Detecting psycho-anomalies on the world-wide web: current tools and challenges

DOMDOUZIS, Konstantinos <http://orcid.org/0000-0003-3679-3527>

Detecting Psycho-Anomalies on the World-Wide Web: Current Tools and Challenges

Konstantinos Domdouzis*
Department of Computing
Sheffield Hallam University
Sheffield
United Kingdom

ABSTRACT

The rise of the use of Social Media and the overall progress of technology has unfortunately opened new ways for criminals such as paedophiles, serial killers and rapists to exploit the powers that the technology offers in order to lure potential victims. It is of great need to be able to detect extreme criminal behaviours on the World-Wide Web and take measures to protect the general public from the effects of such behaviours. The aim of this chapter is to examine the current data analysis tools and technologies that are used to detect extreme online criminal behaviour and the challenges that exist associated with the use of these technologies. Specific emphasis is given to extreme criminal behaviours such as paedophilia and serial killing as these are considered the most dangerous behaviours. A number of conclusions are drawn in relation to the use and challenges of technological means in order to face such criminal behaviours.

* Corresponding Author address
Email: K.Domdouzis@shu.ac.uk

INTRODUCTION

The rise of Social Media and the expansion of the Internet in general created new opportunities but also led to many challenges in relation to how information is handled on the web. Especially for the field of criminology, it has opened new doors to the evolvement of crime and how this evolvement should be handled. The huge mass of information that is updated constantly and daily hides traps that threaten the security of citizens but also hides patterns that can reveal new types and frequencies of criminal behaviours. The worst case of criminal behaviours are those associated with psychopathy, Antisocial Personality Disorder (APD), serial killing, and paedophilia. Psychopathy is a personality disorder which is characterised by callousness and absence of empathy. Psychopathy is different from Antisocial Personality Disorder. Most psychopathic offenders meet criteria for APD, however most offenders with APD are not characterised by psychopathic traits (Shepherd et al. 2018).

Trolling is an antisocial behaviour which characterises Internet culture. The trolling behaviour is characterised by the initiation of aggressive arguments (Klempka & Stimson 2013) and the posting of inflammatory messages in order to provoke other people [Gammon, 2014]. Many people feel attracted to such type of behaviour (Craker & March 2016). Usually, Internet "trolls" are made up of people who fit the profile of narcissists, sadist, or psychopaths. These people like to hurt other people. "Gamergate" is an example of Internet trolling. It refers to the cyberbullying realised by a group of men in the gaming community, who cyber-bullied women who publicly criticized video game culture (Brewer 2019). Narcissism, Machiavellianism, psychopathy, and sadism have been associated with trolling behaviours and it is interesting to explore whether these traits are associated with trolling behaviours on online dating sites and location-based real-time dating (LBRTD) applications. The threat of being trolled in a LBRTD website is a common concert (Weiss, 2015).

Antisocial personality disorders are mental health conditions which are characterised by impulsive and often criminal behaviour. APD is characterised by variations on its severity and it can range from occasional bad behaviour to committing serious crimes (NHS 2018). Cyber bullies take advantage of Information and Communication Technologies (ICTs) in order to exhibit antisocial behaviour. The reasons for this behaviour are revenge, prejudice and intolerance, guilt, shame, pride and anger (Hoff & Mitchell 2009; Jones et al. 2011). Cyberbullying includes sending, posting, or sharing negative and harmful content. The most common places for cyberbullying include Social Media, text messaging, instant messaging (via email provider services and social media messaging) and email (stopbullying.gov 2018). In the past, there have been many incidents of cyberbullying. Jessica Logan was an Ohio high school senior who committed suicide after texting a nude picture of herself to her boyfriend. When they broke up, he sent the photo to everyone in the school and as a result, Jessica was harassed for months by her school peers. Jessica's grades dropped and she started skipping school, eventually killing herself (PureSight 2018).

The Internet has made prevalent a new type of sexual offender, the Internet-facilitated Rapists (IFR). IFRs use various online platforms to solicit their victims, such as dating sites, Internet chat rooms and social media sites. These platforms allow trust to be built-up quickly between the offender and the potential victim through the exchange of disclosures before face-to-face meetings. It is very possible that the victims engage in risky behaviours such as meeting to a private residence (Almond et al. 2017). Australian serial rapist Andrew James Benn used the social media (eg. Facebook, Tinder, Snapchat) in order to lure potential victims. He sexually assaulted 14 women that he met on social media in the Hunter Valley region of Sydney over a period of five years (Rigney 2018). Benn would check random female Facebook profiles searching for victims and trying to befriend them. After befriending them, he would try to gain their confidence and then arrange to meet in person before attacking them. In one occasion, Benn sent a Facebook friend request to a 17-year-old girl and he spent weeks chatting to her until the teenager agreed to meet him. The teenage girl told Benn that she had never been on a date before. He took her to a boat ramp at Morpeth in Maitland where he raped her in the back seat. He then drove her to a remote cemetery where he raped her again (Australian Associated Press 2018).

The fragmented and layered structure of the Internet enhances criminal activity as there is no centralised government body to enforce criminal laws in certain countries (Stalans & Finn 2016). A number of serial killers use the Internet and it is expected that in the future, serial killers will not leave even their house. A global cyberattack using hacking tools developed by the US National Security Agency (NSA) paralyzed the NHS, hot FedEx and infected computers in 150 counties (MacIntyre, 2017). Hackers have been spreading the "ransomware" called WannaCry. It is delivered via emails which trick the recipient into opening attachments and releasing the malware onto their system. The NHS trusts that were infected by the WannaCry ransomware were characterised by two types of disruptions. The first disruption is that the NHS staff were locked out of devices while the medical equipment and devices were also locked or isolated from the trusts' IT systems. The operation of the trusts' radiology and pathology departments was disrupted as the trusts relied on the equipment and devices for diagnostic imaging and for testing blood and tissue samples (Morse 2017).

Paedophiles also use the Internet in order to groom children. There are different types of online paedophiles. There are the people who view child pornography (online voyeurs), people who make and distribute child pornography and child sex abusers (online predators) (Corriveau & Greco 2018). Paedophiles can use social media (eg. Facebook) in order to try to communicate with potential victims. The example of the daughter of a social worker who was groomed by a paedophile through Facebook is a typical example of the dangers that the Web poses for children. The paedophile in the specific example managed to befriend other girls at the school of the social worker's daughter and when he came to request the specific girl, they had 32 mutual friends. The picture on the profile of the paedophile was a blurred picture of a teenager wearing what appeared to be a school uniform. The fact that they had 32 mutual friends "encouraged" the girl to accept his friendship request (Hannah H 2017). Usually, paedophiles that use the Internet to seduce potential victims initiate talk of sex, show to the victim pornography or they ask the victim to perform sexual acts with the intention that the victim's sexual arousal will restrict any inhibitions in relation to the engagement to sexual activity (Lanning 2002). In September 2017, a British court convicted Paul Leighton, 32, of raping a 1-year-old girl though he was approximately 6,000 km away from her. Leighton posed as a teenage girl and managed to convince a 14-year-old uncle to send him nude pictures of him. He then used these pictures in order to blackmail the boy into raping his niece (French 2017).

This chapter focuses on the tools used for the detection of extreme criminal activities through the World-Wide Web. These tools include data mining tools such as data mining algorithms. The use of the Internet and especially the Social Media have made the detection and profiling of criminal activities more complex. The chapter also focuses on the challenges faced related to the detection of these activities. A number of conclusions are drawn in relation to the use of specific tools for the detection of extreme criminal activities and the challenges faced in online crime tracing.

TOOLS FOR DETECTING EXTREME CRIMINAL ACTIVITIES

Thomas Hargrove is a homicide archivist. For the past eight years, he has collected municipal records of murders dating back to 1976. Hargrove uses code that he wrote in order to search his archive and specifically to detect statistical anomalies starting from the more ordinary murders, such as gang fights, robberies, or brawls. Hargrove has focused on how to use his code in order to detect serial killers. The code forms the basis of the Murder Accountability Project (MAP) (http://www.murderdata.org/) which includes a website, a database connected to it and a board of nine people, including former detectives, homicide experts and a forensic psychiatrist. The algorithm that is used in the code is based on data aggregation and it collects data for killings related to method, place, time and the victim's sex. The algorithm also considers whether the rate of unsolved murders in a city is high. In August 2010, Hargrove identified a pattern of murders in the city of Gary, Indiana. Hargrove contacted the police department of Gary providing information about the murders and asking them the question whether these murders were the result of the activity of a serial killer. Specifically, he noticed that between 1980 and 2008, fifteen women had been strangled and this could

be the work of a serial killer. The police department of Gary did not seem interested in checking Hargrove's data. Usually, the Department of Justice advises police departments to inform citizens in case there is a serial killer but some police departments hide this information. Four years later, the police department of Hammond (a town next to Gary) found the dead body of a 19-year-old woman in the bathtub of a motel. A guy called Daren Van was arrested. In the next days, the police recovered the bodies of six strangled women found in abandoned buildings just like the pattern described by Hargrove's algorithm (Wilkinson 2017).

The largest database of serial killers is the Radford Serial Killer Database which started by Dr Michael Aamodt back in 1992. Dr Aamodt is an Emeritus Professor of Psychology at Radford University, Virginia. Today, the database is the largest non-governmental serial murderer database globally. It includes 3304 subjects, including serial killers, mass murderers, etc. It also has over 9000 victim profiles and over 500 documents. A number of variables characterise each subject. These variables include background information, victim treatment and data about the committed crimes. The collected data come from websites, books, court documents and government agencies. The data have the form of a case study or statistical information (Florida Gulf Coast University (FGCU) 2019).

Online grooming detection is achieved through the use of data mining. Text mining approaches have been applied to analyse paedophile activities in chat rooms. One of the major data sources for the automatic detection of paedophiles is the chat logs dataset which is provided by the Perverted Justice (PJ) foundation. In this foundation, adults volunteer to enter chat rooms to act like minors. When conversations include sexual solicitations, the volunteers share the chat logs with the foundation and the authorities (Cano Basave et al. 2014).

Pendar (2017) has worked on the distinction of the conversations into those of predators and those of pseudo-victims. He characterises this dataset by implementing supervised and non-parametric classification models (Cano Basave et al. 2014, Pendar 2017). There are two categories of online text with sexually explicit content. The first category involves the interaction of a predator with what this person believes to be a victim. In this case, the victim can be a real underage victim or the predator could interact with a law enforcement officer without knowing it. The other category of online text is the consensual interaction between two adults. Pendar (2017) downloaded a number of text logs from the website 'www.perverted-justice.com'. These logs covered the conversations between 701 sexual predators and what they thought they were their underage victims. There was separation between the chat lines generated by the predator and the chat lines generated by the pseudo-victims. Then from these set of files (called corpus), a training set of files and a test set of files were created. The corpus was used as the test bench of the text categorization techniques. A series of Support-Vector Machines (SVMs) and k-NN classifiers were trained. k-NN classification is based on the representation of a document as a dimensional vector of weights. Each weight represents a measurement of a specific feature in a text. Unigrams, bigrams and trigrams from the training data were used as features. In text categorization and especially in the pre-processing stage, there was filtering of words while dimensionality reduction by feature extraction was achieved. The study showed that it was easy to automatically distinguish between a sexual predator and a pseudo-victim based on the contents of their text chat (Pendar 2017).

Term network is based on the occurrence of a number of key terms in the same text documents. The first step in the creation of a term network is the key term network extraction. The terms can be provided from a given list which is the output of NLP parsers or any key term extraction algorithms (Tseng, 1998). Examples of such algorithms include the use of NLP parsers to extract specific noun phrases (Schneider, 2006), the use of statistics for the extraction of new words, the use of Bayesian text classification from a range of text classification domains (Lee et al. 2011) or a combination of these algorithms. A fast key term extraction algorithm that uses very limited resources can be used. In this case, an objective way for automatic processing has to be specified. Once the extraction of key terms has been realised, term-to-term association across the whole document needs to be calculated through the use of a variety of similarity measures. In this case, the occurrence of each record in the document is recorded as a vector form. A vector element can take the values of 1 or 0 and thus

denoting the presence of a term in a document. The scanning of vectors can reveal the association between two terms. The final stage is the term network generation where a small term network is developed. The size of the network is defined through the real-time provision of a set of terms with similar attributes that need to be matched against the general term relation structure. The produced term network can be used for the visualization and analysis of entity relationships and this can be used in turn for a number of text-mining applications, such as crime exploration from a large amount of crime news or official criminal records (Tseng et al. 2012).

Galán-García et al. (2014) present how supervised learning can be used for the detection of troll profiles in the Twitter social network. They collected a number of Twitter profiles, the users' ID and their timeline tweets till they had 100 tweets containing abbreviations and slangs and they are not re-tweets. The tweets, their time of publication, their language, geo-position and the Twitter client are the data that were collected. The tweet helps clarifying the writing style of its author. The time of publication helps in the identification of the time that the users interact with the Social Media. The language and geo-position also allow data filtering and the clarification of the authorship of the tweet. Furthermore, independently on the devices used by users to access their Twitter account, they use their favourite Twitter client which can be used as an additional filtering mechanism. The next step is the generation of an Attribute Relation File Format (ARFF) file in order to classify profiles based on the writing style of the tweets and by using WEKA (Waikato University Environment for Knowledge Analysis). The performance of different classification algorithms was tested. Such algorithms are the Random Forest algorithm, the J48 algorithm, the K-Nearest Neighbour (KNN) algorithm and the Sequential Minimal Optimization (SMO) algorithm. The evaluation of the capabilities of the proposed approach in order to assign the correct authorship to Twitter profiles was based on the use of a dataset that includes 1900 tweets that correspond to 19 different Twitter accounts. The Random Forest algorithm includes a range of decision trees so that classification accuracy can be improved in the development of each individual classifier. The J48 is an open-source implementation of the C4.5 algorithm. The KNN algorithm analyses the k-nearest neighbours while the Sequential Minimal Optimization (SMO) is an algorithm which is used for the resolution of optimization problems created through the training of Support Vector Machines (SVMs). The proposed methodology has been tested in a real cyberbullying situation in the city of Bilbao. The staff of the school wanted to identify which of the student(s) was/were the authors behind the trolling profile. The profile was named "Gossip" and this profile was used to present indiscretions about a number of students. A large number of tweets (17536) that corresponds to the 92 users that followed the trolling profile were collected. Also, 43 tweets from the trolling profile were collected. Four classifiers were used to analyse the tweets. These classifiers are the SMO-PolyKernel, J48, SMONormalizedPolyKernel and RandomForest. The proposed methodology has been successful in identifying the users behind the trolling profile (Galán-García et al. 2014).

In 2017, Matthew Falder, a Cambridge University graduate, pleaded guilty to 137 charges making him one of the UK's worst paedophiles. In February 2018, he was jailed for 32 years. Falder blackmailed his victims into carrying depraved sexual acts. At least three of his victims attempted suicide. He committed the offences over a period of eight years and while he never met his victims, he blackmailed them over the Internet. Falder was a member of many "virtual communities" of abusers found in the Dark Web. He used the Dark Web to share pictures of violent sexual and physical abuse. He even encouraged one of his victims to rape a young child. Falder accessed encrypted dark web forums using his 'evilmind' and '666devil' accounts in order to control and devastate his victims. His arrest was the result of an organised collaboration of the National Crime Agency (NCA), the UK police, the UK Government Communications Headquarters (GCHQ), the US Homeland Security, the Australian police and Europol (Davies 2018). The case of Matthew Falder shows the dangers of Dark Web. Dark Web forums and file-sharing websites make it extremely difficult for the Law Enforcement Agencies (LEAs) to identify users or proactively shut down websites that promote criminal activity (Broadhurst et al. 2014). Dark Web offers a sense of anonymity to criminal users and criminal networks (Bleakley 2018. To mine the Dark Web for identifying conclusions for criminal activities is an extremely difficult and challenging task based on its vagueness. Stanford University has developed a prototype engine called the Hidden Web Exposer (HiWE) which attempts to mine the

Deep Web using a human-assisted approach. Similar products are the University of California's Infomine, Infoplease and PubMed. Deep Web Monitor is another tool developed by BrightPlanet which can check the entire web for data (Ovenden 2018).

In Nigeria, gender-based violence has been enhanced through the use of Social Media. Makinde et al. (2006) report five cases of gender-based violence associated with the use of social media. The victims met their perpetrator(s) through Facebook. The victims experienced different types of violence from their perpetrators, such as sexual, psychological or even economic violence. In one case, the victim died while two required immediate hospitalisation because of the severity of the trauma. Three of the victims were raped while in the other cases, there was an attempt for rape. Nigeria is characterised by an increment in the cases of gender-based violence initiated through meetings of people through social media platforms (Makinde et al. 2016). In another case, Stephen Port, a 41-year old chef living in East London, was sentenced to life in prison because of the murders of four young men that he lured to his apartment. Port drugged and raped his victims before murdering them. He used a dating site called Grindr in order to select his victims. Once he selected his victims, he arranged to meet them and lure them to his apartment (Simmons 2016). It is clear that the Social Media are the new way for many potential perpetrators to commit crimes. Injadat et al. (2016) performed a survey of data mining techniques used in social media. Examples of such techniques are Genetic Algorithms, Hierarchical Clustering, the k-Means algorithm, Linear Discriminant Analysis (LDA) and the Support Vector Machines (SVMs) (Injadat et al. 2016).

CHALLENGES IN THE IDENTIFICATION OF EXTREME CRIMINAL ACTIVITIES ON THE WORLD-WIDE WEB

The fundamental job of crime analysts and profilers still remains difficult and manual. Specific crime patterns cannot be identified even with the use of automated tools. Even nowadays, the most successful technique but also very time-consuming of identifying crime patterns is the manual review of crime reports and their comparison with past crimes reports. A number of issues are also associated with criminology. These issues are the study of crime prevalence, the methods for achieving crime prevention, the treatment of offenders and issues related to the definition of crime (Wang et al. 2013).

The introduction of Big Data in criminology led to a number of challenges. These challenges are associated with the capabilities of the existing Big Data analysis methods and the limitations of the current data processing systems (Jin et al., 2015). The comprehension of the notion of Big Data is also challenging (Hargittai, 2015), privacy issues related to them (Lazer et al., 2009) and ethical considerations related to mining Big Data (Boyd & Ctawford, 2012). The challenges associated with Big Data can be grouped in three categories. The first category is the data challenges related to data volume, veracity and quality. The second category is process challenges which are related to the processes of data capturing, integration and transformation. The third category is related to the management challenges which are associated with data privacy, security, governance and ethical aspects (Akerkar 2014, Zicari 2014). It is a big challenge to find faults in large quantities of data and it is even harder to correct them. It is also big challenge to develop an efficient, safe platform to share and distribute such data (Chen et al, 2015).

The World-Wide Web is characterised by huge amounts of information that are also diverse. These information exist in different formats such as tables and multimedia files. Furthermore, the integration of information from multiple pages is challenging as multiple pages show similar information in different words or formats and these information are generated from different authors. The information that is included in the Web can be noisy. This is because a typical web page can include many different types of information, out of which only a portion is significant for the website. Also, a large amount of information that exist on the Web are of low quality which can be changed constantly as most websites are dynamic. Moreover, the Web is a virtual society that does not include only information and services but also, virtual communities and organisations. A user can communicate

with users from other parts of the world and the monitoring of such interactions is challenging (Hosseinkhani et al. 2014).

It is vital for forensics experts to be able to extract data from cell phones. However, there is problem in involving cell phone manufacturers in this process due to time, cost and security. Similar problems with data extraction characterise telecommunications equipment, video game consoles and eBook readers. Techniques used to protect the intellectual property of these devices also pose a problem for their forensic analysis. However, all these devices can be used for the realisation of crimes and they may contain important forensic information. Furthermore, the great variety of modern day devices results to time-consuming court investigations. The processing of modern-day devices can take much time (Garfinkel, 2010).

Predictive policing is the application of analytics in order for targets of police intervention to be identified and crime prevention or resolution through the use of statistical predictions to be achieved. The use of predictive policing characterises the shift from reactive policing to proactive policing (Karppi 2018). Predictive policing uses algorithms in order to analyse patterns of criminal behaviour together with past, place and time of crime data in order to enable law enforcement agencies to predict places and times in which crimes can happen (Predpol 2016). However, as Gangadharan (2015) mentions, predictive policing can be discriminatory. The American Civil Liberties Union [ACLU], the Brennan Centre for Justice and a number of other civil rights organisations have all expressed concerns about the risk of the existence of bias when it comes to predictive policing software. Historical data based on police practices can create a feedback loop through which algorithms make decision that can classify in a biased manner which neighbourhoods are 'bad' and which are 'good' (Rieland 2018).

The 2014 EC3 Internet Organised Crime Threat Assessment (iOCTA) report states that the concept of territorially-based investigative approach is in conflict with the borderless nature of cybercrime. In the past, investigations had a national focus with some international connections, the focus has not shifted towards the coordination of international cybercrime operations. The same report describes the need for more efficient legal tools considering the current limitations of the Mutual Legal Assistance Treaty (MLAT) process and the process of further harmonization across the EU (Hayes et al. 2015).

A number of illicit activities take place in the Dark Web such as hiring assassins and the acquisition and delivery of child pornography. The Tor network is characterised by the lack of full-scale search engine functions such as those of Google. The Tor network is also vast. Despite the closure of SilkRoad, many drug dealers have found new ways through the various Deep Web sites. Terrorists also use dark web-based marketplaces in order to trade illegal weapons. The 'Armory' is the most well-known weapons marketplace in the dark web and it includes items ranging from bombs to bullets (Zulkarnine et al. 2016).

The law enforcement agencies often try to check IP addresses for reasons of tracing illegal activities over the Internet. The perpetrators of crimes often allege that their activities are conducted within the boundaries of their home and they consider the search that law enforcement agencies perform as violation of their privacy. Many of the crimes that are realised over the Internet also involve file sharing which allows interactions with third parties. These interactions discredit the legitimacy of claims in relation to privacy. However, it is difficult to fully prove what is violation of privacy and what it is not as the legal system is slow to adapt to the constant and very quick technological changes. As the law enforcement agencies' investigations become more intrusive, privacy advocates worry more about the future of online privacy. An example of how sensitive the topic of online privacy is can be seen in the Playpen cases. Playpen was a child pornography site which was hosted and used on the Dark Web. The Federal Bureau of Investigation (FBI) received a tip from a foreign informant about the host's IP address. Based on this tip, the FBI managed to locate the host and infiltrate the computers of the anonymous users that were using the site. The FBI used a 'Network Investigative Technique' (NIT) in order to send malware to users that were using the site for a short period of time. Once a user accessed the website and downloaded content from it, at the same time, the malware was also downloaded wit the selected content. This malware allowed FBI to acquire the

IP addresses of the anonymous users and as a consequence, their home addresses. Federal courts declared that the use of malware was legal in contrast to online privacy advocates who declared that this was an illegal search (Larson 2017).

CONCLUSIONS

Technology has managed to reveal how dangerous human nature can be. Crime has taken a different form and predators exploit nowadays the capabilities of modern technologies in order to target their potential victims. Paedophiles, trolls, people with antisocial personality disorders and cyberbullies use the World-Wide Web in order to approach innocent victims by manipulating their ways of approach. Data Mining includes a set of algorithms that can help extract information related to crimes and can help in the prediction of these crimes. As technology becomes more complex, the attempts of Law Enforcement Agencies to trace criminals become more difficult. Furthermore, the progress of technology has allowed criminals to secretly share information using very well-hidden technical platforms. There are also ethical issues related to the prediction of crimes. For example, the possibility of the existence of bias can affect the identification of a potential criminal and in this case, an innocent person can be prosecuted instead.

A number of challenges exist in relation to crime tracing on the Web. The increment of the amount of available data has created challenges related to their analysis and the privacy of the public. These privacy issues result to complications in terms of the limits of an investigation led by the Law Enforcement Agencies. Furthermore, there are legal complications that set boundaries to the better and more efficient collaborations between law enforcement agencies across different states. As crime evolves, the technology of criminal tracing and profiling also has to evolve. It also becomes obvious that the Social Media have affected the way human communication is expressed therefore current technology needs to constantly adapt to these dynamic changes.

REFERENCES

Akerkar, R. 2014. *Big data computing*. CRC Press. Florida, United States: Taylor & Francis Group

Almond, L., McManus, M.A., Chatterton, H. 2017. "Internet Facilitated Rape: A Multivariate Model of Offense Behavior." *Journal of Interpersonal Violence*. pp. 1–26

Australian Associated Press. 2018. "Rapist who used Facebook, Tinder and Snapchat to prey on 14 teenage girls is sentenced to 30 years in prison." Accessed 01 November 2018. https://www.dailymail.co.uk/news/article-6069741/Rapist-used-social-media-prey-14-teenage-girls-sentenced-30-years-prison.html

Bleakley, P. 2018. "Watching the watchers: Taskforce Argos and the evidentiary issues involved with infiltrating Dark Web child exploitation networks." *The Police Journal: Theory, Practice and Principles*. 1-16

Boyd, D. and K. Crawford (2012) "Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon", *Information, communication & society*. 15(5):662-679

Brewer, J. 2019. "Dealing With Psychopaths In The Internet Age." Accessed 20 October 2018. https://evolution-institute.org/dealing-with-psychopaths-in-the-internet-age/

Broadhurst, R, Grabosky, P, Alazab, M, Bouhours, B., Chon, S. 2014 "An analysis of the nature of groups engaged in cyber crime." *International Journal of Cyber Criminology*. 8(1): 1–20

Cano Basave, A., Fern´andez, M., Alani, H. 2014. "Detecting child grooming behaviour patterns on social media." *SociInfo 2014: The 6th International Conference on Social Informatics*. 10-13 Nov 2014, Barcelona, Spain

Chen, F., Deng, P., Wan, J., Zhang, D., Vasilakos, A.V., Rong, X. (2015) "Data Mining for the Internet of Things: Literature Review and Challenges." *International Journal of Distributed Sensor Networks.* Volume 2015

Corriveau, P. and C. Greco. 2018. "Online Paedophilia and Cyberspace." Accessed 01December 2018. https://www.inspq.qc.ca/en/sexual-assault/fact-sheets/online-pedophilia-and-cyberspace

Craker, N. and E. March. 2016. "The dark side of Facebook®: The Dark Tetrad, negative social potency, and trolling behaviours." *Personality and Individual Differences*. 102:79–84

Davies, C. 2018. "'Sadistic' paedophile Matthew Falder jailed for 32 years." *The Guardian*. Accessed 14 December 2018. https://www.theguardian.com/technology/2018/feb/19/dark-web-paedophile-matthew-falder-jailed-for-32-years

Florida Gulf Coast University (FGCU). 2019. "Radford / FGCU Serial Killer Database Research Project." Accessed 01 December 2018. http://skdb.fgcu.edu/info.asp

French, L. 2017. "Virtual Case Notes: Raped Through Internet – Swedish Man Convicted in Unprecedented Case." Accessed 02 December 2018. https://www.forensicmag.com/news/2017/12/virtual-case-notes-raped-through-internet-swedish-man-convicted-unprecedented-case

Galán-García, P., de la Puerta, J.G., Gómez, C.L., Santos, I., Bringas, P.G. 2014. "Supervised Machine Learning for the Detection of Troll Profiles in Twitter Social Network: Application to a Real Case of Cyberbullying." In: Herrero Á. et al. (eds) International Joint Conference SOCO'13-CISIS'13-ICEUTE'13. Advances in Intelligent Systems and Computing, vol 239. Springer

Gammon, A. 2014. "Over a quarter of Americans have made malicious online comments." Accessed November 19, 2018. https://today.yougov.com/news/2014/10/20/ over-quarter-americans-admit-malicious-online-comm/

Gangadharan, S. 2015. "Predictive algorithms are not inherently unbiased." *The New York Times*. Accessed November 18, 2015. http://www.nytimes.com/roomfordebate/2015/11/18/can-predictive-policing-be-ethical-and-effective

Garfinkel, S.L. 2010. "Digital forensics research: The next 10 years." *Digital Investigation*. 7: S64-S73

Hannah H. 2017. "INTERNET SAFETY: A MOTHER'S STORY OF HOW A PAEDOPHILE GROOMED HER 11-YEAR-OLD DAUGHTER ONLINE." Independent. Accessed 05 December 2018. https://www.independent.co.uk/life-style/health-and-families/internet-safety-day-hannah-h-mother-paedophile-online-grooming-11-year-old-daughter-facebook-webcam-a7560801.html

Hargittai, E. 2015. "Is bigger always better? Potential biases of big data derived from social network sites." *The ANNALS of the American Academy of Political and Social Science*. 659(1): 63-76

Hayes, B., Jeandesboz, J., Ragazzi, F., Simon, S., Mitsilegas, V. 2015. "The law enforcement challenges of cybercrime: are we really playing catch-up?" *Study for the LIBE Committee*. http://www.europarl.europa.eu/RegData/etudes/STUD/2015/536471/IPOL_STU(2015)536471_EN.pdf

Hoff, D. L., and S.N. Mitchell. 2009. "Cyberbullying: Causes, effects, and remedies". *Journal of Educational Administration.* 47: 652–665

Hosseinkhani, J., Koochakzaei, M., Keikhaee, S., Nniz, J.H. 2014. "Detecting Suspicion Information on the Web Using Crime Data Mining Techniques." *International Journal of Advanced Computer Science and Information Technology (IJACSIT)*. 3(1):32-41

Injadat, MN, Salo, F., Nassif, A.B. 2016. "Data mining techniques in social media: A survey." *Neurocomputing*. 214: 654-670

Jin, X., Wah, B.W., Cheng, X., Wang, Y. 2015. "Significance and challenges of big data research." *Big Data Research* 2(2): 59-64

Jones, S. E., Manstead, A. S. R., Livingstone, A. G. 2011. "Ganging up or sticking together? Group processes and children's responses to text-message bullying." *British Journal of Psychology*. 102: 71–96

Karppi, T. 2018. ""The Computer Said So": On the Ethics, Effectiveness, and Cultural Techniques of Predictive Policing." *Social Media+Society.* April-June 2018: 1–9

Klempka, A., & Stimson, A. 2013. Anonymous communication on the internet and trolling. Masters thesis. Concordia University. Retrieved from: https://comjournal.csp.edu/wpcontent/uploads/sites/16/2013/12/TrollingPaper-Allison-Klempka.pdf

Lanning, KV. 2002. "Compliant child victims: Confronting an uncomfortable reality." *The American Professional Society on the Abuse of Children (APSAC) Advisor*. 14(2)

Larson, E. 2017. "Tracking Criminals with Internet Protocol Addresses: Is Law Enforcement Correctly Identifying Perpetrators?". *North Carolina Journal of Law & Technology*. 18(5):316-358

Lazer. D., Pentland, A., Adamic, L., Aral, S., Baraba´si, A., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., Van Alstyne, M. 2009. "Computational social science." *Science*. 323(5915):721-723

Lee, L. M., Isa, D., Choo, W. O. 2011. "High relevance keyword extraction facility for Bayesian text classification on different domains of varying characteristic." *Expert Systems with Applications*. 39(1): 1147-1155

Makinde, O.A., Odimegwu, C.O., Abdulmalik, J.O., Babalola, S.O., Fawole, O.I. 2016. "Gender-based violence following social media acquaintance in Nigeria." *African Journal of Reproductive Health,* 20(4): 67-76

Mcintyre, D. 2017. "Cyber serial killers can operate online without fear of being caught." Mirror, October 12. Accessed 12 November 2018. https://www.mirror.co.uk/news/real-life-stories/cyber-serial-killers-can-operate-11329277

Morse, A. 2017. "Investigation: WannaCry cyber attack and the NHS." Report by the National Audit Office. Accessed 01 December 2018. https://www.nao.org.uk/wp-content/uploads/2017/10/Investigation-WannaCry-cyber-attack-and-the-NHS.pdf

NHS. 2018. "Antisocial personality disorder." Accessed 23 October 2018. https://www.nhs.uk/conditions/antisocial-personality-disorder/

Ovenden, J. 2018. "Data Mining In The Deep Web." Accessed 10 December 2018. https://channels.theinnovationenterprise.com/articles/data-mining-in-the-deep-web

Pendar, N. 2017. “Toward Spotting the Pedophile Telling victim from predator in text chats.” *IEEE International Conference on Semantic Computing (IEEE ICSC 2017)*. January 30 – February 1 2017. San Diego, United States. pp. 235-241

Predpol. 2016. “How PredPol works: We provide guidance on where and when to patrol.” Accessed September 3, 2016. http://www.predpol. com/how-predpol-works/

PureSight. 2018. “Jessica Logan 1990-2018.” Accessed 27 October 2018. http://www.puresight.com/Real-Life-Stories/jessica-logan-1990-2008.html

Rieland, R. 2018. “Artificial Intelligence Is Now Used to Predict Crime. But Is It Biased?” *The Smithsonian*. https://www.smithsonianmag.com/innovation/artificial-intelligence-is-now-used-predict-crime-is-it-biased-180968337/

Rigney, S. 2018. “Serial rapist Andrew James Benn's troubling search history revealed.” Accessed 01 November 2018. https://www.smh.com.au/national/nsw/serial-rapist-andrew-james-benn-s-troubling-search-history-revealed-20180903-p501jc.html

Schneider, J. W. 2006. “Concept symbols revisited: Naming clusters by parsing and filtering of noun phrases from citation contexts of concept symbols.” *Scientometrics*. 68(3): 573–593

Shepherd, S.M., Campbell, R.E., Ogloff, J.R.P. 2018. “Psychopathy, Antisocial Personality Disorder, and Reconviction in an Australian Sample of Forensic Patients.” *International Journal of Offender Therapy and Comparative Criminology*. 62(3): 609–628

Simmon, S. 2016. “How a Serial Killer used Social Media to attract his victims – and why we should all take note.” *The Independent*. Accessed 05 December 2018. https://www.independent.co.uk/life-style/gadgets-and-tech/how-a-serial-killer-used-social-media-to-attract-his-victims-and-why-we-should-all-take-note-a7449216.html

Stalans, L.J. and M. A. Finn. 2016. “Understanding How the Internet Facilitates Crime and Deviance.” Victims & Offenders. 11:4: 501-508

Tseng, Y. H. 1998. “Multilingual Keyword Extraction for Term Suggestion.” *21st International ACM SIGIR conference on research and development in information retrieval – SIGIR 1998*. Melbourne, Australia. August 24 - 28, 1998. 377-378

Tseng, Y.-H., Ho, Z.-P., Yang, K.-S., Chen, C.-C. 2012. “Mining term networks from text collections for crime investigation.” *Expert Systems with Applications*. 39: 10082-10090

Wang, T., Rudin, C., Wagner, D., Sevieri, R. 2013. “Learning to Detect Patterns of Crime”, In *Proceedings, Part III, Machine Learning and Knowledge Discovery in Databases: European Conference*, ECML PKDD 2013, Prague, Czech Republic, September 23-27, 2013, pp. 515-530

Weiss, S. 2015. “8 creative ways women are calling out online dating trolls, because sometimes blocking and reporting aren't enough.” Accessed 22 October 2018. http://www.bustle.com/articles/80819-8-creative-ways-women-are-calling-out-online-dating-trolls-because-sometimes-blocking-and-reporting-arent

Wilkinson, A. 2017. “The Serial-Killer Detector.” *The New Yorker – Annals of Crimes*. November 27, 2017. Accessed 02 December 2018. https://www.newyorker.com/magazine/2017/11/27/the-serial-killer-detector

Wolak, J. and D. Finkelhor. 2013. “Are Crimes by Online Predators Different From Crimes by Sex Offenders Who Know Youth In-Person?” *Journal of Adolescent Health*. 53(6): 736-741

Zicari, R.V. 2014. “Big Data: Challenges and Opportunities.” In *Big data computing*, edited by Rajendra Akerkar, CRC Press. Florida, United States: Taylor & Francis Group

Zulkarnine, T.; Monk, B.; Frank, R; Mitchell, J.; Davies, G. (2016) ‘Surfacing collaborated networks in dark web to find illicit and criminal content’, *2016 IEEE Conference on Intelligence and Security Informatics (ISI)*, University of Arizona Campus, Tucson, United States, September 28-30, 2016, pp. 109-114