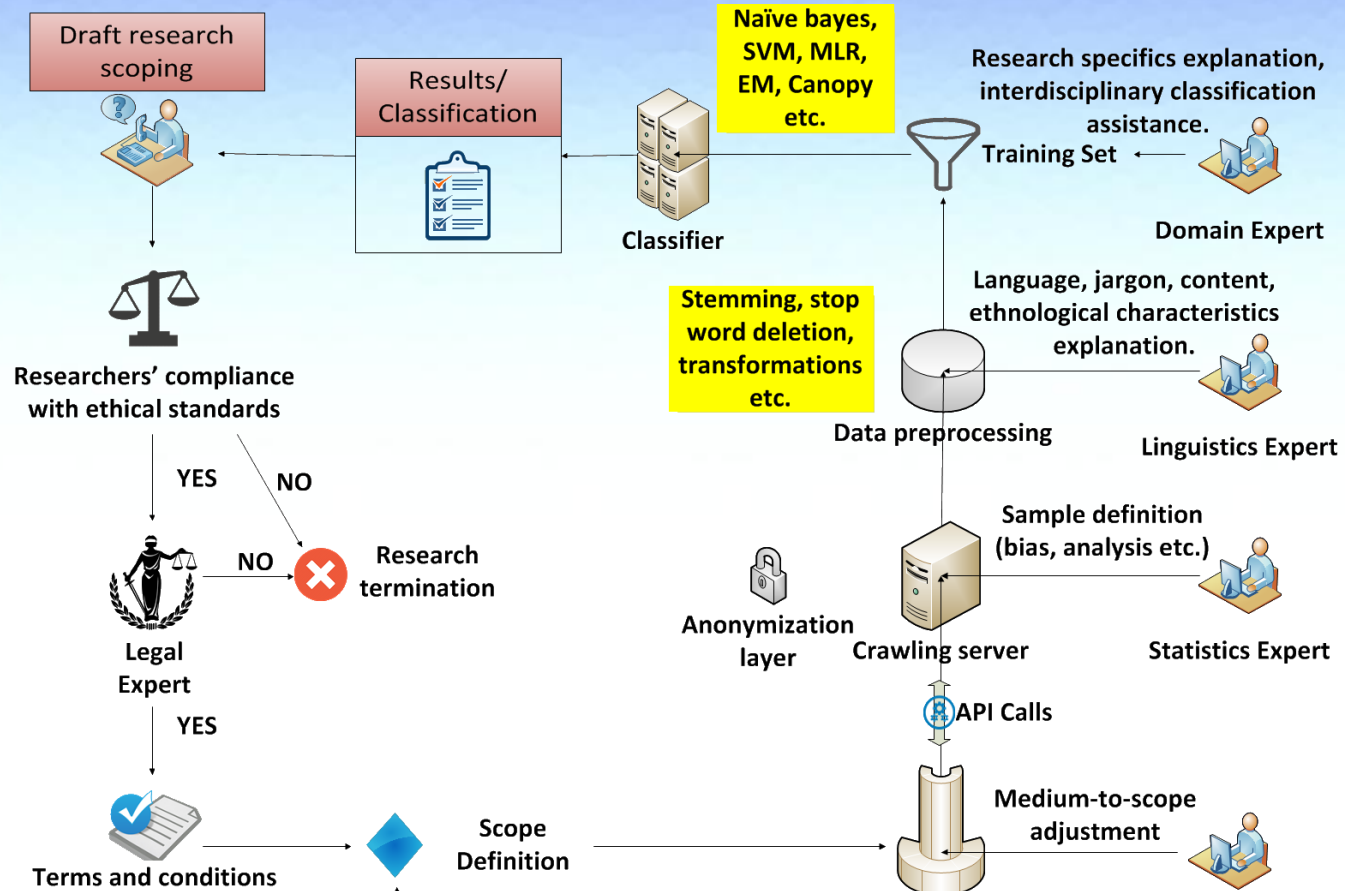
- OSN and Web 2.0 enablers add online content.
- Content can be personalized
 - personalized
 - personalized
 - user/user
- Users are the sharers of the info
- Can content be shared?
 - User behavior
 - User photos
 - Proactive cyber







Delinquent behavior prediction theories




Open Source Intelligence from Web 2.0 Media



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

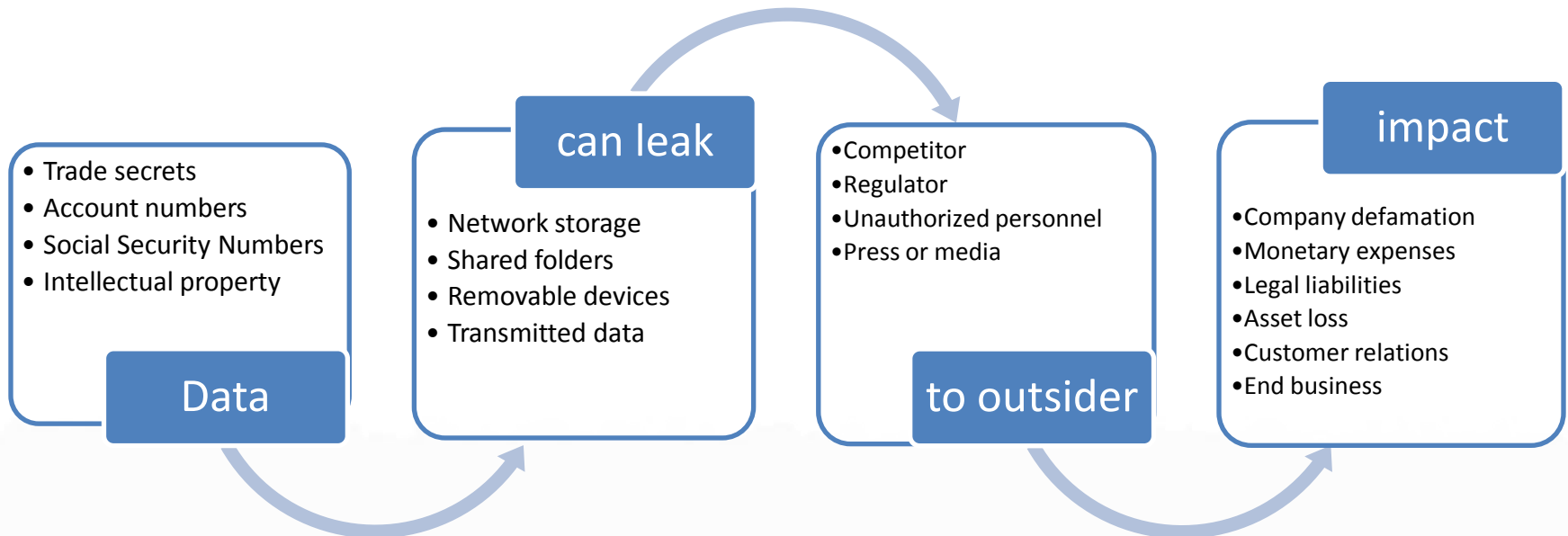
Case 1

Scope: 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.		

Insider Threat

- The insider threat is a severe problem in cyber/corporate security, which originates from persons who:
 - are legitimately given access rights to information systems,
 - misuse privileges and
 - violate security policy.





In a nutshell

Predicting & identifying potential insiders



Researchers' compliance with ethical standards

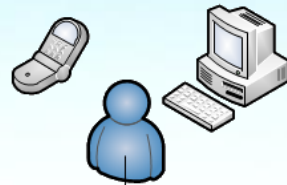
YES



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/oudegree aggregator

Influence valuation



User classification according to categories

Legend

Web 2.0

Medium:

Domain Expert: Psychologist

Twitter



OIA
AUUB

Information Security &
Critical Infrastructure Protection
Laboratory

Category

Loners

Individuals

Known users

Mass Media & Personas

Influence valuation

0 - 90

90 - 283

283 - 1.011

1.011 - 3.604

Klout score

3.55 - 11.07

11.07 - 26.0

26.0 - 50.0

50.0 - 81.99

Usage valuation

0 - 500

500 - 4.500

4.500 - 21.000

21.000 - 56.9000

Case 1: Insider threat prediction based on Narcissism



Narcissistic
behavior
detection

Study: Motive, ego/self-
image, entitlement

Means: Usage Intensity,
Influence valuation,
Klout score

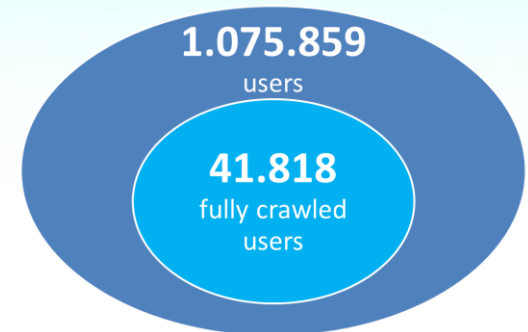
- Individuals tend to transfer offline behavior online.
- Trait of narcissism directly relates to **insider threats, OSN popularity & influence.**
- Utilize graph theoretic tools to perform analysis.
- Valuation of social media **popularity** and **usage intensity.**
- Twitter data to become open.
-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



Dataset: General parameters

- 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

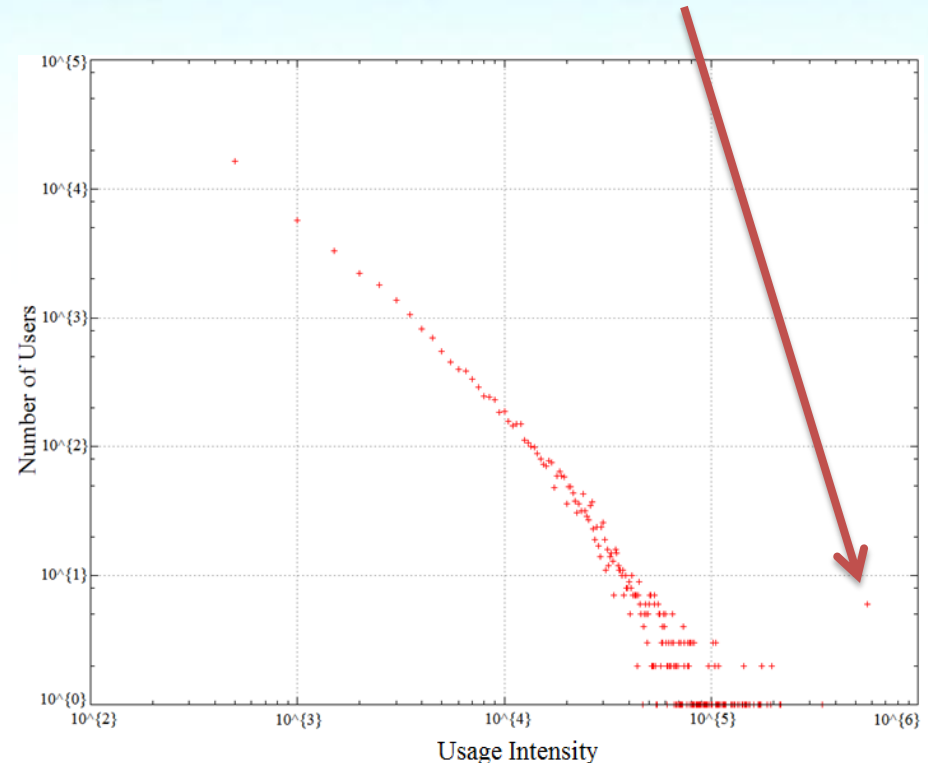
Graph Theoretical approach



- **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 4 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.

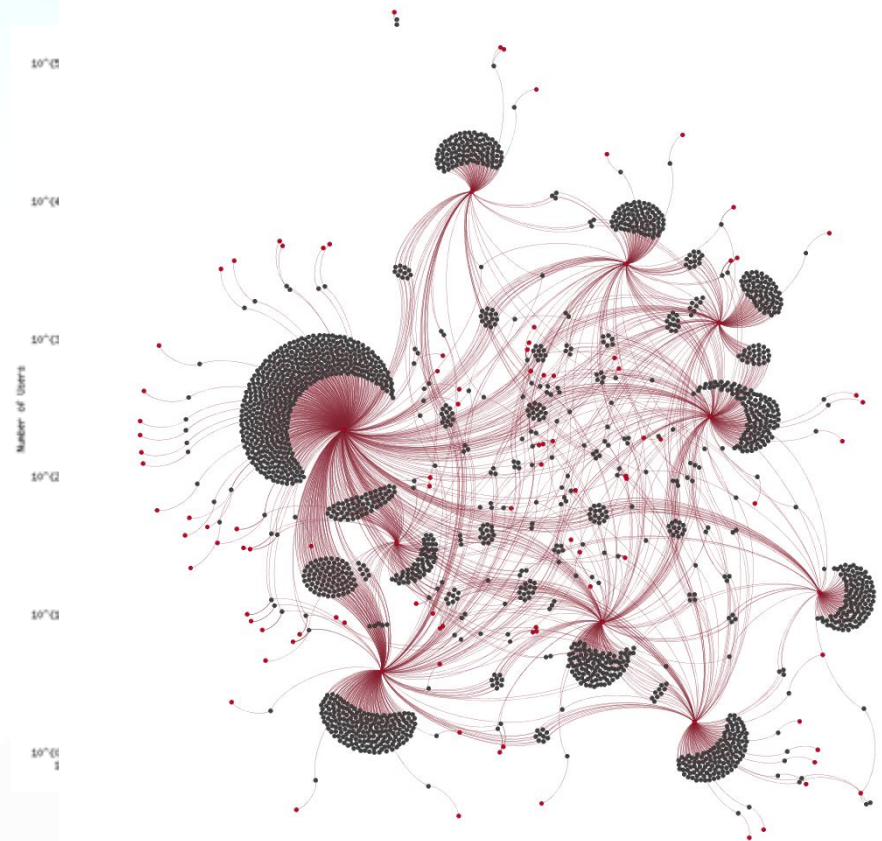
The excessive use of Twitter by persons who are not mass-media or personas could connect to narcissism and identify narcissists, i.e. persons who - inter alia - tend to turn insiders

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

Scope: 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, opportunities of 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

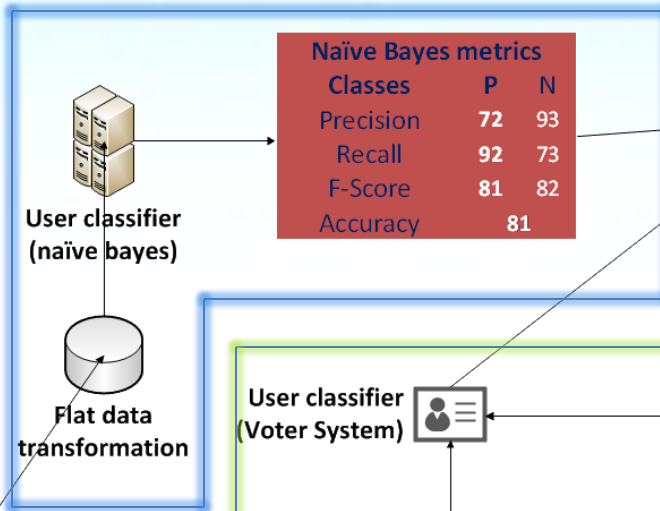
YouTube

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

Data preprocessing

Comment classifier (MLR)

Comments results

Storage

Comments classification path

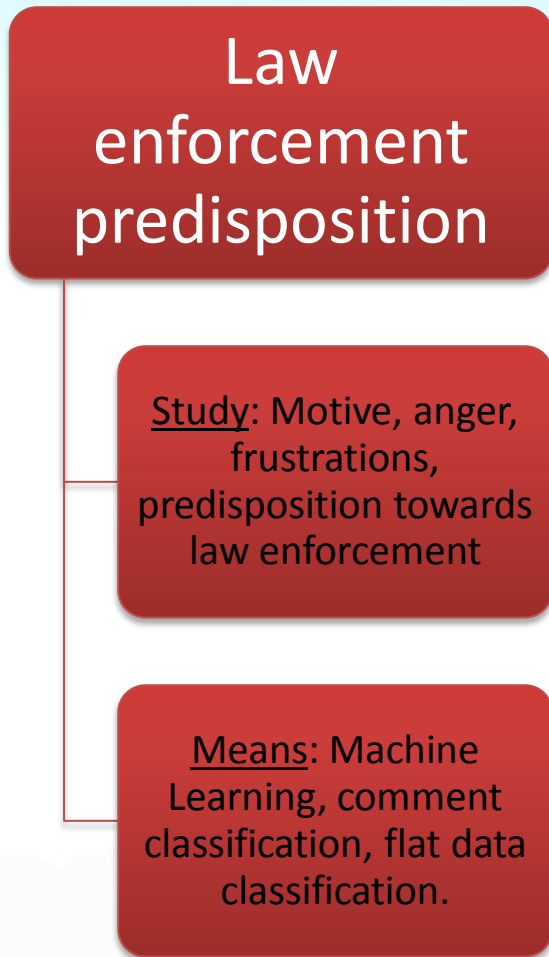
Classifier	Metrics					
	NBM		SVM		LR	
Classes	P	N	P	N	P	N
Precision	71	70	83	77	86	76
Recall	72	68	75	82	74	88
F-Score	71	69	79	79.5	80	81
Accuracy	70		80		81	

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



Information Security & Critical Infrastructure Protection Laboratory

Case 2: Revealing negative attitude towards law enforcement



- Individuals tend to transfer offline behavior online.
- Extract results about users' negative **attitude towards law enforcement and authorities** (government, army, police, hierarchy).
- 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: General parameters



- 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) and
 - com-ment--re-la-ted information (com---ment content, # of likes, # of dislikes).

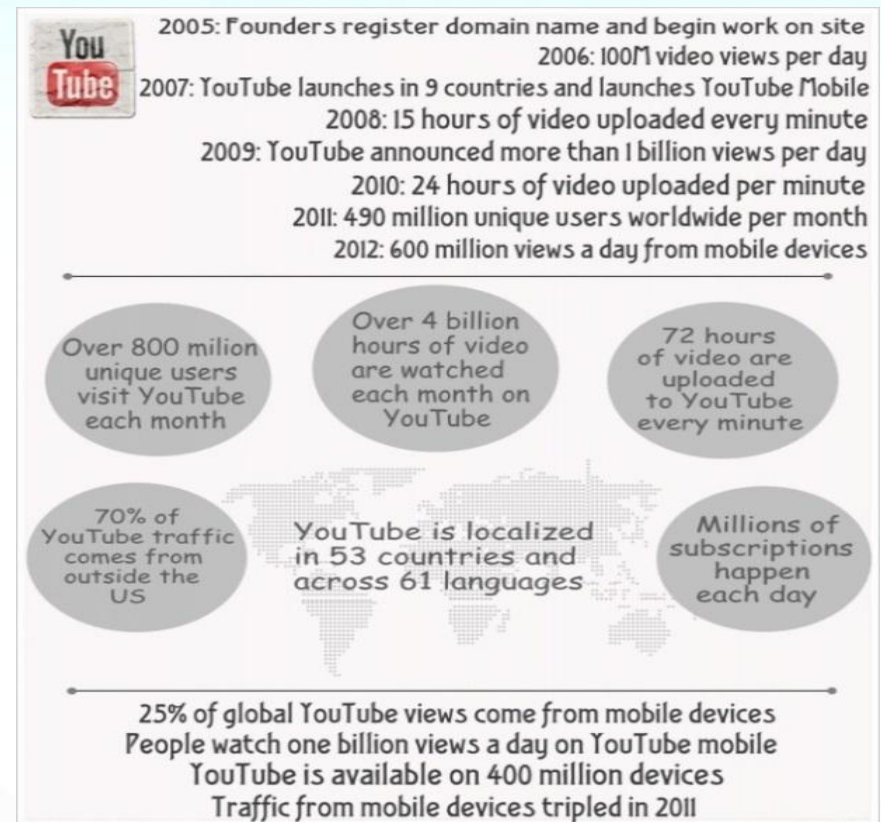


- Ti-me span of collected data covered 7 years (Nov 2005 - Oct 2012).
- A basic anonymization layer added to the col-lec-t-ed data:
 - MD5 hashes instead of usernames.

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)

- 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 comments exist.
- Training sets (greeklish, 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)

- **Comment** classification using:
 - Naï-ve Bayes (NB)
 - Support Vector Machines (SVM)
 - Logistic Re-gression (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.

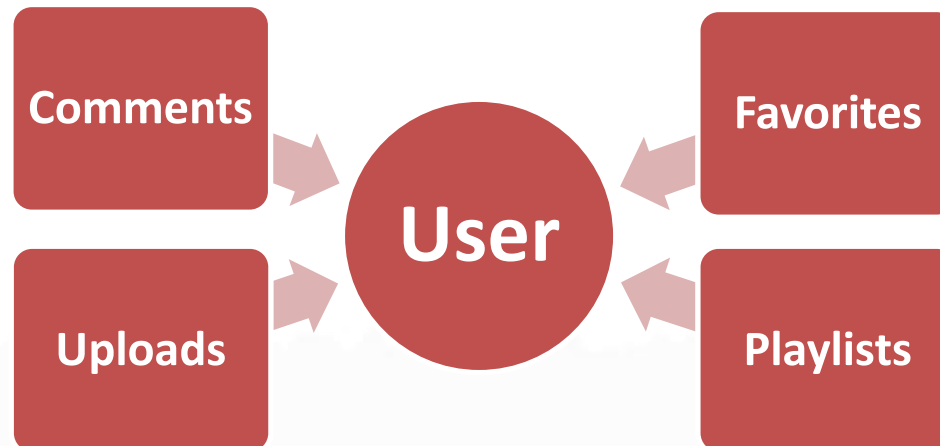
F-Score: Weighted harmonic mean of both metrics.

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)

- **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**).

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

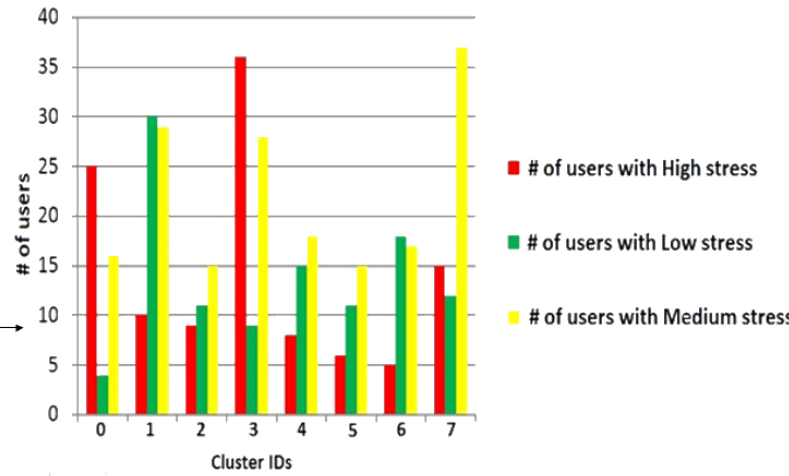
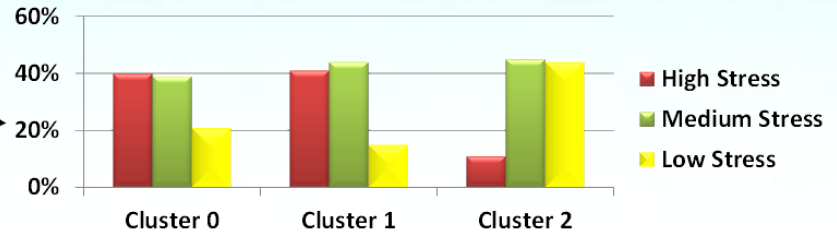
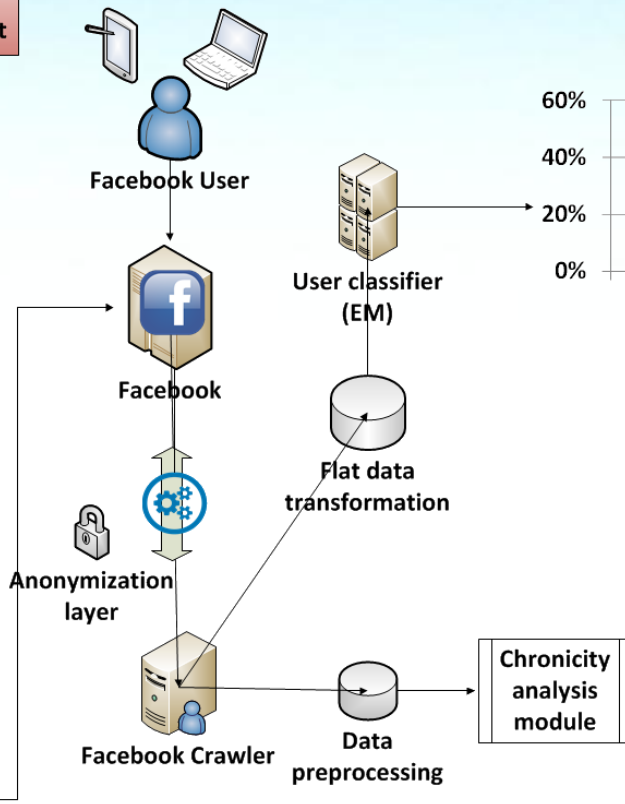
Scope: Detecting stress level usage patterns (overall and over time)

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 & 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

Case 3: Detecting stress level usage pattern (overall and over time)



Stress level detection

Study: User's overall and over time stress level

Means: Machine Learning, flat data classification, chronicity analysis.

- 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 in-hi-bitions**.
- Analysis is based on Social Learning Theory and stress correlations are based on Beck's Anxiety Inventory stress test.



Dataset: General parameters

- 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.

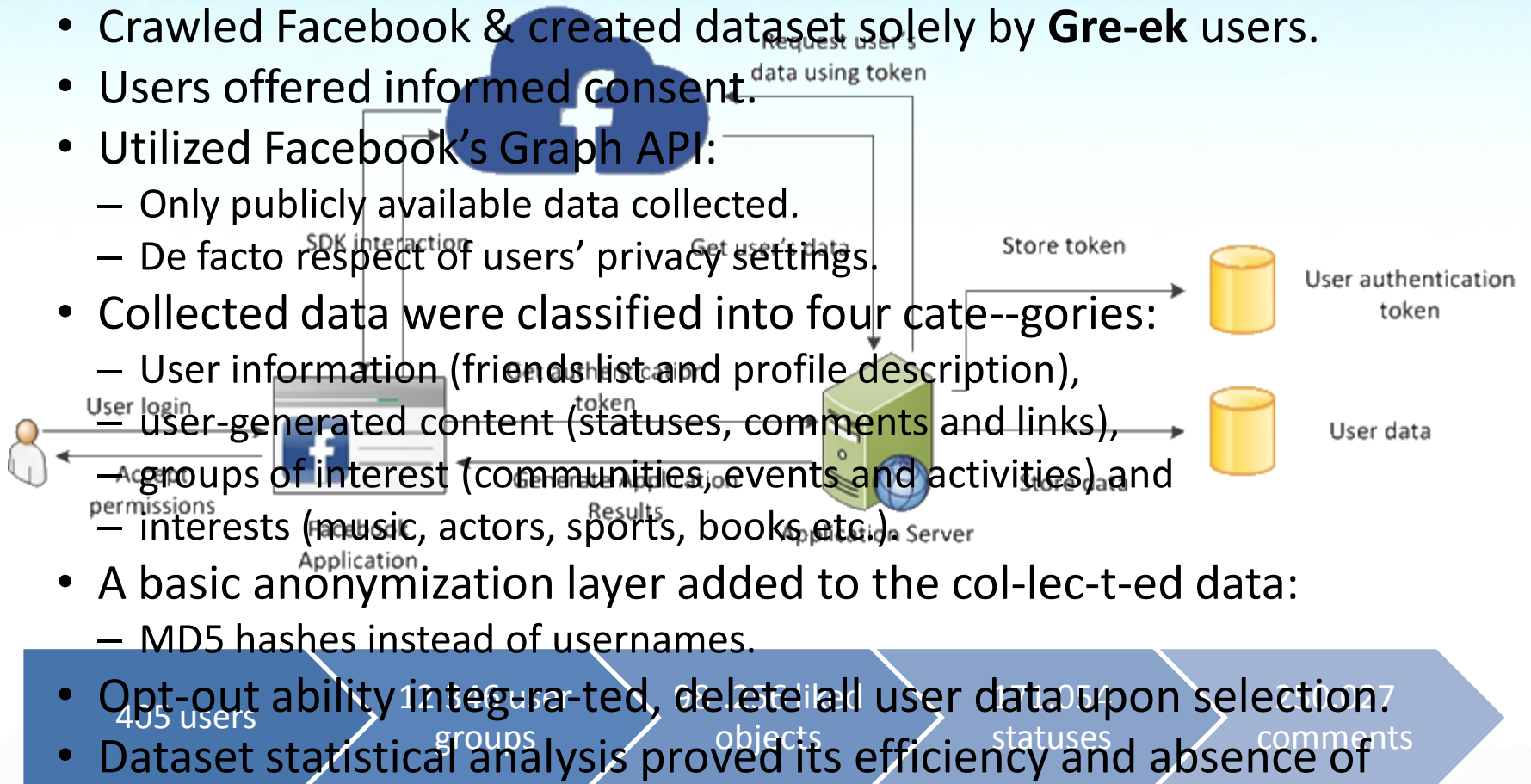
- Collected data were classified into four categories:

- 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.

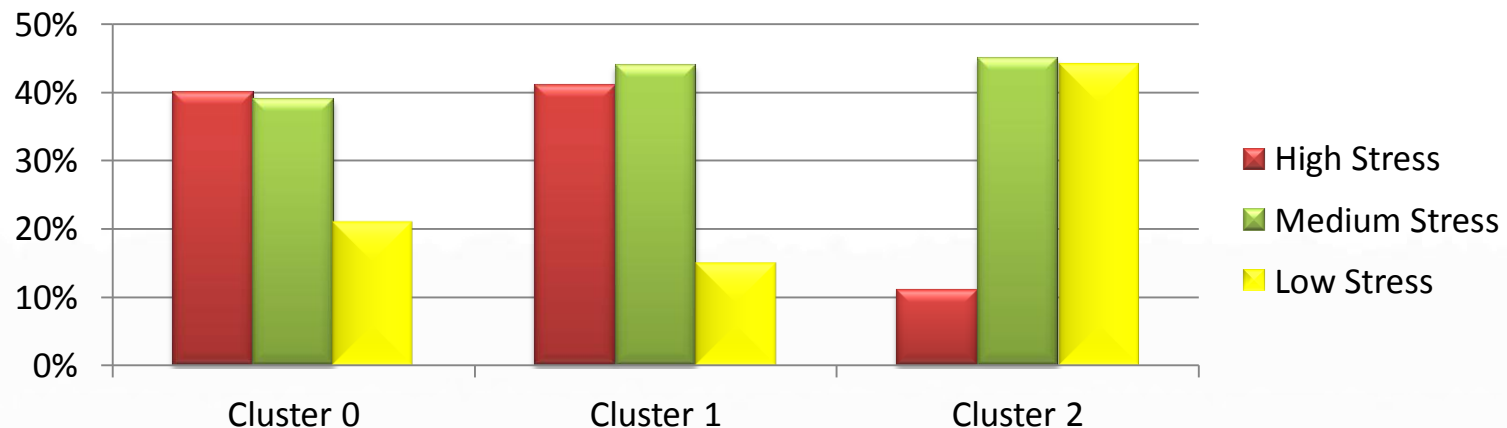
- Dataset statistical analysis proved its efficiency and absence of bias.



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 **statuses**.
- Flat data tuple subjected to stemming process.
- EM algorithm produced 3 clusters:
 - Cluster 0 has too few users.
 - Cluster 1 includes users with high and medium-to-high stress score.
 - Cluster 2 includes users with low and medium-to-low stress score.



Chronicity analysis (indicators over time)



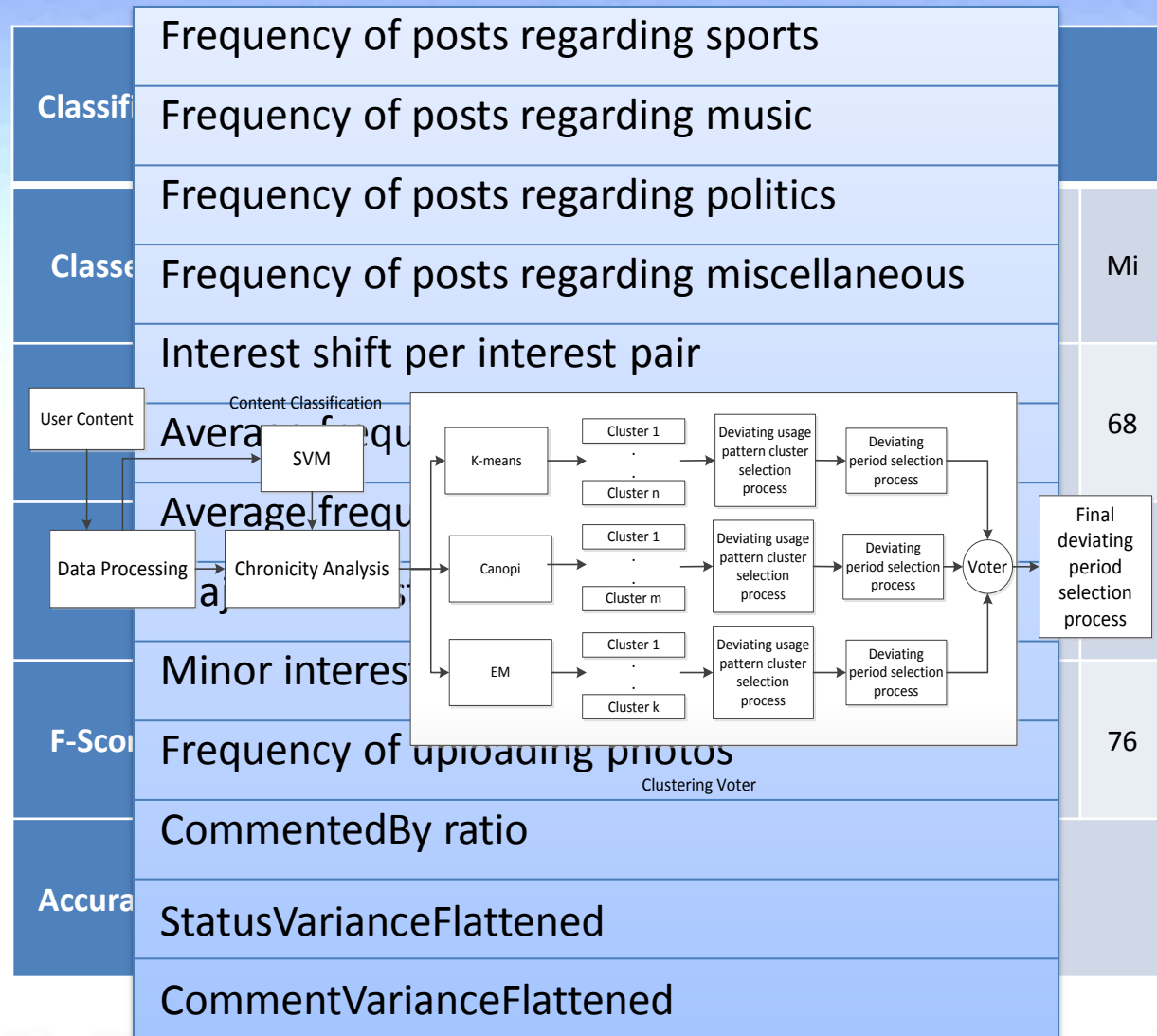
- Goal: **detect differentiations** of OSN usage patterns over **time related** to depicted stress level.
- Split users' usage pattern into time periods (from one day to one month).
 - Time period of one week produced best results.
- Chronicity analysis system consists of 2 modules:
 - Preprocessing data module (responsible for the processing of input data).
 - Usage pattern analysis module (responsible for analyzing usage patterns based on a set of metrics).
- Usage pattern fluctuations depict differentiated medium usage.

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 mis-cellaneous).

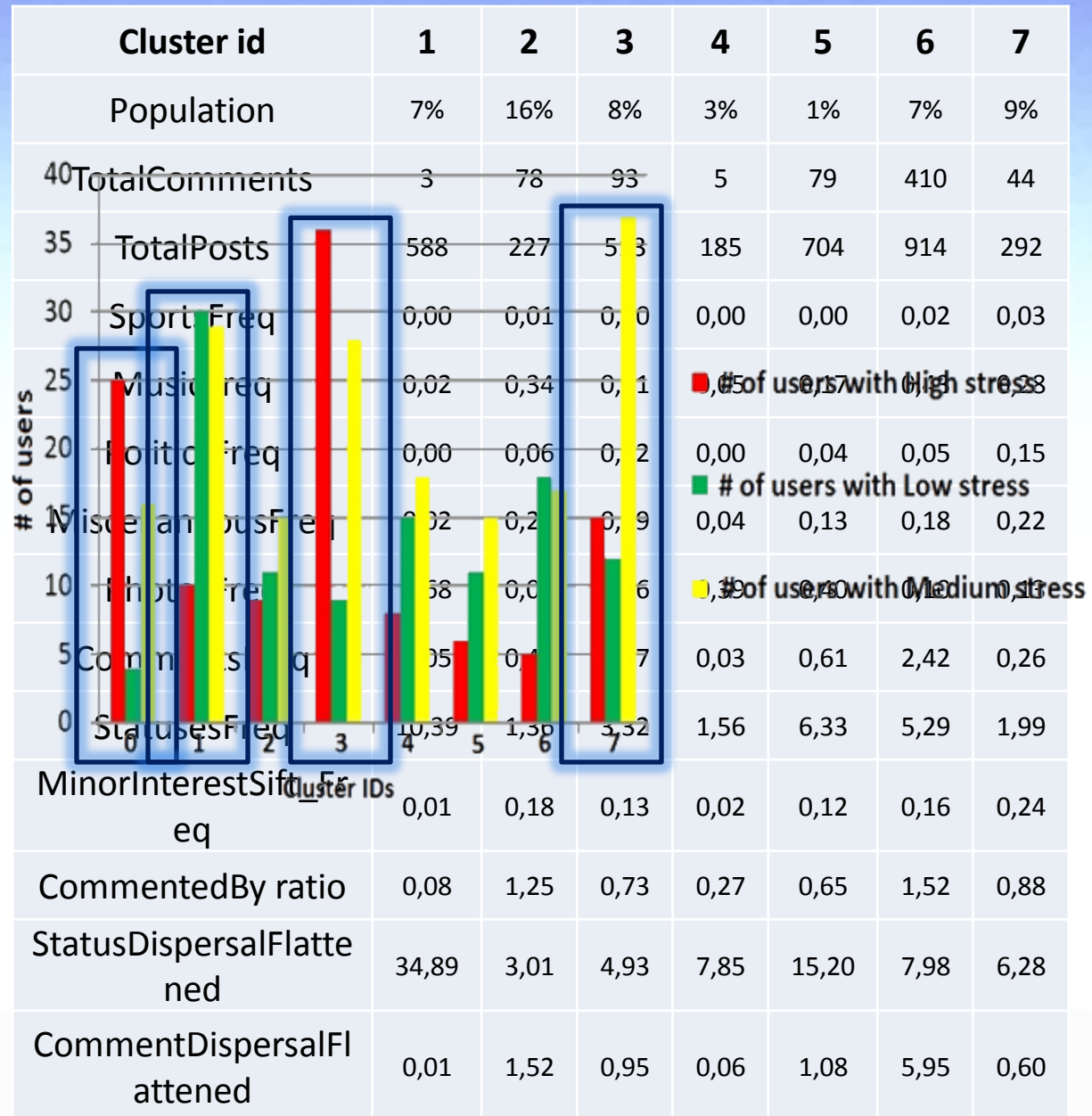
Step 2: Calculate following metrics for each user and time period (metrics developed on an ad-hoc basis according to our observations).

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.




Case 4

Scope: Identifying Political Beliefs



Horror
story

OSINT		OSN: YouTube 	
Tools used for the analysis			
Science		Theory	
Computing		Machine Learning	
		Data Mining	
Political Sociology			

In a nutshell



Identifying political beliefs



Researchers' compliance with ethical standards

YES



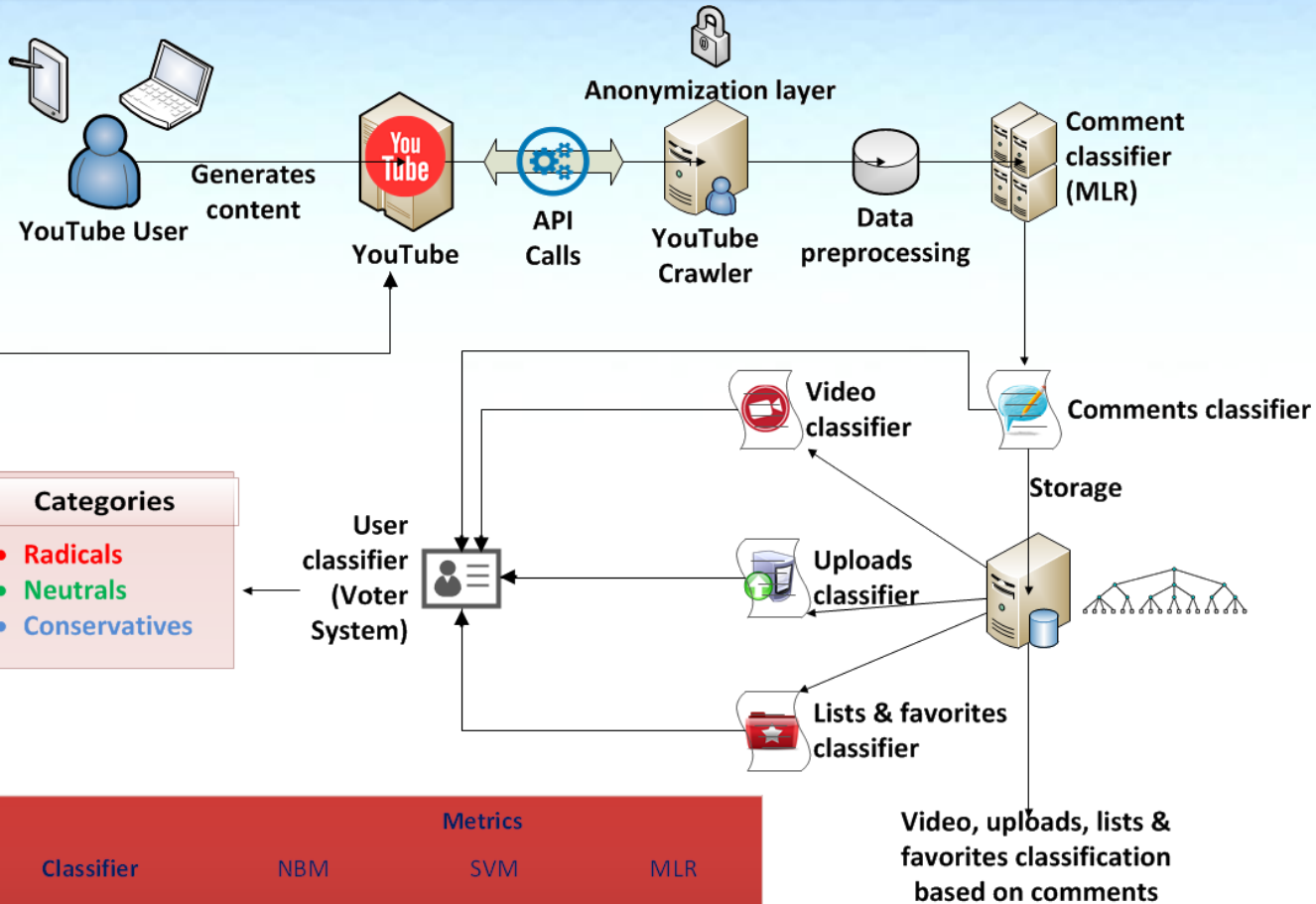
Legal Expert

YES

Highlight social threat
Raise community awareness



Information Security & Critical Infrastructure Protection Laboratory



Classifier	Metrics								
	NBM			SVM			MLR		
Classes	R	N	C	R	N	C	R	N	C
Precision	65	93	55	75	91	74	83	91	77
Recall	83	56	85	80	89	73	77	93	78
F-Score	73	70	60	76	89	73	80	92	77
Accuracy	68			84			87		

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

Case 4: Horror story – Identifying Political Beliefs



Divided loyalty

Study: Motive, ideology, divided/reduced loyalty, predisposition towards law enforcement

Means: Machine Learning, Content Analysis, comment classification

- Same YouTube dataset.
- Political beliefs profiling-clustering.
- Three (indicative, local context based) clusters: **R**adical – **N**eutral – **C**onservative.
- Machine Learning and Content Analysis methods used.
- Massive ethical issues.
- Goal: raise community awareness.



Horror story

Methodology

- Three (indicative) categories: **R**adical, **N**eutral, **C**onservative:
 - Assumptions are local-context-dependent (Greece, 2007-12).
 - Test case consists of an indicative subset of the local community.
 - Analysis reflects the current local political scene.
- Defined (indicative) classes:
 - **R**adical political affiliation: center-left, left, far-left.
 - **N**eutral political affiliation: neutral or non-specified political affiliation disclosed.
 - **C**onservative political affiliation: center-right, right, far-right.
- Com-ments classification:
 - Comments classification performed as text clas-si-fi-ca-tion.
 - Machine trai-n-ed with text examples and the cate-go-ry each one belongs to.
 - As-si-s-tance of field expert (Sociologist).

Analysis of results

- **Comment** classification by:
 - Naï-ve Bayes Multinomial (NBM)
 - Support Vector Machines (SVM)
 - Multinomial Logistic Re-gression (MLR)
- Each classifier's **efficiency** was compared by:
 - Metrics (%): Precision, Recall, F-Score, Ac-cu-ra--cy
- Multinomial Logistic Regression was chosen:
 - MLR classifies appropriately a comment with 87% accuracy.
 - Use of precision, recall and f-score to further examine classifiers' efficiency.

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 /lower recall means less/ more false negative classifi-cations, respectively.

F-Score: Weighted harmo-nic mean of both metrics.

Accuracy: No. of correct clas-si-fi-ca-tions performed by the classifier. Equals to the quotient of good classifica-tions by all the data.

Classifier	Metrics								
	NBM			SVM			MLR		
Classes	R	N	C	R	N	C	R	N	C
Precision	65%	93%	55%	75%	91%	74%	83%	91%	77%
Recall	83%	56%	85%	80%	89%	73%	77%	93%	78%
<u>F-Score</u>	73%	70%	60%	76%	89%	73%	80%	92%	77%
Accuracy	68%			84%			87%		

Basic observations

2% of **comments** demonstrate political affiliation (0.7% Radical, 1.3% Conservative)

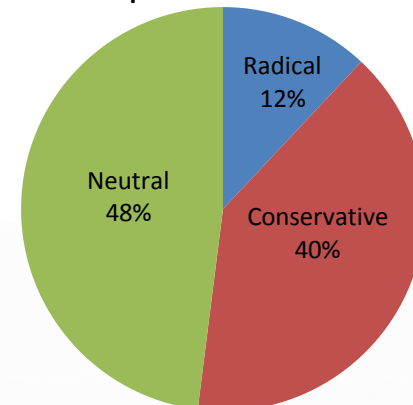
- 2% means that almost **41.000 comments** (of the 2.000.000 collected) include political content.

7% of **videos** classified into a specific category (2% Radical, 5% Conservative)

- 7% means that almost 14.000 videos (of the 200.000 collected) include political content.

12% of **users** ex-press **Radical** political affiliation and 40% **Con-ser--va-ti-ve** affiliation

- 52% means that **6.760 users** reveal - one way or another - their political beliefs.



Some general conclusions

- ✓ Web 2.0 produces vast amounts of **crawable** information and OSINT may transform this information into **intelligence**.
- ✓ OSINT can assist in detecting **narcissistic behavior, predisposition towards law enforcement, divided political loyalty**, etc.
- ✓ OSINT can be a proactive cyber-defense tool and **predict threats, predict delinquent behavior, assist in law enforcement and assess influence opportunities**.
- ✓ OSINT may lead to unwanted **horror stories**.
- ✓ OSINT intrusive nature dictates **limited** use, e.g. security officers selection, critical infrastructure protection, national security.

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-vcy, 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* Theoharidou M., Kotzanikolaou P., Gritzalis D., "Towards a Criticality Analysis Methodology: Redefining Risk Analysis for Critical Infrastructure Protection", in *Proc. of the 3rd IFIP International Conference on Critical Infrastructure Protection (CIP-2009)*, Springer, USA, 2009.