# SOCIAL MEDIA PROFILING: A PANOPTICON OR OMNIOPTICON TOOL?

Lilian Mitrou[1]; Miltiadis Kandias[1]; Vasilis Stavrou[1]; Dimitris Gritzalis[1]

## ABSTRACT

Online Social Networks and Media indicate and incorporate the shift to interpersonal, horizontal, and mutual communication and, thus, information aggregation. Online content (namely YouTube videos and comments or Tweets) along with online relations (friendships, followings, mentions, comments etc.) may be crawled and utilized for a variety of purposes including user/usage profiling and behavior prediction. In our previous research we have demonstrated that it is possible and even potentially trivial (by utilizing a simple personal computer and a broadband internet connection) to extract sensitive, personal information such as political beliefs, psychosocial characteristics (narcissism and predisposition towards law enforcement) etc. about social media users in an automated manner via data crawling, data aggregation, machine learning and graph and content analysis of the collected dataset of YouTube and Twitter Open Source Intelligence. Web 2.0 technological features combined with voluntary exposure to an indefinite audience in social media give rise to traditional surveillance as Government is enabled to connect the dots, combine information about political beliefs and every-day activities and generate mass user profiles on the base of identifying patterns. Despite the lack of centralized control over the Internet, its platforms and applications allow multilevel and latent surveillance, thus pose new risks for individuals by forming new power relations and asymmetries. Our research highlights how Web 2.0 and social media (YouTube and Twitter) may become a *topos* of participatory *panopticism*, an *omniopticon* in which the many watch the many, and can reconstruct sensitive information out of seemingly anonymous data/content. Individuals may be confronted with social exclusion, prejudice and discrimination risks both in their workplace and in their social environment. In our paper we focus on the results of this type of surveillance which facilitates the exculpation of such penetrating and privacy-violating technologies and amplifies the threshold of societal tolerance towards a panopticon-like state of surveillance.

## INTRODUCTION

Web 2.0 technologies have enabled the interaction among users all over the world, the ability to create, redistribute or exchange information and opinions and also the participation in virtual communities, in which they are encouraged to express themselves. With the emergence of Web 2.0 there is a definite change in the role of “average user” who has become a content contributor. All these means of interaction, along with the exchange of user-generated content, are referred as Social Media [1]. The environment of social media is dominated by user-generated content, by information produced, received, disseminated by “non-professionals", in particular, but non only, through SNS. The paradigm of Internet had changed: from static, isolated repositories of information shifted to dynamic, user-driven and participatory sites. Users are now able to interact with other people, create, redistribute or exchange information and opinions, and also express themselves in virtual communities.

[1] Information Security & Critical Infrastructure Protection Laboratory, Dept. of Informatics, Athens University of Economics & Business, Athens, Greece {l.mitrou, kandiasm, stavrouv, dgrit}@aueb.gr

Web 2.0 interactivity and applications allow or enable users to create online profiles, to share personal details and preferences and thus produce persistent, replicable and searchable information [2]. Social media became informal but all-embracing identity management tools, defining access to user-created content via social relationships [3]. Users reveal aspects of their personality and behavior in social media. By revealing their lives users engage in the self-construction of identity [4]. Many users form their personal identity in and through social media.

Users are encouraged by the nature and the structure of social media to reveal information and produce or share content, resulting to a “participatory culture” [5], which supports ordinary users to express their views and disseminate them. Web 2.0 communication is undoubtedly a mass communication but not necessarily “a mass-self communication” as characterized by Castells [6], as the definition of potential receivers is not always “self-directed”. Due to the nature of online social networking, users are exposed to a mass audience, even when they obviously communicate online with a specific audience in mind. When used to convey personal information or preferences, social media augment their users’ visibility, not only to their chosen “friends” but also to other persons (as “friends” of “friends”), agencies and institutions. Social media facilitate information exchange between these entities [7].

Generating content in social media users are generating information flows. Social networks and media indicate and incorporate the shift to interpersonal, horizontal and mutual communication and offer the ability to increase information aggregation. Being “subjects in communication” users make their data available to others, thus becoming “objects of information” and therefore “objects of surveillance” [8]. Several inherent features of Internet (especially Web 2.0) supported technologies and platforms (e.g., digitization, availability, record-ability and persistency of information, public or semi-public nature of profiles and messages) encourage not only new forms of interaction, but also novel surveillance tendencies via behavior and sentiment detection and prediction. Any content uploaded online can be easily accessed, recorded, stored and retrieved, “without respect for social norms of distribution and appropriateness” [9-10].

Such sites and interaction platforms are, by design, destined for users to continually follow digital traces left by their “friends”, “followers” or persons they interact with often by simply consuming or commenting user-generated content. While users’ content visibility increases, the architecture of the majority of Web 2.0 applications allows exposing content to unwanted and/or invisible audiences. Moreover, if “Panopticon” creates a “consciousness of permanent visibility”, Internet technology hides both the possibility of surveillance and the signs of what/who is monitored, although persons living in ubiquitous computing environments and acting in social media contexts could assume (or even accept, if not wish) that they will be monitored by everyone.

If the metaphor of “Panopticon” offers perhaps the ultimate example of unilateral and vertical surveillance, Internet 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. By engaging in social networking activities, the users are becoming objects of a multilateral and constant surveillance.

The conscious or unconscious voluntary exposure of personal information to an indefinite audience gives rise both to the "traditional" and social panopticism [11]. Does Web 2.0 become slightly but definitely an Omniopticon, in which "the many watch the many" [12], an ideal "topos" for "social surveillance" [13-14] and "participatory panopticism" [15]?

## SOCIAL MEDIA PROFILING

User majority participates in social media for various reasons, such as amusement, communication and professional networking. As the number of users participating social media started to increase, research has focused on examining the way they behave in the digital world. One of the most important findings is individuals tend to transfer their offline behavior online [16].

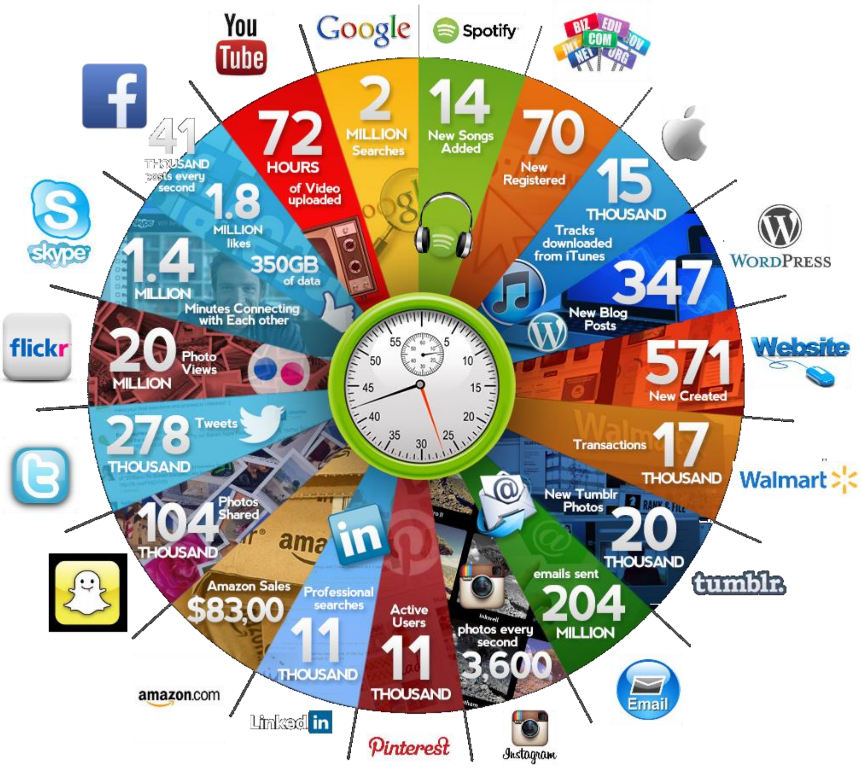

**Fig. 1** What happens in Social Media in 60 seconds (http://socialmediatoday.com/)

As it gets obvious: (a) Web 2.0 contains vast amounts of user-generated content and (b) the growth of ICT has enabled the faster and more efficient processing of such amounts of information using Open Source INTelligence techniques (OSINT). Using such techniques, one may observe that the construction of usage and users patterns is more feasible than ever. OSINT [17] refers to intelligence collected from publicly available sources, such as websites, web-based communities (i.e. social networks, forums, blogs) and publicly available data. Such techniques facilitate the extraction of knowledge that is not easily accessible.

An interesting observation is that, direct access to the target-user's social media data, is not required, since they may be accessible indirectly by crawling either the medium or other users that communicate with the target user. Users' data/User generated content in social media is publicly available, since most of them neglect to use any privacy mechanisms, even if offered, or they are not even aware of or wiling for. Even when users do not fail to utilize the offered privacy mechanisms, it is possible to obtain their data indirectly using the aforementioned ways. Consequently, anyone interested in collecting such data is in position/able to gather personal(ly) identifiable information and analyze it. OSINT processes rely on vast amounts of reliable information. The more reliable the data, the more accurate and intrusive the analysis. It is because of this that the following key points must be taken into consideration, since the quality of the results of the surveillance is directly connected to the quality of the gathered data:

- **Uncovering data location**: It is required to have knowledge of the locations from where the appropriate data can be gathered.
- **Sources preprocessing**: The preprocessing of the useful and the irrelevant sources of information is important, so as to avoid collecting outdated or useless data.
- **Results refining**: After having generated conclusions over the subject of interest, it could be useful to further process the results in order to focus on the required knowledge and provide us with further analysis of the subject's parameters. A process of meta-training on the collected data could reveal secret connections or correlations between the parameters of the dataset.

Web 2.0 and social media offer a valuable source of personal data, which are available for crawling and processing, even, without the user's consent (and knowledge). Although the processing of such data may be used for fair purposes, so as to provide the user with a more personalized user experience, user's privacy may be easily infringed. The knowledge extracted using OSINT may be 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 for purposes such as employee profiling and screening, counter intelligence, protection of critical infrastructures, insider threat prediction or even forensics analysis [18-21].

The behaviour and personality profiling may uncover and analyse psychosocial traits such as introversion, social and personal frustrations, divided loyalty, entitlement/narcissism, and predisposition towards law enforcement, political beliefs and group dynamics. Such traits can be processed in an automated and flexible manner, making it possible to perform psychometric evaluations by utilizing the content a user has made publicly available. Social media have offered a unique opportunity to examine user behavior via them, as in the past, one should have utilized stiff questionnaires and psychometric evaluations, in order to examine the above mentioned personal traits.

In our previous research [22-24], we have highlighted ways that multifaceted information shared/revealed in social media can be utilized, to *predict the behavior of the employees via examining traits of predisposition of delinquent behavior and thus augment existing surveillance methods* [25]. We have extracted conclusions over the users regarding the personality traits of narcissism (via Twitter) and predisposing towards law enforcement (via YouTube), which are common characteristic among insider threats. As we have proved it is further possible to extract the aforementioned conclusions, along with political profiling results, group dynamics analysis and introversion, so as to facilitate the human resources department of an organization and the screening process of new employees.

To highlight the feasibility of implementing the aforementioned cases, we run our crawlers and experiments on limited computing sources. Namely, we used a PC with a Core-2Duo processor (3GHz) and 4GB of RAM. Thus an individual willing to develop profiling algorithms is able to achieve it by solely using her personal computer. It is to be noted that data is gathered from only publicly available sources. No further access is required, from any social medium per se, in order to access private data.

We have applied our methodologies to two popular Social Media, so as to highlight the applicability and the success rate of such techniques. We, also, exploited the available data in Social Media in order to examine the ethical and legal issues that arise from such profiling methods. We have chosen YouTube and Twitter to apply our methodology, as Social Media and collaborative environments are popular among the users and have been used in the battle against the insider threat [26-27]. Also, social media can be used as a means of offering a real-life proof-of-concept of the methodologies applied.

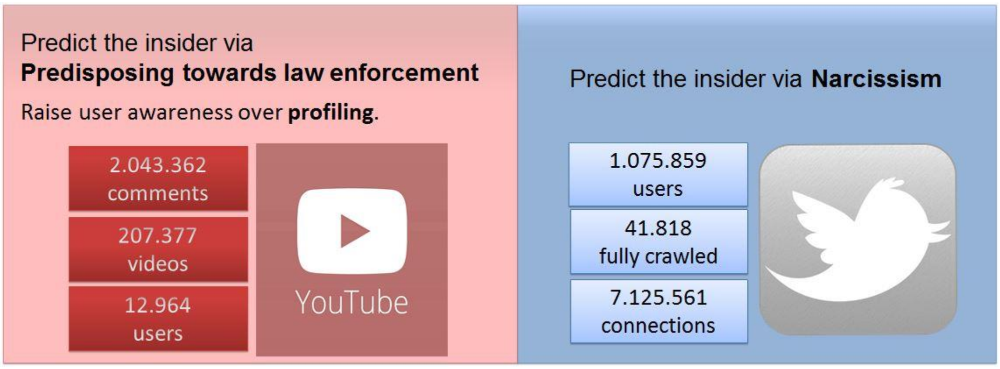

**Fig. 2** Testing environments

## TWITTER: USER PERSONALITY DETECTION THROUGH METADATA ANALYSIS

The methodology applied to Twitter aims at enhancing the prediction front of insider threat. To this end, we focused on a Greek community of Twitter, consisting of 1.075.859 users of which 41.818 are fully crawled, and 7.125.561 connections among them. We used data crawled by Twitter, so as to analyze the collected users in a graph theoretic manner. Additionally, we identified connections between usage patterns and defined which users' behavior could be considered as deviating from the average. The use of such method interferes with the personality and privacy rights of the affected persons. The collected data are used for prediction and deterrence purposes, since we analyze each user under the prism of usage deviation with the tool of graph theory. The results that refer to the Greek Twitter Community indicated that: (a) the majority of the Greek users make very poor use of the medium, (b) there are a lot of normally active users, and (c) very few users are popular.

In order to classify the users to certain taxonomy we utilized three parameters, the user's influence valuation, her klout score and her usage intensity. User's influence is defined as the set of users who are possible candidates to adopt her words by retweeting them. The influential set of each user consists of: (a) followers who directly learn her quotes, (b) her mentioners, (c) re-tweeters who mention her or repeat her word of mouth, even without following her, and (d) the followers of her last two categories, as there is a possibility to learn about her indirectly. The *klout score* is a metric that represents someone's social media influence. Finally, usage intensity is the aggregation of: (a) number of followers, (b) number of followings, (c) number of tweets, (d) number of retweets, (e) number of mentions, (f) number of favorites, and (g) number of lists.

Based on the available data, we spotted a threshold above which the users may become quite influential and perform intense medium usage. Therefore, we defined a specific point where a user turns from a normal one to a "media persona". Research has proved that individuals tend to transfer their offline behavior online. Thus, more extravert individuals tend to form large groups and communicate easier in the territory of social media, while introvert individuals tend to communicate less [16]. Furthermore, research work has connected excessive usage of social media to the personality trait of narcissism [28-30]. Thus, we proposed a general taxonomy of the Twitter users, the data of whom were crawled and analyzed.

According to our findings, the most influential users' influence valuation is between 942 and 3604 and usage valuation is between 21.004 and 569.000. Based on this, users whose sum of the previous values is higher of the threshold of 22.000 are classified in a different category of the taxonomy. Furthermore, the majority of the users with usage valuation above 21.000 are either real life celebrities or news media. This leads us assume that the "normal" users with high scores should belong to a different category. The proposed categories appear on Table 1.

**Table 1** The proposed Twitter user taxonomy

| Category | Influence valuation | Klout score | Usage valuation |
|---|---|---|---|
| Loners | 0 - 90 | 3.55 - 11.07 | 0 - 500 |
| Individuals | 90 - 283 | 11.07 - 26.0 | 500 - 4500 |
| Known users | 283 - 1011 | 26.0 - 50.0 | 4500 - 21000 |
| News Media & Personas | 1011 - 3604 | 50.0 - 81.99 | 21000 - 569000 |

## YOUTUBE: BEHAVIOUR ANALYSIS AND POLITICAL CLASSIFICATION

YouTube can be used as both a story of horror and curiosity. The horror story involves political detection affiliation using the online available data from YouTube and the story of curiosity the detection of potential threats. To this end, we have experimented with an extensive Greek community of YouTube. We have gathered a dataset of 12.964 fully crawled users, 207.377 videos, and 2.043.362 comments.

For a better understanding of the medium we visualized the axis of content of the collected data in the form of a tag cloud (see Fig. 2). Tags "Greece" and "greek" appear frequently in the dataset because the experimentation focuses on a Greek community of YouTube. 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).

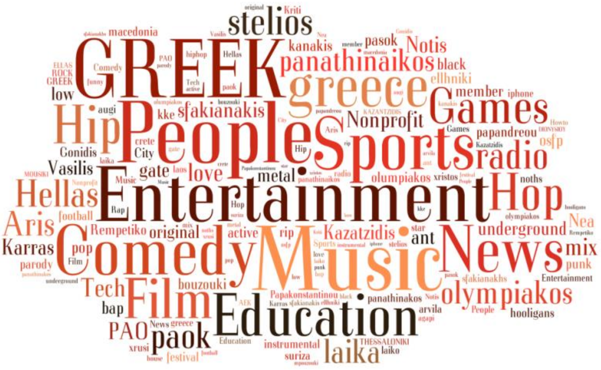

**Fig. 3** Tag cloud of the dataset

To draw conclusions over the *political affiliation* of the users we facilitated machine learning techniques to categorize user-generated content. Machine learning algorithms "learn" from the text examples they receive and construct underlying models that are able to determine the label of any text given as input. Label assignment requires the assistance of an expert, who can distinguish and justify the categories each text belongs to. We performed comment classification using: (a) Naïve Bayes Multinomial [31] (NBM), (b) Support Vector Machines [32] (SVM), and (c) Multinomial Logistic Regression [33] (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 [34]. 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 is why they are combined to produce f-score metric that is the weighted harmonic mean of both metrics.

Thus, we were able to categorize comments into one of the predefined categories, i.e. radicals, conservatives and non-political content according to specific metrics (Table 2). Multinomial Logistic Regression achieves better precision value and SVM better recall value. Multinomial Logistic Regression achieves a slightly better f-score assessment. 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.

**Table 2** 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 | | |

The same methodologies were applied to the dataset, in order to extract conclusions and classify users into predefined categories with respect to another psychosocial trait (i.e., predisposition towards law enforcement). User generated content is analyzed by the same algorithms and users are classified into two categories, those negatively predisposed towards law enforcement and those who are not (Table 3).

**Table 3** Metrics comparison of classification algorithms

| | | | Metrics | | | |
|---|---|---|---|---|---|---|
| **Classifier** | **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 | |

## PREDICTIVE PROFILING OF SOCIAL MEDIA USERS

Information and content flows may be - in fact, have been - utilized for purposes ranging from profiling for targeted advertising (on the basis of analyzing online features and behavior of the users), to personality profiling and behavior prediction. Analyzing user-generated content may be proved useful for personalization, profile management or detecting malicious or even deviating behavior of the user. 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 or the predisposition of the user in relation to law enforcement and authorities.

The examination of datasets collected from YouTube and Twitter indicates that the use of methods like the above mentioned, Panopticon may result to risks that are actually inherent in every kind of profiling. Under the term of profiling we understand methods involving mining of data and automated classification that is likely to assign individuals to particular categories mostly in order to take decisions concerning or affecting them. The respective definition adopted by the Council of Europe Recommendation [35] focuses on the creation and/or use of profiles to evaluate, analyze or predict personal aspects such as performance at work, economic situation, health, personal preferences, or interests, reliability or behavior, location or movements.

The predictability of individual attributes and behavior raises ethical and legal issues. Profiling ends up at treating a person as belonging to a specific category, which in turn indicates what "sort of person" someone is, the category becoming more important than the individual herself. Predictive data mining and profiling (e.g., flagging someone as potential insider threat, offender, inaccurate employee or loan payer) results to classifications that may have considerable implications for an individual's well-being, freedom and other interests and rights [36]. In a micro-social level, data mining of social media content may lead to extended discriminations and prejudice against persons and groups.

A visible risk to consider is the possibility for discrimination in the workplace: Online social media profiles, blogs, tweets, and online fora are increasingly monitored by employers searching for information that may provide insight on employees and prospective hires. Taking into consideration the exponentially growing participation in online social networking sites and social media, it is not surprising that employers are searching for unique information about applicants and employees not found with other selection methods. A broader and potentially less censored or more honest array of information is easily accessible on the Internet [37].

Such findings indicate that the once clear lines between the private and the public, as well as the employee's personal and professional life, are gradually blurring as a result of: (a) the "boundary-crossing technologies" [38], (b) the transformation of workplace structure and ethos through ICT, and (c) the radical changes in communication. The openness and sharing culture that dominates the online social media reflects a population that does not construct communication on the traditional division between private and public contexts. In other words, the more private information has become easily accessible and infinitely shareable and transferable, the more monitoring may extent to private spaces, activities, and time [39-40]. Methods (such the proposed in this paper) allow employers to collect and aggregate information, which reflects behavior of the user and her interaction with other users, in order to produce relevant patterns/profiles and anticipate future behavior and threats.

The security enhancement perspective, against the malevolent insider or other threats, includes the prediction of malevolent behavior that may cause catastrophic results, especially when the security of critical infrastructures is jeopardized. Furthermore, exploitation of user generated content in social media may provide the employer with useful information about the psychological state of the users, traits of their personality or even group analysis in order to maximize the organization's productivity and the group's efficiency. In this context many argue that when one publishes something to all comers it is not reasonable to expect (current and future) employers to respect her privacy and freedom of expression and refrain from judging her based on publicly available information [38]. In particular in the US, recent surveys point out lifestyle concerns among the most common reasons for rejecting candidates [41].

Employers are, actually not prohibited to consider information about a person who is documented in publicly available social media profiles, public posts, or public Twitter accounts. However, both the wide availability of private information, as well as its use beyond the initial context that this information has been produced, may have far reaching effects for the employees' rights and liberties. As Nissenbaum argues [42], the definitive value to be protected by the right to privacy is exactly the "contextual integrity" of a given contextual-self having different behaviors and sharing different information depending on the context. Information gathered through social media analysis is normally not only unintended as application information but often job-irrelevant or, moreover, related to sensitive activities and, consequently, information of the person concerned (religion, political beliefs, etc. [43]). User's data processing, without her consent, could lead to exposure of personal data of hers which could further lead to discriminations, conclusion extraction over sensitive personal characteristics and other inwardly characteristics of her behavior, sentiments and personality and affect her privacy, personality, and dignity.

## SOCIAL NETWORKS DATA MINING AND STATE SURVEILLANCE

Social media and online services with user-generated content (such as Twitter or YouTube, in the case of our research) have made a staggering amount of information available to the government. Governments seek to use this informational goldmine to extract implicit, previously unknown and potentially useful information and to discover or infer previously unknown facts, patterns and correlations [44].

Data mining of social media crawled data may have - obviously - reaching implications when we refer to the surveillance context. Web 2.0 technological features combined with voluntary exposure to an indefinite audience in social media give rise to traditional surveillance as Government is enabled to "connect the dots", combine information and generate mass user profiles on the base of identifying patterns. The privacy/surveillance relevant characteristic/application of user-generated content mining is the determination of correlation between characteristics and patterns and the respective classification of individuals.

The surveillance potential has to be considered also in view of Big Data technologies, which augment(s)/increase(s) knowledge discovery. What makes Big Data technologies and applications, surveillance relevant is not the size as such but exactly the possibility to aggregate and correlate distinct and hidden (massive) data sets. There is no doubt that Governments make extensive use of Big Data for surveillance purposes, such as the case of US Government's PRISM program that involves the US NSA collecting and analyzing foreign communications collected from a range of sources, including social media companies [45].

The observation of the behavior and characteristics of individuals through mining of large quantities of data may infringe fundamental rights, let alone the determination of correlation between every-day activities and political beliefs, between characteristics and patterns and the respective classification of individuals. The risk to stigmatize groups of persons or persons as part of a group life is especially high. As noted by the German Federal Constitutional Court (Rasterfahnung Urteil of the Bundesverfassungsgericht, 04.04.2006, 1 BvR 518/ 02, 23.05.2006), data profiling means a higher risk of becoming the target of further (official) criminal investigations and suffering stigmatization in public life [46].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 [47].

Collecting and processing data about political beliefs is regarded by law as a highly exceptional situation. Many international and national legal instruments 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 should rely on a manifest public interest or the explicit, informed and written consent of the person concerned.

Many agencies ground the legitimacy of collecting and analyzing information gained through data mining methods by pointing to the fact that such data are manifestly made public by the person concerned (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. This argumentation reflects the mainstream theory and jurisprudence in the US, where there is no "reasonable expectation of privacy if data is voluntarily revealed to others" [48] and is evoked to justify large-scale data mining even without ensuring compliance with substantial and procedural safeguards and requirements (concrete and serious suspicion, authorization, warrant, etc.).

Further, information aggregated and discovered through social media mining results from a "fairly opaque process", which raises serious questions about its conformity with the rule of law. Especially the so called pattern-based data mining, when the government investigator develops a model of assumptions about the activities and underlying characteristics of culpable individuals or the indicators of criminal behavior, raises serious concerns [49-50].The 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 also high and obvious. Moreover, 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" [51]. It is quite simple to identify a particular person, even after her key attributes (name, affiliation, address, etc.) have been removed, based on her web history. Anonymity is harder and harder to ensure, as it becomes easier to combine and analyze so-called anonymous or anonymized data to identify (or re-identify) individuals.

## SOCIAL MEDIA: A TOPOS OF PARTICIPATORY PANOPTICISM

Despite the lack of centralized control over the Internet, its platforms and applications allow multilevel and latent surveillance, thus pose new risks for individuals by forming new power relations and asymmetries. As Fuchs argues, on Web 2.0 corporate and state power is exercised through the gathering, combination and assessment of personal data, thus power relations and relationships of communication becoming interlinked [8].

Web 2.0 architecture is encouraging users' actual involvement in social media and empowering a mutual "sharing practice" [4]. Both the Web 2.0 architecture and the users' attitude create a situation that may be characterized as panoptic in design. Social connections, comments, views and preferences (expressed through "likes" or "retweet") are turning into visible, measurable, searchable and correlatable content. This voluntary exposure and engagement with others is regarded as the tool to the practice of "participatory" or "interpersonal" surveillance.

Every social media user can be equally observer and observed, controller and controlled. Moreover, in both cases we have examined (YouTube, Twitter), we have demonstrated that user profiling, along with detection of behavioral patterns, can be achieved using solely limited computer sources and publicly available user data that is easily accessible through the mechanisms provided by the social media per se. Social media users' profiling may become an every-day routine, what Andrejevic has conceived as "peer-to-peer monitoring", understood as the use of surveillance tools by individuals, rather than by agents of institutions public or private, to keep track of one another [52]. In fact, Web 2.0 and social media (YouTube and Twitter) are becoming the *topos* of "participatory panopticism", an Omniopticon in which not only the State and private entities but also "the many" watch "the many". Despite the widespread "exposure" tendencies (a kind of digital "exhibitionism"), which characterizes the social media it seems that users don't share the "total transparency" model [53]. It appears more probable that they do have a false perception of remaining anonymous or being among "friends" and they lack a "perception of audience", i.e. they don't realize that they may be "watched" and analyzed by so many.

Being aware of mass profiling capabilities of persons on the base of their views expressed in social media could have intimidation effects with further impacts on their behavior, the conception of their identity and the exercise of fundamental rights and freedoms such as the freedom of speech [46]. Profiling may indeed entail the risk of formatting and customization of individual behavior that affects her personal autonomy [54]. Extending monitoring to social communication relationships of employees and candidates augments the chances of employers to influence behavior and promote the "well-adjusted employee" [55-56]. Individuals tailor their social identities and aim at controlling others' impressions and opinions of them through behavior and performances within particular audiences [38]. Information gathering about employee performances outside the traditionally conceived work sphere not only increases the dependence on (future) employers but has also a chilling effect on individuals' personality and freedom of expression.

Being deprived from informational privacy, i.e. the capacity to control of the information concerning them, users and thus the capacity for autonomous decision - and choice-making and to maintain a variety of social identities and roles, social media users may feel fear of discrimination and prejudice, which may result to self-censorship and self-oppression. They may sacrifice "Internet participation to segregate their multiple life performance" [41].

In the context of the relation between government and citizens, the fear of widespread and strong data mining and profiling capabilities of the government may affecting individual autonomy by reducing the citizen's willingness to engage in political activities and participate to public sphere and discourse. For this fear to materialize, profiling does not even have to be effective [47].

## ACKNOWLEDGEMENT

This work has been partially funded by the Framework Programme 7 of the European Commission, under project "Public Perception of Security and Privacy: Assessing knowledge, collecting evidence, translating research into action" (PACT).

## REFERENCES

1. Kaplan, A., Haenlein, M.: Users of the world, unite! The challenges and opportunities of Social Media. In: *Business Horizons*, vol. 53, no. 1, pp. 59--68 (2010).
2. Boyd, D.: Social network sites: Public, private, or what. *Knowledge Tree*, vol. *13, no.* 1, pp. 1-7. (2007).
3. Piskopani, A., Mitrou, L.: Facebook: Reconstructing Communication and deconstructing privacy law. In: *4th Mediterranean Conference on Information Systems*, Greece. (2009).
4. Albrechtshund, A.: Online Social Networking as Participatory Surveillance. In: *First Monday* (13). March 2008 (2008).
5. van Dijck, J.: Users like you? Theorizing agency in user-generated content. In: *Media, Culture & Society*, vol. 31, no. 1, pp. 41–58. (2009).
6. Castells, M.: *Communication Power*, Oxford University Press. (2009).
7. Marvwick, A.: The Public Domain: Social Surveillance in Everyday Life. In *Surveillance & Society*, vol. 9, no. 4, pp. 378-393. (2012).
8. Fuchs, C.: New media, Web. 2.0 and Surveillance. In: *Sociology Compass*, vol. 5, no. 2, pp. 134-147. (2011).

9. Farinosi, M.: Beyond the Panopticon Framework: Privacy, Control and User Generated Content. In: A. Esposito et. al. (eds), COST 2011, pp. 180-189. (2011).
10. Lange, P., *Publicly private and privately public: Social networking on YouTube*. (2007).
11. Nevrla, J.: Voluntary Surveillance: Privacy, Identity and the Rise of Social Panopticism in the Twenty-First Century. In: *Commentary - The UNH Journal of Communication Special Issue*, pp. 5-13. (2010).
12. Jurgenson, N.: Review of Ondi Timoner's We Live in Public. In: *Surveillance & Society*, vol. 8, no. 3, pp. 374-378. (2010).
13. Tokunaga, R.: Social networking site or social surveillance site? Understanding the use of interpersonal elect-ronic surveillance in romantic relationships. In: *Computers in Human Behavior*, vol. 27, pp. 705-713. (2011)
14. Marvwick, A.: The Public Domain: Social Surveillance in Everyday Life. In: *Surveillance & Society*, vol. 9, no. 4, pp. 378-393. (2012).
15. Whitaker, R.: *The End of Privacy: How Total Surveillance In Becoming a Reality*. New Press. (1999).
16. Amichai-Hamburger, Y., Vinitzky, G.: Social network use and personality. In: *Computers in Human Behavior*, vol. 26, pp. 1289--1295. (2010).
17. Steele, R.: Open source intelligence. In: *Handbook of Intelligence Studies*, p. 129. (2007).
18. Kandias, M., Mylonas, A., Virvilis, N., Theoharidou, M., Gritzalis, D.: An Insider Threat Prediction Model. In: *7th International Conference on Trust, Privacy and Security in Digital Business*, pp. 26--37, Springer. (2010).
19. Mylonas, A., Kastania, A., Gritzalis, D., Delegate the smartphone user? Security awareness in smartphone platforms. In: *Computers & Security*, vol. 34, pp. 47--66. (2013).
20. Mylonas, A., Meletiadis, V., Mitrou, L., Gritzalis, D.: Smartphone sensor data as digital evidence. In: *Computers & Security*, vol. 38, pp. 51--75. (2013).
21. Shaw, E., Ruby, K., Post, J.: The insider threat to information systems: The psychology of the dangerous insider. In: *Security Awareness Bulletin*, vol. 2, no. 98, pp. 1--10. (1998).
22. Kandias, M., Galbogini, K., Mitrou, L., Gritzalis, D.: Insiders trapped in the mirror reveal themselves in social media. In: *7th International Conference on Network and System Security*, pp. 220--235. Springer. (2013).
23. Kandias, M., Stavrou, V., Bozovic, N., Mitrou, L., Gritzalis, D.: Can we trust this user? Predicting insider's attitude via YouTube usage profiling. In: *10th IEEE International Conference on Autonomic and Trusted Computing*, pp. 347--354, IEEE Press. (2013).
24. Gritzalis, D., Stavrou, V., Kandias, M., Stergiopoulos, G.: Insider Threat: Enhancing BPM through Social Media. In: *6th IFIP International Conference on New Technologies, Mobility and Security*, Springer. (2014).
25. Kandias, M., Stavrou, V., Bosovic, N., Gritzalis, D.: Predicting the insider threat via social media: The YouTube case. In: *12th ACM Workshop on Privacy in the Electronic Society*, pp. 261--266, ACM Press. (2013).
26. Chen, Y., Nyemba, S., Zhang, W., Malin, B.: Leveraging social networks to detect anomalous insider actions in collaborative environments. In: *IEEE International Conference on Intelligence and Security Informatics*, pp. 119--124. IEEE. (2011).
27. Kandias, M., Virvilis, N., Gritzalis, D.: The Insider Threat in Cloud Computing. In: *6th International Conference on Critical Infrastructure Security*, pp. 93--103, Springer. (2013).
28. Skues, J., Williams, B., Wise, L.: The effects of personality traits, self-esteem, loneliness and narcissism on Facebook use among university students. In: *Computers in Human Behavior*. (2012).
29. Buffardi, L., Campbell, W.: Narcissism and social networking web sites. In: *Personality and Social Psychology Bulletin*, vol. 34, no. 10, pp. 1303--1314. (2008).
30. Mehdizadeh, S.: Self-presentation 2.0: Narcissism and self-esteem on Facebook. In: *Cyber-psychology Behavior and Social Networking*, vol. 13, no. 4, pp. 357--364. (2010).
31. McCallum, A., Nigam, K.: A comparison of event models for naive Bayes text classification. In: *Workshop on Learning for Text Categorization*. 752, 41--48. (1998).
32. Joachims, T.: Text categorization with support vector machines: Learning with many relevant features. In: *Machine Learning*. Springer, 137--142. (1998).
33. Anderson, J.: Logistic regression. In: *Handbook of Statistics*, pp. 169-191, North-Holland. (1982).
34. Manning, C., Raghavan, P., Schütze, H.: *Introduction to Information Retrieval*. Cambridge University Press. (2008).
35. Council of Europe Recommendation CM/Rec(2010)13 of the Committee of Ministers to member states on the protection of individuals with regard to automatic processing of personal data in the context of profiling.
36. Kosinski, M., Stillwell, D., Graepel, T.: Private traits and attributes are predictable from digital records of human behavior. In: *Proc. of the National Academy of Sciences*, pp. 5802-5805. (2013).
37. Kandias, M., Mitrou, L., Stavrou, V., Gritzalis, D.: Which side are you on? A new Panopticon vs. privacy. In: *10th International Conference on Security and Cryptography*, pp. 98--110. (2013).
38. Abril-Sánchez, P., Levin, A., Del Riego, A.: Blurred Boundaries: Social Media Privacy and the 21st Century Employee. In: *American Business Law Journal*, vol. 49, no. 1, pp. 63--124. (2012).
39. Fazekas, C.: 1984 is Still Fiction: Electronic Monitoring in the Workplace and US Privacy Law. In: *Duke Law & Technology Review*, pp. 15--15. (2004).
40. Mitrou, L., Karyda, M.: Employees' privacy vs. employers' security: Can they be balanced? In: *Telematics and Informatics*, vol. 23, no. 3, pp. 164--178. (2006).
41. Broughton, A., Higgins, T., Hicks, B., Cox, A.: *Workplaces and Social Networking - The Implications for Employment Relations*. Institute for Employment Studies, UK. (2009).
42. Nissenbaum, H.: Privacy as Contextual Integrity. In: *Washington Law Review*, vol. 79, pp. 119--157. (2004).
43. Davison, K., Maraist, C., Hamilton, R., Bing, M.: To Screen or Not to Screen? Using the Internet for Selection Decisions. In: *Employ Response Rights*, vol. 24, pp. 1--21. (2012).

44. Rubinstein, I.: Big Data: The End of Privacy or a New Beginning? In: *New York University School of Law Public Law & Legal Theory Research Paper Series*, Working paper no. 12-56. (2012).
45. Cumbley, R., Church, P.: Is “Big Data” creepy? In: *Computer Law & Security Review*, vol. 2, no. 9, pp. 601-609). (2013).
46. Cas, I.: 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. (2011).
47. Schermer, B.: The limits of privacy in automated profiling and data mining. In: *Computer Law and Security Review*, vol. 27, pp. 45--52. (2011).
48. Solove, D.: A taxonomy of privacy. In: *University of Pennsylvania Law Review*, vol. 154, no. 3, pp. 477. (2006).
49. Rubinstein, I., Lee R., Schwartz. P.: Data mining and internet profiling: Emerging regulatory and technological approaches. In: *The University of Chicago Law Review*, pp. 261-285. (2008).
50. Dempsey, J.., Flint L.: Commercial data and national security. In: *George Washington University Law Review*, vol. 72, pp. 1459. (2013).
51. Hildebrandt, M.: Who is profiling who? Invisible visibility. In: *Reinventing Data Protection*, pp. 239-252. (2009).
52. Andrejevic, M.: The work of watching one another: Lateral surveillance, risk, and governance. In: *Surveillance & Society*, vol. 2, no. 4, pp. 479–497. (2005).
53. Brin, D.: *The transparent society: Will technology force us to choose between privacy and freedom*. New York. (1999).
54. Dumortier, F.: Facebook and Risks of “De-contextualization” of Information. In: *Data Protection in a Profiled World*, pp. 119-137. (2010).
55. Simitis, S.: Reconsidering the premises of labour law: Prolegomena to an EU regulation on the protection of employees’ personal data. In: *European Law Journal*, vol. 5, pp. 45-62. (1999).
56. Gritzalis, D., Kandias, M., Stavrou, V., Mitrou, L.: The Social Media in the History of Information: Privacy violations and security mechanisms. In: *History of Information Conference*, Greece. (2014).