Marie Sklodowska Curie,
Research and Innovation Staff
Exchange (RISE)

European Commission

Horizon 2020
European Union funding
for Research & Innovation

ENhancing seCurity and privAcy in the Social wEb: a user-centered approach for the protection of minors

ENCASE
ENhancing seCurity and privAcy in the Social wEb

WP3 - Human and societal aspects of security and privacy in the social web
Deliverable D3.2 “Development of accurate and sophisticated sentiment analysis approaches”

Editor(s):	Agathi Galani, Evangelos Kotsifakos (LST)
Author(s):	Vaia Moustaka (AUTH), Petros Papagiannis, Antonis Papasavva (CUT)
Dissemination Level:	Public
Nature:	Report
Version:	0.6

ENCASE Project Profile

Contract Number	691025
Acronym	ENCASE
Title	ENhancing seCurity and privacy in the Social wEb: a user-centered approach for the protection of minors
Start Date	Jan 1st, 2016
Duration	48 Months

Partners

	Cyprus University of Technology	Cyprus
	University College London	United Kingdom
	Aristotle University	Greece
	Universita Degli Studi, Roma Tre	Italy
	Telefonica Investigacion Y Desarrollo SA	Spain
	SignalGenerix Ltd	Cyprus
	Cyprus Research and Innovation Center, Ltd	Cyprus
	L S Tech Ltd	United Kingdom

Document History

AUTHORS

(LST)	Agathi Galani, Evangelos Kotsifakos
(AUTH)	Vaia Moustaka
(CUT)	Petros Papagiannis, Antonis Papasavva

VERSIONS

Version	Date	Author	Remarks
0.1	10.06.2018	LST	Initial Table of Contents and first draft of text
0.2	25.06.2018	AUTH	AUTH contribution
0.3	29.06.2018	CUT	CUT contribution
0.4	30.06.2018	LST	Proposed final version
0.5	05.07.2018	LST	Minor improvements
0.6	06.07.2018	CUT	Final version

Executive Summary

This deliverable, D3.2 “Development of accurate and sophisticated sentiment analysis approaches”, refers to the activities that were carried out as part of tasks 3.2 and 3.3 of the Work Package 3 (WP3) of the ENCASE project.

The purpose of Task 3.2 was to study ways to improve user experience and user behavior when faced with security and privacy risks. The usability and user-experience of existing security and privacy systems for OSNs has been evaluated using query based techniques (questionnaires, interviews, focus groups) and through usability studies (observations, eye-tracking studies).

The purpose of Task 3.3 was to complement the user studies from T3.2 with sentiment analysis of OSN content. The understanding of how people feel about things they are talking about, needs the development of accurate and sophisticated sentiment analysis approaches. Existing challenges, such as the recognition of sarcastic or ironic content, should be addressed for a successful capturing of people’s sentiments. The adoption of such approaches can contribute in solving important open issues (e.g. the timely recognition of people with suicidal tendencies or criminal behavior) and can significantly contribute to the identification and management of critical situations.

A long literature review has been conducted regarding sentiment analysis and articles related to Twitter analysis were recorded and analysed in order to investigate the basic principles of Twitter operation and to compare their outcomes with our own first findings. A statistical analysis of OSWINDS’ dataset was conducted and a report that sums up our findings is included in this deliverable.

Furthermore, developments on the parental console regarding the traffic blocking options are presented in this deliverable as well.

Table of Contents

List of Figures

List of Tables

1. Introduction

In recent years, offensive, abusive and hateful language, sexism, racism and other types of aggressive and cyberbullying behavior have been manifesting with increased frequency, and in many online social media platforms. Bullying and aggression against social media users have grown significantly, causing serious consequences to victims of all demographics. Nowadays, cyberbullying affects more than half of young social media users worldwide, suffering from prolonged and/or coordinated digital harassment. Also, tools and technologies geared to understand and mitigate it are scarce and mostly ineffective. In fact, past scientific work focused on studying these forms in popular media, such as Facebook and Twitter.

In this deliverable we present a literature review on sentiment analysis and on extracting user attributes from OSN and we also present an analysis of OSWINDS dataset focusing on the age of the users. Two related articles based on our research have been published.

2. Literature Review

2.1. Sentiment analysis

We now review related work on studying/detecting offensive, abusive, aggressive or bullying content on social media sources. Chen et al. aim to detect offensive content, as well as, potential offensive users based on YouTube comments. Both Yahoo Finance [9, 24] and Yahoo Answers [18] have been used as a source of information for detecting hate and/or abusive content. More specifically, [18] studied a Community-based Question-Answering (CQA) site and finds that users tend to flag abusive content in an overwhelmingly correct way.

Cyberbullying has also attracted a lot of attention lately, for instance Chatzakou et al.(2017) , Hosseinmardi et al (2014) and Hosseinmardi et al. (2015) focus on Twitter, Ask.fm, and Instagram, respectively, to detect existing bullying cases out of text sources. Chatzakou et al (2017) considers a variety of features, i.e., user, text, and network-based, to distinguish bullies and aggressors from typical Twitter users. In addition to text sources, Hosseinmardi et al (2015) also tries to associate an image’s topic (e.g., drugs, celebrity, sports, etc.) with cyberbullying events. In [8], the cyberbullying phenomenon is further decomposed to specific sensitive topics, i.e., race, culture, sexuality, and intelligence, by analyzing YouTube comments extracted from controversial videos.

A study of specific cyberbullying cases, e.g., threats and insults, is also conducted in Van et al. (2015) by considering Dutch posts extracted from Ask.fm. Apart from cyberbullying, they also study specific user behaviors: harasser, victim, and bystander-defender or bystander-assistant who support the victim or the harasser, respectively. In follow-up work, the authors exploit Twitter messages to detect bullying cases which are specifically related to the gender bullying phenomenon. Finally, in Dadvar et al. (2014), YouTube users are characterized based on a “bulliness” score. The rise of cyberbullying, and abusive incidents in general, is also evident in online game communities. Since these communities are widely used by people of all ages, such a phenomenon has attracted the interest of the research community. For instance, Kwak et al. (2015) studies cyberbullying and toxic behaviors in team competition online games in an effort to detect, prevent, and counter-act toxic behavior. Fox et al. (2014) investigates the prevalence of sexism in online game communities Finding personality traits, demographic variables, and levels of game-play predicted sexist attitudes towards women who play video games. Overall, previous work considers various attributes to distinguish between normal and abusive behavior, like text-based attributes, e.g., URLs and Bag of Words (BoW), lexicon-based (offensive word dictionary), user/activity based attributes, e.g., number of friends/followers and users’ account age.

2.2. Extracting user behavior and attributes from OSN data

Over the past few years, several techniques have been proposed to measure and detect offensive or abusive content / behavior on platforms like Instagram (Hosseinmardi et al. 2015), YouTube (Chen et al. 2012), 4chan (Hine et al. 2017), Yahoo Finance (Djuric, 2015), and Yahoo Answers (Kayes et al. (2015)). Chen et al. (2012) use both textual and structural features (e.g., ratio of imperative sentences, adjective and adverbs as offensive words) to predict a user’s aptitude in producing offensive content in YouTube comments, while Djuric et al. rely on word embedding to distinguish abusive comments on Yahoo Finance. Nobara et al. (2016) perform hate speech detection on Yahoo Finance and News data, using supervised learning classification. Kayes et al. (2015) and that users tend to flag abusive content posted on Yahoo Answers in an overwhelmingly correct way (as confirmed by human annotators). Also, some users significantly deviate from community norms, posting a large amount of content that is flagged as abusive. Through careful feature extraction, they also show it is possible to use machine learning methods to predict which users will be suspended. Dinakar et al. [16] detect cyberbullying by decomposing it into detection of sensitive topics. They collect YouTube comments from controversial videos, use manual annotation to characterize them, and perform a bag-of-words driven text classification. Hee et al. study linguistic characteristics in cyberbullying-related content extracted from Ask.fm, aiming to detect fine-grained types of cyberbullying, such as threats and insults. Besides the victim and harasser, they also identify bystander-defenders and bystander-assistants, who support, respectively, the victim or the harasser. Hosseinmardi et al. (2015) study images posted on Instagram and their associated comments to detect and distinguish between cyberaggression and cyberbullying. Finally, authors in Saravanaraj et al (2016), present an approach for detecting bullying words in tweets, as well as demographics about bullies such as their age and gender. Previous work often used features such as punctuation, URLs, part-of-speech, n-grams, Bag of Words (BoW), as well as lexical features relying on dictionaries of offensive words, and user-based features such as user’s membership duration activity, number of friends/followers, etc. Different supervised approaches have been used for detection: Nobat et al. uses a regression model, whereas [Dinakar et al, Hee et al.] rely on other methods like Naive Bayes, Support Vector Machines (SVM), and Decision Trees (J48). By contrast, Hosseinmardi et al. (2014) use a graph-based approach based on likes and comments to build bipartite graphs and identify negative behavior. A similar graph-based approach is also used in Hosseinmardi et al (2015).

Sentiment analysis of text can also contribute useful features in detecting offensive or abusive content. For instance, Nahar et al. use sentiment scores of data collected from Kongregate (online gaming site), Slashdot, and MySpace. They use a probabilistic sentiment analysis approach to distinguish between bullies and non-bullies, and rank the most in-uential users based on a predatorvictim graph built from exchanged messages. Xu et al. rely on sentiment to identify victims on Twitter who pose high risk to themselves or others. Apart from using positive and negative sentiments, they consider specific emotions such as anger, embarrassment, and sadness. Finally, Patch et al. studies the presence of such emotions (anger, sadness, fear) in bullying instances on Twitter.

3. User behavior and OSN sentiment analysis

Despite the fact that OSN is a useful tool for stakeholders (e.g., local authorities, companies, etc.) that exploit the data produced, many privacy and security concerns arise. Individuals increasingly register and share personal information (such as date of birth, email address, telephone number, home address, photos, videos, etc.) on OSN and their content can be used in many ways exposing them to danger. In many cases human activity, sentiment and opinion of individuals on OSN are recorded and analyzed in their absence (Luo et al, Martinez et al., Rizzo et al.). For the purpose of investigating the privacy and security concerns raised in the SC context, at this point, we will clarify the terms of ―privacy‖ and ―security‖, which are often confused. Privacy concerns the protection of individual’s personal information from the illegal disclosure and use by third malicious parties and is directly related to the individual's online behavior and privacy preferences (Martinez et al., Zhang et al., Patsakis et al. Individuals' belief that their privacy is more protected than that of others, and the degree of their trust in other users compromise their privacy (Bergström et al., Baek et al.).

According to Zhang et al. individual’s privacy on OSN consists of:

1. Individual’s identity anonymity: concerns the protection of the user's identity, so that it is not easily detected on the Internet.

2. Individual’s personal space privacy: refers to access control to a user's profile, in particular information and content posted on it.

3. Individual’s communication privacy: concerns the protection of information related to the connection network (e.g., IP address, location etc.) and the user's navigation activities (e.g., friends, messages sent etc., online preferences etc.).

On the other hand, security refers to the protection of OSN users from threats caused either by inside attackers (i.e. other OSN users) or by external attackers (i.e., individuals who do not participate but can commit attacks on the OSN system) who exploit the unawareness and naivety of their potential victims (Zhang et al.).

Many research efforts have focused on identifying and dealing with risks and threats affecting OSN. According to Fire et al., OSN threats can be divided into the following four main categories.

1. Classic threats: threats that occurred when the Internet was created and spread, and referred as malware, phishing, spam or cross-site scripting attacks. Although these threats have been addressed in the past, due to the spread of OSN, they are becoming more viral and spreading through their users and their friends.

2. Modern threats: threats related to OSN and target the individuals’ personal information and the personal information of their friends. Information and location leakage, fake profiles, identity clone attacks and face recognition are just some of these threats.

3. Combination threats: threats which are the combination of classic and modern threats to create more effective threats.

4. Threats targeting children: threats directed exclusively at children and adolescents. Online predators, cyberbullying and children’s risky behaviors when communicate online with strangers and publish private information and photos on OSN are the most risky of these threats.

OSN users are also exposed to risks by their share multimedia content, many of which are indirect or often ignored by the majority of them. The most dangerous from these risks are: i) multimedia content, ii) lack of policies, iii) platform vulnerabilities and iv) open access. The individual’s sensitive and personal content is stored, daily, as multimedia files on OSN, which are software platforms vulnerable to the bugs and malicious third parties. Additionally, the lack of policies to govern every possible privacy issue or to allow fine-grained user customization and the existing —freemium‖ model, which allows individuals to register quite easily, contribute to the creation of multiple and false accounts complicating the detection of malicious actors (Patsakis et al).

The most peculiar and dangerous threats mentioned above are threats targeting children. These threats, which can be extended to adults, are usually caused by psychological factors and occur both in real life and in online life. Online predators and cyber-bullying attacks are booming nowadays. Adults or minors in order to satisfy their fantasies and to erase their frustration and anger, often, sexually harass or intimidate their potential victims (Fire et al). Parents cannot fully protect their children whose critical ability and online defense on OSN are minimal, while in many cases adults are sharing sensitive personal information and photos on OSN regarding to their children, exposing them to privacy and security risks (Minkus et al.). The Canadian Centre for Child Protection[1] has revealed that children under 12 years old were depicted in 78.30% of the images and videos assessed by their team. Furthermore, recent surveys have revealed that cyber-bullying[2] occurs mainly through OSN, while more than 82% of online sex crimes related to sexual predators[3] and online sexual offenses originate from OSN that predators use to gain insight into their victims. As these threats greatly affect children’s behavior and psychology, they can have disastrous and irreversible effects, such as in the cases of Amanda Michelle Todd and Rebecca Ann Sedwick, both of whom committed suicide after being cyber-bullied on Facebook (Fire et al., Minkus et al.).

3.1. User’s behavior on OSNs

The Social Web (Web 3.0) has improved users’ online experience by offering intelligent, interactive and personalized services, dynamic applications, and machine to machine (M2M) communication. Online social networks (OSNs) such as Facebook, Twitter, Instagram, LinkedIn, etc. provide their users with the opportunity to build, maintain and enrich their personal and professional networks

[1] https://www.protectchildren.ca/app/en/

[2] http://enough.org/stats_cyberbullying

[3] http://www.kidslivesafe.com/child-safety/online-predators-and-cyberbullyingstatistics

and social relationships, to spend their free time in online activities and to share content and ideas with other. Often, users with common interests or "friends" develop relationships among themselves and exchange information related to their professional or private life through OSNs. According to IBM-Big Data Hub[4] and Business Insider’s Intelligence Report[5] more than 22 million individuals visit LinkedIn every day, 32 million Tweets are released per day and the “like” button in Facebook is pressed 2.7 billion times every day across the web.

The Social Web and OSNs, despite the limitless opportunities that offer to users' online life, are prone to various privacy and security vulnerabilities risks, while their users are often suspicious or naïve during their use (LaOrden et al., 2010). Recent studies revealed that privacy concerns regarding website personalization have grown significantly between 2002 and 2008 (Anton et al., 2010). Privacy risks associated with several current and prominent personalization trends, such as social-based personalization, behavioral profiling, and location-based personalization as well as user attitudes towards privacy and personalization were analyzed by Toch et al. (2012). According to Cranor (2003) many users feel uncomfortable being online "watched", while Turow et al. (2009) pointed out that 66% of Americans react against to their interests' recording and personalized advertisements and this attitude is consistent across age groups and gender. On the other hand, cyber-attacks that have so far had a limited effect now have a huge distributed effect through OSNs due to their “freemium model” (Alqatawna et al., 2017; Patsakis et al., 2014). Plenty of new privacy risks and security threats have been appeared on OSNs that are proven to affect their use and users’ behavior on them (Alqatawna et al., 2017; Patsakis et al., 2014, Fire et al. 2014; Moustaka et al., 2018).

The study and deciphering of online user behavior are very important for the design and development of appropriate methods and tools (e.g., applications, software, etc.) for protecting privacy and security in OSNs. In this context, Jin et al. (2013) conducted a literature review aiming at understanding user behavior in regard to the connectivity and interaction between users, by analyzing a) several types of social graphs, b) traffic activity by monitoring the network records, c) behavior of mobile users by studying the activities on mobile platforms, and d) malicious behavior by analyzing the security threats. According to Ellingsen et al. (2016), users’ decisions and actions are depend on their personalities and motivated by their expectations, while their attitudes are significantly influenced by OSNs. Although there are several empirical studies concerning the behavior of users in social networks with the purpose of investigating the factors that affect it, a literature review summarizing these studies is still missing. A brief review, which intends to fill this gap with the exploitation of primary studies published the last years in scientific journals and proceedings of international conferences, is presented below.

[4] http://www.ibmbigdatahub.com/gallery/quick-facts-and-stats-big-data

[5] http://www.businessinsider.com/social-network-big-data-lens-2014-7

The users’ behavior on OSNs is often unpredictable and reflect both their privacy concerns and their personalities. Consolvo et al. (2005) revealed that OSNs user are more willing to share vague information than specific personal information, while Acquisti and Gross (2005) highlighted that incomplete information, bounded rationality, and systematic psychological deviations from rationality are the main challenges in privacy decision-making. Knijnenburg et al. (2013) claimed that users' disclosure behavior is in fact multi-dimensional as different people have different tendencies to disclose various types of information, while Norberg et al. (2007) have considered that most OSNs users share content and personal information much more freely than expected based on their attitudes. Ball et al. (2015), in their study, found that users’ habits were determined to have the strongest influence on their practices and information sharing activities, while the awareness was not significantly influencing them, by assessing the influence of users’ personal information sharing awareness (PISA) on their habits (PISH) and practices (PISP) and by comparing the three constructs between OSNs.

In terms of gender, surveys’ outcomes differ. A few scholars have claimed that gender does not have an effect on users’ practices (Furnell, 2008; Levy & Ramim, 2009), while other scholars such as Fogel and Nehmad (2009) argued that gender affects users’ online personal information sharing practices. Kisekka et al. (2013) adopting the Communication Privacy Management (CPM) theory and using a sample size of 488 adult Facebook mobile phone users have investigated the differential impact of age on the extent of Private Information Disclosure (PID). Their findings have revealed the following: a) the likelihood of PID was less for three groups of users: females, users who use smartphones to access their accounts, and users with more than one active OSN account, b) the usability affected older and younger adult users differently, and c) an increase in social networking involvement does not increase the likelihood of PID. Hinduja & Patchin (2008) studying the characteristics of typical cyber-bullying victims and offenders have found that gender and race did not significantly differentiate respondent victimization or offending. On the contrary, computer proficiency and time spent on-line were positively related to both cyber-bullying victimization and offending. The same conclusion came also from the study of Smith et al. (2008), who ascertained that being a “cybervictim”, but not a “cyber-bully”, is correlated with internet use.

OSNs offer their users the option to manage the information they reveal and protect their privacy through their privacy settings (Kuczerawy & Coudert, 2011). Several studies have been conducted to investigate the suitability and reliability of privacy settings, as well as users’ behavior regarding their proper use (Li et al., 2015; Kuczerawy & Coudert, 2011; Hugl, 2011;). Madejski et al. (2011), evaluating the actual preferences and behavior of Facebook users, found that there is a lower limit of the inconsistencies between users' sharing intentions and their privacy settings. Additionally, Netter et al. (2014) studying the OSNs’ privacy settings with the use of a novel approach based on profiles content of Facebook users, have indicated a mismatch between perceived, preferred, and actual settings due the lack of users’ awareness. Finally, Aljohani et al. (2016), conducting a survey on OSNs users' privacy settings and information disclosure and investigating users' behavior on Facebook, Twitter, Instagram, and Snapchat, found that there is information leak at different levels between OSNs, and in fact socio-demographic factors such as, age, gender and education influence information disclosure and privacy settings use. Therefore, the existence of privacy settings does not guarantee users pro on OSNs, but their proper use is required, which depends on users’ online behavior determined by their personalities.

The users' behavior when they are confronted with privacy risks on OSNs, as expected considering the above analysis, varies. Shin (2010), in his study, examined security, trust, and privacy concerns with regard to OSNs among consumers, developing a novel model of trust-based OSNs acceptance. His findings have shown that: a) users are concerned about the vulnerability of security and privacy breaches when they use OSNs, b) perceived security and perceived privacy are directly associated with trust in OSNs use, and c) perceived security affects much more the users' attitude than perceived privacy. According to Saridakis et al. (2015), who investigated how user' online activity and perceptions of personal information security on OSNs are related to their online victimization, has revealed that the latter is affected positively by: a) high OSNs use, b) low perceived risk, and c) high risk propensity; and negatively by: a) high perceived control over information, and b) high computer efficacy.

An exploratory study regarding common experiences of online privacy-related panic and users’ reactions on frequently occurring privacy violations was conducted by Angulo & Ortlieb (2015). Specifically, by utilizing the allegory of a *privacy panic button*, (see section 5) they investigated users’ expectations and mental models of appropriate mechanisms that could lead these users to a solution, calming their distress, and preventing similar situations from happening in the future. The conduct of a survey (*n* = 549) and user semi-structured interviews (*n* = 16) led to the identification of 18 different scenarios of privacy panic situations. The findings have shown that victims’ topmost concerns included possible harm to their finances or fear of embarrassment, as well as third-parties knowing things that might not be of their business. The cases of account hijacking and personal data leakage were among the most notable self-reported panic stories, while incidents involving regrets when sharing content online were found to be experienced most frequently. Furthermore, scenarios related to the online data loss, the mobile device loss, or falling pray of identity theft also were at the top of users’ worries. The study also revealed that, if a service provider offers a hypothetical *privacy panic button*, users expect that the assistance provided will be immediate, uncomplicated, actionable, and in-place.

3.2. Teenagers behaviour and attitude on OSNs

According to Pew Research Center’s Relationships Survey (2015), 73% of teenagers or millennial teens have access to smartphones and more than half of teens have access to a tablet, while the 87% have desktop or laptop. It is true that, teenagers constitute the 25% of Facebook users and the 34% of Instagram users (Statista, 2015). Teenagers usually use OSNs to communicate, connect and remain in contact with others (Gross & Acquisti, 2005; Charnigo & Barnett-Ellis, 2007; Acquisti & Gross, 2006), while sometimes they use them for their self-presentation and self-identity (Doster, 2013). Charnigo & Barnett-Ellis (2007) claimed that Facebook is used for dating, while Acquisti & Gross (2006) reported that students do not use Facebook for this purpose.

Despite their young age, teenagers are concerned about online security and are aware of the existing risks on OSNs. Many studies have discussed teenagers' perceptions for online privacy and their behaviors in OSNs (Stutzman, 2006; Youn, 2005; Fogel & Nehmad, 2008). The Stutzman's (2006) study showed that college students agreed that it is important for them to protect their identity information while the same students on average rated it was okay if their friends, family, or classmates accessed their social networking profile. However, on average they rated as ''neutral" the item about strangers accessing their social networking profile. A national survey conducted by the Annenberg Public Policy Center reported that teens aged 13 to 17, who are not covered by COPPA (1998), were more open to providing their information to Web sites for an incentive than were children aged 10 to 12 (Turow & Nir, 2000). In addition, Youn (2005), conducting a survey based on 326 high school students (>13 years old) sample, revealed that a higher level of risk perception of information disclosure led to less willingness to provide information. When teenagers perceived more benefits from information disclosure, they were more willing to provide information. Since teenagers were less likely to give out their information, they tended to engage in several risk-reducing strategies such as falsifying information, providing incomplete information, or going to alternative websites that do not ask for personal information. Fogel & Nehmad (2008), by exploring risk taking, trust, and privacy concerns with regard to OSNs, among 205 college students (17-32 years old), found that individuals with profiles on OSNs have greater risk taking attitudes than those who do not, while greater risk taking attitudes exist among men than women. Greater percentages of men than women display their phone numbers and home addresses on OSNs.

Many scholars have also investigated the patterns of information revelation in OSNs and their privacy implications. Gross & Acquisti (2005), by analyzing the online behavior of more than 4,000 Carnegie Mellon University students on Facebook, have found that only a limited percentage of users change the highly permeable privacy preferences, as well as users expose themselves to various cyber risks facilitating third parties to create digital dossiers of their behavior. Another study by Acquisti & Gross (2006) revealed that privacy concerns of Facebook participants for strangers knowing their schedule of classes and their place of residence's address are not related to the likelihood of their providing this information on OSNs websites. Among the 16% of the participants who expressed the highest privacy concerns for a stranger knowing their schedule and where they lived, even so 22% provided at least their home address and 40% provided their schedule of classes. Moreover, as found, a significant fraction of minors circumvents the COPPA law and states false ages, putting both lying and truthful minors at risk as strangers not only can discover more minors, but can also build more extensive profiles than what would be the case in a world without an age restriction (Dey at al., 2013). In respect of cyber-bullying phenomenon, Kowalski and Limber (2007) have observed that the most of middle school students who have been cyber-bullied was to do nothing, while Whittaker and Kowalski (2015) have found that the majority of college students, who have been cyber-bullied, blocked the attackers from OSNs and reported them. Finally, Dehue et al. (2008) have highlighted that "*Youngsters mostly react to cyber-bullying by pretending to ignore it, by really ignoring it, or by bullying the bully*".

With regard to some special categories of young people, a qualitative study conducted by Velden & Emam (2012) examined the privacy concerns and behaviors of teenage patients (12-18 years old) when using social media and have revealed that the majority of teenage patients do not disclose their personal health information on social media. The findings showed that OSNs, and mainly the Facebook, play an important role in the social life of teenage patients as they enable young patients to be "regular" teenagers. Nevertheless, the most teenage patients do not use social media to come into contact with others with similar conditions and they do not use the Internet to find health information about their diagnosis. Their online privacy behavior is an expression of their need for self-definition and self-protection.

The completion of the review led to the conclusion that the behavior of users on OSNs is multidimensional and is mainly determined by: *i) psychological (personal) factors* (e.g., level of user's education, habits, self-esteem, self-presentation, personality, etc.), *ii) demographic factors* (e.g., age, gender, etc.), and *iii) socio-political factors* (e.g., legislation related to privacy and security protection, level of public education, city's or country's culture, etc.).

4. Dataset

4.1. Data collection

The particular dataset was collected by the OSWINDS group (AUTH) during the period of July-September 2017 and includes geo-located data (tweets) from the New York region.

Data understanding, pre-processing and basic analysis and assumptions took place and new charts related to statistical analysis of the dataset -CDF and CCDF charts of tweets per user- were analyzed.

Due to the deviation of the number of tweets of some users from the description of the dataset, the tweets' IDs were checked and verified.

4.2. Age classification on Twitter

Introduction

Online Social Networks (OSNs), the so-called "Internet of People" (Miranda et al., 2015), record almost all human activities, offering the possibility of extracting patterns that are useful for multiple purposes in various fields of Science (Moustaka et al., 2018). Lampos et al. (2016) classified the OSNs users based on their socioeconomic status using Twitter data; Hossain et al. (2016) discovered and compared alcohol consumption patterns in a large urban area (New York City) and a more suburban and rural area (Monroe County) by exploiting fine-grained localization of activities and home locations from Twitter data; while Zhagheni et al. (2014) have attempted to infer international and internal migration patterns utilizing Twitter data. In many cases, OSNs were used for the purposes of investigating and classifying latent demographics attributes and studying human behavior, leading to the acquisition of valuable knowledge and facilitating decision-making and design and implementation of new policies (e.g., online security, government, public health, transport, etc.) and applications in advertising, recommendation and personalization (Rao et al., 2010; Flekova et al., 2016; Siswanto & Khodra, 2013; Patsakis et al., 2014; Kisseka et al. 2013; Efstathiades, et al. 2015; Gkatziaki et al., 2017).

Of particular interest is the recent study of Cesare et al. (2017), which include a literature review on existing approaches to automated detection of demographic characteristics of OSNs users. The exploitation of 60 selected studies focused on different OSNs platforms resulted in Table 1, which summarizes the following: a) the data analytics methods, b) the used metadata, c) the OSNs platforms, and d) extracted traits. In total, 39 (65%) studies focused on Twitter, 2 on Facebook, 2 on Livejournal, 1 on Yelp, 1 on YouTube, and 1 on Pinterest. The rest of the studies focused on other OSNs (e.g., Netlog, Fotolog) and blogs. As it turns out, Twitter holds a prominent position among OSNs as it offers: a) flexibility, as a user can track someone else's post without being friends, b) real-time update, c) ability to harvest huge amounts of data through its APIs, and iv) potentiality for future situations prediction (Bright et al., 2014; Hutchinson, 2016). With regard to data analytics methods, 44 (73%) studies investigated supervised or semi-supervised machine learning methods, 11 (18%) used raw or adjusted data matching, 3 (5%) used facial evaluation (human or automated), and 1 (2%) used unsupervised learning.

As revealed by the findings of the literature review the users' behavior in OSNs is influenced and determined by personal, demographic and socio-political factors. This work, focusing on demographic factors, aims to explore and identify the exact age and gender of Twitter users by analyzing the content they generate during social networking activities. Unlike the 29 studies of the literature review of Cesare et al. (2017) that predict the age of OSNs users using text features and supervised learning methods, our work aims to combine and exploit text-, user- and network-based features on Twitter and unsupervised learning methods for the purpose of identifying age and gender of Twitter users. Certainly, as several studies have shown, the prediction of age is far more laborious compared to gender prediction (Tuli, 2015; Nguyen et al. 2011; Nguyen et al., 2013; Rosenthal & McKeown, 2011; Siswnato & Khodra, 2013; Nguyen et al. 2014).

In the framework of this work, a Twitter dataset which include geo-located data for a specific location was selected, with the purpose of research conducting, and designing and developing a novel methodology for age detection and classification based on OSNs content analysis. Since our research is in progress, the dataset's statistical analysis that has been completed is presented herein.

Table 1. Literature Review Findings (Cesare et al., 2017)

Citation	Method	Metadata	Platform	Trait
Al Zamal, F., Liu, W., & Ruths, D. (2012). Homophily and Latent Attribute Inference: Inferring Latent Attributes of Twitter Users from Neighbors. In *Proceedings of the 6th International Conference on Weblogs and Social Media (ICWSM),* 270.	Supervised learning : SVM classification	User tweets, neighbor tweets	Twitter	Age, Gender
Alowibdi, J. S., Buy, U. A., & Yu, P. (2013). Empirical Evaluation of Profile Characteristics for Gender Classification on Twitter. In *Proceedings of the 12th International Conference on Machine Learning and Applications (ICMLA 2013),* 365–369.	Supervised learning: Naïve Bayes/Decision-Tree Hybrid	Profile colors, user name, user tweets	Twitter	Gender

https://doi.org/10.1109/ICMLA.2013.74	(NB-Tree).			
An, J., & Weber, I. (2016). #greysanatomy vs. #yankees: Demographics and Hashtag Use on Twitter. In *Proceedings of the 10th International Conference on Weblogs and Social Media (IWCSM)*, 523–526.	Facial recognition: automated	Profile image, User descriptions	Twitter	Age (validated), Gender (validated), Race/ethnicity (not validated)
Argamon, S., Koppel, M., Pennebaker, J. W., & Schler, J. (2009). Automatically Profiling the Author of an Anonymous Text. *Communications of the ACM*, 52(2), 119–123. https://doi.org/10.1145/1461928.1461959	Supervised learning: Bayesian multinomial regression	User posts	Various blogging platforms	Age, Gender
Asoh, Hideki, Ikeda, Kazushi, & Ono, Chihiro. (2012). A Fast and Simple Method for Profiling a Population of Twitter Users. In *The Third International Workshop on Mining Ubiquitous and Social Environment*. Bristol, UK.	Adjusted data matching w/Bayesian estimation	User tweets	Twitter	Age (distribution), Gender
Bamman, D., Eisenstein, J., & Schnoebelen, T. (2014). Gender Identity and Lexical Variation in Social Media. Journal of Sociolinguistics, 18(2), 135– 160. https://doi.org/10.1111/josl.12080/abstract	Supervised learning: expectation maximization framework	User tweets	Twitter	Gender
Benton, A., Raman, A., & Dredze, M. 2016. Learning Multiview Embeddings of Twitter Users. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Berlin, Germany, 14–19.	Supervised learning : SVM classification	User tweets, Neighbor tweets	Twitter	Gender
Beretta, V., Maccagnola, D., Cribbin, T., & Messina, E. (2015). An Interactive Method for Inferring Demographic Attributes in Twitter. In Proceedings of the 26th ACM Conference on Hypertext & Social Media (HT ’15), 113–122. https://doi.org/10.1145/2700171.2791031	Data matching/ Supervised learning: SVM classification	User name, User tweets	Twitter	Age, Gender
Bergsma, S., Dredze, M., Van Durme, B., Wilson, T., & Yarowsky, D. (2013). Broadly Improving User Classification via Communication-Based Name and Location Clustering on Twitter. In Proceedings of the 2013 North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Hlt-Naacl), (June), 1010–1019.	Supervised learning: SVM classification	User location, User name,	Twitter	Gender, Race/Ethnicity
Burger, J. D., Henderson, J., Kim, G., & Zarrella, G. (2011). Discriminating Gender on Twitter. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, 1301–1309. https://doi.org/10.1007/s00256-005-0933-8	Supervised learning: Winnow	User tweets, User names, Screen handles, User description	Twitter	Gender
Chang, J., Rosenn, I., Backstrom, L., & Marlow, C. (2010). epluribus: Ethnicity on Social networks. In *Proceedings of the Fourth International Conference on Weblogs and Social Media (ICWSM),* 18–25.	Adjusted data matching w/Bayesian estimation	User names User names, profile	Facebook	Race/Ethnicity
Culotta, A., Ravi, N. K., & Cutler, J. (2015). Predicting the Demographics of Twitter Users from Website Traffic Data. Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, 72–78.	Supervised learning: OLS regression	Following relationship	Twitter	Gender, Race/Ethnicity
Chen, X., Wang, Y., Agichtein, E., & Wang, F. (2015). A comparative study of demographic attribute inference in twitter. In *Proceedings of the Ninth International Conference on Weblogs and Social Media (ICWSM),* 590-593.	Supervised learning: SVM	Images, User descriptions, Neighborhood info	Twitter	Gender, Race/Ethnicity

Filippova, K. (2012). User Demographics and Language in an Implicit Social Network. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning,* 1478–1488.	Supervised learning: Maximum entropy	Posts and social environment	YouTube	Gender
Fink, C., Kopecky, J., & Morawski, M. (2012). Inferring Gender from the Content of Tweets: A Region Specific Example. In *Proceedings of the 6th International Conference on Weblogs and Social Media (ICWSM*), 459–462.	Supervised learning: SVM classification	User tweets	Twitter	Gender
Gadiya, M., & Jain, S. V. (2016). A Study on Gender Prediction using Online Social Images. *International Research Journal of Engineering and Technology*, *3*(2), 1300–1307.	Supervised learning: SVM classification	Images	Pinterest	Gender
Goswami, S., Sarkar, S., & Rustagi, M. (2009). Stylometric Analysis of Bloggers’ Age and Gender. In *Proceedings of the Third International Conference on Weblogs and Social Media (ICWSM)*, 214-2017.	Supervised learning: Naïve Bayes	User posts	Blogger	Age, Gender
Hofstra, B., Corten, S., Van Tubergen F., Ellison, N. (2016, April). “Segregation in Social Networks: A Novel Approach using Facebook.” Paper presented at the *International Sunbelt Social Network Conference (Sunbelt 2016)*, Newport Beach, CA.	Data matching /Supervised learning	User names	Facebook	Gender, Race/Ethnicity
Ikeda, D., Takamura, H., & Okumura, M. (2008). Semi-Supervised Learning for Blog Classification. In *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence,* 1156–1161	Supervised learning: ASO algorithm	User posts	Blogs	Age, Gender
J. Alowibdi, U. Buy and P. Yu, "Language Independent Gender Classification on Twitter", *In Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (ASONAM'13) Niagara Falls, Canada, 2013	Supervised learning: Naïve Bayes/ Decision-Tree Hybrid (NB-Tree).	Profile colors	Twitter	Gender
Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private Traits and Attributes are Predictable from Digital Records of Human Behavior. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(15), 5802–5. http://doi.org/10.1073/pnas.1218772110	Supervised learning: OLS/logistic regression	User 'likes' (could be translated to follows)	Facebook	Age, Gender
Liu, W., & Ruths, D. (2013). What’s in a Name? Using First Names as Features for Gender Inference in Twitter. In *Analyzing Microtext: Papers from the 2013 AAAI Spring Symposium,* 10–16.	Supervised learning: SVM	User names, user tweets	Twitter	Gender
Longley, P. A., Adnan, M., & Lansley, G. (2015). The Geotemporal Demographics of Twitter Usage. *Environment and Planning A*, *47*(2), 465–484. https://doi.org/10.1068/a130122p	Data matching	User names	Twitter	Gender, Ethnicity, Age
Ludu, P. S. (2014). Inferring gender of a Twitter user using celebrities it follows. arXiv Preprint arXiv:1405.6667. Retrieved from http://arxiv.org/abs/1405.6667	Supervised classification (SVM algorithm)	User tweets, User follows	Twitter	Gender
Mandel, B., Culotta, A., Boulahania, J., Stark, D., Lewis, B., & Rodrigue, J. (2012). A Demographic Analysis of Online Sentiment during Hurricane Irene. In *Proceedings of the Second Workshop on Language in Social Media (LSM 2012),* 27-36.	Data matching	User names	Twitter	Gender

Marquardt, J., Farnadi, G., Vasudevan, G., Moens, M. F., Davalos, S., Teredesai, A., & De Cock, M. (2014). Age and Gender Identification in Social Media. In *Proceedings of CLEF 2014 Evaluation Labs*, 1129–1136.	Supervised learning: Logistic regression	Tweet content	Twitter	Gender, Age
McCormick, T. H., Lee, H., Cesare, N., Shojaie, A., & Spiro, E. S. (2015). Using Twitter for Demographic and Social Science Research: Tools for Data Collection and Processing. Sociological Methods & Research, 0049124115605339.http://doi.org/10.1177/0049124115605339	Facial evaluation: Human	Profile photos	Twitter	Age, Gender, Race/Ethnicity
Mechti, S., Jaoua, M., & Belguith, L. H. (2014). Machine Learning for Classifying Authors of Anonymous Tweets, Blogs, Reviews and Social Media: *Notebook for PAN at CLEF 2014. CEUR Workshop Proceedings*, 1137–1142.	Decision table	User tweets	Twitter	Age/Gender
Miller, Z., Dickinson, B., & Hu, W. (2012). Gender Prediction on Twitter Using Stream Algorithms with N-Gram Character Features. *International Journal of Intelligence Science, 2*(24), 143–148. https://doi.org/10.4236/ijis.2012.224019	Supervised learning: Naïve Bayes and Perceptron	User tweets	Twitter	Gender
Peersman, C., Daelemans, W., & Van Vaerenbergh, L. (2011). Predicting age and gender in online social networks. In *Proceedings of the International Conference on Information and Knowledge Management*, 37–44. https://doi.org/10.1145/2065023.2065035	Supervised learning: SVM classification	User posts	Netlog (Belgian social networking site)	Age, Gender
Pennacchiotti, M. (2011). Democrats, Republicans and Starbucks Aficionados: User Classification in Twitter. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 430–438. https://doi.org/10.1145/2020408.2020477	Supervised learning: Gradient Boosted Decision Trees	Name, location, description, user tweets	Twitter	Race/Ethnicity
Pennacchiotti, M., & Popescu, A. (2011). A Machine Learning Approach to Twitter User Classification. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, 281–288.	Supervised learning: Gradient Boosted Decision Trees	User name, profile photo, friends/followers, date of creation, user tweets	Twitter	Race/Ethnicity
Rao, D., Paul, M. J., Fink, C., Yarowsky, D., Oates, T., & Coppersmith, G. (2011). Hierarchical Bayesian Models for Latent Attribute Detection in Social Media. In *Proceedings of the Fifth International Conference on Weblogs and Social Media (ICWSM)*, 598–601.	Semi-supervised classification (Hierarchical Bayesian models)	User posts, user names	Facebook	Gender/Ethnicity
Rao, D., Yarowsky, D., Shreevats, A., & Gupta, M. (2010). Classifying latent user attributes in Twitter. Proceedings of the 2nd International Workshop on Search and Mining User-Generated Contents, 37–44. http://doi.org/10.1145/1871985.1871993	Supervised learning: SVM classification	User tweets	Twitter	Gender
Reddy, S., Wellesley, M. A., Knight, K., & Marina del Rey, C. A. (2016). Obfuscating gender in social media writing. In *Proceedings of the 1st Workshop on Natural Language Processing and Computational Social Science* (pp. 17–26). Retrieved from http://www.aclweb.org/anthology/W/W16/W16-56.pdf#page=29	Supervised learning: Logistic regression	User posts	Twitter/Yelp	Gender
Rosenthal, S., & McKeown, K. (2011). Age prediction in blogs: A study of style, content, and online behavior in pre-and post-social media generations. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1, 763–772.	Supervised learning: Logistic regression	Post content and network info	Livejournal	Age

Rustagi, M., Prasath, R. R., Goswami, S., & Sarkar, S. (2009). Learning Age and Gender of Blogger from Stylistic Variation. In Pattern Recognition and Machine Intelligence (pp. 205–212). Springer, Berlin, Heidelberg.	Supervised learning: Naïve Bayes	User posts	A variety of blogging platforms	Age, Gender
Santosh, K., Joshi, A., Gupta, M., & Varma, V. (2014). Exploiting Wikipedia Categorization for Predicting Age and Gender of Blog Authors. In *Proceedings of the UMAP 2014 Posters, Demonstrations and Late-Breaking Results*, 33-36.	Supervised learning: K-nearest neighbors and SVM	User posts	Undisclosed blog	Age, Gender
Sap, Maarten, Eichstaedt, Johannes, Kern, Margaret L., Stillwell, David, Kosinski, Michal, Ungar, Lyle H., & Schwartz, H. Andrew. (2014). Developing Age and Gender Predictive Lexica over Social Media. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1146–1151.	Supervised learning: SVM, OLS regression	User posts	Facebook/ Twitter/Blogs	Age, Gender
Schler, J., Koppel, M., Argamon, S., & Pennebaker, J. W. (2006). Effects of Age and Gender on Blogging. In *AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs*, 199–205.	Supervised learning (multi-class real winnow	User posts	Blogs	Age, Gender
Siswanto, E., & Khodra, M. L. (2013). Predicting Latent Attributes of Twitter User by Employing Lexical Features. In *Proceedings of the 2013 International Conference on Information Technology and Electrical Engineering (ICITEE),* 176–180. https://doi.org/10.1109/ICITEED.2013.6676234	Supervised learning: SVM classification	User name, user tweet	Twitter	Age
Sloan, L., Morgan, J., Burnap, P., & Williams, M. (2015). Who tweets? deriving the demographic characteristics of age, occupation and social class from twitter user meta-data. PLoS ONE, 10(3), 1–20.	Data matching	User descriptions	Twitter	Age
Sloan, L., Morgan, J., Housley, W., Williams, M., Edwards, A., & Burnap, P. (2013). Knowing the TweetersDeriving Sociologically Relevant Demographics from Twitter. Sociological Research Online, 18(3).	Data matching	User names, user tweets, user locations	Twitter	Gender
Smith, J. (2014). Gender Prediction in Social Media. arXiv Preprint. Retrieved from http://arxiv.org/abs/1407.2147	Data matching	User name	Fotolog	Gender
Tuli, G. (2015). Modeling and Twitter-based Surveillance of Smoking Contagion. Virginia Polytechnic Institute and State University. Unpublished Dissertation. Retrieved from https://vtechworks.lib.vt.edu/handle/10919/64426	Supervised learning: SVM classification	User tweets	Twitter	Age (above/under 18)
Vicente, M., Batista, F., & Carvalho, J. P. (2015). Twitter Gender Classification Using User Unstructured Information. In *Proceedings of the IEEE International Conference on Fuzzy Systems*, 1-7	Unsupervised learning: fuzzy c-means clustering	User names	Twitter	Gender
Volkova, S., Bachrach, Y., Armstrong, M., & Sharma, V. (2015). Inferring Latent User Properties from Texts Published in Social Media. In *Proceedings of the Twenty-Ninth Conference on Artificial Intelligence (AAAI),* 4296–4297.	Supervised learning: log-linear regression	User tweets	Twitter	Age, Gender, Race/ethnicity
Wang, Y., Xiao, Y., Ma, C., & Xiao, Z. (2016). Improving Users’ Demographic Prediction via the Videos They Talk About. In *Proc. of the Conference on Empirical Methods in Natural Language Processing (EMNLP16),* 1359-1368.	Bayesian estimation	Keyword mentions	Weibo (Chinese platform similar to Twitter)	Gender, Age
Zagheni, E., Garimella, V. R. K., Ingmar, W., & State, B. (2014). Inferring International and Internal Migration Patterns from Twitter Data. *In Proceedings of the 23rd International Conference on World Wide Web*, 1-6	Facial evaluation: automated	Profile photos	Twitter	Age, Gender

Zhang, C., & Zhang, P. (2010). Predicting Gender from Blog Posts. Technical Report. University of Massachusetts Amherst, USA.	Supervised learning: SVM	User posts	A variety of blogging platforms	Gender
Mislove, A., Lehmann, S., & Ahn, Y. (2011). Understanding the Demographics of Twitter Users. In *Proceedings of the 5th International Conference on Weblogs and Social Media (ICWSM 11).* Barcelona, Spain.	Data matching/ Adjusted data matching w/Bayesian Estimation	User names, User location	Twitter	Gender, Race/Ethnicity
Mohammady, E., & Culotta, A. (2014). Using County Demographics to Infer Attributes of Twitter Users. In ACL 2014 Joint Workshop on Social Dynamics and Personal Attributes in Social Media Proceedings of the Workshop Baltimore , Maryland , USA (pp. 7–17).	Supervised learning: OLS regression	User location, user tweets, User name	Twitter	Race/ethnicity
Mueller, J., & Stumme, G. (2016). Gender Inference using Statistical Name Characteristics in Twitter. 5th ASE International Conference on Social Informatics (SocInfo 2016), Union, NJ, USA, August 15-17, 2016. Proceedings, 47:1--47:8.	Supervised learning: SVM classification	User names (not handles)	Twitter	Gender
Mukherjee, A., & Liu, B. (2010). Improving Gender Classification of Blog Authors. *In Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, 207–217)	Supervised learning: SVM regression	User posts	A variety of blogging platforms	Gender
Nguyen, D., Gravel, R., Trieschnigg, D., & Meder, T. (2013). “How old do You Think I Am?”: A Study of Language and Age in Twitter. *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media*, Cambridge, Massachusetts, USA, 439–448.	Supervised learning: OLS and logistic regression	User tweets	Twitter	Age (category, numeric, life stage)
Nguyen, D., Gravel, R., Trieschnigg, D., & Meder, T. (2013). TweetGenie: Automatic Age Prediction from Tweets. *ACM SIGWEB Newsletter*, 1–6. https://doi.org/10.1145/2528272.2528276	Supervised learning: Logistic regression	User tweets	Twitter	Age
Nguyen, D., Smith, N., & Rosé, C. (2011). Author Age Prediction from Text using Linear Regression. In *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities,* 115–123.	Supervised learning: OLS regression	User posts	blogger/online breast cancer forums	Age
Nguyen, T., Phung, D., Adams, B., & Venkatesh, S. (2011). Prediction of age, sentiment, and connectivity from social media text. In International Conference on Web Information Systems Engineering (pp. 227–240). Springer. Retrieved from http://link.springer.com/10.1007%2F978-3-642-24434-6_17	Supervised learning: Logistic regression	User posts	Livejournal	Age
Nowson, S., & Oberlander, J. (2006). The Identity of Bloggers: Openness and Gender in Personal Weblogs. In *AAAI Spring Symposium: Computational approaches to analyzing weblogs*, 163–167.	Supervised learning: SVM classification	User posts	Blogs	Gender
Oktay, H., Firat, a, & Ertem, Z. (2014). Demographic breakdown of Twitter users: An analysis based on names. In *Proceedings of the ASE BIGDATA/SOCIALCOM/CYBERSECURITY Conference*, 1–11.	Adjusted data matching w/Bayesian estimation	User names	Twitter	Age, Race/Ethnicity

Methodology

This section includes the description of the used dataset and the statistical analysis which was performed to ensure the dataset's appropriateness and to design the research path of work.

The OSWINDS’ Dataset

Description: A dataset collected by the OSWINDS[6] group was exploited for the design and development needs of our methodology. The dataset collection was conducted during the period of July-September 2017 (20/7/2017-26/9/2017) and includes geo-located data (tweets) from the New York region. The dataset provides information on the activities of **10,332 unique users** on Twitter, including the following elements for each user:

a) statuses count
b) tweets (whole .json)
c) number of followers and their IDs
d) number of friends and their IDs
e) language
f) date

Collection Method: The dataset collection was carried out using a suitable crawler that was developed in Python language and exploited the Twitter Streaming API[7] which gives access to 1% of all tweets. Initially, random data was collected for a specified location, and 10,332 unique user IDs were extracted. Then, friends and followers of the users were searched for, and finally around 15K friends and followers were found. Subsequently, the last 200 tweets for each user were gathered and this step took place twice. Thus, about 400 tweets per user were collected. Finally, all the doubles were removed and the final dataset was emerged.

Statistical Analysis

The statistical analysis carried out with the aim of extracting and checking the validity of generic metrics, consists of three different stages. These stages and their respective results are outlined below.

1st Stage

The number of unique users was measured, and the maximum values of friends, followers, statuses and tweets were found. The dataset contains the activity of 10,332 Twitter users and **3,657,396**

[6] http://oswinds.csd.auth.gr/

[7] https://dev.twitter.com/streaming/overview

tweets. The maximum number of friends, followers, tweets and statuses corresponding to a random user are:

14,371 friends

18,932 followers

2,380,119 statuses and

898 tweets

The **average number of tweets per user** was found to be equal to **354**.

The total number of **geo-tagged tweets** is 389,693 tweets (~ 94 geo-tagged tweets per user) and corresponds to 10.7% of total number of tweets. These geo-tagged tweets were published by 4,149 different users (40.16% of total number of users).

2nd Stage

The next step was to calculate the cumulative probability distribution for the available values (friends, followers, statuses and tweets) and to draw the corresponding Cumulative Distribution Function (CDF) charts that are depicted in Figures 1 & 2.

The CDF charts which are depicted in Fig. 1 & Fig. 2 (a) have revealed the following:

About **90.7%** (9,372) of users have fewer or equal to **2,000 friends**

About **87.43%** (9,033) of users have fewer or equal to **2,500 followers**

About **97.5%** (10,074) of users have created fewer or equal to **100,000 statuses**

As shown, the above numbers of friends, followers and statuses represent the vast majority of users. Comparing these values to the corresponding maximum values, there is a large deviation. Another finding is that few users exceed the above values and achieve the maximum values found in the sample. With respect to users’ followers, Kwak et al. (2010) have demonstrated that Twitter followers follow a non-power-law distribution.

The CDF chart related to users’ tweet (Fig. 2 (b)) differs from other charts. This is divided into four distinct intervals as follows:

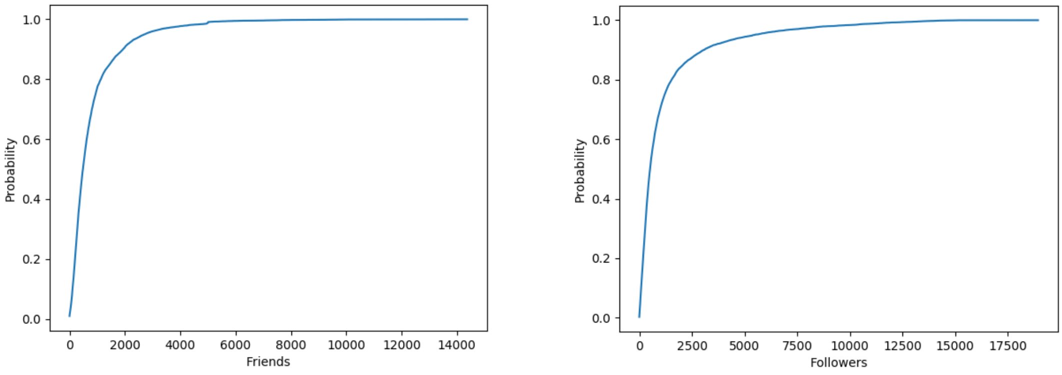
[0, 200] tweets with 577 (**5.58%**) users

(200, 400] tweets with 7,303 (**70.68%**) users

(400, 600] tweets with 2,214 (**21.43%**) users and

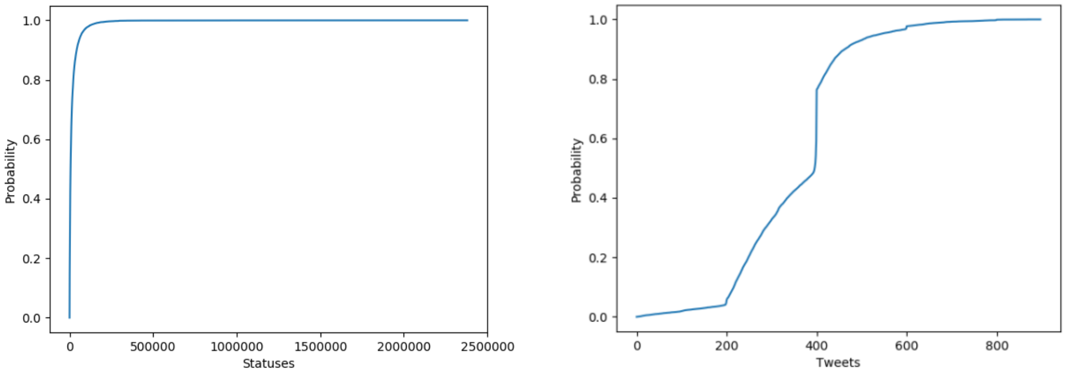
(600, 898] tweets with 238 (**2.3%**) users

As expected, due to the collection process, the majority of users are found in the interval (200, 400], followed by users correspond to the interval [0, 200]. Users with less than 200 tweets are probably those who do not publish many tweets and use Twitter to be informed of the latest news or events. On the contrary, users with equal or more than 400 tweets appear to create often tweets and their content can be replicated through re-tweets. According to Kwak (2010), the propagation of a re-tweet depends mainly on the popularity of the original tweet itself and not on the number of followers of the user.



(a) (b)

Figure 1: CDF charts of Friends and Followers



(a) (b)

Figure 2: CDF charts of Statuses and Tweets

A) 3rd Stage

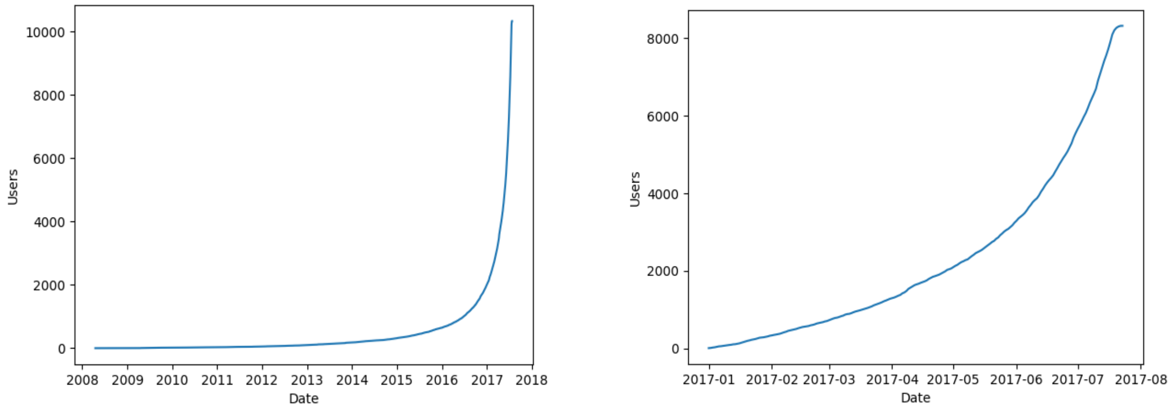
Subsequently, we identified for each user the date on which he/she published his/her first tweet and we calculated the cumulative frequency distribution per year, which is depicted in Fig. 3 (a). As depicted, the majority of users (80.51%) published their first tweet in 2017. The highest participation rate of users is also shown in 2017, when about 41 users / day publish their first tweet. In detail, our findings for the years 2015-2017 are as follows:

2017: 8,318 new users in 204 days → 80.51%

2016: 1,368 new users in 335 days → 13.24%

2015: 329 new users in 211 days → 3.18%

As shown in Fig. 3(b), around 4,000 users have published their first tweets in the first half of 2017. This number almost doubles in the following two months, since the last day of the sample is the 23th of July, 2017 in which 8,318 users correspond.



(a) Unique Users per Year

(b) Unique Users per Month

Figure 3: CDF charts of Unique Users based on their first Tweet

The number of total tweets per user in 2017 is depicted in Fig. 4. In first five months of 2017, the number of tweets per user published daily ranges between 2 and 4 and is almost constant. Then, the number of tweets per user is growing dynamically and in the middle of July, presents the first peak which corresponds to 14 tweets per user. Subsequently, the number of tweets per user is reduced to the original levels, rising again, and shows the second peak corresponding to 14 tweets per user in the end of September.

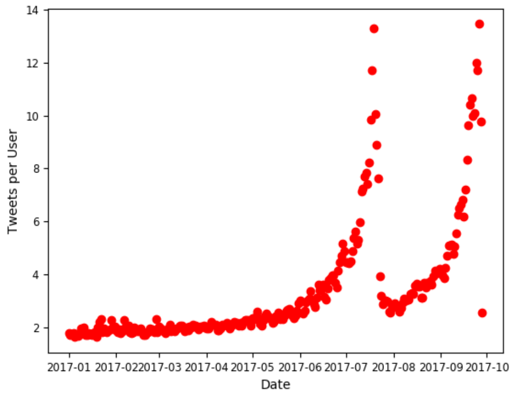


Figure 4: Tweets per User per Month

Data Cleaning: Identification & removal of professional accounts

The statistical analysis led us to understand the OSWINDS' dataset and validate its suitability for the purposes of our research. The algorithm developed to identify and extract the ages of Twitter users provided the first results and allowed us to understand the need to distinguish the accounts of organizations and companies from the accounts of individuals. The separation of accounts was performed using previous studies (Efstathiades et al., 2015; Yang et al., 2013; Lee et al., 2010; Stringhini et al., 2010; Wang, 2010) and combining the following three criteria: a) number of followers, b) fofo ratio (i.e., which is the ratio of the number of an account's friends to its followers), and c) reputation score. The findings revealed that 9718 (94.06%) accounts correspond to individuals, while 614 (5.94%) accounts are professional.

Our work in Task 3.3 has resulted in two publications

V. Moustaka, Z. Theodosiou, A. Vakali, A. Kounoudes. 2018. Smart cities at risk!: privacy and security threats borderlines from social networking in cities. In the 2018 Web Conference Companion (WWW' 18 Companion), April 23-27, 2018, Lyon, France, ACM, New York, NY, 6 pages. DOI: https://doi.org/10.1145/3184558.3191516

This article investigates the security and privacy issues of OSNs in the Smart Cities context and proposed a novel model which specifies the relationships between privacy and security threats.

V. Moustaka, Z. Theodosiou, A. Vakali, A. Kounoudes., L.-G. Anthopoulos. 2018. Enhancing Social Networking in Smart Cities: Privacy and Security Borderlines (under review for publication in Journal of Technological Forecasting and Social Change, Elsevier)

This article deals with security and privacy issues aiming to answer:

RQ1: What vulnerabilities lie behind individuals activities in OSNs in SC that threaten individuals privacy and security?

RQ2: What are the individuals' behavioral patterns in OSNs that could be exploited by SC stakeholders, with the purpose of adopting and applying appropriate policies aiming at strengthening protection of individuals during social networking and encouraging their participation in SC?

Future Work

The future work includes the following: a) the creation of a ground truth dataset, b) the extraction of the Twitter users' age groups, and c) the validation of these age groups using the ground truth dataset.

The algorithm that classifies twitter users according to their age will be used in our software to detect the kind of conversation that is taking place between the minor and the other OSN users.

5. Application on the parental console

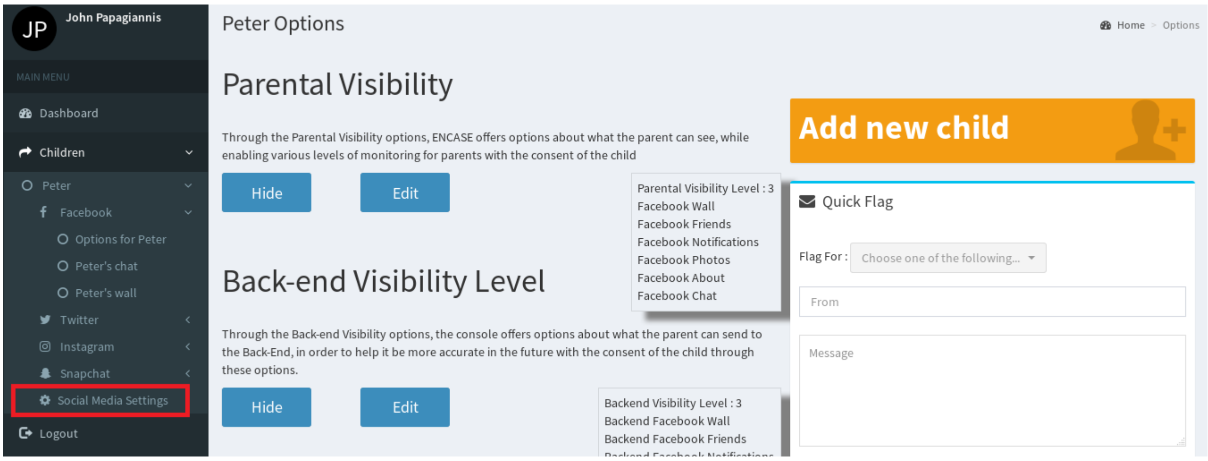
The work of Angulo & Ortlieb (2015) is one of the very few tries which aims to unveil and understand common online privacy panic situations. The authors presented an exploratory study on common experiences of online privacy-related panic and on users’ reactions to frequently occurring privacy incidents. By using the metaphor of a privacy panic button, they have investigated users’ expectations and mental models of suitable help mechanisms that could lead these users towards a solution, calming their distress, and preventing similar episodes from happening in the future. Through user semi-structured interviews (n = 16) and a survey (n = 549), they have identified 18 scenarios of privacy panic situations. The results have shown that victims’ topmost worries included possible harm to their finances or fear of embarrassment, as well as third-parties knowing things that might not be of their business. Among the most memorable self-reported panic stories were cases of account hijacking and ‘leakage’ of personal data, while incidents involving regrets when sharing content online were found to be experienced most frequently. However, scenarios related to the loss of online data, the loss of a mobile device, or falling pray of identity theft also were at the top of users’ concerns. Their findings also indicate that, in the case a service provider were to offer a hypothetical privacy panic button; users would expect that the help provided is immediate, uncomplicated, actionable, and in-place.

Based on these findings we integrate a traffic blocking option in the parental console, as described below.

Traffic blocking in case of critical situations

The parental Console is the tool with which the parent or educator is able to monitor the activity of the minor in a fine grained way. Based on the studies of Task 3.2, we set requirements in order to implement tools that will be deployed in the Parental Console in order for the parent or educator to identify and manage critical situations.

The guardian of the minor is able to use tools in order to block any outgoing and incoming traffic between the child and a specified website. The guardian can use the “Social Media Settings” button that he can find under the tools toolbar of his console in order to manage any critical situation. This button is shown in a red square in Figure 5.



Peter Options

Parental Visibility

Through the Parental Visibility options, ENCASE offers options about what the parent can see, while enabling various levels of monitoring for parents with the consent of the child

Parental Visibility Level : 3
Facebook Wall
Facebook Friends
Facebook Notifications
Facebook Photos
Facebook About
Facebook Chat

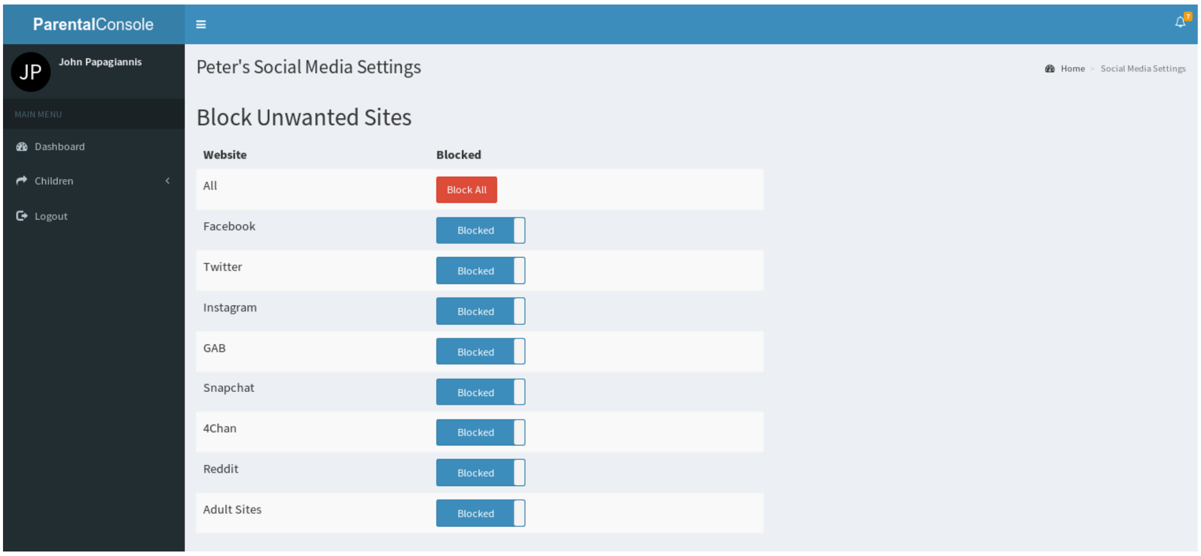
Back-end Visibility Level

Through the Back-end Visibility options, the console offers options about what the parent can send to the Back-End, in order to help it be more accurate in the future with the consent of the child through these options.

Backend Visibility Level : 3
Backend Facebook Wall
Backend Facebook Friends

Figure 5. Additional button to manage the incoming and outgoing traffic

This button was added during the development of the Parental console, so that the parent can manage the incoming and outgoing traffic between a specific minor and a website. Also the parent can block all internet traffic with the push of a button. These tools aim for the parent to instantly block any unwanted website if he senses that something is wrong. In figure 6 below, the options given to the guardian are shown.



Peter's Social Media Settings

Block Unwanted Sites

Website	Blocked
All	Block All
Facebook	Blocked
Twitter	Blocked
Instagram	Blocked
GAB	Blocked
Snapchat	Blocked
4Chan	Blocked
Reddit	Blocked
Adult Sites	Blocked

Figure 6. Traffic blocking options

The guardian has the ability to block any access to the internet. In addition, he can select to block specific website access and traffic. When the parent makes a selection, the url of the blocked site is sent through a web socket to the IWP database. During incoming and outgoing traffic, the IWP will check if that minor has any records in that database. In case any traffic comes from that site, then the IWP will just drop the contents of that traffic and the child will be redirected to a site explaining that this content is blocked by ENCASE.

Future work includes the creation of a text input box where the guardian of the child can enter the site that he wishes to block the traffic for. All the input of the guardians and educators will be sent to the Back-End of our infrastructure in order for the Back-End to know that the site with that specific name was blocked. This way, the administrators of the system will be able to get feedback of any harmful sites that were not included in the “Traffic Blocking” options of the parent. In addition, through the add-on, the minor will be able to see a list of the sites that he has no access to. In addition, in this page the parent will be able to read our research on what to do in case of a critical situation like password loss, account high jacked, extreme malicious behavior activity, etc.

6. Conclusions

The main work described in this deliverable includes the study and analysis of users’ behavior and experience when they faced security and privacy risks on OSNs. Towards this end more than 70 research articles published the last years in scientific journals and proceedings of international conferences were selected and analysed and several aspects related to the user behavior and experience, security threats and privacy risks, and privacy leakage on OSNs were investigated.

The results show that the behavior of individuals in OSNs is quite difficult to be adequately clarified and predicted, while is determined mainly by psychological (personal), demographic factors and socio-political factors. It is a fact that there are only a few studies on children’s behavior and attitude on OSNs (e.g., Youtube, Instagram, etc). Finally, the disclosure of children's information on OSNs is mainly dependent on their parents and the closed family circle (e.g., siblings, relatives, etc).

Studying the work of Angulo J. & Ortlieb M. (2015) “WTH..!?!” Experiences, reactions, and expectations related to online privacy panic situations”, we identified the users’ topmost concerns regarding their privacy on OSNs and we are implementing a traffic blocking option in our parental console.

Future Work includes the optimization of the algorithm used to detect ages, the expansion of the ground truth dataset, the extraction of the Twitter users' age groups by exploiting text-, user- and network-based features and using unsupervised learning methods and the verification and validation of these age groups using the ground truth dataset.

Our work in Task 3.2 has resulted in two publications

1. V. Moustaka, Z. Theodosiou, A. Vakali, A. Kounoudes. 2018. Smart cities at risk!: privacy and security threats borderlines from social networking in cities. In the 2018 Web Conference Companion (WWW’ 18 Companion), April 23-27, 2018, Lyon, France, ACM, New York, NY, 6 pages. DOI: https://doi.org/10.1145/3184558.3191516
2. V. Moustaka, Z. Theodosiou, A. Vakali, A. Kounoudes., L.-G. Anthopoulos. 2018. Enhancing Social Networking in Smart Cities: Privacy and Security Borderlines (under review for publication in Journal of Technological Forecasting and Social Change, Elsevier)

Age classification will be used in ENCASE to be able to detect the kind of conversation that is taking place between a kid and another OSN user.

7. References

1. A. Bergström. 2015. Online privacy concerns: A broad approach to understanding the concerns of different groups for different uses. Computers in Human Behavior 53, Dec. 2015, 419-426. DOI: https://doi.org/10.1016/j.chb.2015.07.025
2. A. Martínez-Balleste, P.-A. Pérez-Martínez, and A. Solanas. 2013. The Pursuit of Citizens' Privacy: A Privacy-Aware Smart City Is Possible. IEEE Communications Magazine 51, 6 (June 2013), 136-141. DOI: http://dx.doi.org/10.1109/MCOM.2013.6525606
3. A. Saravanaraj, J. I. Sheeba, and S. Pradeep Devaneyan. Automatic Detection of Cyberbullying from Tw`itter. IJCSITS, 6, 2016.
4. Acquisti A., & Gross R. (2006). "Imagined communities: Awareness, information sharing, and privacy on the Facebook". In Proceedings of 6th International Conference on Privacy Enhancing Technologies (PET'06), Cambridge, UK, June 28-30, 2006. DOI: https://doi.org/10.1007/11957454_3
5. Alqatawna J., Madain A., Al-Zoubi A.M., Al-Sayyed R. (2017). "Online Social Networks Security: Threats, Attacks, and Future Directions". In: Taha N., Al-Sayyed R., Alqatawna J., Rodan A. (eds) Social Media Shaping e-Publishing and Academia. Springer, Cham
6. Angulo J. & Ortlieb M., (2015). "WTH..!?!: Experiences, reactions, and expectations related to online privacy panic situations", In the Symposium on Usable Privacy and Security (SOUPS) 2015, July 22-24, 2015, Ottawa, Canada.
7. Anton, A.-I., Earp, J.-B., Young, J.-D. (2010). "How internet users' privacy concerns have evolved since 2002. IEEE Security & Privacy, Vol. 8, No. 1, pp. 21-27. DOI: https://doi.org/10.1109/MSP.2010.38
8. Ball A.-L., Ramim M.-M., Levy Y. (2015). "Examining users' personal information sharing awareness, habits, and practices in social networking sites and e-learning systems". Online Journal of Applied Knowledge Management, Vol. 3, No. 1, pp. 108-207. Retrieved June 2017 from http://www.iiakm.org/ojakm/articles/2015/volume3_1/OJAKM_Volume3_1pp180-207.pdf
9. C. Nobata, J. Tetreault, A. Thomas, Y. Mehdad, and Y. Chang. Abusive Language Detection in Online User Content. In WWW, 2016
10. C. Patsakis, A. Zigomitros, A. Papageorgiou and A. Solanas. 2014. Privacy and Security for Multimedia Content shared on OSNs: Issues and Countermeasures, Computer Journal 58, 4, 518-535. DOI: https://doi.org/10.1093/comjnl/bxu066
11. C. Van Hee, E. Lefever, B. Verhoeven, J. Mennes, B. Desmet, G. De Pauw, W. Daelemans, and V. Hoste. Automatic detection and prevention of cyberbullying. In HUSO, pages 13–18, 2015
12. C. Zhang and J. Sun. 2010. Privacy and Security for Online Social Networks: Challenges and Opportunities. IEEE Network 24, 4 (July-August 2010). DOI: 10.1109/MNET.2010.5510913
13. Cesare N., Grant C., Nsoesie E.-O., 2017. "Detection of User Demographics on Social Media: A Review of Methods and Recommendations for Best Practices". Available at:

https://arxiv.org/abs/1702.01807

14. Charnigo L., & Barnett-Ellis P. (2007). "Checking out Facebook.com: The impact of a digital trend on academic libraries". Information Technology and Libraries, Vol. 26, pp. 23-34. DOI: https://doi.org/10.6017/ital.v26i1.3286
15. Children's Online Privacy Protection Act (COPPA). (1998). Accessed June 2017: https://www.ftc.gov/enforcement/rules/rulemaking-regulatory-reform-proceedings/childrens-online-privacy-protection-rule
16. Consolvo S., Smith I.-E., Matthews T., LaMarca A., Tabert J., Powledge P. (2005). "Location disclosure to social relations: Why, when, & what people want to share". In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05), Portland, Oregon, USA, April 02-07, 2005, pp. 81–90. ACM. DOI: https://doi.org/10.1145/1054972.1054985
17. Cranor L.-F. (2003). "I didn't buy it for myself: privacy and ecommerce personalization". In the 2003 ACM workshop on privacy in the electronic society, pp. 111–117 ACM Press, Washington, DC
18. D. Chatzakou, N. Kourtellis, J. Blackburn, E. De Cristofaro, G. Stringhini, and A. Vakali. Mean Birds: Detecting Aggression and Bullying on Twitter. In WebSci 2017
19. Dehue F., Bolman C., and Völlink T. (2008). "Cyber-bullying: Youngsters' Experiences and Parental Perception" CyberPsychology & Behavior. April 2008, Vol. 11, No. 2, pp. 217-223. DOI: https://doi.org/10.1089/cpb.2007.0008
20. Dey R., Ding Y., Ross K.-W. (2013). "Profiling High-School Students with Facebook: How Online Privacy Laws Can Actually Increase Minors' Risk". In Proceedings of the 2013 Conference on Internet Measurement Conference, Barcelona, Spain, Oct. 23-25, 2013, pp. 405-416. DOI: https://doi.org/10.1145/2504730.2504733
21. Efstathiades, H., D. Antoniades, G. Pallis and M.- D. Dikaiakos. 2015. "Identification of Key Locations based on Online Social Network Activity". In *Proceedings of 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining.* DOI: https://doi.org/10.1145/2808797.2808877
22. Ellingsen D.-M., Leknes S., Loseth G., Wessberg J., Olausson H. (2016). "The Neurobiology Shaping Affective Touch: Expectation, Motivation, and Meaning in the Multisensory Context". Front. Psychol. DOI: https://doi.org/10.3389/fpsyg.2015.01986
23. F. Luo, G. Cao, K. Mulligan and X. Li. 2016. Explore spatiotemporal and demographic characteristics of human mobility via Twitter: A case study of Chicago. Applied Geography 70, May 2016, 11-25. DOI: http://dx.doi.org/10.1016/j.apgeog.2016.03.001
24. Fire M., Goldschmidt R., Elovici Y. (2014). "Online Social Networks: Threats and Solutions", IEEE Communication Surveys & Tutorials, Vol. 16, No. 4, Fourth Quarter 2014, pp. 2019-2036
25. Flekova L., Jordan Carpenter and Salvatore Giorgi, Lyle Ungar and Daniel Preot¸iuc-Pietro, 2016. "Analyzing Biases in Human Perception of User Age and Gender from Text". In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pp. 843–854, Berlin, Germany.
26. Fogel J. & Nehmad E. (2009). "Internet social networking communities: Risk taking, trust, and privacy concerns". Computers in Human Behavior, Vol. 25, No. 1, pp. 153-160. DOI: https://doi.org/10.1016/j.chb.2008.08.006

27. Furnell S. (2008). "End-user security culture a lesson that will never be learnt?", Computer Fraud & Security, Vol. 2008, No. 4, pp. 6-9. DOI: https://doi.org/10.1016/S1361-3723(08)70064-2
28. G. E. Hine, J. Onaolapo, E. De Cristofaro, N. Kourtellis, I. Leontiadis, R. Samaras, G. Stringhini, and J. Blackburn. Kek, Cucks, and God Emperor Trump: A Measurement Study of 4chan's Politically Incorrect Forum and Its Effects on the Web. In ICWSM, 2017
29. G. Rizzo, R. Meo, R.-G. Pensa, G. Falcone, and R. Troncy. 2016. Shaping City Neighborhoods Leveraging Crowd Sensor. Information Systems, July 2016. DOI: http://dx.doi.org/10.1016/j.engappai.2012.05.005
30. Gkatziaki, V., Giatsoglou, M., Chatzakou, D., Vakali, A., 2017. "DynamiCITY: Revealing city dynamics from citizens social media broadcasts", *Information Systems*. DOI: 10.1016/j.is.2017.07.007
31. Gross R. & Acquisti A. (2005). "Information Revelation and Privacy in Online Social Networks (The Facebook case)". In the Proceedings of the 2005 ACM workshop on Privacy in the Electronic Society (WPES'05), Alexandria, Virginia, USA, November 7, 2005. DOI: https://doi.org/10.1145/1102199.1102214
32. H. Hosseinmardi, R. Han, Q. Lv, S. Mishra, and A. Ghasemianlangroodi. Towards understanding cyberbullying behavior in a semi-anonymous social network. In IEEE/ACM ASONAM, 2014.
33. H. Hosseinmardi, S. A. Mattson, R. I. Rafiq, R. Han, Q. Lv, and S. Mishra. Analyzing Labeled Cyberbullying Incidents on the Instagram Social Network. In SocInfo, 2015.
34. H. Kwak, J. Blackburn, and S. Han. Exploring Cyberbullying and Other Toxic Behavior in Team Competition Online Games. In CHI, 2015
35. Hinduja S. & Patchin J.-W. (2008). "Cyber-bullying: an exploratory analysis of factors related to offending and victimization". Deviant Behavior, Vol. 29, pp. 129-156. DOI: 10.1080/01639620701457816
36. Hossain N., Tianran Hu, Roghayeh Feizi, Ann Marie White, Jiebo Luo, Henry Kautz, 2016. "Inferring Fine-grained Details on User Activities and Home Location from Social Media: Detecting Drinking-While-Tweeting Patterns in Communities". Available at: https://arxiv.org/abs/1603.03181
37. Hugl U. (2011). "Reviewing person's value of privacy of online social networking". Internet Research, Vol. 21, No. 4, pp. 384-407. DOI: http://dx.doi.org/10.1108/10662241111158290
38. I. Kayes, N. Kourtellis, D. Quercia, A. Iamnitchi, and F. Bonchi. The Social World of Content Abusers in Community Question Answering. In WWW, 2015
39. J. A. Patch. Detecting bullying on Twitter using emotion lexicons. 2015
40. J. Fox and W. Y. Tang. Sexism in online video games: The role of conformity to masculine norms and social dominance orientation . Computers in Human Behavior, 33, 2014.
41. J.-M. Xu, X. Zhu, and A. Bellmore. Fast Learning for Sentiment Analysis on Bullying. In WISDOM, 2012
42. Jin T., Chen Y., Wang T., Pan H., Vasilakos A.-V. (2013). "Understanding User Behavior in Online Social Networks: A Survey". IEEE Communication Magazine, Vol. 51, No. 9, pp. 144-150. DOI: https://doi.org/10.1109/MCOM.2013.6588663

43. K. Dinakar, R. Reichart, and H. Lieberman. Modeling the detection of Textual Cyberbullying. The Social Mobile Web, 11, 2011.
44. Kisekka V., Bagchi-Sen S., Raghav Rao H. (2013). “Extent of private information disclosure on online social networks: An exploration of Facebook mobile phone users”. Computers in Human Behavior, Vol. 29, (2013), pp. 2722–2729. DOI: https://doi.org/10.1016/j.chb.2013.07.023
45. Kisekka, V., Bagchi-Sen, S., Raghav Rao, H., 2013. “Extent of private information disclosure on online social networks: An exploration of Facebook mobile phone users. *Computers in Human Behavior*, 29 (2013), 2722–2729. DOI: https://doi.org/10.1016/j.chb.2013.07.023
46. Knijnenburg B.-P., Kobsa A., Jin H. (2013). “Dimensionality of information disclosure behavior”. International Journal of Human Computer Studies, Vol. 71, No. 12, pp.1144-1162. DOI: https://doi.org/10.1016/j.ijhcs.2013.06.003
47. Kowalski, R. M., & Limber, S. P. (2007). “Electronic bullying among middle school Students”. Journal of Adolescent Health, Vol. 41, No.6, pp. 22–S30. DOI:10.1016/j.jadohealth.2007.08.017
48. Kwak H., Lee C., Park H., and Moon S., 2010. “What is Twitter, a Social Network or a News Media?”. In *Proceedings of the 19th international conference on World Wide Web (WWW’10),* pp. 591-600. DOI: https://doi.org/10.1145/1772690.1772751
49. Lampos V., Aletras N., Geyti J.K., Zou B., Cox I.J., 2016. “Inferring the Socioeconomic Status of Social Media Users Based on Behaviour and Language”. *In: Ferro N. et al. (eds) Advances in Information Retrieval. ECIR 2016. Lecture Notes in Computer Science*, vol 9626. Springer, Cham
50. Laorden C., Sanz B., Alvarez G., Bringas P. (2012). “A threat model approach to threats and vulnerabilities in on-line social networks”. Computational Intelligence in Security for Information Systems, pp. 135-142.
51. Lee K., J. Caverlee, and S. Webb, 2010. “Uncovering Social Spammers: Social Honeypots + Machine Learning”. In *ACM SIGIR Conference (SIGIR)*, 2010.
52. Levy Y., & Ramim, M.-M. (2009). “Initial development of a learners’ ratified acceptance of multibiometrics intentions model (RAMIM)”. Interdisciplinary Journal of E-Learning and objects, Vol. 5, pp. 378-319. http://nsuworks.nova.edu/gscis_facarticles/31
53. Li Y., Li Y., Yan Q., Deng R.-H. (2015). “Privacy leakage analysis in online social networks”. Computers & Security, Vol. 49, pp. 239-254. DOI: https://doi.org/10.1016/j.cose.2014.10.012
54. M. Dadvar, D. Trieschnigg, and F. Jong. Experts and machines against bullies: A hybrid approach to detect cyberbullies. In Canadian AI, 2014
55. M. Fire, R. Goldschmidt and Y. Elovici. 2014. Online Social Networks: Threats and Solutions, IEEE Communication Surveys & Tutorials 16, 4, Fourth Quarter 2014, 2019-2036
56. Madejski M., Johnson M., Bellovin S.-M. (2011). “The Failure of Online Social Network Privacy Settings”. Columbia University Academic Commons, DOI: https://doi.org/10.7916/D8NG4ZJ1
57. Miranda J., Mäkitalo N., Garcia–Alonso J., Berrocal J., Mikkonen T., Canal C., Murillo J.-M., 2015. From the Internet of Things to the Internet of People. *IEEE Internet Computing* 19, 2

(Mar.–Apr. 2015), 40–47. DOI: https://doi.org/10.1109/MIC.2015.24

58. Moustaka, V., Theodosiou, Z., Vakali, A., Kounoudes, A., 2018. “Smart cities at risk! : privacy and security threats borderlines from social networking in cities”. 2018 Web Conference Companion (WWW’ 18 Companion). ACM. DOI: https://doi.org/10.1145/3184558.3191516
59. Moustaka,V., Theodosiou, Z., Vakali, A., Kounoudes, A., (2018). “Smart cities at risk! : privacy and security threats borderlines from social networking in cities”. In the 2018 Web Conference Companion (WWW’ 18 Companion), April 23-27, 2018, Lyon, France, ACM, New York, NY. DOI: https://doi.org/10.1145/3184558.3191516
60. N. Djuric, J. Zhou, R. Morris, M. Grbovic, V. Radosavljevic, and N. Bhamidipati. Hate Speech Detection with Comment Embeddings. In WWW, 2015
61. Nguyen D., D. Trieschnigg, A. Seza, D. R. Gravel, T. Meder, and F. Jong, 2014. “Why gender and age prediction from tweets is hard: Lessons from a crowdsourcing experiment,” in *In Proceedings of COLING 2014*.
62. Nguyen D., N. A. Smith, and C. P. Rosé, 2011. “Author age prediction from text using linear regression,” in *Proceedings of the 5th ACLHLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, pp. 115–123.
63. Nguyen D.-P., R. Gravel, R. B. Trieschnigg, and T. Meder, 2013. “‘ How old do you think I am?’ A study of language and age in Twitter,”
64. Patsakis C., Zigomitros A., Papageorgiou A., Solanas A. (2014). “Privacy and Security for Multimedia Content shared on OSNs: Issues and Countermeasures”, Computer Journal, Vol. 58, No. 4, pp. 518-535. DOI: https://doi.org/10.1093/comjnl/bxu066
65. Patsakis, C., Zigomitros, A., Papageorgiou, A. and Solanas, A., 2014. « Privacy and Security for Multimedia Content shared on OSNSs: Issues and Countermeasures”, *Computer Journal*, 58 (4), 518-535. DOI: https://doi.org/10.1093/comjnl/bxu066
66. Pew Research Center’s Relationships Survey 2015. (2015). “Teens, Social Media & Technology Overview 2015”. Retrieved June 2017 from: http://www.pewinternet.org/2015/04/09/teens-social-media-technology-2015/pi_2015-04-09_teensandtech_07/
67. Rao D., Yarowsky D., Shreevats A., Gupta M., 2010. “Classifying Latent User Attributes in Twitter”. In *Proceedings of the 2nd international workshop on Search and mining user-generated contents (SMUC '10),* pp. 37-44. DOI: https://doi.org/10.1145/1871985.1871993
68. Rosenthal S. and K. McKeown, 2011. “Age prediction in blogs: A study of style, content, and online behavior in pre-and post-social media generations,” in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, 2011, pp. 763–772.
69. Siswanto E. and M. L. Khodra, 2013. “Predicting latent attributes of Twitter user by employing lexical features,” in *2013 International Conference on Information Technology and Electrical Engineering (ICITEE)*, pp. 176–180.
70. Statista. (2015). “Age distribution of active social media users worldwide as of 3rd quarter 2014, by platform”. Retrieved June 2017 from: https://www.statista.com/statistics/274829/age-distribution-of-active-social-media-users-worldwide-by-platform/

71. Stringhini G., S. Barbara, C. Kruegel, and G. Vigna, 2010. “Detecting Spammers On Social Networks”. In *Annual Computer Security Applications Conference (ACSAC’10)*
72. Stutzman F. (2006). “An evaluation of identity-sharing behavior in social network Communities”. iDMAa Journal, Vol. 3, No. 1, pp. 1-7. Retrieved April 2017 from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.91.617&rep=rep1&type=pdf
73. Sundar S.-S. & Marathe S.-S. (2010). “Personalization versus customization: The importance of agency, privacy, and power usage”. Human Communication Research, Vol. 36, No. 3, pp. 298–322. DOI: http://dx.doi.org/10.1111/j.1468-2958.2010.01377.x.
74. T. Minkus, K. Liu, and K.-W. Ross. 2015. Children Seen But Not Heard: When Parents Compromise Children’s Online Privacy. In Proceedings of the 24th International Conference on World Wide Web (WWW’15), 776-786. DOI: https://doi.org/10.1145/2736277.2741124
75. Toch E., Wang Y., Cranor L.-F. (2012). “Personalization and privacy: a survey of privacy risks and remedies in personalization-based systems”. User Modeling and User-Adapted Interaction, Vol. 22, No. 1, pp. 203-220. DOI: 10.1007/s11257-011-9110-z
76. Toch, E., Wang, Y., Cranor, L.-F., (2012). “Personalization and privacy: a survey of privacy risks and remedies in personalization-based systems”. User Model User-Adap Inter 22, pp. 203–220. DOI: 10.1007/s11257-011-9110-z
77. Tuli, G., 2015. “Modeling and Twitter-based Surveillance of Smoking Contagion,” Dissertation, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.
78. Turow J., King J., Hoofnagle C.-J., Bleakley A., Hennessy M. (2009). “Americans reject tailored advertising and three activities that enable it”. Retrieved from: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1478214.
79. Van der Velden M. & El Emam K. (2013). “Not all my friends need to know”: a qualitative study of teenage patients, privacy, and social media”. J Am Med Inform Assoc, Vol. 20, No. 1, pp. 16-24. DOI: https://doi.org/10.1136/amiajnl-2012-000949
80. Wang X., 2010. “Don’t follow me: spam detecting in Twitter”. In *Int’l Conference on Security and Cryptography (SECRYPT)*
81. Whittaker E. & Kowalski R.-M. (2015). “Cyber-bullying via Social Media”. Journal of School Violence, Vol. 14, No.1. DOI: http://dx.doi.org/10.1080/15388220.2014.949377
82. Y. Chen, Y. Zhou, S. Zhu, and H. Xu. Detecting Offensive Language in Social Media to Protect Adolescent Online Safety. In PASSAT and SocialCom, 2012.
83. Y.-M. Baek, E-M Kim, and Y. Bae. 2014. My privacy is okay, but theirs is endangered: Why comparative optimism matters in online privacy concerns Computers in Human Behavior 31, Feb. 2014, 48-56. DOI: https://doi.org/10.1016/j.chb.2013.10.010
84. Yang C., R. Harkreader, and G. Gu, 2013. “Empirical evaluation and new design for fighting evolving twitter spammers”. *IEEE Transactions on Information Forensics and Security,* 8(8):1280–1293, Aug 2013.
85. Youn S. (2005). “Teenagers' Perceptions of Online Privacy and Coping Behaviors: A Risk–Benefit Appraisal Approach”. Journal of Broadcasting & Electronic Media, Vol. 49, No.1, pp.86-110, DOI: 10.1207/s15506878jobem4901_6
86. Zagheni, E., Garimella, V. R. K., Ingmar, W., & State, B., 2014. “Inferring International and Internal Migration Patterns from Twitter Data”. In *Proceedings of the 23rd International*

Conference on World Wide Web. Available at: https://ingmarweber.de/wp-content/uploads/2014/02/Inferring-International-and-Internal-Migration-Patterns-from-Twitter-Data.pdf