Which Side Are You On?
A New Panopticon vs. Privacy

Miltiadis Kandias, Lilian Mitrou, Vasilis Stavrou and Dimitris Gritzalis
Information Security & Critical Infrastructure Protection Research Laboratory,
Dept. of Informatics, Athens University of Economics & Business, 76 Patission Ave., GR-10434, Athens, Greece

Keywords: Awareness, Panopticon, Privacy, Social Media, Surveillance, User Profiling, YouTube.

Abstract: Social media and Web 2.0 have enabled internet users to contribute online content, which may be crawled and utilized for a variety of reasons, from personalized advertising to behaviour prediction/profiling. One negative case scenario is the political affiliation profiling. Our hypothesis is that this scenario is nowadays realistic, applicable to social media, and violates civil rights, privacy and freedom. To demonstrate this, we developed a horror story, i.e., a Panopticon method, in order to reveal this threat and contribute in raising the social awareness over it. The Panopticon relies on data/opinion mining techniques; hence it classifies comments, videos and playlists, collected from the popular social medium YouTube. Afterwards, it aggregates these classifications in order to decide over the users' political affiliation. The experimental test case of the Panopticon is an extensive Greek community of YouTube users. In order to demonstrate our case, we performed an extensive graph theoretical and content analysis of the collected dataset and show how and what kind of personal data (e.g. political attitude) can be derived via data mining on publicly available YouTube data. Then, we provide the reader with an analysis of the legal means that are available today, to a citizen or a society as a whole, so as to effectively be prevented from such a threat.

1 INTRODUCTION

The exponential growth of Information and Communication Technologies (ICT) and the rapid explosion of social media have contributed to substantive changes in the social dimensions of information sharing and (mis)use. However, several inherent features of Internet (and especially Web 2.0) supported technologies and platforms (e.g., digitization, availability, recordability and persistency of information, public or semi-public nature of profiles and messages, etc.) encourage not only new forms of interaction but also surveillance behaviours and tendencies (Tokunaga, 2011). ICT have often been accused for facilitating surveillance via CCTV or even the Internet (Brignall, 2002).

The common conception of surveillance is this of a hierarchical system of power between the observer and the observed, represented in metaphors, such as the "Panopticon" of J. Bentham, i.e. a theoretical prison structure, an "ideal" prison building designed in a way that allows observing of the prisoners from a central location at all times. The observed subject is never sure of whether or not she is under surveillance. Foucault, who elaborated extensively on the modern implications of the Panopticon, emphasized that the conscious and permanent visibility assures the automatic functioning of power (Foucault, 1975). The Panopticon creates "a consciousness of permanent visibility as a form of power, where no bars, chains and heavy locks are necessary for domination, anymore" (Almer, 2012).

Is the Internet surveillant in the way of a Panopticon? The metaphor of Panopticon offers perhaps the ultimate example of unilateral and vertical surveillance, while social networks and media indicate and incorporate the shift to interpersonal, horizontal, and mutual information aggregation and surveillance. However, despite the lack of centralized control over the Internet, its platforms and applications allow multilevel and latent surveillance, thus pose new risks for the rights of the individuals by forming new power relations and asymmetries. Surveillance and surveillors remain invisible: The technology hides both the possibility of surveillance and the signs of what/who is monitored (Uteck, 2009); (Fuchs, 2011), although persons living in future ubiquitous computing environments can - in antithesis to the

Kandias M., Mitrou L., Stavrou V. and Gritzalis D..
Which Side Are You On? - A New Panopticon vs. Privacy.
DOI: 10.5220/0004516500980110
In *Proceedings of the 10th International Conference on Security and Cryptography* (SECRYPT-2013), pages 98-110
ISBN: 978-989-8565-73-0

classical "Panopticon" - assume (or even accept, if not wish) that they will be monitored.

May Web 2.0 become an Omniopticon, in which "the many watch the many"? (Jurgenson, 2010). Is the "social" and "participatory network" (Beyer et al., 2008) the ideal "topos" for "social surveillance" (Tokunaga, 2011); (Marwick, 2012) and "participatory panopticism" (Whitaker, 1999) By being subject of communication and engaging in social networking activities the users are becoming objects of a lateral surveillance (Fuchs, 2011). In social media users monitor each other. Moreover, such sites and interaction platforms are, by design, destined for users to continually digital traces left by their "friends" or persons they interact with - often by simply consuming or commenting user-generated content.

Nowadays, several often rely on user generated content in social media, which is the most popular type of information transmitted through internet, as users are able to express themselves, inform others, republish/redistribute news and opinions of others, thus they form their personal identity in the digital world. All these activities produce user generated information flows. These flows may be - in fact have been - utilized for purposes ranging from profiling for targeted advertising (on the basis of analysing online features and behaviour of the users), to personality profiling and behaviour prediction.

Along with consumer and configuration based offerings, exploitation/use of user generated data has contributed to the shaping of the "Open Source Intelligence" (Gibson, 2004). This data may be used for both, the social good/in the interest of the society (e.g. Forensics), or in a way that infringes fundamental rights and liberties or private interests (e.g. social engineering, discriminations, cyber bullying, etc.) (Gritzalis, 2001); (Lambrinoudakis, 2003); (Marias, 2007); (Mitrou et al., 2003); (Spinellis et al., 1999).

The voluntary exposure of personal information to an indefinite audience gives rise both to the traditional and social panopticism (Nevrla, 2010). Surveillance of user generated content and information flows takes place between organisational entities and individuals and between individual users (Marwick, 2012). Governments' interest on information gained through data aggregation and mining of social media is easily understandable, as law enforcement and crime prevention may require "connecting the dots" and combining information about political beliefs and every-day activities.

However, such information aggregation and profiling of political beliefs and affiliations may result to a "nightmare" for a democratic State, especially in the case that such practices concern a large thus disproportional) number of citizens/netizens. Users' political profiling implicates the right to decisional and informational privacy and may have a chilling effect on the exercise of freedom of expression.

In order to prove our hypothesis we have developed a proof-of-concept panopticon and applied it on real-life data. We crawled the YouTube social medium and created a dataset that consists solely of Greek users, hence a Greek YouTube community. We examined the data (comments, uploads, playlists, favourites, and subscriptions) using text classification techniques via comments classification. The panopticon can predict the political affiliation of a video and then predict the political affiliation expressed in a list of videos. The method applies the above on users' comments, uploaded videos, favourite videos and playlists, so as to aggregate the results and extract a conclusion over the users' political affiliation.

We have decided to define three categories of broad political affiliations, namely Radical, Neutral and Conservative. These assumptions are context-dependent, given that our experimental test case is a real-life Greek community. Thus, in order to reflect the recent political/historical context in Greece, we define the following pairing: Radical political affiliation refers to centre-left, left, and far-left political beliefs, Neutral political affiliation refers to non-political content, whereas Conservative political affiliation refers to centre-right, right and far-right political beliefs. The definition of the above categories has no impact on the results of the analysis.

The paper is organized as follows: In section 2 we review the existing literature. In section 3 we describe the panopticon methodology and the testing environment. In section 4 we demonstrate when and how a political affiliation of a YouTube user can be revealed. In section 5 we provide a detailed statistical results evaluation and analysis. In section 6 we highlight the social threats that can emerge from a malevolent exploitation of a panopticon, along with an analysis of the legal means that are available today to a citizen or a society in order to avoid such a case. Finally, in Section 7 we conclude and refer to our plans for future work.

2 RELATED WORK

The advent of Web 2.0 has contributed in the transformation of the average user from a passive reader into a content contributor. Web 2.0 and social media have, in particular, become a valuable source of per-

sonal data, which are available for crawling and processing without the user's consent. The rise of social media usage has challenged and directed researchers towards opinion mining and sentiment analysis (Pang and Lee, 2008).

Opinion mining and sentiment analysis constitute computational techniques in social computing (King et al, 2009). As presented by King et al., social computing is a computing paradigm that involves multidisciplinary approach in analysing and modelling social behaviour on different media and platforms to produce intelligence and interactive platform results. One may collect and process the available data, so as to draw conclusions about a user mood (Choudhury and Counts, 2012). Choudhury and Counts present and explore ways that expressions of human moods can be measured, inferred and expressed from social media activity. As a result, user and usage profiling and conclusion extraction from content processing are, today, more feasible and valuable than ever.

Several methods have been utilized in order to process online data and materialize the above mentioned threat. These methods include user behaviour characterization in online social networks (Benevenuto et al., 2008), as well as analysis of the relationship between users' gratifications and offline political/civic participation (Park et al., 2009). Park et al. examine the aforementioned relationship through Facebook Groups and their research indicated the four needs for using Facebook groups. The analysis of the relationship between users' needs and civic and political participation indicated that informational uses were more correlated to civic and political action than to recreational uses.

Users often appear not to be aware of the fact that their data are being processed for various reasons, such as consumer behaviour analysis, personalized advertisement, opinion mining, user and usage profiling, etc. Automated user profiling (Balduzzi et al., 2010) and opinion mining may be used for malevolent purposes, in order to extract conclusions over a crowd of users. Balduzzi et al. utilized the ability of querying a social network for registered e-mail addresses in order to highlight it as a threat rather than a feature. Furthermore, they identified more than 1.2 million user profiles associated with the collected addresses in eight social networks, such as Facebook, MySpace and Twitter. They also proposed a number of mitigation techniques to protect the user's privacy. Such techniques include CAPTCHA, limiting information exposure, rate-limiting queries to prohibit automated crawling and raising user awareness. In order to raise awareness the attack was applied on a realistic environment, consisting of a social networks group. Graph theoretic analysis has, also, been utilized in order to examine narcissistic behaviour of Twitter users (Kandias et al., 2013).

Jakobsson et al., as well as Ratkiewicz (Jakobsson et al., 2008); (Jakobsson and Ratkiewicz, 2006), have also conducted research on a realistic social media environment regarding online fraud experiments, such as phishing, with respect to users' privacy. Such realistic approaches are proposed as a reliable way to estimate the success rate of an attack in the real-world and a means of raising user awareness over the potential threat.

3 METHODOLOGY

In this paper we have experimented with an extensive Greek community of YouTube. We present a panopticon political affiliation detection method, in order to raise user awareness over political profiling via social media. Furthermore, we present our findings related to political profiling as a proof-of-concept. The twofold purpose of this research is to (a) raise users' awareness over political profiling, and (b) highlight the social threat of processing users' online available data for discriminative purposes.

3.1 Data Crawling

In order to collect our dataset, we crawled YouTube using its REST-based API, which simplifies and accelerates the procedure. Each API request includes parameters, such as the number of results and the developer key. We have chosen to use a developer key, as the crawler is less likely to be flagged as malevolent for quota violations. Thus, we managed to send more API requests and accelerate the process of data collection. When quota violation emerges, YouTube rejects all API calls for 10 minutes to "reset" quota. Regarding the number of results, YouTube poses a limit of 1000 results/request. However, the limit turned out to be much lower (50 results/request); otherwise, YouTube API kept returning error codes. During the process of data crawling we collected only publicly available data and respected quote limitations posed by YouTube, so as not to cause even the mildest harm to YouTube's infrastructure.

Crawling was initiated by a set of a few Greek users. In order to crawl more users, we ran a breadth-first search on user subscribers and on the users who have commented on the crawled videos. During the data collection process a user was added to the crawling queue only if she had added a Greek location to her profile or had a profile description written

in Greek.

The gathered data were classified into three categories: (a) user-related information, e.g., her profile, uploaded videos, subscriptions, favorite videos, and playlists, (b) video-related information, e.g., video's license, the number of likes and dislikes it has received, its category and tags, and (c) comment-related information, e.g., the content of the comment and the number of likes and dislikes received. The collected data include: (a) 12.964 users, (b) 207.377 videos, and (c) 2.043.362 comments. The time span of the collected data covered 7 years (Nov. 2005 - Oct. 2012).

We added to the collected data an anonymisation layer. In specific, usernames have been replaced with MD5 hashes, so as to eliminate possible connections between collected data and real life users. Each user is processed as a hash value, so it hardly feasible for the results to be reversed. Thus, single real life users cannot be detected.

It is in principle possible, though, to reverse this process by using indirect means, such as searching for specific comments in search engines (e.g. Google hacking), or by utilizing Open Source Intelligence techniques.

3.2 Graph-theoretic Approach

The forms of interactions in YouTube are easily noticed. Each user can subscribe to other users, so as to receive notifications about the recently generated content. Furthermore, users are able to comment on videos. These types of interaction are considered relationships between users and represent ties in the network graph of the collected dataset. In this section we present a graph analysis, so as to identify characteristics of users' behaviour in YouTube.

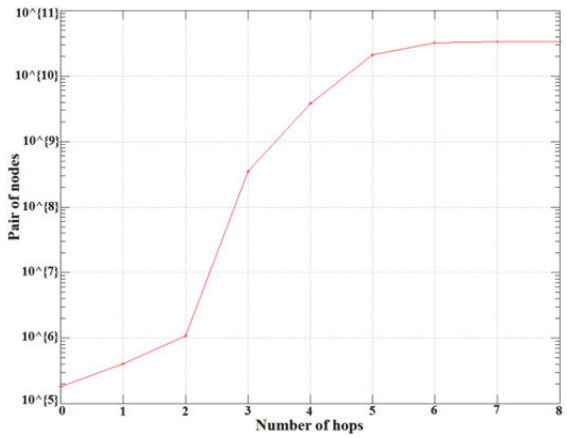

Figure 1: Small world phenomenon.

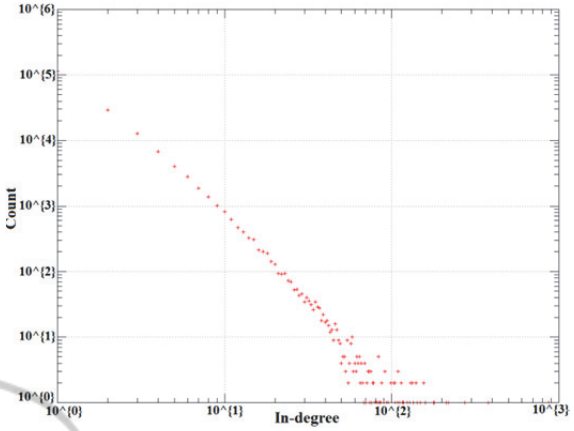

Figure 2: Indegree distribution.

The main conclusions from this analysis are:

(a). The *small world* phenomenon does apply to the collected Greek community. This is depicted in Fig. 1, where we calculated the effective diameter of the graph, i.e. every user of the community is 6 hops away from everyone else (Watts and Strogatz, 1998).

(b). A *small group of users have the most subscribers,* while the rest of the users have considerably fewer subscribers (Fig. 2). Most nodes (users) have a low number of ingoing ties. A small fraction of the nodes have a big number of ingoing ones. Also, a small number of nodes (users) have a lot of outgoing ties. The rest of the nodes have fewer outgoing ties (Fig. 3). Higher outdegree means more subscribers to the user, or intense comment activity on a video.

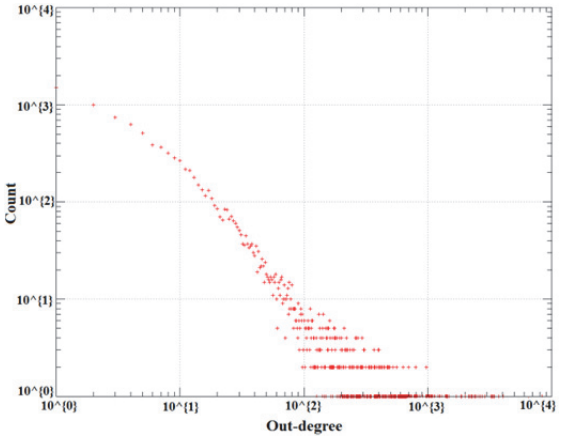

Figure 3: Outdegree distribution.

User indegree value indicates the number of users who subscribe to the user or comment to her uploaded videos, while outdegree value is the num-

ber of users to whom she subscribes or comments. Both indegree and outdegree distributions tend to have heavier tails (Costa et al., 2007), (Barabasi, 2005) for the biggest values of indegree and outdegree values, respectively. Fig. 2.shows the indegree distribution of the graph, while Fig. 3 shows the outdegree distribution.

(c). *Users join YouTube to participate*. Fig. 4 represents the group of nodes formed in the graph. There is one large and strongly connected component consisting of approximately 175.000 users, 2 small connected components with approximately 20 users, and 3.795 consisting of one user. Thus, most nodes in the graph have an outgoing tie to another node. Only a considerably small number of nodes have no outgoing ties and is inactive.

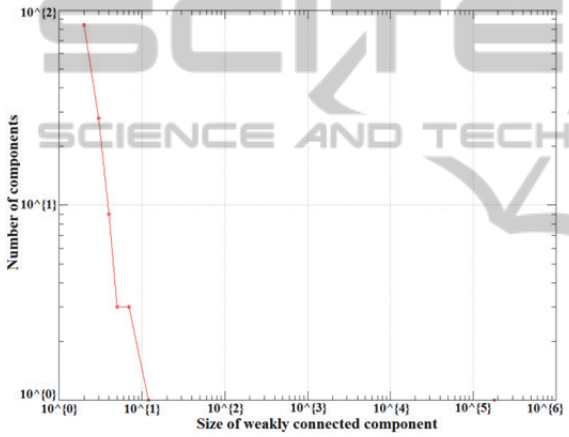

Figure 4: Group of nodes.

3.3 Tag Cloud Description

For better observing the axis of content of the collected data, we visualized the results in the form of a tag cloud. This is demonstrated in Fig. 5.

Figure 5: Tag cloud of the dataset.

Tags “Greece” and “greek” appear frequently in the dataset because the experimentation focuses on a Greek community of YouTube. The majority of the tag cloud tags are Greek words written in Latin (i.e. “greeklish). We have transformed the Greek tags to greeklish, in order to deal with duplicates of a word (one in Greek and one in greeklish).

The majority of videos are related to *music* and *entertainment*. The next topic that can be found on the collected YouTube video tags is *sports*. Several tags containing Greek sports teams’ names are also shown in the tag cloud. One may also notice political content in the tag cloud (i.e., tags with the names of the Greek major political parties).

3.4 Panopticon and Youtube

Our experimentation was carried out in a real environment (YouTube). We have followed this approach so as to offer a real-life proof-of-concept of a panopticon and contribute in the international debate over the issue. We exploited the results, so as to enhance user privacy and raise user awareness, without disrespecting users’ permissions of profile access.

According to official statistics depicted in Fig. 6 (www.youtube.com/yt/press/statistics.html) YouTube is a popular social medium. Furthermore, it grows exponentially along with user generated content that it hosts. YouTube is characterized by emotional-driven responses in its comments because of videos’ emotional content. Audio-visual stimuli, combined with the anonymity offered by usernames, appear to enable users to express their feelings and opinions, regarding the content of a video. Also, YouTube users are able to interact with each other through video comments or subscriptions. Even though it is not essential for the users to have formed a real life bond, they can interact on the basis of common interests, views, hobbies, or political affiliations.

Our observations indicate that users tend to participate in the medium and generate personalized content. YouTube videos and comments contain political characteristics as presented in the tag cloud. Thus, we formed the hypothesis that political affiliation may be extracted via content analysis. Based on our observations, we consider that:

(a) YouTube often contains political content.

(b) Users often feel free to express their opinions, especially when it comes to politics (because of the anonymity they assume and the emotional content of the medium).

(c) Most users join YouTube to participate, so one can reasonably expect that they will reveal, inter alia, their personal data.

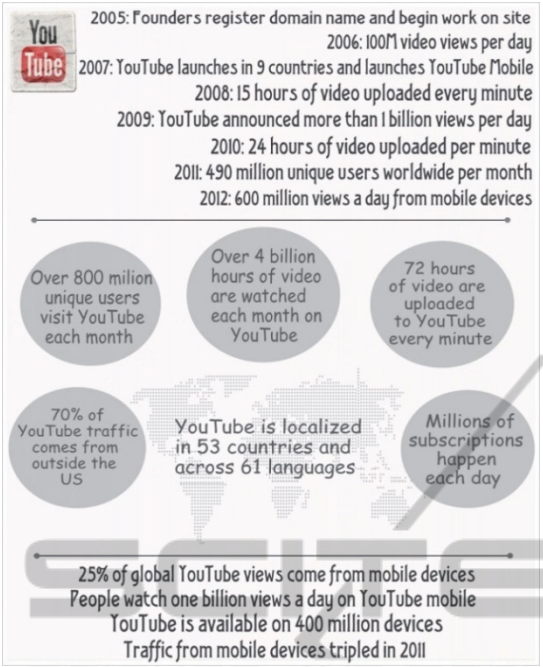

Figure 6: YouTube penetration.

3.5 Drawing Conclusions

We demonstrate that one can identify the political affiliation of a user via the political beliefs expressed within the comments of her videos. The reason why videos are examined is because video is YouTube's basic module. Since we cannot process the video itself, we draw a conclusion for the video through its comments. We detect the political affiliation expressed in a comment by performing text classification into three main categories: (a) category R, which contains expressions related to radical affiliation, (b) category C, which contains expressions related to conservative affiliation and (c) category N, which contains all the comments that hold a neutral political stance or have no political content.

Text classification uses machine learning techniques to classify a comment in the appropriate category. Assigning a comment into one of the categories is equivalent to the fact that the comment contains the respective political affiliation its category depicts. An alternative would be to create a vocabulary including words of each category and scan each comment to detect specific words. Machine learning leads to a more reliable result than a simple word existence check. Also, text classification performs better than scanning lists of words in a vocabulary.

Comment classification enables to extract conclusions for a video's political affiliation. The conclusion drawn helps us to classify any video into one of the defined categories of political affiliations. So, by assigning a comment into a category implies that the conclusion drawn for the comment is the political affiliation expressed in the category. The same applies to a list of videos, such as favorite videos and playlists. Having the category in which a video falls into, a conclusion can be drawn for the political affiliation expressed in the list. Being able to classify user's content, we may extract conclusions for user's comments, uploaded videos, favourite videos and playlists. This way we can draw a final conclusion about user's political affiliation.

4 PANOPTICON

We store the crawled data in a relational database for further analysis. The first step of the process is to train a classifier that will be used to classify comments into one of the three categories of political affiliation (radical, neutral, conservative). Comment classification is performed as text classification (Sebastiani, 2002), which uses machine learning techniques to train the system and decide in which category a text falls into. The machine is trained by having as input text examples and the category the examples belong to. Label assignment requires the assistance of an expert, who can distinguish and justify the categories each text belongs to.

We formed a training set so as to perform comment classification. By studying the collected comments, we noticed that a significant percentage of the comments are written in the Greek language. Another characteristic of the Greek YouTube community is that users write Greek words using Latin alphabet in their communication ("greeklish"). This is the dominant way of writing in Greek YouTube (the 51% of our dataset's comments are written in greeklish). Most users prefer to use greeklish, instead of Greek, because they do not care about correct spelling of their writings.

The appearance of those two different types of writing in comments has led us to pick two different approaches in comment classification, i.e., analyze them as two different languages. Another issue is due to the use of both greeklish and Greek. In order to mitigate this problem we have chosen to merge these training sets into one and train only one classifier. Forming Greek and greeklish training sets requires the selection of comments from the database and proper label assignment for each one of them, based on the category it belongs to. We consulted a domain expert (i.e., Sociologist), who could assign and justify the chosen labels on the training sets. Thus we

created a reliable classification mechanism. We chose 300 comments from each category (R,C,N) of the training set for each language. The expert contributed by assigning a category label to each comment.

Apart from the training set, we also created a testing set, which is required to evaluate the efficiency of the resulting classifier. The testing set contains pre-labeled data that are fed to the machine to check if the initial assigned label of each comment is equal to the one predicted by the machine. The testing set labels were also assigned by the domain expert.

We performed comment classification using: (a) Naïve Bayes Mutlinomial (McCallum and Nigam, 1998) (NBM), (b) Support Vector Machines (Joachims, 1998) (SVM), and (c) Multinomial Logistic Regression (Anderson, 1982) (MLR), so as to compare the results and pick the most efficient classifier. We compared each classifier's efficiency based on the metrics of precision, recall, f-measure and accuracy (Manning et al., 2008).

Accuracy measures the number of correct classifications performed by the classifier. Precision measures the classifier's exactness. Higher and lower precision means less and more false positive classifications (the comment is said to be related to the category incorrectly) respectively. Recall measures the classifier's completeness. Higher and lower recall means less and more false negative classifications (the comment is not assigned as related to a category, but it should be) respectively. Precision and recall are increased at the expense of each other. That's the reason why they are combined to produce f-score metric which is the weighted harmonic mean of both metrics.

Table 1 presents each classifier's efficiency, based on accuracy, precision, recall, and f-score metrics. Multinomial Logistic Regression and Support Vector Machines achieve the highest accuracy. The accuracy metric is high due to the dominant number of politically neutral comments. Precision and recall are proper metrics to evaluate each classifier (Manning et al., 2008).

Table 1: Metrics comparison of classification algorithms.

	Metrics								
Classifier	**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		

Multinomial Logistic Regression achieves better precision value and SVM better recall value. Multinomial Logistic Regression achieves a slightly better f-score assesment. Support Vector Machines and Multinomial Logistic Regression achieve similar results regarding both recall and precision metrics. As a result, we chose Multinomial Logistic Regression because of the better f-score value achieved for each one of the categories.

4.1 Video Classification

Regarding the extraction of the political affiliation expressed in a video, we studied each video, based on its comments, classified to one of the three categories. Also, we know the number of likes/dislikes each comment has received. Likes and dislikes represent the acceptability a comment has from the audience, so it may be an indication of the comment's importance to the overall video's result. Thus, a comment that receives a significant number of likes should be treated differently than a comment with no likes, as the first one is acknowledged as important by more users. This assumption has been confirmed by the data mining process. Subsequently, in order to extract a conclusion for the video, we take into consideration only comments that belong either to category R or C. Neutral comments are ignored.

Each comment importance is measured via its number of likes and dislikes. In order to come to a video overall result we utilize two sums, one for category R and one for C. For every comment that belongs to categories R or C we add the following quantity to the respective aggregation:

$$1 + \{(likes/total\_likes) - (dislikes/total\_dislikes)\}$$

The quantity added to each sum shows that a comment that has received more *likes* than *dislikes* should affect the overall score more than a comment with more *dislikes* than *likes*. Finally, the category with the larger sum is the category that represents video's political affiliation.Table 2 illustrates the procedure described above.

Table 2: Example of video classification decision.

	Video "Example"		
Comment	**Political affiliation**	**Likes**	**Dislikes**
#1	R	90	10
#2	C	15	20
#3	R	30	5
#4	N	5	2
#5	R	10	3
Total		150	40

Sum R equals to: $(1 + 90/150 - 10/40) + (1 + 30/150 - 5/40) + (1 + 10/150 - 3/40) = 4.1$, whereas Sum C equals to $(1 + 15/150 - 20/40) =$

0.6. Sum R≥ C, so video "Example" is classified to category R and expresses radical political affiliation.

4.2 List Classification

The procedure followed to extract a conclusion about a list of videos is similar to the above mentioned video method. The only difference is that we utilize videos instead of comments. The two sums are also applied, one for category R and one for C. In this case, instead of *likes* and *dislikes* we used the video ones. In the end, the category with the greater sum is the result for the list's political affiliation. This procedure is applied to the "favourite videos" list, as well as to the other playlists that the user may have created.

4.3 User Classification

A user political affiliation can be identified based on the category she is assigned to. The procedure, as shown in Fig. 7, takes into account a user's comments, her uploaded videos, her favourite videos, and her playlists. A user is able to: (a) write a comment to express her feelings or her opinion, (b) upload a video (the content may have a distinctive meaning for her), (c) add a video to her favourites list (it may have an emotional or intellectual meaning for her), and (d) create a playlist and add videos to it.

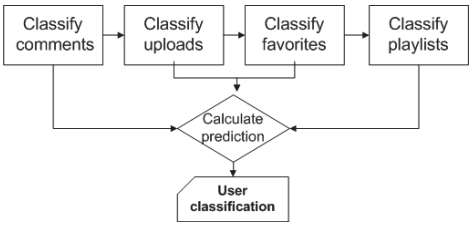

Figure 7: User classification process.

Based on these observations one may look for indications of political beliefs within the user generated content. For the needs of our experimentation, we have defined ad-hoc weights for each of the cases we examine (Table 3). Each phase of the process generates a result on the category each user belongs to. In comment classification, the result is based on the number of political comments that exhibit the highest aggregation. If the comments classified in category R are more than those classified in category C, then the result is that user tends to be Radical. The results on uploaded videos, favourite videos, and playlists are extracted as described in the list result extraction process. User comments are the most important factor to decide of a user's political affiliation (Table 3).

Table 3: Ad-hoc weights of each element.

	Comment	Upload	Favourite	Playlist
Weight	3	2	2	1

Regarding the aggregation of the results, we utilize two sums, one for category R and the other for C. Comments, videos, and lists classified as neutral do not contribute to the aggregation. The sub-results are appropriately weighted and added to the final sums, in order to extract the final result. An example of this procedure appears in Table 4.

Table 4: User classification example.

	User "Example"	
	Political beliefs	Weight
Comments	R	3
Uploaded videos	N	2
Favorite videos	R	2
Playlists	C	1

Sum R equals to 3 + 2 = 5, while Sum C equals to 1. Sum R ≥ Sum C, which implies that the user belongs to category R. A user is classified to categories R, or C, if there is at least one political comment or political video detected to her content. A user may not express a political stance via her comments or uploaded videos; however, she may have added a video with political content to her playlists. The result for the user will be that she belongs either to category R or C, depending on the political thesis expressed in the video. The weights are defined on an ad-hoc basis. The weight of each result could be better determined after a meta-training process.

5 STATISTICAL ANALYSIS

Based on the analysis, the 2% of the collected comments exhibit a clear political affiliation (0.7% was classified as R, while 1.3% was classified as C). On the contrary, 7% of the videos were classified to one of these categories, 2% as R and 5% as C. On the other hand, 50% of the users have been found to clearly express, at least once, their political affiliation. Out of them, the 12% of the dataset has been found to express Radical affiliation and 40% Conservative.

Regarding users classified as radicals, we found that - on average - the 20% of their comments has a political position expressed. Also, they tend to prefer the Greek alphabet (i.e., 54% of their comments are

written in Greek, 33% in greeklish, and 13% use both Greek and Latin alphabet).

On the other hand, users classified as conservatives tend to prefer the greeklish way of expression, namely 55% of the comments are in greeklish, 35% written using the Greek alphabet and 10% use both Greek and Latin alphabet. In table 5 the average number of characters that a comment consists of is depicted.

Comments written in greeklish tend to be shorter and more aggressive. On the contrary, comments written in Greek tend to be larger, more explanatory, and polite. Another finding is that the more aggressive a comment, the more misspelled it is.

Table 5: Average number of characters in a comment.

Alphabet	Average no. of characters (R)	Average no. of characters (C)
Greek	294	179
Greeklish	245	344
Both	329	227

Regarding the license assigned to each video (Typical YouTube or Creative Commons), the 7% of the videos are published under the Creative Commons license. A 55% of these videos were uploaded by users classified as Radicals, 10% by Conservatives and 35% by Neutrals.

Radicals tend to massively comment on the same videos. We found that these videos have unequivocal political content, namely political events, music, incidents of police brutality, etc. Moreover, the videos that radicals tend to add to their favourites are mainly documentaries and political music clips. Conservatives tend to share mainly conspiracy-based videos, as well ones with a nationalistic content.

6 PANOPTICON AND THE LAW

The use of this kind of methods may result to problems that are actually inherent in every kind of profiling. In brief, these methods may be regarded as a kind of (behavioural) profiling on the Internet, in the meaning of collecting data (recording, storing, tracking) and searching it for identifying patterns (Castelluccia et al., 2011). Such profiling methods interfere with the right to informational privacy and are associated with discrimination risks.

The observation of the behaviour and characteristics of individuals through mining of large quantities of data may infringe fundamental rights, let alone the determination of correlation between characteristics and patterns and the respective classification of individuals. A major threat for privacy rights derives from the fact that profiling methods can generate sensitive information "out of seemingly trivial and/or even anonymous data" (Hildebrandt, 2009).

By studying user's uploads it is possible to extract information related to the content, especially when it refers to areas such as political affiliation. Furthermore, a user is possible to have a private profile, however her comments could be collected from crawling random videos. Thus, a limited profile can be build based on those comments. The predominant rationales for acquiring knowledge about the political opinions and the relative sentiments seems to be either (political) research purposes or the goal of reducing risks both in the private and the public sector. However, personal data that are, by their nature, particularly sensitive and vulnerable to abuse, deserve specific protection.

Collecting and processing data about political beliefs is regarded by law as a highly exceptional situation. Many international and national laws prohibit explicitly the processing of personal data revealing political opinions (e.g. Art. 8 of the European Data Protection Directive and Art. 6 of the Convention 108 of the Council of Europe). Derogating from the prohibition on processing this "sensitive category" of data is allowed if done by a law that lays down the specific purposes and subject to suitable safeguards. Such derogations rely on a manifest public interest or the explicit, informed and written consent of the person concerned.

However, in European data protection law derogation is sometimes allowed also in the cases that "the processing relates to data which are manifestly made public by the data subject" (Art. 8, §2e of the European Data Protection Directive), which is the case if people generate content or comment on other users' content in social networks or media using their real identity and aiming at expressing their opinions publicly. According to the American theory and jurisprudence there is no "reasonable expectation of privacy if data is voluntarily revealed to others" (Solove, 2006). It is "apparent", according to this theory, that one cannot retain a reasonable expectation of privacy in the case of YouTube, videos, likes, and comments left open to the public (Henderson, 2012).

By generating content in social media users are generating information flows and aggregations. Providers and Online Social Networks encourage - also through the default settings - "producers" (Bruns , 2006) to publish personal information and enable anyone accessing this information thus actively contributing to shaping social media as an attractive

product (Ziegele and Quiring, 2011). Does self-exposure in social media amount to freely and consciously chosen privacy abandonment?

YouTube offers several privacy options to users. These privacy options encompass: (a) creation of private channel/profile, which disables access to user's channel where her profile is available, (b) creation of private videos, which enables users to share them with a limited number of viewers (up to 50 persons) after inviting them, (c) creation of private video lists, which applies to favourite videos and playlists and disable a playlist from being publicly available, and (d) potentially disclose user's activity, e.g., comments, subscriptions or favourite videos. These options may protect user's actions from being tracked in order to get information about the video she likes, comments on, or the users she is subscribed to. However, we should take into consideration individual's general inertia toward default terms (Mitrou, 2009); (Mitrou, 2003); (Lambrinoudakis et al., 2003). Moreover, it seems that the majority of users choose to disclose their personal data to as many users as possible, although average users do not have a clear idea about the actual reach of information they reveal or they underestimate the possible reach of their profiles visibility (Ziegele and Quiring, 2011).

Users are losing control over their data and the use thereof, as they are becoming detectable and "correlatable". The combination of all this information provides a powerful tool for the accurate profiling of users. Moreover, it is quite simple to identify a particular person, even after her key attributes (name, affiliation, address) have been removed, based on her web history (Castelluccia et al., 2011).

However, even if individuals are profiled in a pseudonimised way they may be adversely influenced (Schermer, 2011); (Spinellis, 1999). Informational privacy protects individuals against practices that erode individual freedom, their capacity for self-determination, and their autonomy to engage in relationships and foster social appearance. If individuals fear that information pertaining to them might lead to false incrimination, reprisals or manipulation of their data, they would probably hesitate to engage in communication and participatory activities (Mitrou, 2010). The autonomy fostered by informational privacy generates collective benefits because it promotes "reasoned participation in the governance of the community" (Cohen, 2000).

Risks of misuse and errors arising out of the aggregation and data mining of a large amount of data made public for other purposes are manifest. According to the German Federal Constitutional Court the "cataloguing" of the personality through the connection of personal data for the purpose of creating profiles and patterns is not permitted (Judgment of the Bundesverfassungsgericht, 4 April 2006, 1 BvR 518/ 02, 23.05.2006). A mass profiling of persons on the base of their views expressed in social media could have intimidation effects with further impacts on their behaviour, the conception of their identity and the exercise of fundamental rights and freedoms such as the freedom of speech (Cas, 2011). Fear of discrimination and prejudice may result to self-censorship and self-oppression (Fazekas, 2004). Indeed, while profiling risks are usually conceived as threats to informational privacy we should point out the - eventually more - significant and actual risk of discrimination (Gutwirth and Hert, 2008). The safeguards relating to the use of personal information aim - among others. if not principally - at preventing discrimination against persons because of their opinions, beliefs, health or social status. Studies conveyed how profiling and the widespread collection and aggregation of personal information increase social injustice and generate even further discrimination against political or ethnical minorities or traditionally disadvantaged groups (Mitrou, 2010).

Individuals may be confronted with major problems both in their workplace and in their social environment. Employers or rigid micro-societies could demonstrate marginalizing behaviour against persons because of their deviating political affiliation. There are a lot of historical examples of people who have been side-lined by the hegemonic attitude of society. One should not look for numerous examples in order to evaluate this thesis: Victor Hugo's "Les Miserables" (Section X: The Bishop in the presence of an unknown light) is the most representative evidence towards this result.

If we may generalize the above mentioned consideration to a macro environment, consequences to the deviating from the average political affiliation could lead to mass social exclusion, prejudice and discriminations. Such minorities may even be considered de facto delinquent and face a social stigma. In the context of a totalitarian/authoritarian regime, implementation of such methods could lead to massive violation of civil and human rights or even threat the life of specific individuals.

7 CONCLUSIONS AND RESEARCH

In this paper we dealt with the possibility of a social

threat that is based on user generated content exploitation and leads to political affiliation profiling; namely a new panopticon of the digital era. Political beliefs and affiliation have been a cause for social marginalization, prejudice, and discrimination, especially in totalitarian and authoritarian regimes. Thus, we bring this issue to the fore and contribute to the debate and awareness raising. A user might want to protect personal information other than political affiliation, namely information related to sexual orientation, racial discrimination or even the health condition of the user regardless of the national scope. Such an improper information disclosure could be easily conducted via expantion of the Panopticon methodology on the condition that domain experts of each case are available to interp-ret the collected data and train an appropriate model.

In order to prove and highlight the above mentioned we developed a panopticon methodology that is able to materialize this threat. During our research we collected a number of 12.964 users, 207.377 videos and 2.043.362 comments from YouTube. Afterwards, we conducted content and graph theoretic analysis of the dataset in order to verify that it is possible to extract conclusions over users' political affiliation. Our results confirmed the initial hypothesis that YouTube is a social medium that can support the study of users' political affiliation, namely audio-visual stimuli along with the feeling of anonymity enables users to express their political beliefs, even the most extreme ones.

The panopticon needs the contribution of a field specialist in order to assign category labels (Radical, Neutral or Conservative) to each comment of the training set. Then, a machine is trained in classifying YouTube comments to these categories. We experimented with three algorithms, i.e., Naïve Bayes Multinomial, Support Vector Machines, and Multinomial Logistic Regression. Comparison of each classifier's efficiency was based on the metrics of precision, recall, f-score and accuracy and indicated that the MLR algorithm is the most appropriate because of the better f-score value achieved for each of the categories. F-score is a combination of recall and precision metrics. Classifying comments to these categories enables the panopticon to classify playlists, lists of favourites and uploads, thus it manages to classify users to the above mentioned categories.

Furthermore, we carried out a series of statistics regarding our dataset. In specific we quoted characteristics and demographics of the radical and conservative users that we located in our data. Alongside with these characteristics, we highlighted possible consequences of an alleged implementation of the described panopticon method. Regardless of the scope of the implementation, the resulting threats include working place discriminations, social prejudice or even stigma and marginalization of the victims. These phenomena could be identified even in a democratic and stable society, not to mention the threats one could face in a military or totalitarian regime. Thus, we adopted a pro-privacy attitude and included a legal point of view in our analysis, along with the emergence of the demand for raising social awareness over this threat and the necessity for institutionalization of digital rights.

For future work we plan on further studying the panopticon and recognize more aspects of this social threat. We intend on spreading our research on other social media and study the phenomenon under the prism of different tools and methodologies along with optimization of our weight factors. Finally, we plan on proposing optimized methods for online privacy and anonymity with particular interest on the field of social media.

ACKNOWLEDGEMENTS

Authors would like to thank N. Bosovic for his help in the process of data mining, K. Galbogini for her contribution in the graph theoretic analysis of the dataset and V. Rongas for his help with Sociology and political affiliation analysis.

This work was supported in part by the Solo (56 NEW_B_2012) project, funded by the Hellenic General Secretariat for Research & Technology.

REFERENCES

Allmer, T., 2012. *Towards a Critical Theory of Surveillance in Informational Capitalism*, Frankfurt am Main: P. Lang.

Anderson, J., 1982. Logistic regression, In *Handbook of Statistics*, pp. 169-191, North-Holland, USA.

Balduzzi, M., Platzer, C., Holz, T., Kirda, E., Balzarotti, D., Kruegel, C., 2010. Abusing social networks for automated user profiling, In *Recent Advances in Intrusion Detection*, pp. 422-441. Springer..

Barabasi, A., 2005. The origin of bursts and heavy tails in human dynamics, In *Nature*, vol. 435, no. 7039, pp. 207-211.

Benevenuto, F., Rodrigues, T., Cha, M., Almeida, V., 2009. Characterizing user behaviour in online social networks, In *Proc. of the 9th ACM SIGCOMM Conference on Internet Measurement*, pp. 49-62, ACM.

Beyer A., Kirchner M., Kreuzberger G., Schmeling J., 2008. Privacy im Social Web - Zum kompetenten Um-

gang mit personlichen Daten im Web 2, *Datenschutz und Datensicherung (DuD*) 9/2008, pp. 597-600

Brignall, T., 2002. The new panopticon: The internet viewed as a structure of social control, In *Theory and Science,* vol. 3, no. 1, pp. 335–348.

Bruns, A., 2006. Towards produsage: Futures for user-led content production, In *Proc. of Cultural Attitudes towards Communication and Technology*, pp. 275-284.

Cas, I., 2011. Ubiquitous Computing, Privacy and Data Protection: Options and limitations to reconcile the unprecedented contradictions, In *Computers, Privacy and Data Protection: An Element of Choice*, pp. 139-170, Springer.

Castelluccia, C., Druschel, P., Hübner, S., Pasic, A., Preneel, B., Tschofenig, H., 2011. Privacy, Accountability and Trust-Challenges and Opportunities, *Technical Report*, ENISA.

Cohen, J., 2000. Examined Lives: Informational Privacy and the subject as object, In *Stanford Law Review*, vol. 52, pp. 1373-1438.

Costa, L., Rodrigues, F., Travieso, G., Boas, P., 2007. Characterization of complex networks: A survey of measurements, In *Advances in Physics*, vol. 56, no. 1, pp. 167-242.

De Choudhury, M., Counts, S., 2012. The nature of emotional expression in social media: measurement, inference and utility, In *2012 Human Computer Interaction Consortium* (HCIC) Workshop.

Fazekas, C., 2004. 1984 is Still Fiction: Electronic Monitoring in the Workplace and US Privacy Law, In *Duke Law & Technology Review*, pp. 15-25.

Foucault, M., 1975. Surveiller et punir: Naissance de la prison, Paris: Gallimard. Alan Sheridan (Trans.), Discipline and punish: The birth of the prison: Penguin.

Fuchs, C., 2011.New Media, Web 2.0 and Surveillance. In *Sociology Compass* 5/2 (2011), pp. 134–147

Gibson, S., 2004. Open source intelligence, In *The RUSI Journal*, vol. 149, no. 1, pp. 16-22.

Gritzalis, D., 2001. A digital seal solution for deploying trust on commercial transactions, In *Information Management & Computer Security Journal*, vol. 9, no. 2, pp. 71-79, MCB University Press.

Gutwirth, S., De Hert, P., 2008. Regulating profiling in a democratic constitutional State, In *Profiling the European citizen cross-disciplinary perspectives*, pp. 271-302, Springer.

Henderson, S., 2012. Expectations of Privacy in Social Media, In *Mississippi College L. Rev.*, pp. 31 (Symposium Edition)). Available at:
http://works.bepress.com/stephen_henderson/10

Hildebrandt, M., 2009. Who is profiling who? Invisible visibility, In *Reinventing Data Protection*, pp. 239-252.

Jakobsson, M., Finn, P., Johnson, N., 2008. Why and how to perform fraud experiments, In IEEE *Security & Privacy*, vol. 6, no. 2, pp. 66-68.

Jakobsson, M., Ratkiewicz, J., 2006. Designing ethical phishing experiments: a study of (ROT13) rOnl query features, In *Proc. of the 15th International Conference on World Wide Web*, pp. 513-522, ACM.

Joachims, T., 1998. Text categorization with support vector machines: Learning with many relevant features, In *Machine learning: ECML-98*, pp. 137-142.

Jurgenson, N., 2010. Review of Ondi Timoner's We Live in Public. *Surveillance & Society* 8(3), pp. 374-378. http://www.surveillance-and-society.org

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.

King, I., Li, J., Chan, K., 2009. A brief survey of computational approaches in social computing, In *International Joint Conference on Neural Networks*, pp. 1625-1632.

Lambrinoudakis, C., Gritzalis, D., Tsoumas V., Karyda, M., Ikonomopoulos, S., 2003. Secure Electronic Voting: The current landscape, In *Secure Electronic Voting*, Gritzalis, D. (Ed.), pp. 101-122, Springer.

Manning, C., Raghavan, P., Schütze, H., 2008. *Introduction to Information Retrieval*, Cambridge University Press.

Marias, J., Dritsas, S., Theoharidou, M., Mallios, J. Gritzalis, D., 2007. SIP vulnerabilities and antispit mechanisms assessment, In *Proc. of the 16th IEEE International Conference on Computer Communications and Networks*, pp. 597-604, IEEE Press.

Marvwick, A., 2012. The Public Domain: Social Surveillance in Everyday Life. In *Surveillance & Society* 9 (4). 2012, pp. 378-393

McCallum, A., Nigam, K., 1998. A comparison of event models for naive Bayes text classification, In *Workshop on learning for text categorization*, vol. 752, pp. 41-48.

Mitrou, L., Gritzalis, D., Katsikas, S., Quirchmayr, G., 2003. Electronic voting: Constitutional and legal requirements, and their technical implications, In *Secure Electronic Voting*, pp. 43-60, Springer.

Mitrou, L., 2009. The Commodification of the Individual in the Internet Era: Informational self-determination or "self-alienation", In *Proc. of the 8th International Conference of Computer Ethics Philosophical Enquiry* (CEPE '09), pp. 466-485.

Mitrou, L., 2010. The impact of communications data retention on fundamental rights and democracy: The case of the EU Data Retention Directive, In Haggerty/-Samatas, pp. 127-147.

Nevrla, J. 2010, Voluntary Surveillance: Privacy, Identity and the Rise of Social Panopticism in the Twenty-First Century, In *Commentary - The UNH Journal of Communication Special Issue*, 2009-10, pp. 5-13

Pang, B., Lee, L., 2008. Opinion mining and sentiment analysis, in *Foundations and Trends in Information Retrieval*, vol. 2, nos. 1-2, pp. 1–135.

Park, N., Kee, K., Valenzuela, S., 2009. Being immersed in social networking environment: Facebook groups, uses and gratifications, and social outcomes, In *CyberPsychology & Behaviour,* vol. 12, no. 6, pp. 729-733.

Schermer, B., 2011. The limits of privacy in automated profiling and data mining, In *Computer Law and Security Review*, vol. 27, pp. 45-52.

Sebastiani, F., 2002. Machine learning in automated text categorization, In *ACM Computing Surveys*, vol. 34, no. 1, pp. 1-47.

Solove, D., 2006. A taxonomy of privacy, In *University of Pennsylvania Law Review*, vol. 154, no. 3, pp. 477.

Spinellis, D., Gritzalis, S., Iliadis, J., Gritzalis, D., Katsikas, S., 1999. Trusted Third Party services for deploying secure telemedical applications over the web, In *Computers & Security*, vol. 18, no. 7, pp. 627-639.

Tokunaga, R. 2011, Social networking site or social surveillance site? Understanding the use of interpersonal electronic surveillance in romantic relationships. In *Computers in Human Behaviour* 27, pp 705-713.

Uteck, A., 2009, Ubiquitous Computing and Spatial Privacy, In Kerr I., et al (eds.), *Lessons from the identity trail: Anonymity, privacy and identity in a networked society*, Oxford University Press 2009, pp. 83-102.

Watts, D., Strogatz, S., 1998. The small world problem, In *Collective Dynamics of Small-World Networks,* vol. 393, pp. 440-442.

Whitaker, R., 1999. *The End of Privacy: How Total Surveillance In Becoming a Reality*. New Press, 1999.

Ziegele, M., Quiring, O., 2011. Privacy in social network sites, In *Privacy Online: Perspectives on privacy and self-disclosure in the Social Web*, pp. 175-189, Springer.