On Combating Online Radicalisation: A Framework for Cybercrime Investigations

Mariam A. Nouh

Mansfield College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas Term 2019

This thesis is dedicated to
My loving parents
Prof. Adnan Nouh & Badia Nadra
I owe everything to you!

Acknowledgements

This journey would not have been possible without the support of many individuals. To all those who contributed in many ways to the success of this journey and made it an unforgettable experience, I am forever indebted.

I would like to express my sincere gratitude to my supervisors Prof Michael Goldsmith and Dr Jason R C Nurse for their continuous support, patience, motivation, and immense knowledge. I am forever grateful to you.

Thank you to Prof Andrew Martin and Prof Adam Joinson for agreeing to be my examiners. Thanks also to Dr Helena Webb, Dr Ning Wang, and Dr Gianluca Stringhini for their valuable feedback during the interviews for Transfer and Confirmation of status. Special thanks to Dr Emilio Ferrara for his valuable advice and guidance during my internship at the USC-Information Sciences Institute.

I am indebted to my research team for providing a stimulating and fun-filled environment. It has been a pleasure to be part of such a wonderful team. My thanks go in particular to Jass, who always offered advice and guidance, to my peers, Mered, Louise, Alastair, Mary, Ari, Jan, and Bushra, for their friendship and encouragement. It has been a wonderful journey to study with you.

I am also grateful to the wider CDT in Cyber Security community. I'm lucky to have been part of this brilliant diverse community. Big thank you to Maureen, David, and Katherine for all the support they provided.

I take this opportunity to sincerely acknowledge King Abdulaziz City for Science and Technology (KACST) and the Ministry of Higher Education for financially supporting and sponsoring my studies.

The road to my DPhil started with training from many mentors over the years, all of whom have inspired me to reach this point. In particular, I'd like to express my special appreciation to Mourad Debbabi and Anas Alfaris, both have played a key role in my training and shaping my research skills.

Indeed this has been a challenging journey that I could not have completed without the support from family and friends. My heartfelt thanks to my Mom and Dad for the support you have always given me to pursue my dreams. Without your encouragement and love, I would not be where I am today. I am also grateful to my siblings for their love, care, moral support, and countless visits to lift my spirit. I owe it all to you!

To my dear friends, my support system, I'm so grateful for all those times you stood by me, for tolerating my complaints, providing motivation, fun, and laughter. Thank you for always being there, unconditionally. In particular, a big shout out to Shada, you have always been my partner in crime and an integral part of my support system. My special appreciation to Hala, Jallelah, Moe, Mohsin, and Lina for the friendship, encouragement, and pushing me to get through the last mile.

Thanks are due to numerous coffee shops in Oxford, where big chunks of this thesis were written, for supplying me with the caffeine I needed and in many occasions extending their closing time to allow me to wrap up sections!

Finally, to everyone who has touched my life. You know who you are. Thank you!

Abstract

"The Internet is the crime scene of the 21st century."

The complexity of cybercrimes is constantly increasing with advanced tools, attack vectors, and Modus Operandi adopted by offenders every day. Criminals have easy access to advanced technical abilities that they need to carry their attacks, using what is called crime-as-a-service, from the dark web and online black markets. Similarly, the nature of cybercrimes has generated multitudes of data introduced by the "cyber" aspect of these crimes, which makes the process of identifying evidence similar to searching for a needle in a haystack. To aid law enforcement to better detect, analyse, and understand the threat landscape posed by cyber-criminals, research into the area of cybercrime intelligence has flourished. Law enforcement faces numerous challenges when policing cybercrimes. The methods and processes they use when dealing with traditional crimes do not necessarily apply in the cyber world. Additionally, criminals are usually technologically-aware and one step ahead of the police. Furthermore, current tools created to support law enforcement to better police cybercrimes more often conflict with how they are used to operate, and are too complex, thus making them difficult to adopt.

In this thesis, we aim to design and develop a cybercrime intelligence framework for law enforcement that provides decision support to detect and analyse the behaviour of cyber-criminals. To do so, we need to better understand the cyber-criminal ecosystem, as well as understand the current capabilities of law enforcement agencies, and the challenges they face when policing cybercrimes. We achieve this through semi-structured interviews conducted with professionals and law enforcement agents investigating cybercrimes. From there, we define a framework to aid them in addressing some of the challenges they face. Moreover, the cybercrime landscape varies considerably in regards to the type of crime and what they target. Some crimes target computers and systems while others target the human. As there has been considerable research focusing on analysing cybercrimes that target systems such as (Malware, Hacking, DDOS), the focus on the crimes that target the human (e.g., cyber-bullying, online radicalisation) has recently become more evident. In this research, we focus on the area of online radicalisation and utilise our framework to better understand the properties of radical propaganda and develop methods to defend against its spread. We focus on the ISIS group aiming to identify

measures to automatically detect radical content and activities in social media. We identify several signals, including textual, psychological and behavioural, that together allow for the identification of radical messages, using methods such as natural language processing, social network analysis, and machine learning. Our findings can be utilised as signals for detecting online radicalisation activities by law enforcement and social media platforms to help keep the online world safe.

Statement of Originality

This thesis is written in accordance with the regulations for the degree of Doctor of Philosophy. The thesis has been composed by myself and has not been submitted in any previous application for any degree. The research within this thesis has been undertaken by myself. Parts of the thesis have been published or submitted as papers, and the list can be found in Section 1.5.

Contents

List of Figures

List of Tables

List of Abbreviations

AUC . . . . . Area Under the Curve.

BOW . . . . . Bag of Words.

HITS . . . . . Hypertext Induced Topic.

NLP . . . . . Natural Language Processing.

NN . . . . . Neural Networks.

RF . . . . . Random Forest.

ROC . . . . . Receiver Operator Characteristic.

SNA . . . . . Social Network Analysis.

TF-IDF . . . . Term Frequency-Inverse Document Frequency.

API . . . . . Application Programming Interface.

AR . . . . . American Renaissance.

ISIS . . . . . Islamic State of Iraq and Syria.

ALT-RIGHT . Alternative Right.

DDOS . . . . . Distributed Denial of Service.

CSCW . . . . Computer-Supported Cooperative Work.

OSINT . . . . Open Source Intelligence.

SNA . . . . . Social Network Analysis.

RL . . . . . Regional-Level.

NL . . . . . National-Level.

LL . . . . . Local-Level.

CCINT . . . . CyberCrime INTelligence framework.

NCA . . . . . National Crime Agency.

ROCU . . . . . Regional Organised Crime Units.

RAT . . . . . Remote Access Trojan.

IM RAT . . . Imminent Monitor RAT.

OSN . . . . . Online Social Network.

LDA . . . . . Latent Dirichlet Allocation.

GIFCT . . . . Global Internet Forum to Counter Terrorism.

LIWC . . . . . Linguistic Inquiry and Word Count.

TENE . . . . . Terrorism and Extremism Network Extractor.

MO . . . . . Modus Operandi.

DWFP . . . . Dark Web Forum Portal.

EC3 . . . . . European Cybercrime Center.

MDS . . . . . Multidimensional Scaling.

GSJ . . . . . Global Salafi Jihad.

1 Introduction

Cybercrime is defined as any crime that is facilitated or committed using a computer, network, or hardware device [1]. Initially, it emerged as a threat to individuals and organisations; now it also impacts on entire countries. Experts in the criminology field have reported that the existence of organised cybercriminals in the online world is growing rapidly [2]. According to the Cyber Security Breaches Survey [3], around a third (32%) of businesses in the UK report having cyber security breaches or attacks in 2019. In one minute on the internet, $2.9 million is lost to cybercrimes[1]. These figures show that there is an increase in scale and numbers of cybercrime incidents, which is often correlated with a lack of proper technological response and support from law enforcement [4].

The area of cybercrime poses a low risk and high return for cybercriminals since typically they carry out their operations remotely, equipped with tools that provide them with anonymity and encryption. This makes detecting and identifying them even more challenging for law enforcement. Therefore, many traditional criminals have migrated to the cybercrime world, even if they lack the technological advancement, by making use of the online cybercrime marketplace or what is called cybercrime-as-a-service [5].

[1] RiskIQ Report: https://www.riskiq.com/infographic/evil-internet-minute-2019/

The threat of cybercrime is evolving as a tool for terrorism and serious and organised crimes. The convergence of cyber and terrorism introduces new threats and attack victors. Terrorist groups are exploiting emerging online platforms and adopting new technologies to facilitate online communication and distribution of their propaganda [6]. As these threats continue to evolve, online platforms and law enforcement need to innovate and collaborate in order to stay ahead and effectively respond to these threats.

1.1 Motivation

Cybercrime Intelligence — Cybercrime investigators are facing numerous difficulties trying to keep pace with the increasing numbers of incidents and the evolution of techniques used by cybercriminals. Previous studies suggest that the methods and processes typically used by law enforcement when investigating traditional crimes do not necessarily apply in the cyber world [7, 8]. Thus, there is a need to adopt a different process and build a new set of skills and knowledge to be able to mitigate these technologically advanced crimes. This is essential mainly as the cybercriminals are usually technologically-aware and are constantly adapting and developing new tools to allow them to stay ahead of law enforcement investigations [9–11]. Even if they do not have the appropriate technical knowledge to commit the crime, thanks to several online markets that provide crime-as-a-service, any advanced technical skill is now available for anyone willing to pay.

Moreover, from the technology side, current tools created to support law enforcement to better police cybercrimes more often conflict with how they are used to operate, and are too complex, thus making them difficult to adopt [4, 12]. As a result, the area of cybercrime intelligence has attracted the attention of the research community intending to aid law enforcement to better detect, analyse, and understand the threat landscape posed by online cybercriminals. Cybercriminals use the Internet as their crime scene since it provides them with ease of communication, wider recruitment possibilities, and opportunities to form a partnerships with other national and international criminal groups [13, 14]. This often results in them

leaving several "crumbs" that collectively produce a digital footprint for each group of cybercriminals. Previous research has studied these footprints to gain a better understanding of the characteristics of these cybercriminal groups [15–18]. For example, this could be by adopting techniques such as content analysis. This is used to interpret meaning from text and extract inferences, to profile criminals, or by uncovering the leadership and organisational structure of a given criminal organised group using social network analysis methods [19], which is the process of investigating social structures using network and graph theories.

Furthermore, while there are many research efforts focused on understanding and analysing the tactics used by cybercriminals [11, 20, 21], very little research exists that focuses on understanding the processes, challenges, and needs of law enforcement and cybercrime investigators. Therefore, there is a need for new research that focuses on understanding the needs and challenges faced by practitioners.

Online Radicalization — The rise of Online Social Networks (OSN) has facilitated a wide application of their data as sensors for information to solve different problems. For example, Twitter data has been used for predicting election results, detecting the spread of flu epidemics, and as a source for finding eye-witnesses during criminal incidents and crises [22], [23]. This phenomenon is possible due to the great overlap between our online and offline worlds. Such a seamless shift between both worlds has also affected the modus operandi of cybercriminals and extremist groups [24]. They have benefited tremendously from the Internet and OSN platforms, as they provide them with opportunities to spread their propaganda, widen their reach for victims, and facilitate potential recruitment opportunities. For instance, recent studies show that the Internet and social media played an important role in the increased number of violent right-wing extremism activities [25]. Similarly, extremist groups such as Al-Qaeda and the Islamic State of Iraq and Syria (ISIS) have used social media to spread and promote their propaganda, which inspired the Boston Marathon bombers in 2010 [26].

To limit the reach of cyber-terrorists, several private and governmental organizations are policing online content and utilising big data technologies to minimize the

damage and counter the spread of such information. For example, the UK launched a Counter-Terrorism Internet Referral Unit in 2010 aiming to remove unlawful Internet content and it supports the police in investigating terrorist and radicalizing activities online. The unit reports that among the most frequently referred links were those coming from several OSNs, such as Facebook and Twitter [24]. Similarly, several OSNs are constantly working on detecting and removing users promoting unlawful and extremist content. In 2018, Twitter announced that they suspended over 1.2 million accounts that featured terrorist content [27].

Realizing the danger of violent extremism and radicalization and how it is becoming a major challenge to societies worldwide, many researchers have attempted to study the behaviour of pro-extremist users online. Looking at existing literature, we find that several existing studies incorporate methods to identify distinguishing characteristics that can aid in automatic detection of these users [28, 29]. However, many of them depend on performing a keyword-based textual analysis which, if used alone, may have several shortcomings, such as producing a large number of false positives and having a high dependency on the data being studied. Additionally, it can be easily evaded if users adapt their writing styles using automated tools.

1.2 Scope

As in any research project, it is important to draw boundaries around the scope of the project, and this is defined as follows:

UK Centric — The majority of the research in the literature that focuses on studying investigators' needs and the challenges they face is based in the United States [30–32]. Very little published literature has investigated elsewhere in the world, including the UK [33]. Similarly, the majority attempted to use surveys with close-ended questions as a means to collect data. One advantage of using this approach is that responses may be collected from a large number of participants. However, a disadvantage is that it limits the ability to capture insights and understand the reasons behind the chosen answers [31]. Additionally, related literature studied the views of particular types of participants dealing with

cybercrimes (e.g., local police officers). Very few looked at this issue from multiple viewpoints of different participant types (i.e., local officers, regional and national cybercrime units). In our study (presented in Chapter 4) we address these gaps by interviewing participants dealing with cybercrimes from the government-sector (including local and regional units) as well as from the private sector. We aim to identify the needs and challenges faced by cybercrime investigators and practitioners using detailed, in-depth interviews. In these expert interviews, we not only focus on a specific aspect of the investigation, such as digital forensics, but we also look at the holistic process followed when investigating and therefore touch on several issues related to technical and human aspects.

Cybercrimes — The topic of cybercrime is quite diverse and varies considerably in terms of methods, motives, and victims, based on the type of crime under investigation. While there is a considerable amount of research that focuses on studying cybercrimes that target networks and systems, such as malware, DDoS, and cyber-fraud [34–36], research is sparse related to cybercrimes that target the human's mindset. In recent years, an evolving threat that targets people's hearts and minds has grown. These crimes cause online harm and have ideological, political, or psychological motives. Such threats include the use of online resources to manipulate the public for political or social gain, hate crimes, cyber-bullying, trolling, the spread of fake news and violent and extremist content and activities. Starting from Chapter 6, we focus our topic of cybercrime on the area of online extremist content and activities as this area still lacks a proper and effective response [37].

Extremist Groups — Several terrorist and extremist groups have benefited tremendously from technology such as the Internet and Online Social Network (OSN) platforms as this provides them with opportunities to spread their propaganda, widens their reach for victims, and facilitates potential recruitment opportunities. The Islamic State of Iraq and Syria (ISIS) is one of the leading groups to heavily utilise digital media and propaganda to promote their views and recruit sympathisers [38]. In the study presented in Chapter 6 and Chapter 7, we focus on studying strategies used by the ISIS extremist group with the aim of analysing their

propaganda narrative in order to identify measures to automatically detect the spread of radical content and activities online. We then compare the learned signals and strategies across different extremist groups in Chapter 8 to ascertain if these signals are a group focused or whether they are shared across different extremist narratives. Through the understanding of how the extremist narrative is constructed and distributed we can design and build systems to aid law enforcement and social network platforms effectively detect and mitigate the spread of extremist content.

1.3 Research Questions

In the course of this research project, we aim to answer several research questions that relate to the two sides of the crime: defenders (investigators) and offenders (criminals). From the defenders' side, we define the following research questions.

RQ1. What are the main challenges currently being faced by cybercrime investigators?

RQ2. What are the requirements for cybercrime intelligence systems that would be appropriate and helpful for investigators?

Before identifying the challenges that are faced by investigators, we need to understand their existing capabilities and the processes and workflows they follow during investigations. To address the above two questions, we start by exploring the current literature relating to the needs and challenges of cybercrime investigators. Then, we design our study where we interview practitioners from different organisations to capture their requirements and areas where improvements are needed in terms of processes and tools. This leads to the design of a cybercrime intelligence framework.

From the offenders' side, we focus on the topic of online extremism and radicalisation, where we apply our framework to study properties of extremist narrative and build systems to detect the spread of online radicalisation. As such, we define the following questions.

RQ3. Can we identify hidden properties and signals that describe extremist strategies in crafting their online propaganda?

RQ4. Can we detect the spread of extremist narrative in social media?

RQ5. To what extent are extremist narrative properties shared across multiple extremist groups?

To address the above questions, we focus on the ISIS extremist group to understand their methods and strategies in crafting their propaganda material. Using our framework, we identify a set of properties that are extracted from ISIS propaganda. Then we conduct a study using Twitter, to understand the behaviour of ISIS supporters and evaluate our approach for detecting the extremist narrative. Finally, we end by comparing the identified properties across different extremist groups, mainly Al-Qaeda and Alternative-Right (Alt-Right) groups.

1.4 Contributions

The main contributions of this thesis are presented in the following chapters of work:

1. Chapter 4: **Cybercrime Investigations: Process and Challenges.** We study the current practices followed by law enforcement, and more generally security intelligence companies when investigating cybercrimes and gathering intelligence. Using detailed in-depth expert interviews, we identify the general characteristics related to the investigation process of cybercrimes. This allows for the identification of empirical socio-technical challenges currently faced and areas where new technologies, processes, and workflows are necessary. Identifying these challenges led to formulating user-driven recommendations for designing and building processes and tools to improve the day-to-day operations of cybercrime investigators. In these expert interviews, we not only focus on a specific aspect of the investigation, such as digital forensics, but we also look at the holistic process followed when investigating and therefore touch on several issues related to technical and human aspects.

2. Chapter 5: **The CCNIT Framework.** Based on our evaluation of the literature and our understanding of the processes and challenges faced by experts in the field, we propose a cybercrime intelligence framework named

CCINT. The framework is designed to fill gaps identified in the literature and adhere to the gathered requirements and design guidelines recommended for intelligence systems. We describe the different components of the framework, which include a conceptual model that aids investigators in the process of collecting effective intelligence, and an operational model that focuses on developing analytical tools.

3. Chapter 6: **Properties of Radical Propaganda.** In this chapter, we focus our study on the problem of online radicalisation. More precisely, we apply methods from the CCINT framework and utilise computational techniques to analyse online extremist propaganda and reveal their underlying tactics. We explore the use of new methods to uncover syntactic, semantic, and psychological properties associated with the extremist discourse.

4. Chapter 7: **Detection of Online Radicalisation.** After identifying properties of extremist propaganda, we design and build a system to detect online supporters of extremist groups in social networks. We build different models to detect extremist narrative and behaviour on Twitter and show how the results are consistent across different datasets. This system can be useful for OSN platforms and law enforcement to counter extremist groups' online propaganda.

5. Chapter 8: **Comparison of Radical Groups.** We compare the properties of propaganda published by different radical groups and identify the similarities between them. This is important in order to understand what properties are global across different extremist groups, and what properties are local and linked to a particular group's culture and ideology. This allows us to highlight the potential wider applicability of our system proposed in Chapter 7.

1.5 Academic Publications and Awards

During our research, we have published a number of peer-reviewed papers and achieved several awards. These are listed below with one article currently under submission.

1. **M. Nouh** and J. R. C. Nurse, *Identifying Key-Players in Online Activist Groups on the Facebook Social Network.* Proceedings of the IEEE International Conference on Data Mining Workshop (ICDMW), pp.969-978. 2015

2. **M. Nouh**, J. R. C. Nurse and M. Goldsmith, *Towards Designing a Multipurpose Cybercrime Intelligence Framework.* Proceedings of the European Intelligence and Security Informatics Conference (EISIC), pp.60-67. 2016

3. **M. Nouh**, J. R. C. Nurse, Helena Webb, and M. Goldsmith, *Cybercrime Investigators are Users Too! Understanding the Socio-Technical Challenges Faced by Law Enforcement.* Proceedings of the Workshop on Usable Security (USEC) at the Network and Distributed Systems Security Symposium (NDSS), Internet Society, 2019

4. **M. Nouh**, J. R. C. Nurse, and M. Goldsmith, *Understanding the Radical Mind: Identifying Signals to Detect Extremist Content on Twitter.* Proceedings of the IEEE International Conference on Intelligence and Security Informatics (IEEE ISI), 2019

5. **Under Submission. M. Nouh**, J. R. C. Nurse, and M. Goldsmith, On Understanding Extremist Narrative: Applying Computational Methods to Compare Properties of Extremist Propaganda. IEEE Access 2020

6. **Award.** *3rd place award* at the Grace Hopper Conference (GHC) ACM Student Research Competition (SRC), graduate category for research titled *CCINT: The CyberCrime Intelligence Framework for Detecting Online Radical Content.* 2017

7. **Award.** *Distinguished Poster Presentation award* at the Network and Distributed System Security Symposium (NDSS) for the research paper titled *Cybercrime Investigators are Users Too! Understanding the Socio-Technical Challenges Faced by Law Enforcement.* 2019

1.6 Thesis Roadmap

An illustration of the structure of this thesis and the relationship between research contributions and produced publications is presented in Figure 1.1.

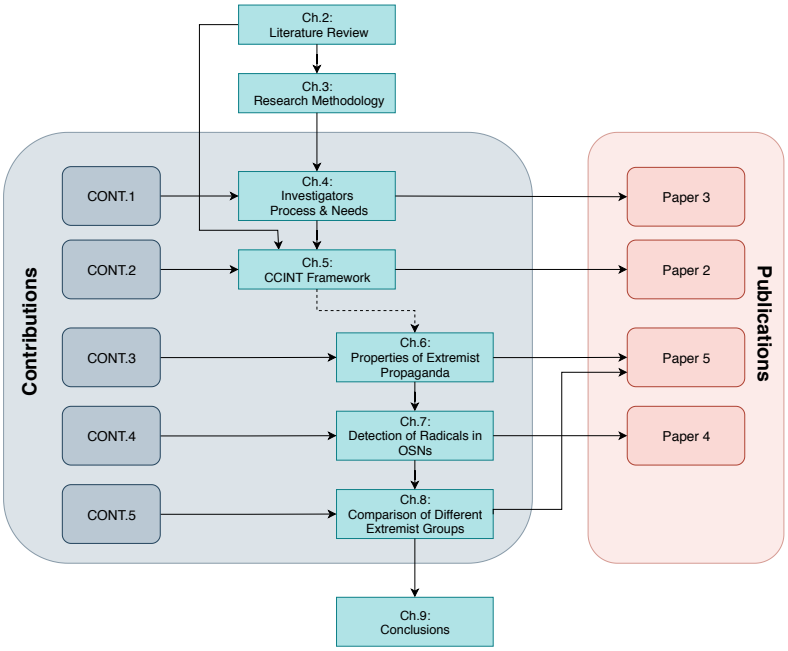


Figure 1.1: Thesis Roadmap

Chapter 2 introduces the necessary background related to the area of cybercrimes by explaining the definition and different taxonomies of cybercrimes. Then, we present literature related to intelligence gathering and the existing requirements for intelligence frameworks. We then explore existing literature related to the topic of understanding and detecting online radicalisation.

In Chapter 3, we present an overview of the general research methodology adopted in conducting this research. This explains the mixed (qualitative and quantitative) methods we adopt to answer the defined research questions. Chapters 4, 5, 6, 7, and 8 cover the main contributions of this work, as summarised in

Section 1.4. Finally, we conclude this thesis in Chapter 9, where we summarise the key findings of the research and the main lessons learned; we also discuss the limitations of this research and propose a set of future research directions.

2
Literature Review

In this chapter, we present a review of related literature and key concepts relevant to the research presented in this thesis. First, we understand the necessary background related to the field of cybercrime. Then, we discuss the literature related to intelligence gathering, before exploring existing work on requirements for intelligence frameworks. Since we later scope on the topic of online radicalisation, we outline the existing related studies.

2.1 Cybercrime Definitions and Taxonomies

In order to understand the state-of-the-art research in the area of cybercrime, it is important to define what we regard as cybercrime and cybercriminals. There have been several arguments in the literature over the exact definition of cybercrime with no single universal definition [39]. Similarly, some articles use generic references such as computer-crime, Internet-crime, digital-crime, and most widely used cybercrime (or cyber-crime), as well as terms references to specific forms of crimes such as cyber-terrorism and cyber-stalking [40]. The European Commission proposed the following definition for cybercrime: "criminal acts committed using electronic communications networks and information systems or against such networks and

systems" [39]. The definition incorporates crimes that were facilitated by computers and those that were committed against them.

Some articles in the literature define cybercrime as any crime that involves computers or networks, others define it as purely digital crimes, or traditional crimes which are enhanced through the use of digital technology [41]. Initially, previous research drew a distinction between crimes, where the computers are the target of the crime (hacking, spam, terrorism-related offences), and crimes where computers act as tools to commit the crime (child pornography, phishing) [40]. David Wall considers cybercrime to be one of three types: (1) Crime in the machine, (2) crime using the machine, and (3) crime against the machine [42]. Crime in the machine relates to the content of computers which include for example the trade of pornographic material, crime using the machine relates to any crimes committed using networked computers such as targeting victims by phishing emails, finally crimes against the machine covers the integrity of computers and networks such as hacking and planting of viruses and Trojans.

Furthermore, Brar et al. [43] attempts to provide a taxonomy of cybercrimes in the context of the three security principles (confidentiality, integrity, and availability). They break-down cybercrimes to four types: cyber-violence (e.g., cyber-terrorism, cyber-stalking), cyber-peddler (e.g., cyber-fraud, cyber-activism), cyber-trespass (e.g., cyber-theft, cyber-pornography), and cyber-squatting (e.g., attackers illegally register the trademark of legitimate owners to deny them).

Another classification of cybercrimes that is adopted by the UK Home Office, is the break-down to computer-enabled and computer-dependant crimes. The former includes traditional crimes that can be enhanced in scale and reach using computers and networks, while the latter includes crimes that can only be committed using computers and networks [44]. The matrix in Table 2.1 gives examples for each of these categories based on the level of opportunity and type of crime. For example, a phishing attack is considered a computer-dependant crime and a crime using the machine. On the other hand, distribution of viruses and hacking are classified as computer-enabled crime and a crime against the machine.

Table 2.1: The Matrix of Cybercrimes: Level of Opportunity by Type of Crime (with examples) [45]

Crime Types	Crime against machines/ Integrity related	Crime using machines/ Computer-related	Crimes in the machine/ Content-related
Opportunities	Harmful/ Trespass	Acquisition/ (Theft/ Deception)	Obscenity/ Violence
Cyber-Enabled Crimes - Hybrid cybercrime New opportunities For traditional crime	- Hacking - Viruses - Hactivism	- Multiple large-scale frauds - Trade secret theft - ID Theft	- Online Sex trade - General Hate speech - Child abuse
Cyber-Dependent Crimes - True Cybercrime New opportunities for new types of crime	- Spams - Denial of Service - Information Warfare - Parasitic Computing	- Intellectual Property - Online Gambling - E-auction scams - Phishing	- Cyber-sex - Cyber-pimping - Online Grooming - Organised Bomb talk/ Drug talk - Targeted hate speech

This section points out the variations in definitions and taxonomies that exist in the current literature. Such variations contribute to the challenge of establishing proper international efforts to counter cybercrimes.

2.2 Cybercrime Intelligence Literature

We classify existing work in the area of cybercrime intelligence into three main research tracks presented in Figure 2.1 and detailed in the following subsections.

2.2.1 Detection of Cybercrimes

This section reviews the literature related to cyber-dependent crimes, the methods and data sources being employed. As the majority of online content is lawful; the challenge is to detect outlier criminal-related content and behaviour. Several methods are reported in the literature to address this issue [17, 46, 47]. Some focus on studying textual content through the use of Natural Language Processing (NLP) techniques in order to achieve different tasks, such as authorship profiling and analysis; text classification; and sentiment analysis [17]. Others adopt different methods such as Social Network Analysis (SNA) to study the relations between criminals and properties of the communities they form [46]; information extraction to extract intelligence from large amounts of text [47]; and machine learning to classify/cluster users' accounts and identify outliers automatically [28].

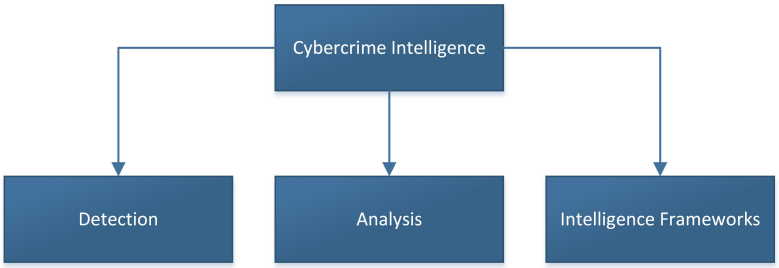


Figure 2.1: Classification of cybercrime intelligence research tracks

One of the tough challenges for all researchers in this domain is access to reliable data. Previous studies have focused on different types of online data, such as online forums and social media. Similarly, some used data gathered from different news websites and online auctions to detect fraudulent behaviours [48]. When it comes to methods for evaluation, most of the existing research either lacks a proper evaluation method, performs manual evaluation [46], depends on expert evaluation [49], or uses statistical measures [50].

The literature review shows that one of the heavily studied research problems is the detection of online spam. One method that is employed the most to achieve this task is machine learning [51, 52]. McCord et al. [51] present a traditional machine learning classifier to detect spam users in Twitter. They used a combination of content-based and user-based features, and compare the performance of four different classifiers. For evaluation they use the standard metrics for measuring the usefulness of the detection scheme (confusion matrix). In a similar study, Wang [53] proposed spam detection classifier for Twitter based on content and graph features. His results showed that the Bayesian classifier generated the best performance. He used classic evaluation metrics (i.e., precision, recall, f-measure) to compare the performance of different classification methods. Looking at a different social network platform, Beeutel et al. [54] study the problem of spammers performing "Page Like" actions in order to make a profit. They propose a method called CopyCatch, aimed at detecting ill-gotten Page-Likes on Facebook by analysing the social network of users and Pages and the times at which the Likes-actions were created. They use two algorithms, an iterative algorithm and an approximate MapReduce implementation. Stringhini at

al., [55] showed the possibility to automatically identify spammers accounts using honey-profiles (similar to honeypots concept) in several social networks (Twitter, Facebook, MySpace). To evaluate their method they submitted the identified profiles to Twitter, and Twitter verified these accounts and deleted around 15,857 spam profiles, while 75 profiles were reported by Twitter as false positives.

Moreover, another type of threat that emerged is what is called social-bots. Several research efforts [52, 56–58] studied the identification of malicious bots (software agents imitating humans). These bots are able to mimic characteristics related to content, network, sentiment, and temporal patterns of activity. Ferrara et al. [56] study the incorporation of behavioural features to detect social bots in Twitter. Everett et. al., [57] investigated how easily bots can deceive humans to the extent of believing that an automatically-generated text was written by a human. They identified a set of factors that contribute to how convincing the text is. In other work Thomas et al. [52] infiltrated the underground market of fraudulent accounts. By collaborating with Twitter, they were able to investigate fraudulent accounts and develop a classification algorithm to retroactively detect millions of fraudulent accounts that were sold by the underground marketplace they infiltrated. For the classification they used features such as account-naming patterns, form-submission timing, and sign-up flow events. Precision and recall were used to validating the proposed model. Moreover, Wang et al. [58] studied the use of crowd-sourcing for detecting Sybils (fake accounts) in online social networks, more specifically Facebook. They analyse the detection accuracy using workers from Amazon Mechanical Turk called "turkers" and experts.

Furthermore, Yang et al. [59] perform an empirical study to study cybercriminal communities in Twitter. They focus on developing criminal-account detection algorithm based on social relationships and semantic coordination. Their study focuses on identifying profiles that post malicious URLs as a starting point for identifying communities, then they look at social relations between these accounts. They evaluate their algorithm on two datasets and determine the effectiveness of the algorithm based on the number of correctly inferred criminal accounts and malicious

accounts. Savari et al. [46] built a social network from a seed of publicly leaked email addresses of criminals. They identified Facebook accounts associated with these emails and constructed their social graph. Through several SNA methods they identified multiple criminal communities and profile groups on Facebook. They manually validated the results to evaluate the effectiveness of their method.

2.2.2 Analysis of Cybercriminals

Analysing the criminal content online allow us to gain further insight into the structure and behaviour of cybercriminal groups. The work of Stringhini et al. [60] sheds light on the relations and interactions between different actors involved in the spam ecosystem. The authors investigate the relations between email harvesters, botmasters, and spammers, where the analysis is based on correlating their behaviour based on indirect measurements. Thus, whilst insightful, a limitation of their work is apparent since no exact figures are calculated and only indirect measurements are considered. Moreover, Almaatouq et al. [61] analyse the behaviour of spam accounts on Twitter. Spammers are categorised based on their behaviour into two different categories: compromised accounts and fraudulent accounts. The authors then analyse the profile properties of the accounts, and their social interactions including following and mention-behaviour.

Garg et al. [62]. study the organisation structure of criminals in three underground online forums. They analyse the communities within these forums and compare topics of communication used by these communities using topic modeling. They found that most communities specialise in specific crime types. Additionally, they identified central members from each community using different SNA centrality measures. Finally, they investigated the effects of removing misbehaving criminals on the criminal network, and found that this had a positive impact on the network. Results generalisation is a limitation of this research, similar analysis should be repeated on larger dataset and different types of underground forums.

Chen et al. [63] present a general framework for data-mining of criminal and terrorist behaviour. Four categories of crime data-mining techniques were identified,

namely entity extraction, association, prediction, and pattern visualisation. Three tasks were performed using those techniques. First, they extracted named entities (e.g., person names, addresses) from police reports. Second, they detected deceptive criminal identities within police database using string comparators to measure similarities between strings. Finally, they worked on identifying subgroups and key members. Hierarchical clustering was applied to identify subgroups and centrality measures was used to detect key members in each group. Their work showed that using SNA concepts is suitable for studying such terrorist networks. Choudhary et al. [64] surveyed existing literature on counter terrorism and social network analysis. They found that SNA is one of the most successful methods for counter terrorism in social networks. This is given that the most studied problems in this domain are related to identifying key-players, finding behaviour patterns, community discovery, and disrupting terrorist networks.

Similarly, Qin et al. [65] studied the Global Salafi Jihad (GSJ) terrorist network by applying techniques such as web structural mining, social network analysis, and statistical analysis. The GSJ dataset was collected from several sources including transcripts of court proceedings and several people statements. For each identified terrorist they calculated degree, betweenness, and closeness centrality measures. Then they used statistical analysis methods (e.g., link density, average node degree, degree distribution) to study the organisation of the terrorist group and the overall topological properties of the GSJ network. Finally, they applied PageRank algorithm (adopted from the Web structural mining area) to study the communication patterns in the GSJ network and identify the influential leaders within local terrorist groups. This allows them to understand the hierarchy of the GSJ network. They evaluated the results using knowledge obtained from domain experts.

Xu et al. [66] proposed algorithms to identify the strong ties between nodes in a criminal network. They used shortest-path algorithms, priority-first-search (PFS) and two-tree PFS. To evaluate the effectiveness of the algorithm in uncovering crime investigative leads, they compared the PFS algorithms with the typical approach used by crime investigators. Xu and Chen [67] developed CrimeNet framework to aid

law enforcement automatically extract criminal network knowledge. The framework has four stages, which include network creation, partitioning, structural analysis, and visualization. The framework allows for fast identification of central members in the network, detection of subgroups, and extraction of interaction patterns between these subgroups. They used methods such as social network analysis, hierarchical clustering, and multidimensional scaling. For evaluating the system performance, they performed a controlled lab experiment where they recruited 30 university students with the task of detecting subgroups, identify interaction patterns, and finally identify central members within the groups.

One of the very first initiatives to link the disciplines of network analysis and crime intelligence is the work of Sparrow [68]. He used network analytic techniques to identify vulnerabilities in criminal organisations. Lu et al. [16] empirically study a hacker community called "Shadowcrew", which is a known group for committing identity theft and credit-card fraud. The authors study questions related to the hacker network centralization, leadership and their influence on the group, and the existence of different subgroups within the community. The data was collected from newspapers, journals and law reviews that had a keyword match for "Shadowcrew". The methods used for the analysis was based on social graph analysis and the leaders were identified using centrality measures. This study has a couple of limitations: It is based on analysis of a single group, which means it can not be generalized. Additionally, the network was built based on data gathered from text documents, which does not capture the behaviour of the group.

As observed in the related literature, most existing work uses SNA related measures to analyse cybercriminal groups. They mainly focus on detecting communities and influential members within the group by analysing the social structure and their interactions. Very few consider the behavioural and temporal factors when analysing these groups.

2.2.3 Intelligence Frameworks

Several previous efforts have worked on developing frameworks and tools to support law enforcement by providing intelligence to facilitate the detection and analysis of cybercrimes. Some of these frameworks focus on providing web crawling capabilities for collecting criminal-related content [69–71]. Zhang et al. [69] introduced the first version of the Dark Web Forum Portal (DWFP). The system supports several functions including data acquisition from different online forums, forum browsing and searching, multiple language translation and network visualisation. The system was later enhanced in a newer version, where several limitations were improved [72].

Moreover, Mei et al. [70] present a semi-automated web-crawler for collecting extremist content using sentiment analysis. The system uses a decision tree that classifies the web pages into a set of classes by combining methods of web-crawler, parts-of-speech tagging, and sentiment analysis. The content is classified into: content with extremist sentiment (pro-extremist class); news sources (neutral class); government or anti-extremist organisations (anti-extremist class); and content unrelated to extremism.

The work of Bouchard et al. [71] presents a web-crawler called the Terrorism and Extremism Network Extractor (TENE). The aim of TENE is to collect information about extremist activities online and help differentiate between extremist websites and other similar websites. The crawler starts at a user-specified webpage then analyses the content and further follows any hyperlinks in the page. In order to add the webpage to the analysis, it has to contain a set of user-defined keywords. TENE extracts around 200 characters before and after the user-defined keywords in order to determine the context in which the keywords were used. Although the context extraction is done automatically, the analysis of the context is performed manually. The COPLINK system [73, 74] was designed to aid law enforcement in extracting information from police reports and provide an environment for information-management in the intelligence domain. The system uses data-mining techniques to build a concept space of objects and entities and their associations,

as well as social network analysis to study the relations between them. In addition, the system provides visualisation functionality.

Furthermore, CrimeNet Explorer [67] is a framework for discovering criminal-network knowledge that incorporates both structural analysis and visualisation methods. Similar to COPLINK, the framework uses data gathered from crime incident reports. The framework includes four main steps, network creation, network partition, structural analysis, and network visualisation using multidimensional scaling. Limitations of the CrimeNet Explorer framework include the use of "concept space" to create the network, as this approach is fairly simplistic [67]. Additionally, the framework only focuses on analysing networks of people (criminals) and does not look at networks of people and entities (e.g., places, weapons). Concept space method is used for network creation, and hierarchical clustering for network partitioning. In the structural analysis stage subgroups are detected, interaction patterns between these groups are identified, and the central members of these subgroups are detected using a set of centrality measures. For network visualisation multidimensional scaling (MDS), which is usually used to visualise distance and similarity levels, is used to visualise the social networks.

The Isis toolkit [75] provides law enforcement with the ability to analyse digital personas in cybercrime investigations. The main features supported by the toolkit are establishing a stylistic language fingerprint, establishing the age and gender of the person behind the persona, and finally establishing interaction patterns between a set of digital personas. The toolkit combines techniques from corpus-based natural-language analysis and authorship attribution. It presents the results in a visualisation view, but it is based on a fairly simple, chart-like visualisations with no support for user interaction. Furthermore, the toolkit is able to detect deceptive personas (users with masquerading behaviour) with high degree of accuracy. Jigsaw [76] is a visual analytic system that supports investigative analysis. It provides visual representations of information extracted from textual documents to aid analysts in better understanding the documents. The analysis is based on extracting entities (e.g., person, place, date) and identifying connections and

relationships between them. Two entities are connected if they appear together in one or more documents. The Jigsaw system consist of multiple views that provide the user with different perspectives of the data. Although the Jigsaw system provides rich visualisations and supports user interaction, it does not provide any sophisticated data-mining or analytic capabilities.

Chen et al. [49] propose a semi-automated methodology for collecting and analysing Dark Web information. The methodology consists of collecting, analysing and visualising information gathered from multiple web sources. The data collection is based on keyword search and websites of terrorist organisations. They also used the back-link search function of Google to identify websites that point to the terrorist organisations sites. Unrelated sites such as news and government websites are manually filtered out, which resulted in total of 39 websites to be included in the analysis phase. The analysis consists of clustering websites based on similarity measure calculated based on the number of hyperlinks in website A pointing to website B, and vice versa. Moreover, websites classification was performed based on their affiliation with a specific terrorist group. However, the classification was performed manually by reviewing the content of each website, which is not efficient and is error-prone especially for large datasets. Two types were adopted for visualisation, these are MDS and snowflake. The proposed methodology was evaluated by a terrorism expert.

There are other frameworks that have more specialised purposes that can be applied to the cybercrime-intelligence domain. For example, TwitterHitter [77] which is focused on spatio-temporal analysis of Twitter data, such as the geo-spatial information that includes time zone, and latitude/longitude of the twitter user. The framework supports name/alias keyword search, which then provides a spatio-temporal record of the target user's activity. Additionally, the framework provides relationship investigation analysis of users in a particular region, key players identification and detection of existing communities. TwitterHitter provides analysts with a map view that shows the location of hot-spots where people are tweeting about a particular topic that match a given keyword. EVILCOHORT [78] is a

system that detects malicious online accounts that are controlled by cybercriminals. The system relies on detecting criminal accounts by identifying the connection points (IP addresses) used to access them. Typically, these IP addresses correspond to bots controlled by the criminal. By identifying communities of accounts that are accessed by a common set of computers (botnets) they were able to pro-actively detect these malicious accounts even before they spread the malicious content (e.g., spam).

Moreover, there are several commercial and open-source tools available that support open-source intelligence gathering. For example, Maltego [79] combines intelligence gathering and forensics of open online information. It offers mining and information gathering capabilities, with the ability to visualise the information in a graph-like format. This allows the analyst to infer relations that may be otherwise hidden. The tool uses the concept of entities and run different transformations on them. An entity could be anything from websites, IP addresses, people, companies, etc. and a transformation is a code that transform one entity to another. For example, through a transformation a website entity may be transformed to the IP address associated with that website. This will be visualised as new child entity of the original website. Another example is Wynyard Advanced Crime Analytics Platform [80], an off the shelf solution to help law enforcement reveal actionable intelligence hidden in the data. The platform supports time, link, geo-spatial, and mobile-device analysis.

2.2.4 Outstanding Research Challenges

The previous section discussed literature related to cybercrime intelligence focusing on the detection and analysis of cybercrimes. We found that different techniques are adopted to detect online cybercriminals such as machine learning and data mining. In addition, a variety of analysis methods are used such as content analysis and sentiment analysis to analyse their behaviour.

Furthermore, in order to evaluate the existing intelligence frameworks we compare them against a set of criteria identified from our review of the literature. Table 2.2 presents a summary of this comparison. The frameworks included in the table

were selected based on the year of publication and number of citations, as well as the relevance to our topic. The columns in Table 2.2 outlines the comparison criteria: these are the data-sources used by the framework; type of cybercrime they target; whether they aim at detection or analysis; support for data acquisition or only operating on previously collected datasets; the types of analysis they use; and finally whether or not the framework supports visual analytics. If this criterion is satisfied, we also look at support for feedback and interaction from the analyst.

Through our analysis of the literature, we identify several gaps which future research efforts should aim to fill. As observed in Table 2.2, none of the previously developed frameworks cover all three stages related to detection, analysis, and visualisation, and very few cut across both the detection and analysis stages. Thus, most frameworks are focused on a *single functionality* in order to solve a particular problem. This approach limits the investigator abilities as it narrows their view when working on a particular case. Similarly, nearly all of the frameworks we examined focus on only a *single data-source* (very few papers combine different data-types), which make the framework dependant on a particular platform (e.g., Twitter) and the methods used are not generalisable to other data sources and types. Additionally, in many cases cyber criminal groups make use of multiple platforms to plan their crimes, thus, a platform-focused framework may fail in providing the investigator with the whole picture.

Another area that is worth mentioning, although considered outside our scope, is visual analytics for cybercrime. We find that most of the existing frameworks either do not support any form of visualisation or provide very abstract and static visualisations. Those who do provide visual representation of the analysis results do not necessarily support user interaction with the framework. In order for the investigator to be able to make sense out of the data and results they need to have a way to inject their own expertise and hypothesis into the framework analytics.

Table 2.2: Review of existing intelligence frameworks

Ref	Framework description	Data Sources	Types of cybercrime	Detection	Analysis	Data acquisition	Content analysis	Social network analysis	Time analysis	Spatial analysis	Sentiment analysis	Visual analysis	Analyst feedback
[61]	Analysis and detection of microblogging spam accounts	Twitter	Spam	✓	✓	✗	✓	✓	✗	✗	✗	✗	
[78]	EVILCOHORT framework for detecting malicious online accounts	Generic online sources	General	✓	✗	✓	✗	✓	✗	✗		✗	
[70]	A sentiment analysis based web crawler for extremist content	Web pages	Terrorism	✓	✗	✓	✓	✗	✗	✗	✓	✗	
[71]	Terrorism and Extremism Network Extractor (TENE)	Web pages	Terrorism	✓	✗	✓	✓	✓	✗	✗	✗	✗	
[60]	Studies the relations between botmasters, spammers, and email harvesters	Emails	Spam	✗	✓	✗	✗	✗	✓	✓	✗	✗	
[81]	Models the social media reaction to terrorist attacks	Twitter	Terrorism	✗	✓	✗	✓	✓	✓	✗	✓	✗	
[75]	Isis toolkit, identifies individual or group identities hiding behind multiple personas	Chat messages	General	✗	✓	✗	✓	✗	✓	✗	✗	✓	✗
[77]	TwitterHitter a framework for harvesting Insight from volunteered geographic information	Twitter	General	✗	✓	✓	✓	✓	✓	✓	✗	✓	✓
[16]	Studies a hacker network called (Shadowcrew) using social network measures	News articles	Cyber-fraud	✗	✓	✗	✓	✓	✗	✗	✗	✗	
[69]	Dark Web Forum Portal that provides access to extremist web forums	Online Forums	Terrorism	✗	✓	✓	✓	✓	✗	✗	✗	✓	✗
[76]	Jigsaw visual analytic system to support investigative analysis	Offline reports	General	✗	✓	✗	✓	✓	✓	✗	✗	✓	✓
[49]	Collection, analysis, and visualization of dark web pages	Web pages	Terrorism	✗	✓	✓	✓	✓	✗	✗	✗	✓	✗
[67]	CrimeNet a framework for discovering criminal network knowledge	Offline police reports	General	✗	✓	✗	✓	✓	✗	✗	✗	✓	✓
[74]	COPLINK framework for criminal network analysis and visualization	Offline police reports	Terrorism	✗	✓	✗	✓	✓	✗	✗	✗	✓	✗
[63]	A general framework for crime data mining	Offline police reports	General	✗	✓	✗	✓	✓	✗	✗	✗	✗	

2.3 Requirements for Intelligence Frameworks

Requirements Engineering is the process of defining the requirements of a system, documenting these requirements, and maintaining them. It provides a framework to identify stakeholders, understand what a system should do, and how it will be used [82]. Two main types of requirements are typically identified, functional and non-functional requirements. Functional requirements deal with concerns related to the main functionality and behaviour of the system, while non-functional requirements describe the non-behavioural aspects of a system such as usability and performance.

Very few of the surveyed literature listed requirements for building intelligence systems. However, many of them focus on distinguishing the cognitive processes that underlie the intelligence analysis process and the associated cognitive biases [83–85]. Such processes could directly help inform the appropriate framework requirements. In order to understand the cognitive processes and biases it is first important to know what are the main steps included within the intelligence analysis process. Previous research proposed different cybercrime investigation models to guide law enforcement through the investigation process [4, 86, 87]. Although these models may seem different as they use different terminology to define the models' activities, most of them have similar processes.

At its core, the intelligence cycle consist of six main steps [87]: Direction, Collection, Processing, Analysis, Dissemination, and Feedback. The Direction step focus on identifying what are the questions to be answered, and plan the best course of action to be followed. The Collection step deals with collecting information and data (overtly and covertly) to help answer the identified questions. The Processing step is when the heterogeneous data is reformatted into a common format for future analysis. The Analysis step is the heart of the intelligence cycle where the processed data is fit together in order to find answers and extract intelligence. The Dissemination process is where the analysed data "intelligence" is shared with the intended stakeholders. Finally, the Feedback step is where the stakeholders give

their views on the intelligence by either accepting it or coming back with more questions to answer by repeating the process.

Elm et al. [83] describe the cognitive process for intelligence analysis as three interacting cognitive functions: (1) Down-Collect which focus on the extraction of an essential, representative sample from available data, (2) Conflict and Collaboration focus on the construction of interpretations of the findings from the Down-Collect process, and (3) Hypothesis Exploration focus on the construction of coherent stories/hypotheses to explain the interpretations of the findings. Furthermore, they identified five requirements for designing effective decision support system for intelligence analysis. Observability, the ability to gather insight and allow the analyst to see sequences and evolution over time, patterns and relationships. Directability, the ability to direct/redirect resources, and priorities as situations change. Teamwork with agents, the ability to coordinate, collaborate, and synchronise activity across agents. Directed attention, the ability to re-orient focus which allows the team of analysts to work in a coordinated manner. Finally, resilience, having failure strategies, and being able to anticipate and adapt to surprise, and error.

Previous related literature [84, 85] has developed design guidelines that act as recommendations for system designers in order to minimize the occurrence of cognitive biases. Hillemann et al. [84] carried out a requirements elicitation process through interviews, workshops and meetings with intelligence analysts in law enforcement in order to identify their needs, tasks they perform, and current issues with existing techniques. In many situations the prior beliefs and experiences (i.e., biases) of the decision-maker (the analyst) influence his decisions, which may lead to incorrect results. To avoid such issues, information should be presented in a way that minimises cognitive biases and supports the sense-making process. Typically, analysts deal with large volumes of data with the objective of extracting insights and actionable intelligence to make informed decisions. Cognitive biases may impact any step from the collection of raw data up to reaching actionable intelligence [84]. From these interviews, requirements were gathered and relevant biases that affect them were identified. The following biases have been recognised

in the literature in the context of intelligence analysis: 1) Confirmation bias, the most well known cognitive bias in decision making [84], in which the analyst selects information that match his initial expectation and hypothesis. 2) Evidence evaluation bias, which is related to persistent impressions on discredited evidence, evidence of uncertain accuracy, and missing information [85]. 3) Cause and effect perception biases, in which the analyst is looking for causal patterns and rejecting explanations of randomness or errors in the observed events [88].

Furthermore, existing research in the field of visual analytics and sense-making has shed the light on problems that designers need to address in their designs in order to support sense-making for information analysts. Wong et al. [89] identified three key problems that impact the intelligence gathering process: (i) Blackholes problem: the problem of representing missing data, (ii) Keyhole problem: the problem of being able to access and view only a small part of a large dataset or only part of a problem; and (iii) Brown worms problem: the problem of representing misleading or deceptive data. These three problems should be taken into account when designing systems for intelligence analysis.

In summary, the elicited requirements for cybercrime intelligence frameworks should include as part of its functional requirements support for the six steps of the intelligence cycle previously described. Additionally, as functional requirements the framework should allow collection of data and evidence, and allow the user to form connections and interpretations of the gathered evidence. Finally, the framework should be designed to mitigate the cognitive biases as much as possible. For example, as a non-functional requirement the system should have a clear separate view listing the gathered evidence, ranked based on their accuracy, to keep the analyst focused on tangible evidence and avoid internal biases.

2.4 Understanding Online Radicalisation

In this section, we focus on the topic of online radicalisation and review existing literature in the area. But first we need to define what we mean by online radicalisation. The U.S. Department of Justice defines online radicalisation as

"the process by which an individual is introduced to an ideological message and belief system that encourages movement from mainstream beliefs toward extreme views, primarily through the use of online media, including social networks such as Facebook, Twitter, and YouTube" [90]. There are several incidents of violent extremism that took place world-wide in the past few year. They stem from different ideological movements, such as religion (e.g., ISIS) and white nationalism (e.g., Alternative-Right). These groups utilise the Internet and online platforms to radicalise individuals and gather supporters.

According to experts, individuals are not radicalised by viewing a single video or propaganda, but it takes time and it occur gradually through constant exposure to extremist views. Realising the danger caused by the spread of violent extremism and radicalisation content online, and how it is becoming a major challenge to societies worldwide, many researchers have attempted to study the online behaviour of a number of extremist and terrorist groups.

The Islamic State in Iraq and Syria (ISIS) is one of the leading terrorist groups that utilise social media platforms and invest huge resources in creating digital media to spread their ideology and propaganda [38]. This is achieved by regular publications of a number of propaganda magazines, namely, Dabiq and Rumiyah, as well as a number of high-end quality videos and images promoting their activities [91]. Previous research looked at analysing the propaganda distributed by ISIS online and focused on understanding the narrative and logic used in these magazines [38, 91, 92]. For example, Ingram [91, 92] studies the strategic differences between Inspire magazine published by Al-Qaeda group and Dabiq magazine published by ISIS. He looks at how images, narrative, and counter-narrative are used to affect the readers' perceptions and polarise their support. Additionally, Allendorfer et al. [38] examined the strategies used by ISIS propaganda videos and compared it with those published by the U.S. Department of State to counter that propaganda narrative in order to evaluate how the structure and content of videos may be affecting the recipients.

2.4.1 Process of Online Radicalisation

Several research efforts attempted to study the process towards radicalisation in the online world. Torok et al. [93] used a grounded theory approach to develop an explanatory model for the radicalisation process utilising concepts of psychiatric power. Their findings show that the process typically starts with social isolation of individuals. This isolation seem to be self imposed as individuals tend to spend long time engaging with radical content. This leads to the concept of homophily, the tendency to interact and associate with similar others. Through constant interaction with like-minded people, an individual gradually strengthen their mindset and reach more extreme levels. Similarly, they start to feel as being part of a group with strong group identity which lead to group polarisation. In psychology, group polarisation occurs when discussion leads the group to adopt actions that are more extreme than the initial actions of the individual group members [94].

The National Police Service Agency of the Netherlands developed a model to describe the phases a "jihadist" may pass through before committing an act of terrorism [95]. These four sequential phases of radicalism are: 1) Preliminary phase where the individual experience a crisis of confidence in the authorities. 2) Social alienation phase where the individual feel alienated from the rest of the society, which increases their susceptibility for the extremist ideology. 3) Jihadization phase where the individual have strong involvement with radical groups and willingness to support them. Finally, 4) Jihad extremism phase is the last phase where the individual is prepared to use violence and commit terrorist acts.

Furthermore, since the process to radicalisation involve a lot of psychological and personality changes, an interesting angle to analyse propaganda material is to look at the psychological properties conveyed through the text. This is typically referred to as psycholinguistics, where one examines how the use of the language can be indicative of different psychological and personality states [96]. Examples of such properties include the OCEAN model [97] that include five components (big-five): Openness, Consciousness, Extraversion, Agreeableness, and Neuroticism. In addition, emotions, sensitivity, and personal drives are all examples of properties

that can be inferred from written communication. The social psychologist, Michael Billing, is one of the advocates of this approach, building on the theory that the choice of language is reflective of one's psychology [98]. A number of studies in the literature have demonstrated the potential and effectiveness of extracting psychological and personality properties from online communications [99–101].

One of the most used tools to automate the process of analysing text and extracting such properties is the Linguistic Inquiry and Word Count (LIWC) tool [102]. LIWC includes a semantic dictionary to measure properties for different categories such as cognitive, emotional, and personal drives. This approach has been used in the literature to study the behaviour of different groups and to predict their personality and psychological states [103–105]. For instance, it was used to predict depression [104], identify emotional expression in cancer narratives [103]. In addition, more recently it has been applied to identify different properties of extremist groups to uncover their targets as well as the intentions behind their recruitment campaigns [106].

2.4.2 Detection of Online Radicalisation

In the recent years, there has been an increase in online accounts advocating and supporting extremist groups such as alt-right and ISIS [107, 108]. This phenomenon has attracted researchers to study their online existence, and research ways to automatically detect these accounts and limit their spread. Ashcroft et al. [28] make an attempt to automatically detect Jihadist messages on Twitter. They adopt a machine-learning method to classify tweets as ISIS supporters or not. They focus on English tweets that contain a reference to a set of predefined English hashtags related to ISIS. Three different classes of features are used, including stylometric features, temporal features and sentiment features. However, one of the main limitations of their approach is that it is highly dependant on the data.

Rowe et al. [29] focused on studying Europe-based Twitter accounts in order to understand what happens before, during, and after they exhibit pro-ISIS behaviour. They define such behaviour as sharing of pro-ISIS content and/or using pro-ISIS

terms. To achieve this, they use term-based approach such that a user is considered to exhibit a radicalisation behaviour, if he uses more pro-ISIS terms than anti-ISIS terms. While such approach seems effective in distinguishing radicalised users, it is unable to properly deal with lexical ambiguity (i.e., polysemy). Furthermore, Kaati et al. [109] focused on detecting twitter users who are involved with "Media Mujahideen", a jihadist group who distribute propaganda content online. They used machine learning approach using a combination of data-dependent and data-independent features. Similar to [29] they used textual features as well as temporal features to classify tweets, and twitter accounts. The experiment reported in this work was based on a limited set of twitter accounts, which makes it difficult to generalise the results for a more complex and realistic scenario.

Weir et al. [110] attempted to classify crawled web-pages to "pro-extremist", "neutral", and "anti-extremist" using a combination of language analysis and data mining techniques. They used two different approaches on the same dataset, one based on keywords, part of speech tagging, and sentiment while the other approach was based on quantitative frequency analysis of syntactic features extracted from the web-pages. The reported results show that the approach based on frequency analysis yielded higher classification accuracy. Klausen et al. [111] used a different approach to detect pro-extremist accounts on twitter that is not based on textual features. They developed a behavioural model that is used to identify new extremist accounts and predict if it will be suspended by twitter for extremist activity. They used logistic regression using network-based features together with account-based features extracted from the user profile.

A number of recent works have looked at identifying different sensors for detecting potential online radicals or pro-extremist users [112–115]. For instance, the authors in [112, 113] looked at signals for detecting lone-wolf terrorists and violent extremists online. They suggest that those lone-actor terrorists leak online traces on their intent which can be utilised to predict attacks and prevent them in a proactive manner. Similarly, Cohen et al. [113] suggest there are a set of linguistic markers that can be used to detect radical violence in social media. In a recent study by

Williams et al. [114], they found that there is an association between aggregated online open source data and aggregated police-recorded crime data. In this research they used Twitter as the online source of data and the premise is that tweets can be used as sensors reporting on crimes occurring in a user's local environment. Furthermore, Alsaedi et al. [115] present a framework for detecting real-world events from Arabic Twitter posts. They use an integrated classification-clustering system that groups tweets discussing similar topics. The framework does not look at behavioural features and mainly considers textual, and spatio-temporal features. Finally, we will look into more literature related to detection of online radicalisation and present a comparison with our proposed approach in Chapter 7.

2.5 Summary

In this chapter, we presented an overview of the existing literature relevant to the key topics discussed in this thesis. We have also highlighted the research gaps in cybercrime intelligence and online radicalisation. We will describe how some of these gaps are addressed throughout this thesis.

3
Research Methodology

3.1 Introduction

The research conducted in this thesis has a multi-disciplinary nature as it cuts across different disciplines, including computer science, criminology, behaviour science, and psychology. We adopt mixed methods to answer the different research questions outlined in Chapter 1. In this chapter, we provide a high-level description of the core research methods used throughout this thesis. However, the detailed methodological steps followed are included in each subsequent chapter.

3.2 Qualitative Methods

Data Collection — Questionnaires, surveys, and interviews are different methods to collect data from users. Due to the sensitive topic of this research, it is difficult to gather rich detailed information from experts in the field of cybercrime intelligence through general means (e.g., surveys). They are not likely to participate and share detailed views without us gaining their trust and confidence. Also, we intend to collect detailed data to capture practitioners experiences and needs, which can only be achieved through direct communication. Therefore, we chose to collect information by a combination of questionnaire and interviews. The questionnaire was used first to gather generic information about the participant's role and expertise

within the field of cybercrime. Additionally, it acted as a means for us to assess their applicability to participate in the interview process. We then conducted face-to-face semi-structured interviews with open-ended questions, to allow the interviewees to elaborate on their experiences when dealing with different incidents, provide lengthy and descriptive answers, and make them express their views in their own terms. This study is presented in Chapter 4.

Data Analysis — All interviews were audio-recorded and later transcribed to facilitate the analysis. For the extraction of qualitative findings, we adopted a thematic analysis [116] approach. The thematic analysis focuses on identifying common themes within the data. A theme is defined as a recurrent feature or topic describing particular perceptions and experiences relevant to the research questions [117]. It allows for theoretical freedom, as it provides a flexible and useful research tool, which can yield a rich and in-depth account of the data [118]. We adopted thematic analyses with a mixture of deductive and inductive approaches to identify themes and analyse the interview data [117]. While the deductive approach allows us to guide the flow of analysis according to our aims and predefined generic themes from the literature, inductive analysis allows us to incorporate additional themes that are new and arise from the data [116]. This allows us to generate findings from the bottom up to be inductive and identify patterns based on participants' expressed reasoning. This is in contrast to using rigid pre-set criteria, which may or may not map to participants' perspectives. We used a specialised qualitative analysis software package (i.e., NVivo[1].) to perform the coding and support our analyses.

Coding — We followed the popular approach outlined by Ritchie et al. [119]. We began by familiarising ourselves with the data by parsing through the transcripts in detail. This enabled initial patterns to emerge. We then started coding by assigning annotations to reoccurring ideas. This was an iterative process where, at the end, we grouped related codes into themes. Moreover, since participants hold different roles and are associated with different types of organisations, they had some unique views.

[1]https://www.qsrinternational.com/nvivo/home

As such, we also included themes that were salient to a particular organisations and roles. By including these deviant cases [120], our themes should represent the data.

Reliability — Since the coding process was carried out by a single coder, to mitigate any misinterpretation and biases, we performed a number of iterations of the coding process spread over a long period. This allowed us to refine the coding and the produced themes, thus increasing the robustness of the process. Furthermore, the sample size of participants in the study presented in Chapter 4 is in line with in-depth interview guidelines [121]. As explained by Crouch and McKenzie, an in-depth interview study is one which "scrutinises the dynamic qualities of a situation (rather than elucidating the proportionate relationships among its constituents)" [121]. As such, having a small number of participants when conducting an exploratory in-depth study has little bearing on the reliability of the findings.

3.3 Quantitative Methods

Different types of quantitative measures were used in this research. In this section, we describe the key measures used and provide justification for their selection.

Questionnaire Analysis — Quantitative methods were used to analyse questionnaire data that were collected from participants in the study documented in Chapter 4. The questionnaire included categorical data which was analysed using simple descriptive statistical methods. Also, it included a set of assertions to collect participants' views. The Likert scale [122] is considered an efficient method for collecting data related to participants' attitudes [123]. As such, it was used to capture participants' level of agreement with a set of statements and requirements gathered from the literature. The scale categories include: 1 *strongly disagree*, 2 *disagree*, 3 *neutral*, 4 *agree*, and 5 *strongly agree*. To avoid forced choices, it is important to include a neutral response so as not to coerce an answer. We calculated the percentages of participants' attitudes towards each of these statements.

Measures — Several quantitative measures were used throughout this research. This included similarity measures, distance measures, and significance tests. In the following section, we define each of these measures.

Cosine Similarity. In the study described in Chapter 6, we compare topics of different propaganda magazines in order to identify how similar their topics are. Many measures can be used to measure the similarity between two documents. We opted to use cosine similarity as it has been identified in the literature as being most appropriate for information retrieval and text mining tasks. It measures the cosine of the angle between two vector representations of the documents. The smaller the angle, the higher the cosine similarity [124]. An important property of this measure is that it gives a similarity score for two documents based on how similar their subject matter is, regardless of the length of the documents. Thus, this makes it suitable for our task of comparing variable length documents.

Manhattan Distance. In Chapter 7, we describe an experiment that aimed to detect radical tweets. As part of the experiment, we measured the distance between a user's psychological profile and a high dimensional radical profile. In such a case, Manhattan distance (L1 norm) is preferable to other types of measures such as Euclidean distance, due to its applicability for high dimensional data [125].

Significance Tests — These tests were used to assess whether a statistically significant difference exists between multiple groups. In Chapter 8, we studied the properties of extremist propaganda published by different groups (AL-Qaeda and Alt-Right). We compared these properties against the ISIS propaganda and a control group (news articles). To determine if the differences were statistically significant, we used appropriate significance tests. We examined certain characteristics of the different groups to select a suitable test. These were the number of data records in each group, the variance of each group, and number of groups to compare. In case of comparing two groups, if both had the same variance and had the same number of records we performed an equal variance (independent) t-test [126]. However, if both had a variable number of records and different variance, then we performed an unequal variance t-test. If the samples were not drawn from a normal distribution, we used the non-parametric statistical test Mann-Whitney U test [127]. In case of comparing three or more groups, the Kruskal–Wallis H test [128] were employed.

3.4 Computational Analyses

Methods of online research, in the context of social research, can be divided into obtrusive and unobtrusive research [129]. Obtrusive research, also called reactive research, is when the researcher has direct contact with research subjects. It includes surveys, interviews, focus-groups, and field observations. Meanwhile, unobtrusive research is when the researcher has no contact with the research subjects. It includes two main types: secondary data analysis methods and content analysis. In our research documented in Chapters 6, 7, and 8, we use unobtrusive content analysis methods, which include quantitative web content analysis and social media network analysis. Such analysis is performed through different methods, including Natural Language Processing (NLP), Machine Learning, and Social Network Analysis (SNA). In the next section, we describe the main methodological concepts we use.

3.4.1 Natural Language Processing

NLP explores how computers can be used to understand and manipulate natural language text to accomplish a desired task [130]. NLP consists of a sub-field of artificial intelligence, information engineering, and computer science. Applications of NLP include machine translation, sentiment analysis, content analysis, and summarisation. To computationally analyse text, we need to transform it to a representation that can be processed by different algorithms. In the following, we describe the main techniques we used in this research.

Bag of Words (BOW) — A BoW model is a simple and effective representation of text that reflects the number of word occurrences. Using this approach, we focus on the histogram of the words in the text rather than where they occur in the text. The intuition here is that if two documents include a similar set of words, then they are considered similar [131].

TF-IDF — Short for Term Frequency-Inverse Document Frequency, which is a statistical method used to measure how important a word is to a document in a corpus. The TF-IDF value increases in proportion to the number of times a word appears in the document and is offset by the number of documents in the

corpus that contain the word; this allows for adjustments as some words, such as articles and prepositions, generally appear more frequently. As a result, common words in a small group of documents will have higher TF-IDF numbers [132]. Nonetheless, TF-IDF does not account for the semantic relations of the words; this is addressed by the following method.

Word Embedding — It is a real-valued vector representation of text such that words that have the same meaning will have a similar representation. Vector values are learned using neural networks [133] and require a large amount of text. There are many pre-trained embeddings available that are already trained on a large corpus, which can be used to seed the model and update later based on a specialised corpus. Different algorithms can be used to learn a word embedding. In the following, we describe the most common techniques and highlight their advantages and disadvantages.

Word2Vec— [134] is a statistical method developed at Google for learning word embedding from text corpus. It uses skip-gram model [135], which is an efficient method that does not involve dense matrix multiplications for learning vector representations of words from unstructured text. It has become the de-facto standard for developing pre-trained word embedding. As demonstrated in [134], semantic information is preserved when performing operations on words vectors. A popular example used to illustrate this feature is: *King - Man + Woman*, which results in *Queen*. This method is considered the most effective in preserving such semantic relationships between words. However, it requires a large amount of training corpus.

FastText— [136] is an extension of the skip-gram model [135] developed at Facebook, to learn representations for character n-grams. This method represents words by the sum of their character n-gram vectors. As such, it can capture sub-word information. This feature makes this approach capable of handling out-of-vocabulary words (words that did not appear in the training data), by constructing the representations from its character n-grams. However, studies looking at comparing the representation of the semantic relationship between words, found that skip-gram Word2Vec models perform better in capturing such relations.

As a result, after testing both models on our corpus, we opted to use Word2Vec model to build a semantic representation of the radical propaganda as it better captured the semantic relations of our corpus.

3.4.2 Machine Learning

Machine Learning can be defined as the process of automated detection of meaningful patterns in the data [137]. Machine learning models are categorised as follows: supervised (labelled data) and unsupervised models (unlabelled data). In this research, we use supervised machine learning methods and perform classification tasks to identify radicalisation activities. In the following, we describe the main algorithms and evaluation metrics we use.

Classification Algorithms — Random Forest and Neural Networks algorithms have been identified in the literature to produce the best performances for text classification tasks [138, 139]. *Random Forest* (RF) [140] is an ensemble learning method that creates multiple decision trees using bootstrapped datasets. Using a majority-wins model, it reduces the risk of error from an individual tree. Thus, it outperforms decision trees as it has a hierarchical structure. This allows RF to be able to model non-linear decision boundaries. RF is scalable and is robust to outliers. *Neural Networks* (NN) are known to perform well when applied to tasks related to image recognition and natural language processing [138, 141]. However, they usually tend to require large amounts of data to train.

Evaluation Metrics — We use multiple evaluation metrics to assess the performance of the classification model. *Accuracy.* looks at the overall performance of the model by measuring the sum of true positives and true negatives divided by the total number of observations. *Recall* is considered a measure of effectiveness in retrieving positive classes [142]. It helps when the cost of a false negative is high. For example, it measures how many radical tweets we can detect out of all the radical tweets in the data. *Precision*, on the other hand, is a measure of purity as it considers the number of retrieved positive classes that are labelled as positive class [142]. It helps when the costs of false positives are high. For instance, it

measures how many radical tweets we can detect without falsely accusing anyone. Thus, if we identify every single tweet as radical, we will expose all radical tweets and obtain high recall, but at the same time, we will call everyone in the population a radical and obtain low precision. *F-measure* is the average of both precision and recall. When predicting the probability of a binary outcome, the Receiver Operating Characteristic curve (ROC curve) is a useful tool to evaluate the performance of the model. It plots false-positive rate against true-positive rate. We use the *area under the ROC curve* (AUC) to summarize the performance of the model.

3.4.3 Social Network Analysis (SNA)

In graph theory, centrality measures identify the most important vertices in a graph [143]. In analysing social networks for identifying a user's behaviour and their influence level, several SNA metrics have been deemed useful. SNA metrics can be divided into node-level metrics (e.g, degree centrality), and network-level metrics (e.g., network density). SNA is an important tool that facilitates the identification of properties related to criminal and terrorist networks using traces of collected communications. In the following, we give a brief description of the main metrics used in this research.

Degree Centrality — Measures the number of direct links a given node has. The higher the score of a node, the more connected the node is in the network. In a directed graph, it is also useful to measure the in-degree and out-degree to identify the direction of influence/information flow. For example, in the context of a retweet network, users with high in-degree values represent the role of information generators, while users with high out-degree values correspond to information distributor role. While degree centrality focuses on the number of direct connections a node has, it does not consider the global localisation the node has in the network and does not take into consideration how powerful those connections are in the graph. Another centrality measure that focuses on the influence of the connected nodes is Eigenvector Centrality which is described below.

Eigenvector Centrality — A node that has a high Eigenvector score is one that is adjacent to nodes that themselves have high scores [143]. This centrality is useful to identify which nodes can spread information to many other nodes quickly.

Betweenness Centrality — This is a measure that indicates if a node has an intermediary role in the network (i.e., broker role) by looking at all the shortest paths that pass through a particular node. In other words, nodes with high betweenness centrality mean that they have the most control over information flow in the network [144].

PageRank — This is a variant of Eigenvector centrality [145], which can be used to uncover influential nodes whose reach extends beyond their direct connections. One main difference to Eigenvector centrality is that it considers the link weight and direction when calculating the score [146]. Therefore, it is more suitable for directed graphs.

HITS — Hypertext Induced Topic Selection also called *Authorities and Hubs* [147]. Similar to PageRank, it is used in directed graphs to capture the importance of nodes. Nodes have high authority values if they are pointed to by many nodes with high hub values. Similarly, nodes have high hub values if they point to nodes with high authority values. In an information network, this measure indicates how valuable the information carried by each node is.

3.5 Ethics

Formal ethical approval was obtained from the Central University Research Ethics Committee at the University of Oxford. Besides, careful adherence to research ethics guidelines were followed before, during and after the study, the reporting of results, and the storage of the collected data.

In the expert interviews study (Chapter 4), the process of participants' consent took place at the beginning of each study. Participants were presented with an information sheet that described the aim of the study, information on the handling and anonymisation of their data, processes for withdrawal of their data, and contact details should they have any concerns. Data related to the participating

individuals and organisations were anonymised in any published results, as per participants' requests.

Furthermore, in our radicalisation study, all user-related data were anonymised. All collected data related to extremist material was carefully handled and stored. Given the sensitive nature of the topic, we followed the university guidance on research involving security-sensitive research material, and measures have been put in place to consider the psychological safety of the main researcher.

3.6 Summary

In this chapter, we outlined our high-level methodology. We justified the usage of a range of quantitative and qualitative methods. We also outlined the computational methods used to perform content analysis. Finally, we described the ethical considerations followed in this research.

4

Understanding Cybercrime Investigations: Process, Needs, & Challenges

4.1 Introduction

Cybercrime investigators are facing numerous difficulties trying to keep the pace with the increase in numbers and the evolution of techniques used by cybercriminals. Previous studies suggest that the methods and processes typically used by law enforcement when investigating traditional crimes do not necessarily apply in the cyber world [7, 8]. Thus, there is a need to adopt a different process and build a new set of skills and knowledge in order to be able to mitigate these technologically advanced crimes. This is essential mainly as the cybercriminals are usually technologically-aware and are constantly adapting and developing new tools to allow them to stay ahead of law enforcement investigations [9–11].

While there are many research efforts focused on understanding and analysing the tactics used by cybercriminals [11, 20, 21], very little research exists that focuses on understanding the processes, challenges, and needs of law enforcement. With the majority of the research literature focusing on challenges related to the digital forensics side of cybercrime investigations [148–150], few focus on studying the challenges related to the holistic process of cybercrime investigations. Additionally, current cybercrime literature, especially those considering the technical side, tend to

primarily focus on proposing and developing new solutions and tools that researchers believe are required by cybercrime investigators. However, little research in the literature focuses on understanding the needs and challenges that are actually being faced by the investigators. The novelty of this study is that it attempts to bridge that gap, thereby setting a better foundation for future research in the field.

The majority of the literature attempted to use surveys with close-ended questions as a means to collect data related to how police officers deal with cybercrimes [30–32]. One advantage of using this approach is collecting answers from a large number of participants. However, a disadvantage is that it limits the ability to capture insights and understand the reasons behind the chosen answers. Therefore, we decided to conduct a qualitative study using a combination of direct questionnaires and semi-structured interviews. The questionnaires were used to gather generic information about the participants' role and expertise within the field of cybercrime, while the interviews gathered in-depth data.

Additionally, related literature studied the views of a particular type of participants dealing with cybercrimes (e.g., local police officers). Very few looked at this issue from multiple viewpoints of different participant types (i.e., local officers, regional and national cybercrime units). In this study, we address this gap by interviewing participants dealing with cybercrimes from the government-sector (including local and regional units) as well as from the private sector. We interviewed a total of ten experts who hold varying roles and responsibilities, including operational and managerial positions. This gives us diversity in the collected viewpoints and richness of data.

This chapter makes the following contributions to the cybercrime, socio-technical, and usable-security fields:

1. Establishes an empirical understanding of some of the key processes used by cybercrime investigators within government (e.g., police and law enforcement) and private (security intelligence) sectors.

2. Identifies a set of outstanding and important challenges (practical, procedural and usability-based) faced by investigators while combating cybercrimes and gathering intelligence. This is useful in terms of directing future research towards more user-centric approaches, practices and systems for investigators.

3. Presents recommendations for areas of improvement associated with the processes of intelligence sharing, reporting cybercrimes, skills and training, and improving the usability of cybercrime systems.

4.2 Methods

In order to gain an understanding of the different socio-technical challenges and needs associated with investigating cybercrime incidents, multiple qualitative methods can be used. Questionnaires, surveys, and interviews are all possible methods that can be used to achieve this [119]. However, our aim is to understand the problem in-depth and also the needs from multiple stakeholders' perspectives, such as government (law enforcement agents), and private sectors (security intelligence and consultancy companies). Additionally, due to the sensitive nature of this research, it is difficult to gather such detailed information from experts in the field by general means (e.g., surveys) as they are not likely to participate without us gaining their trust and confidence using direct communication. Another viable option is to rely on observations, which would provide rich data, however, there are many difficulties inherent in observing this kind of activity directly. Therefore, we rely on participants' reports via a combination of direct questionnaires and interviews to collect our data.

4.2.1 Study Design

We designed the study to be conducted in two phases: first, we used a questionnaire to gather generic information about the participants' role and expertise within the field of cybercrime. This acts as a means for us to assess their applicability to participate in the interview process and be familiar with their area of expertise. Second, we conducted face-to-face, semi-structured interviews using open-ended

questions. This allowed the interviewees to elaborate on their experiences when dealing with different cybercrime incidents, provide lengthy and descriptive answers, whilst also allowing us to ask probing and clarifying questions where needed.

The questionnaire focused on collecting information to give us a general overview of the participants' role, expertise, and technical capabilities. In addition, we used the questionnaire to capture participants' views regarding a set of assertions related to the process of investigating cybercrimes. We used a five point Likert Scale to capture how strongly they agree/disagree with each statement. The final set of questions and assertions used are summarised in Tables 4.1 and 4.2.

After collecting the questionnaire responses, we proceeded with conducting the face-to-face, semi-structured interviews. These interviews ranged from 45-min to 1.5 hours long, depending on the availability of the participant and the level of detail they were able to provide. We formed a set of predefined questions informed by our review of the literature to guide the interview, but we also asked some probing questions when needed [151]. All participants were provided an information sheet describing the project prior to participating. All interviews were audio recorded and then manually transcribed producing transcripts for each participant discussion. Additionally, to ensure the validity of the questionnaire and the interview questions, we discussed them with a subject matter expert and incorporated appropriate feedback before conducting the study.

4.2.2 Interview Questions Design

When designing the interview questions we focused on four main themes to capture the socio-technical aspects relevant to fighting cybercrimes, as described below. These themes emerged from our understanding of the literature and aims to gather details regarding the incident, the process, the people, and the technology. An overview of the guiding questions are summarised in Table 4.3 and are described below.

Table 4.1: Questionnaire Sample

#	**Questions**
Q.1	What is your current position?
Q.2	Describe your role within your organisation
Q.3	How do you rate your experience with the topic of cybercrime? [1 No experience, 2 Little experience (1-2 years), 3 Moderate (3 -5 years), 4 Expert (5+ years)]
Q.4	Which types of cyber-crime have you dealt with? [Select all that apply] 4.1 Offences against the confidentiality, integrity and availability of computer data and systems] 4.2 Computer-related offences 4.3 Content-related offences 4.4 Copyright-related or Intellectual-property offences 4.5 Cyber Terrorism and Cyber-warfare 4.6 Others
Q.5	How would you describe your involvement in investigating and countering these crimes? [Select all that apply] 5.1 Detection of cybercriminals 5.2 Analytics of cybercrimes 5.3 Mitigation of cyber-crimes 5.4 Policy formulation 5.5 Derive approaches to address these crimes 5.6 Others
Q.6	Technical Expertise 6.1 Intelligence gathering from open online sources (OSINT) 6.2 Visual Analytics 6.3 Social Network Analysis 6.4 Machine Learning 6.5 Data Mining 6.6 Others
Q.7	Have you held any managerial/leadership position in operations, investigating, and countering cybercrimes?

Theme 1: Cybercrime Incidents

The focus of this theme was to understand which types of cybercrime incidents were dealt with by the respective organisation, the volume of incidents reported, and whether they had the capacity and resources to investigate all reported incidents. If not all incidents are investigated, we wanted to identify the filtering or scoring mechanisms they used to determine which incidents to investigate.

Table 4.2: Questionnaire - Assertions

A#	**Assertions: Do you agree/ disagree, with the below statements 5 Strongly Agree, 4 Agree, 3 Neutral, 2 Disagree, 1 Strongly Disagree**
A.1	A complete intelligence gathering life cycle should consist of five stages: data acquisition, detection, analysis, visualisation, feedback from the analyst
A.2	Visual presentations of data within current intelligence tools do not have adequate interactivity to support data exploration
A.3	Data visualisation is an essential method for communicating findings to colleagues, and documenting decisions.
A.4	The investigation of large volume of data, and the discovery of patterns within data, can be enabled through visual data presentation.
A.5	When formulating/testing hypotheses analysts search for patterns or trends from data.
A.6	Analysts usually uses a series of filters to reduce data and find relevant intelligence.
A.7	Keyword-based search and analysis techniques are not enough to identify the intent of the text (message/post).
A.8	Intelligence gathering tools should provide the analyst with the ability to distinguish between different levels of uncertainty in the data (e.g., assign confidence scores to each data item)
A.9	Intelligence gathering tools should be platform-independent.
A.10	Intelligence gathering tools should be scalable and able to handle big-data.
A.11	Analytical process consist of Six steps: 1-Problem definition, 2-Hypotheses generation, 3-Information collection, 4-Hypotheses evaluation, 5- Selecting the most likely hypothesis, 6-Continuous monitoring of new information.
A.12	[if agree to A.11] Do current tools provide support for each of these steps.

Table 4.3: Interview Questions

Role	What are your role(s) within the organisation?
Cybercrime Incidents	1. Do you deal with cyber-crime investigations in your department? 2. What types of incidents/investigations do you deal with? 3. How many incidents are reported (volume)? 4. Is it reported by individuals or business? 5. What types of cybercrimes are most often reported? 6. Are all reported crimes investigated or subset?
Investigation Team	1. How big is the team that work on a given cyber-crime incident? 2. What are the main roles included in the team?Is there a hierarchy? [e.g., lead investigator, analyst, data collector, etc.] 3. What percentage of the team if any have cyber-security expertise/background?
Investigation Process	1. Differences in process of investigating traditional and cyber crimes? 2. How does the initiation of any investigation occur? 3. Is it reactive / proactive? Or both? 4. What is the process you follow when investigating a cybercrime? 5. Would you describe the process as iterative? 6. Do you use top-down or bottom-up approach? Or both? What is used when? Is it dependant on crime-type? 7. What are the tasks associated with the investigation process? 8. Which tasks are most time consuming? 9. Are the tasks dependent on the crime type? If yes, can you give examples? 10. What types of evidence do you look for in an investigation? What is considered an evidence in OSINT investigation? 11. During an investigation, do you use the any of the following techniques: a. Social Network Analysis b. Time analysis c. Sentiment analysis d. Textual analysis e. Entity extraction [e.g., identify persons, addresses, places] f. Topic modelling g. Data mining h. Machine learning 12. What are current major challenges you face when investigating a cybercrime?
Tool Support	1. What types of tools do you use to investigate these crimes? 2. Why did you choose the set of tools that you have? 3. Do these tools work well together? 4. Do you use open online sources to gather intelligence? 5. Are automated analysis tools more appropriate for investigating cybercrimes? 6. Do current tools used support user's feedback and interactions with the framework? 7. Do they support collaboration? a. Team members sharing analysis/investigation? b. Collaborative investigation sessions (users work on a shared case scenario). 8. Any of the existing frameworks/tools you use provide support for all the five stages: [data acquisition, detection, analysis, visualisation, and feedback]. 9. If the investigation consist of manual processes, what are they and do you think it can benefit from automation? 10. What capabilities you want the current tools to provide and they do not?

Finally, the method by which victims report cybercrimes and how such interaction occur was explored.

Theme 2: Investigation Process

After understanding the types of incidents investigated, we then explored the actual process used to investigate a given incident. Under this theme, we collected information on the different tasks performed, if the investigation process was iterative, and whether it was a reactive or proactive process. Understanding the details of the investigation process will allow us to map the different steps performed and identify, where applicable, how each step can benefit from technology adoption.

Theme 3: Investigation Team

This theme sought to understand the dynamics within the investigation team. How big is the team that works on any single case? What role does each team member typically have? Are they all located in the same location? Such questions will allow us to ascertain if there is a need for processes, systems or tools to support collaborative investigation sessions. This relates to the field of Computer-Supported Cooperative Work (CSCW), which focuses on describing how collaborative activities and cooperation can be supported by computer systems [152].

Theme 4: Tool Support

This theme covered the different tools currently being used by the investigation teams. The aim was to understand what advantages were gained from using these tools, the perceived usability of the tools, their availability, and effectiveness at supporting the investigation tasks. We also asked about the perceived limitations of the tools, which is important to be able to enhance these limitations and provide better support to investigators.

4.2.3 Recruitment of Experts

Since our aim is to gather the perspectives of practitioners working in various organisations who deal with investigating and mitigating cybercrimes, we targeted a population from both government and private sectors. Participants were recruited based on their knowledge and experience in investigating cybercrimes and gathering intelligence. Government sector organisations that handle cybercrimes have a hierarchical structure that consists of multiple levels, including local cybercrime units, regional cybercrime units, and national cybercrime units. We aimed at recruiting participants from each of these levels.

We used a snowball sampling approach [153] where we asked initial participants to recommend candidates from their network. The recruited participants held different roles; some possessed technical experience while others held more managerial positions. This diverse sample of experiences and backgrounds was important to capture different perspectives on the topic.

Reaching out to professionals working in this field is a difficult task given the sensitivity of the topic. Many of the individuals we contacted were very busy and could not afford the time to conduct a full face-to-face interview. In total, we interviewed ten experts, six from UK law enforcement, and four from the private sector, one of whom had ten years of experience working for law enforcement before moving to the private sector. Due to the sensitive nature of the topic, all the participating organisations and individuals requested to remain anonymous.

4.2.4 Data Analysis

The data collected from the questionnaires consist of open-ended text (e.g., participant's role) and categorical data (e.g., types of cybercrimes). The categorical data was analysed using descriptive statistics by calculating the frequencies for each category. The results from the questionnaires played a role in guiding the discussion during the interview session to areas that matched the participant's experience. Likert scale results for the assertion statements were analysed by

calculating the mode of the most frequent responses and measured the distribution of responses i.e. (% that agree, disagree).

To analyse the interviews data, thematic analysis was chosen as it best fits the exploratory nature of this study as described in Chapter 3. We began our analysis with initial coding using a deductive approach by coding the data to a pre-defined set of codes relevant to our research interest. These codes correspond to the four themes described earlier, i.e., cybercrime incidents, investigation process, investigation team, and technology. Then, through an iterative analysis approach, the coder immerses herself in the data and assign new codes where appropriate to the recurring ideas. Using this approach makes the findings more robust as it allows us to test emerging themes against new data.

An example of our coding process is shown in the quote below, where the participant was asked about their impression of utilising automated tools in their investigation of different cybercrimes. This question falls under the technology and tool support generic theme.

"I think fully automated tools need to be introduced into our work. Purely because you know budget cuts. So there's going to be no more resources. So we need these tools, scripts, data mining tools to go through everything we see."

We initially assigned the code *impression of automated tools* to highlight the participant's views on the topic during the first coding round. Then, in our subsequent iterations, we assigned the code *drivers for automation* and *resources availability* as they also mentioned the lack of resources and budget cuts as reasons for their views. Overall, we identified 35 codes in our data, that were then refined and grouped into themes.

4.3 Results

Through the analysis of the interview data, we attempt to understand different aspects connected to the process followed when investigating different cybercrimes. Several themes appeared across the interview data that show how private and public sectors carry out their investigations. In this section, we first present

the demographics of interview participants. Second, we discuss results from the collected questionnaires. These give us an idea regarding the expertise of the sampled participants. Third, we report results from the interview analysis relating to the different identified themes.

4.3.1 Participants Demographics

We interviewed 10 participants who are experts in cybercrime investigations. The participants work in the government (6 participants) and private sector (4 participants). All participants are based in the United Kingdom. Of the participants, 4 are intelligence analysts; 2 are intelligence researchers; 1 is a sergeant; 1 is a detective; 2 are law enforcement agents dealing with cybercrimes (1 is currently working as a consultant). The participants are associated with 5 different organisations: 2 private sector, and 3 government sector. The government sector organisations tackled cybercrime at different levels, including local, and regional forces. This allowed us to collect the viewpoints from experts working at tackling cybercrimes from different levels. The private sector organisations deal with security investigations and intelligence gathering from open and closed sources to provide clients with threat intelligence.

Table 4.4: Participant Groups

Group ID	**Roles**	**Organisation Type**
G1	Intelligence analysts and researchers	Private and Ex-Government
G2	Cybercrime officers and detectives	Government
G3	Senior intelligence analysts and researchers	Government
G4	Senior cybercrime officers and detectives	Private and Government

To ensure proper anonymisation and to allow for better protection of our participants' identities, we grouped them into 4 groups based on their roles and level of experience. Table 4.4 summarises the different groups by providing the Group ID (GID), the job roles of participants, and the type of organisation they

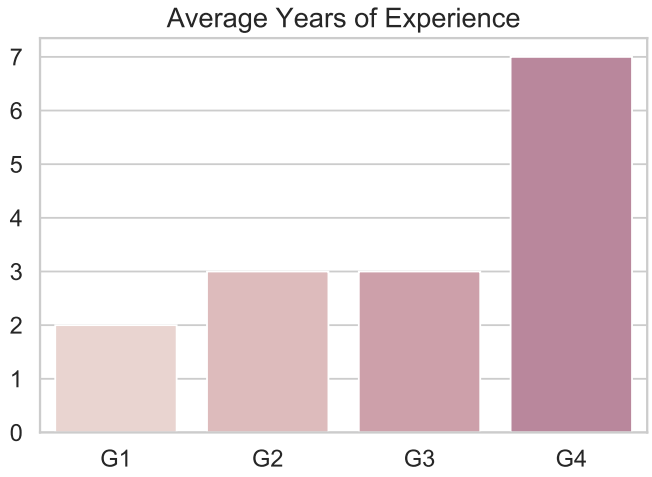


Figure 4.1: Participants' average years of experience with cybercrimes per group

have worked in (i.e., private, or government). We note that each group consist of a minimum of 2 members.

4.3.2 Questionnaires Results

The questionnaires were designed to collect data that can paint a picture of the participants' experience, role, types of cybercrimes they investigated, and their technical capabilities. Additionally, it was used as a tool to obtain the participants' views on a set of assertion statements related to cybercrime investigations.

Years of Experience— The average years of experience participants had was 3.5 years working in investigating cybercrimes. The minimum was 1 year and the maximum was up to 10 years of experience. They held roles that ranged from intelligence analysts, consultants, lead analysts, researchers, and investigation officers. This generates multiple viewpoints and allows us to gather requirements from different angles (Figure 4.1).

Types of Cybercrimes— A list of different types of cybercrimes were presented to the participants and we asked them to select the crimes that they have experience dealing with. The results were distributed between cyber-enabled and cyber-dependent crimes, with the majority having experience dealing with cyber-fraud, spam, hacktivism, and child pornography. Additionally, participants noted

that after the 2017 WannaCry ransomware, the numbers of reported ransomware spiked and became the most reported cybercrime (Figure 4.2).

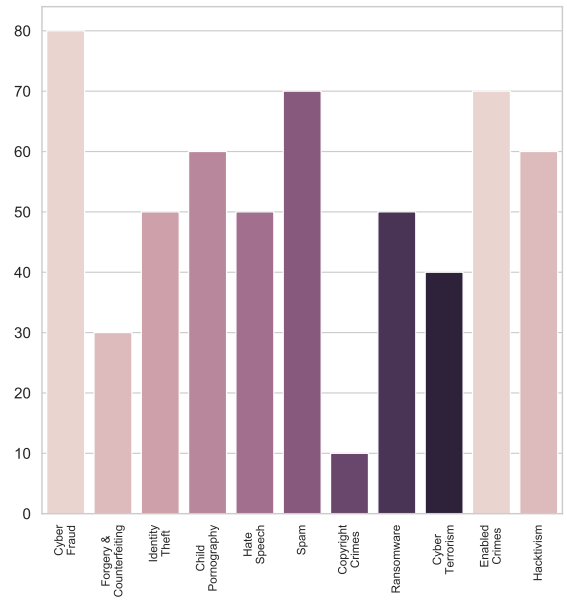


Figure 4.2: Participants' experiences with different types of cybercrimes (%)

Role of Investigation— We wanted to know if participants' involvement with the investigation had an operational, tactical, or managerial nature. We asked them to describe and categorise their role as contributing to the (1) detection phase, (2) analysis phase, (3) mitigation techniques, (4) policy formulation, or (5) derive approaches to address crimes. The majority (around 70%) were mainly involved in the analysis and mitigation of cybercrimes. Additionally, around 50% of participants were involved in the detection phase and in deriving different approaches to address these crimes. Only 30% of participants had a role in formulating cybercrime policies (Figure 4.3). We can see the diverse experiences in the participants sample that will allow us to collect in-depth data from different view points.

Techniques Used— We asked the participants about the techniques and experiences they adopt when investigating cybercrimes. Among the most common techniques were OSINT (Open Source Intelligence) that gathers data from online sources, Social Network Analysis (SNA) to study relationships, links, and interactions between criminals, and visual analytics. None of the participants reported

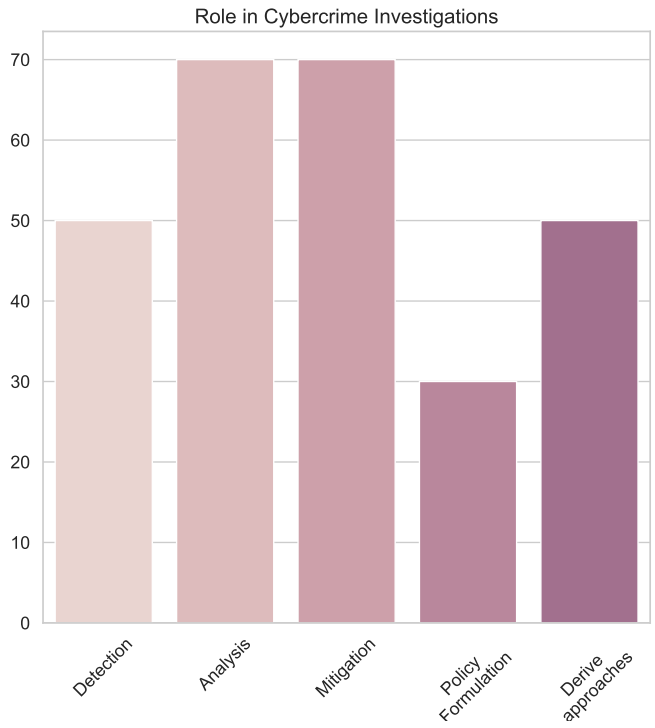


Figure 4.3: Participants involvement with cybercrimes (%)

that they use machine learning approaches in their investigations and only some reported using data mining techniques.

Assertions

The assertion statements are designed to measure participants' agreement with a set of claims inferred from the literature related to the process and technological support of cybercrime investigations. Five assertions were focused on the investigation process *(A.1, A.5, A.6, A.7, A.11)*, and six assertions were focused on the technological support *(A.2, A.3, A.4, A.8, A.9, A.10)*. All eleven assertions and corresponding distribution of responses are presented in Figure 4.4.

Overall, participants agreed with the majority of assertions from the literature, with the exception of four assertions that prompted some level of disagreement.

Both *A.1* and *A.11* from the process-related assertions prompted disagreement from some of the participants. When discussing this with them, they clarified that their objection was not related to the described steps, but rather it was based

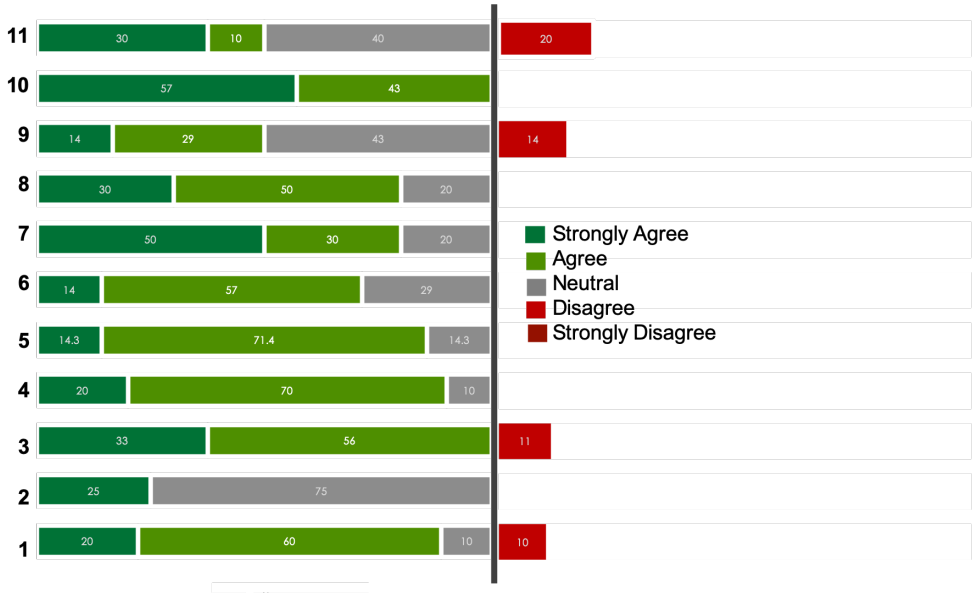


Figure 4.4: Assertions results

on it being depicted as a cycle or an ordered process. From their experience, the investigation process steps are not strictly ordered but are quite interchangeable depending on the situation and incident at hand. For example, they might start with data collection, and then do some analysis before going back to collect additional data, and then do more analysis before visualising the results and obtaining more feedback.

As for the technology related assertions, *A.3* and *A.9* prompted disagreement from some participants. We noticed that there is a relation between those who disagreed and the role/level of experience they hold. For instance, regarding the importance of visualisation (*A.3*), participants with high levels of experience and who hold managerial or leadership positions appreciate the importance of visualisation more that those who have less experience. Furthermore, with regards to having tools that are platform independent (*A.9*), participants that disagreed were more technically oriented. From their experience, they explained that some tools need to be platform-specific to be able to deliver deeper analysis for the task at hand. They expressed that if the tool becomes platform independent, it will lose some of its specialised abilities.

We also note that a number of assertions prompted a neutral response from participants. For instance, more than 40% of participants indicated a neutral response for *A.2, A.9*, and *A.11*. Having the neutral response option is important so as not to force participants reach an answer, when they might not have a strong opinion towards a particular topic. For example, this might be when an assertion is more technically oriented and the participant has a managerial role.

4.3.3 Cybercrime Incidents

We asked participants about the types of cybercrime incidents they usually deal with. Participants from the private sector reported that their clients are usually interested to know if they get any mentions on the DarkNet. For example, if people are buying and selling related hacked accounts, personally identifiable information and credit card information. Moreover, private sector participants reported that they tend to monitor social media platforms to detect any potential threats to their clients, identify the groups that are involved in these threats, and then relay that information back to the client. For example, they cover incidents related to activism, hacktivism, or public order matters that could impact on the day-to-day operations of the client.

On the other hand, government sector participants reported that they would deal with cyber-dependent and cyber-enabled crimes. Although they are meant to only focus on cyber-dependent crimes, they also assist with cyber-enabled crimes mainly due to the current structure and limited capacity of the police force. On the cyber-enabled side, they look at crimes such as fraud and financial offences and selling/buying of drugs from dark-web markets. On the cyber-dependent side, they cover a number of incidents: (1) Viruses, Ransomware, Malware, and Trojans, (2) Distributed Denial of Service (DDOS) attacks, (3) DDOS Extortion, where attackers demand extortion fees to stop the attack, (4) Hacking offences which include hacking servers (e.g., data breaches), hacking websites (e.g., website defacement), hacking personal accounts (e.g., emails), and hacking phones.

Participants reported that among the most reported cybercrimes are Ransomware. One participant related this to the urgency effect that a Ransomware

introduces.

"I think another thing about Ransomware is it tends to have an urgent impact. All of a sudden you find your computer system is locked and a lot of people don't know why. Whereas with some of the other leaks or hacks where actually it's a lot slower and information is pulled more slowly, then companies are probably able to deal with it internally and much more low key because their operation is still able to carry on. Whereas with Ransomware it's an abrupt halt to all their systems. [G2]"

When investigating a cybercrime incident, participants reported that there are several factors that affect how this process will be carried out and which cybercrime unit will deal with the incident. A cybercrime incident might be dealt with at a local-level (LL) by a local cybercrime force unit, at a regional-level (RL) by a regional cybercrime unit, or at a national-level (NL) by the national crime agency.

"We broke things down between what might be dealt with by local cybercrime force unit, a regional cybercrime unit, and then the National Crime Agency, each of those tiers kind of elements will have some kind of criteria. And we used to have criteria at regional level that says we can only take so many jobs on because it takes a long time to investigate them and a lot of resources. Unfortunately a lot of the sentences, even if you get it that far, are tiny. So in that regards, I know that there would be financial limits on losses set, impact, who the victim was, is there a national infrastructure kind of element to it. There are hundreds of thousands of reports in action fraud every day and there is not the resources to tackle them. So lots of them, it's based on: is there a viable lead here to take it forward." [G1]

Based on this works analysis, the main factors that affect how a cybercrime incident will be handled include: the type of crime committed (cyber-enabled, or cyber-dependant), the victim type (individual, small/medium business, large enterprises), type of potential suspects (hacker or organised crime group), what

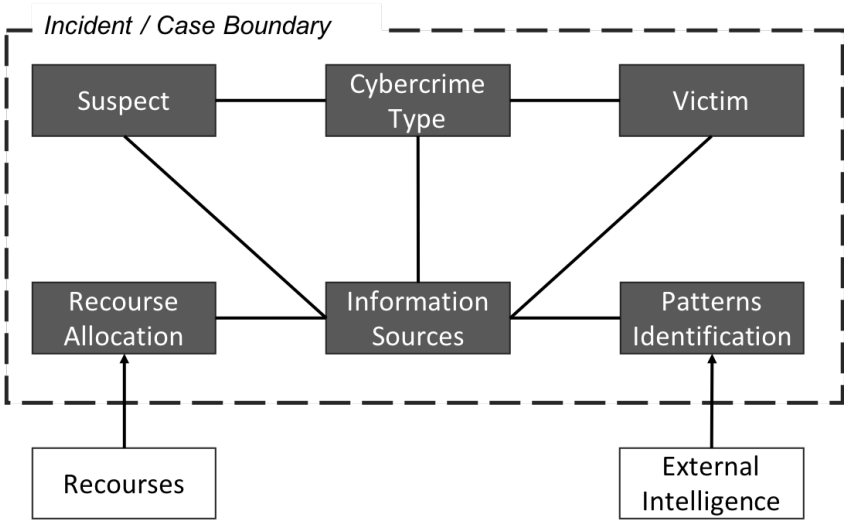


Figure 4.5: Investigation Factors

information or intelligence is available (investigation leads), and existing resources available (human and technological resources). We map these factors and show the relationship between them in Figure 4.5.

4.3.4 Understanding the Investigation Process

When asked about the process of intelligence gathering, the majority of the participants agreed with the six steps of the intelligence gathering cycle described in the literature [86]. However, the initiation of the intelligence gathering process differs between the private sector and government sector participants. This is due to the difference in nature between the two sectors. In the case of the private sector, typically they work with a client and start by collecting and understanding the client's requirements. Thus, they are able to scope down the task of gathering specific intelligence to answer clear requirements. The process tends to take a top-down approach where they formulate a hypothesis and then collect data to prove it.

On the other hand, with law enforcement, the investigation and intelligence gathering are initiated when they receive a report that someone has been a victim of a cybercrime. In such a case, the intelligence gathering process begins with a bottom-up approach. This means they start from the collected data, analyse it, and then try to formulate a hypothesis of what happened based on the analysed

data. Although the majority of participants from the government sector emphasised this "bottom-up" approach, one participant stated that they tend to use mixed approaches. They may start with a hypothesis or initial belief "bias" based on their experience and then look at the data to either prove or dismiss the hypothesis.

"What I've done over time is, you will have your own kind of biases or stereotypes in your head which guide you. Not in like negative stereotypes but ones that help you. You've just learnt from how you've gone before."[G3]

When we asked participants about the differences between investigating traditional crimes and cybercrimes, we received a diverse set of opinions. Some felt that for cyber-enabled crimes, there are no differences and the same kind of detective tactics would work. To them, it is still just a crime that has been committed regardless of the means used. Thus, the investigation process is basically the same.

"The fundamental thing is the majority of crime associated with fraud or cyber is enabled crime. In other words, it's traditional crime that is just leveraged through a digital device. Nothings changed, it's the existing crime that has been committed for centuries if not thousands of years."[G2]

On the other hand, others felt that there are key differences that the cyber aspect brings to the investigation. The main differences mentioned by participants relate to the globalisation of cybercrimes, which restricts how far the investigation may go due to jurisdictional issues. Another key difference mentioned by participants is that cybercrimes, compared to traditional crimes, tend to have less investigation leads 90% of the time. This may be due to the sophistication of cybercriminals in covering their online traces, or due to the lack of cyber awareness of the victims. For example, they may accidentally destroy digital traces and evidence, or report the wrong information to the police.

We asked participants about the use of OSINT during their investigations, and the general consensus across participants from different organisations was that OSINT is an important tool for intelligence gathering.

"Open Source Intelligence is a massive tool for policing, and it's used across the board, not just by us, it's used for all sorts of things, so it is a technique that detectives are familiar with."[G4]

Participants stated that its importance arises from the fact that suspects spend a lot of time online and have a large presence either on social media or on the internet in general. One participant stated that OSINT is not only valuable to the cybercrimes investigations, but even traditional crimes would have benefited from using online sources to help with the investigation.

"With some of the historic crimes we investigate, I think we should have been doing more open source research side.".. "I think open source intelligence has got a much bigger role to play than any more traditional crimes and can be much more successful, we can get a lot better results"[G2]

4.3.5 Collaboration and Investigation Teams

Participants across different organisations described similar team structure and dynamics when investigating cybercrimes and gathering intelligence. Typically, a given case will have a lead investigator and may have the assistance of an analyst and a researcher to support the investigation. Researchers are responsible for conducting regular monitoring of different sources and preparing intelligence briefs for analysts/investigators or clients. Analysts are responsible for developing the intelligence products and identifying trends related to specific expertise and themes in cybercrime. For instance, this might involve taking the volume of data and making it presentable and useful either to the lead investigator (and the court later

on) or to the client. Depending on the size of the department and the available resources, the team may also include technical staff who examine devices, whether this is for the victim or suspect. Once these are reviewed, all the case data goes to the analyst who tries to evaluate it, challenge it, and pick it apart.

We asked participants about the typical technical and cyber experiences of people working in each role. Most mentioned that investigators and analysts do not necessarily come with previous technical knowledge or backgrounds, however, they would learn on the job the technical skills needed. This phenomenon was true across all types of organisations, be it private sector or government and police force.

"Some before they become cybercrime investigators, they were all probably 6 to 12-year detectives, so they've been dealing with burglaries and things like that. But there were a couple that had programming backgrounds, and one had worked in information security."... " The basis of it was getting good detectives and good research and analyst skills and then try to give them enough technical knowledge to do what they need to do."[G1]

Moreover, we wanted to understand how the collaboration across departments and/or organisations happens when investigating a specific case. In the context of the private sector, this was easily done online, through emails, or face-to-face meetings with people from different departments. Depending on the size of the organisation, teams would either all be based in the same office or city, or they may be working with teams based overseas. Either way, they reported that information sharing within and across teams was quite seamless and did not present a challenge to their investigation process. Overall therefore, these represented usable and efficient workflows.

On the other hand, participants from the government sector had a different and more challenging user experience. The flow of information and intelligence sharing from different cybercrime units and agencies was a challenge. Participants from different cybercrime force units mentioned that they often have difficulties connecting information from a case they are investigating with possible linked cases

in other forces.

"So very little intelligence sharing happens within cybercrime. ... Getting stuff up from forces to regional units and vice versa is a lot more difficult. ... Say you had an offence down south [placeName] and you had an offence up north [placeName] and they were connected, whether it be like a suspect email address, the chances of connecting those two offences, depending on what level it came in, could never happen basically."[G3]

This is particularly true for information sharing between units at different levels (local force and regional force) as well as between force units at the same local level. However, the information sharing is currently better between force units at the regional level. Moreover, one participant mentioned that sometimes such linkage of cases may happen by chance.

"Because you have such a workload that you'd like to be most of the time now identifying that something connected to something else is purely based of an individual's memory as opposed to a system alerting. So those are the kind of next steps we need."[G3]

Another participant expressed the need for a connected system across forces that is able to calculate such associations between relevant cases.

"If we had the ability to share our data from our jobs and their jobs and we had the analytical ability to be able to pick out and go hang on a minute, their, our job links to that one in [placeName]. Sometimes that linkage happens by virtue of us having conversations but not you know automated."[G4]

4.3.6 Technological Capabilities and Tool Support

Within the general theme of technological tools and capabilities, participants reported that multiple tools are usually used during investigations. Each of these tools is used to achieve a specific task, such as data collection, analysis, or visualisation. When asked about the interoperability and interactions between these tools, the majority stated that these tools do not work well together. Instead, it requires significant manual effort to extract the output of one tool and map it into another tool for further analysis.

"They're not necessarily as good at each part of the investigation. Certainly, most of them have a focus at one point of the investigation cycle. And then the rest might be add-ons which are getting better but do not necessarily, see the job all the way through as good as you'd like."[G1]

Moreover, participants stated that sometimes they work on different systems and networks, and this is due to corporate regulations meaning that some tools are only allowed to reside within public systems. Thus, different tools will be forced to sit on different networks.

"They very much work in isolation, in most cases you would use a tool, and say it was going to be the main tool that I was using to build court products, it is great for visualisation but there is very little in terms of labelling, exporting, producing reports. So I might do all my working outside that tool, narrow it down to the picture I want to show and then manually map that in the tool."[G3]

Participants from the government sector also mentioned that there is currently no national mandate for the selection of tools to be used within forces. Therefore, each police force operates a different system and process, which adds to the challenge of sharing knowledge and intelligence between forces. This suggests a much larger issue with regards to creating an efficient, effective and ultimately widely usable

cybercrime investigation process.

"Tools are not mandated nationally to each force, and if it was a national infrastructure for a lot of these things then that would be really helpful. It's generally on a force by force basis."[G4]

We asked participants from the government sector about how the current police systems are being used to report cybercrimes. Participants expressed that the current police reporting system for cybercrimes (called Action Fraud (AF) [154]) suffers from a number of issues. AF is the UK's national fraud and cybercrime reporting centre, which was initially designed for reporting fraud crimes specifically. The system was not designed for recording cybercrimes and thus generates ambiguity for the AF system users (police officers or self-reporting victims) when reporting these crimes. This sometimes leads to offences being recorded under the wrong category.

"It was not really built to deal with cybercrime, and so they had real issues in kind of categorising offences, and when they get reported you have the victim knowledge being able to accurately report it, and then we have the call taker knowledge."[G4]

Moreover, many details specific to cybercrime data are not properly recorded and stored. This is due to the lack of structured methods to collect the necessary cyber-related information, such as IP addresses, bitcoin wallets and usernames, which makes reports collected from different victims differ in the level of detail. This has an effect on how these reports are being searched and eventually leads to missing links between crimes that may share similar details, such as email or IP address.

"The police reporting systems themselves are not even necessarily catered to record cybercrime offences. They have to be kind of pushed in under a different

category sometimes and systems might be set up to record things like home addresses, vehicle indexes and everything like that, but very little systems are set up to record email addresses, IPs and everything like that."[G3]

Additionally, participants mentioned that there may be a shortage in the number of call-takers within Action Fraud, which has an effect on the number of cybercrimes that are assigned to different forces. Participants also mentioned complications related to the limited level of cyber knowledge and experience in dealing with cybercrimes that initial call-takers have. This leads them to misdirect crimes to the wrong department which causes delays in responding to the crime. Also, one of the participants perceived a link between how prevalent cybercrimes are in the news to the number of crimes that are communicated to them from the call-takers.

"One of the issues that we've identified within our force is when a victim calls the Police, they get a really patchy response as to how it's dealt with because our call takers have a lot of experience of dealing with reports of rapes and assaults but when it comes to cybercrimes they do not really know what to do so it often gets misdirected or delayed."[G2]

We talked with participants about the implications of using technology to automate certain investigation processes. Many of them had some concerns regarding the automation effects on the process, and expressed a preference for performing investigation tasks manually. This was either due to a lack of technical skills to re-validate previously obtained results, or due to the fear of missing some evidence that is relevant to the context they are investigating. Nevertheless, participants from both private and government sectors emphasised that automating certain processes may be necessary, but it should be done under human control and guidance.

" Automated tools can be useful for finding new data. I still think, and may be it's a cultural thing, but it still requires very human element to it. I know

the context of what we're looking for and I can interpret the messages and the chats that I'm reading and pick up different things and bring it all together. I might see something in one investigation, actually I realise is relevant for another one."[G1]

Including the human (user) in the process is particularly important in order for them to understand the results produced by the tool, and to be able to trace back the process from the input to the produced output. This is important because investigators need to explain their findings to clients and to be able to defend it in court.

"You do not want to use a tool and just feel like you are just clicking a button" ... "If you are stood in court you want to be able to explain what is happening behind that tool."[G3]

We asked participants about the technical capabilities and desirable tool features they think will benefit the way they conduct their investigations and gather intelligence. Government sector participants all agreed that they need a substantial upgrade in their IT infrastructure to be able to cope with the sophisticated cybercrimes they investigate. This was not an issue for the private sector. Additionally, both groups described the need for more user-centred tools to allow them to easily do their jobs, tools with better support for data visualisations, and tools that can aid in analysing bulk amounts of online data to cope with the evolving online ecosystem.

"We could be so much better with better technology, share data, analyse data, be proactive with data. So an improvement in technology would certainly assist us for the future. I think we do things perhaps the long way and over time we could improve on that."[G4]

In summary, this section highlighted the main findings gathered from the experts' interviews. These are mainly spread across topics related to the process of

cybercrime policing, the collaboration between investigation teams, and technological capabilities of the investigators.

4.4 Key Challenges Faced by Cybercrime Investigators

The literature from the criminology field discusses a number of challenges facing the police when investigating crimes. For example, some of the discussed challenges are related to police patrol and organising hot-spot policing in order to minimise the number of street crimes [155]. Moreover, Ratcliffe [156] argues that there is a need for a new shift in policing from relying on old knowledge, which is relating to the criminal activity that is typically collected in traditional crimes, to a new knowledge, which is focused on the crime event that is collected from public open source information.

While some of the challenges that are inherent within the traditional crimes field transcend to cybercrimes, a number of new challenges emerge that are linked to the cyber-world. For example, cybercrimes are borderless by nature, thus there are several legal and jurisdiction issues related to investigating cybercrimes. In traditional crimes, the offender has a physical link to the crime scene which makes the investigation and detention easier. However, in cybercrimes, this is not the case. Offenders may be in a different place, state, or country from the victim. This makes cooperation across different agencies a necessity.

Through the analysis conducted in this study, we are able to identify some key socio-technical challenges that cybercrime investigators face. These challenges emerged from the themes discussed previously and are grouped into four main topics: (1) Reporting of Cybercrimes, (2) Information Sharing, (3) Tools and IT infrastructure, and (4) Skills and Technical Abilities. We discuss these further below.

4.4.1 Reporting of Cybercrimes

One of the main issues is related to the AF system and how it was initially designed and used. The system users face many ambiguities when recording the reported crimes, which results in miss-categorising cybercrime offences. This leads to

inconsistencies in the level of detail collected from victims, which typically depends on how technologically-aware the victim is and the level of cyber-experience held by the call-taker preparing the report. There is also an increased likelihood of misdirecting the report to the wrong department thus causing delays in addressing the crime. It also increases the time and effort needed to conduct the investigation. Additionally, the lack of structured methods to collect the necessary cyber-related information, has multiple implications on the investigation process, such as missing some connections between possibly related victim reports. Overall, a major review of the questions included in AF was deemed necessary and appropriate training is needed to better equip call-takers to handle cybercrime related reports.

4.4.2 Information Sharing

One of the most significant challenges faced by law enforcement is the lack of centralised coordination of intelligence sharing between forces and agencies working on cybercrimes. This has multiple effects. Firstly, the intelligence products produced by forces may be incorrect since they do not have the big picture and are not fully aware of the crimes which are being reported in different regions. Second, the lack of communication between different local forces results in missing possible connections between different reported cybercrimes. This makes establishing links between cases a real challenge. Based on our results from interviewing participants from different cybercrime force units, there exist a variety in the level of information flow across local, regional, and national levels. Third, following up on specific cybercrime cases and providing status updates either to the victim or the media when requested, is difficult due to the lack of sharing case information. This has a negative effect on the police reputation and image as it may make them look incompetent in the eyes of the public. Moreover, given that cybercrimes have a border-less nature, unlike traditional crimes, they may involve suspects in distributed locations. Therefore, having better communication, coordination, and automated data sharing between different units and forces will have a positive effect on mitigating and responding to these crimes.

4.4.3 Tools and IT Infrastructure

A key challenge associated with investigating cybercrimes is related to the level of IT infrastructure available to police forces and the cyber capabilities of investigators. In recent years, there have been several budget cuts that prevent forces from investing in upgrading the IT infrastructure and acquiring advanced tools. This limits their capabilities and has effects on the efficiency and quality of the conducted investigations. Similarly, investigators expressed that a lot of the investigation time is spent doing manual tasks that may be saved by utilising some automated or semi-automated tools. Examples of such tasks include looking at OSINT data to collect intelligence and analysing terabytes of data from server logs and victim devices.

Furthermore, the lack of a national mandate of tools and systems to be used in cybercrime investigations adds another challenge for collaboration between different cybercrime units. Police forces are generally decentralized in the UK. While having a structure of separate 43 police forces [157] may be fit for policing traditional crimes, this may not be the best structure for delivering effective actions against crimes that are cross-jurisdictions (i.e., cybercrimes). This structure has lead to each cybercrime unit becoming a silo, operating a different set of processes and tools and being unable to interoperate. Similarly, this interoperation issue also exists for the different tools used, where much of the investigation time is spent manually modifying and formatting data to be able to move it between different analytical tools.

4.4.4 Skills and Technical Abilities

Another challenge is related to the cyber skills of the staff. There is a lack in the number of skilled technical personnel working on investigating cybercrimes. The majority are highly experienced investigators who have been working with traditional crimes and are now, because of the increasing numbers of cybercrimes, moving toward the cyber field. Investigators are highly trained in gathering intelligence and investigation tactics but might not be as well-trained in the cyber-world. There are, therefore, open questions pertaining to employee training and expertise. This is also unlikely to be a problem only faced in this geographic area due to the widespread

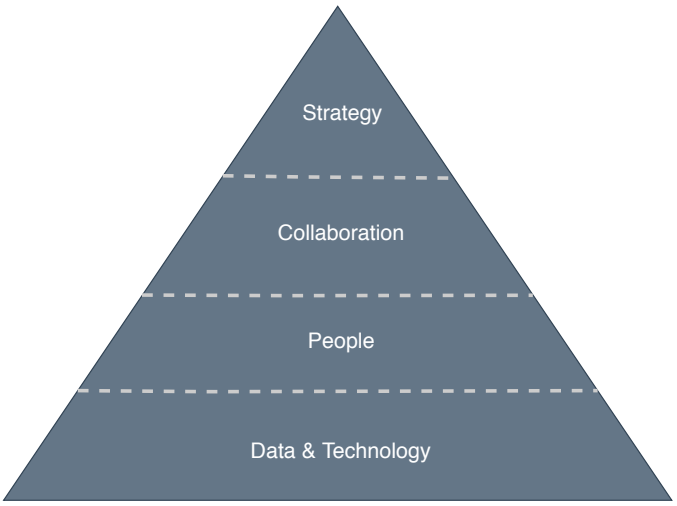


Figure 4.6: Response areas to cybercrime challenges

increase in cybercrime. Potential avenues that may be explored going forward include further training and upskilling for the necessary personnel, but also creating tools that are better geared towards supporting user skills and activities. The ideal case will be the provision of practices and tools that are easy to use and would reduce the learning time for new cybercrime investigators.

4.5 Recommendations

As discussed in the previous section, there are a number of challenges currently faced by cybercrime investigators working in government and private sectors. Some of these challenges are inherent in the area of policing traditional crimes and have transcended to cybercrimes. Others are new challenges that have emerged because of the way the cyber space is designed. To overcome the challenges of combating cybercrimes, we provide a set of recommendations covering four key-areas depicted in Figure 4.6.

4.5.1 Strategy

A clear holistic national strategy to combat cybercrimes is critical to any nation. The latest UK Cyber Security Strategy [158] focuses on delivering three objectives: (1) *Defend* the UK against cyber threats and respond effectively to occurring incidents, (2) *Deter* by making the UK a difficult target for any cyber attack, and (3) *Develop* existing talent in the cyber security industry and invest in innovation and research to overcome future threats. Due to the nature of cybercrimes and the diversity of its threats, actors, and motives, a comprehensive strategy is required to tackle every aspect. The UK Serious and Organised Crime Strategy [159] covers four main objectives: Pursue, Prevent, Protect, and Prepare. Pursue and Prevent are critical to reduce the threat, while Protect and Prepare are focused on reducing vulnerabilities.

While these developed strategies have achieved great advancement with regards to deterring and combatting cybercrimes, we believe that further advancements may also help in winning this battle. Based on the interviews conducted and our empirical study we recommend the following:

- Evaluate the current structure of law enforcement cybercrime units and its effectiveness in addressing cybercrimes. This includes improving the capacity and capabilities within these units.

- Re-engineering the investigation process and workflow to be more suitable to the nature of cybercrimes.

- Review and update the usability, workflow, and current questions included in the national cybercrime reporting systems [e.g., UK Action Fraud]. This should include details to capture new threats, and to standardise the level of details collected from victims in order to have a better picture of the incident and thus direct the incident to the appropriate cybercrime unit to handle.

4.5.2 Collaboration

Collaboration and coordination of intelligence and partnership is required at a national and international level to be able to share intelligence, make effective use of resources, and better combat cybercrimes. International initiatives and treaties are important to tackle cross-jurisdiction cybercrimes. Strengthening those agreements and continuing to develop them with common procedures and policies will lead to effective prosecution and deterrence of cybercrimes.

At a national level, information sharing between different cybercrime agencies is required. Having a central system connected across cybercrime forces will allow a seamless flow of information between local, regional, and national agencies dealing with different threat levels of cybercrimes. To achieve this, further research is needed to facilitate the exchange of intelligence in a secure and usable manner across different systems and cybercrime units. Additionally, it is critical to study and evaluate the best architecture for information sharing across different levels of cybercrime units; common options include a hierarchical architecture (i.e., local unit connected to regional, and regional to national) or a flat architecture (i.e., having each cybercrime force unit connected to the others). Additionally, the design of appropriate access control models is needed to facilitate the availability of intelligence to those who require access to it while maintaining the confidentiality and integrity of that information.

Moreover, collaboration and intelligence sharing is critical not only between different government agencies but also across public and private sectors. Collaboration within the private sector will facilitate the exchange of expertise and knowledge. Establishing frameworks to encourage the private sector to share intelligence related to new threats they face will allow the threat effect on others to be reduced, thus protecting them from becoming a potential victim of similar threats. This can be achieved through a public-private consortium where businesses can share threat information in a secure way, and the government can provide actionable intelligence. Furthermore, such collaboration will allow the government and large enterprises to provide a set of best practices and guidelines on how to

tackle threats, especially to small/medium enterprises who do not have the capacity and resources to protect themselves.

4.5.3 People

Many of the investigators working in cybercrime from private and government sectors come with experience in investigation tactics associated with traditional crimes, and may lack the complex technical skills. One way to address this is to provide more training, capacity building, and awareness in cyber for law enforcement forces (especially at the local-force level). Attracting, recruitment, and retention of highly skilled individual is a must to be able to cope with the constantly evolving cybercrimes. Cybercriminals adapt new technologies and methods to evade detection which makes investigating these crimes technically complex. Therefore, attracting highly skilled people with expertise in cyber security and forensics will lead to more effective combatting of cybercrimes.

Moreover, to proactively address cybercrimes, it is important to focus on educating the general public and provide awareness campaigns on online safety best practices, existing threats and tactics used by cybercriminals to aid in lowering the chances of being victims to cybercrimes.

4.5.4 Data and Technology

The complex nature of cybercrimes and the constant adoption of new methods and tools by cybercriminals to carry out their attacks generate a large digital footprint. Data sources such as the dark web, social media, underground markets, and digital logs are all examples of data sources that would be collected and analysed to gather evidence in cybercrime cases. As described by participants, going through terabytes of data is not an easy task, especially when you do not necessarily know what you are looking for. This makes evidence gathering a difficult and time-consuming task. Making use of automated methods and new technology such as artificial intelligence would help filter out the noise, accelerate the investigation, and uncover patterns and trends from the data.

Current tools used within the investigation process as suggested by participants sometimes lack an important requirement that is related to the transparency of the analytical process. Investigation and intelligence gathering tools should be designed with *human-in-the-loop* concept in mind. As an investigator, he/she needs to be able to understand how the tool is working, and how particular results were reached. This is critical to them as they need to be able to explain their findings and defend them in court as well as in other legal settings. To achieve this, the HCI and usable security research community need to investigate how best to improve the user interaction and involvement with intelligence tools. Additionally, the systems and tools used in each cybercrime force unit should be mandated nationally. This will facilitate a common infrastructure of sharing of analytical capabilities and linked intelligence. Furthermore, this will increase the interoperability between different tools and systems, and reduce the amount of incompatible tools currently being used across different organisations. Finally, an investment in upgrading the current IT infrastructure in different cybercrime force units is necessary, to allow investigators to have the capacity to analyse the vast amounts of data being generated from these crimes.

Many of the investigators working in cybercrime from private and government sectors come with experience in investigation tactics associated with traditional crimes, and may lack the technical skills. Increasing the usability of the cybercrime investigation and intelligence tools is crucial as it can reduce undue burden on the technical abilities of workers. The ideal situation is for tools and workflows to be designed to take advantage of investigators' expertise and support them in cases where this may require bolstering.

4.6 Limitations and Implications

There were some limitations in our study related to the sampling of participants, which we discuss in this section. The size of the interviewed sample that participated in this study is small and does not allow us to draw generalisable conclusions. To identify the main challenges faced by investigators and the process they use, we opted

for depth rather than breadth. This was achieved through the detailed in-person discussion with participants to understand their experience while investigating cybercrimes. Moreover, sample representativeness is another limitation that we acknowledge. As we discussed earlier, we chose a diverse sample from both government and private sectors. However, since the government sector consists of cybercrime units within local police forces, regional, and national levels it was not feasible to interview participants from all units. Similarly, with all 43 police forces in England and Wales have now established a dedicated cybercrime unit, in addition to the 10 regional organised crime units it was not attainable to recruit participants from all units.

This small sample size may affect the obtained results concerning the described processes and challenges. Mainly, since each unit operates a different set of processes, this may influence the identified challenges and needs. Thus, we do not claim to have identified all the challenges faced by investigators. Furthermore, this study was focused on the processes followed within local and regional force units. The national-level perspective is yet to be explored. Nevertheless, our goal was to reach data saturation for discontinuing data collection, i.e., when no new information or themes are being observed in the data [160]. This was achieved for each sample group in relation to the identified themes.

Through this study, we have answered the first two research questions we outlined in Chapter 1. We have created an understanding of the existing capabilities, processes, and workflows followed by practitioners in the field. Similarly, we identified a set of challenges they face and provide a set of recommendations to overcome these challenges. Our work has implications for future research in the field as it defines a set of recommendations and future research directions that can allow for better support to investigators and more effective combating of cybercrimes.

One of the main findings is the need for a common cybercrime framework that unifies the investigation process across different units. This will facilitate the collaboration and exchange of information between different forces. Moreover, better utilisation of new technology, such as Artificial Intelligence (AI) that can digest

the big data associated with cybercrimes is needed. As a result, in the following chapter, we address these issues by proposing the *CCINT framework*.

4.7 Summary

In this chapter, we focused on providing an understanding of the needs and challenges faced by practitioners working on cybercrime investigations. We interviewed participants from both private and government sectors regarding their experience with investigating different cybercrime incidents and the process of gathering intelligence. The interviews were focused on four main themes: (1) investigation process, (2) cybercrime incidents, (3) collaboration and information sharing, and (4) technology. Moreover, we discussed four key areas where improvements are needed to better combat cybercrimes. A clear holistic strategy is critical to improving the current advancement with regards to deterring and combatting cybercrimes. National and international collaboration with government, private and public sectors is needed to facilitate the sharing of intelligence and information across jurisdictions. Investment in education and awareness campaigns are necessary to inform the general public. Finally, the development and utilisation of new tools and technology are needed to overcome the burden associated with analysing big-data generated from different cybercrime incidents.

Having understood the existing landscape of cybercrime investigations and identified key needs and challenges faced by professionals. In the next chapter, we will propose a cybercrime intelligence framework that focuses on developing models, methods, and tools to aid investigators in their battles against cybercrimes.

5

CCINT: The CyberCrime INTelligence Framework

5.1 Introduction

Cybercrimes complexity is increasing with new tools and Modus Operandi (MO) being adopted by offenders to evade detection and achieve their goals. Criminals have easy access to advanced technical abilities that they need to carry out their attacks using what is called crime-as-a-service through the dark web and online black markets. Similarly, the nature of cybercrimes have generated multitudes of data introduced by the "cyber" aspect of these crimes, which makes the process of identifying evidence similar to the process of searching for a needle in a haystack. In order to aid law enforcement better detect, analyse, and understand the threat landscape posed by cyber criminals, research into the area of cybercrime intelligence has flourished.

Cyber criminals use the internet as their crime scene since it provides them with ease of communication, wider recruitment possibilities, and opportunities to form partnerships with other national and international criminal groups [13, 14]. This results in them leaving several "crumbs" that collectively produce a digital footprint for each cyber-criminal. Previous research has studied these

footprints in order to gain better understanding of the characteristics of these cyber-criminal groups [15–18].

Moreover, as presented in Section 2.1 of Chapter 2, most of the taxonomies and classifications of cybercrimes in the literature focus on describing cybercrimes from a single view point, such as the target of the crime (e.g., [40]), the method of the crime as in cyber-enabled or cyber-dependent (e.g., [45]), or the security principle the crime exploits (e.g., [43]). To the best of our knowledge, none of the existing models consider in their classification the wide range of factors that constitute a cybercrime, including the victims' and offenders' backgrounds, motives, incurred losses, and skills.

In order for cybercrime investigators to be able to effectively defend against cybercrimes, they need to understand the motives, methods, targets, and MO used by cyber criminals. In this chapter, we propose the *CCINT Framework*, a CyberCrime INTelligence framework that provides conceptual model and operational framework for investigating cybercrimes. The conceptual model is inspired by how practitioners describe the process they use and the key factors they consider when investigating cybercrimes (insights from Chapter 4), while the operational framework covers the technical-side with the focus on methods and tools to handle the data analysis and investigation.

We will then demonstrate examples of how the model is applied to different examples of cybercrimes. Having a common framework to follow while investigating will help investigators extract effective intelligence and better analyse such crimes. Finally, we will present the design and architecture of the CCINT framework that will aid analysts in the operational aspects of their investigations.

The main contributions of this chapter are outlined as follows:

1. Provide a conceptual model to be used during investigation of cybercrimes inspired by practitioners' experiences. Having a common framework to follow while investigating will help investigators extract effective intelligence and better analyse such crimes.

2. Present a detailed example of how the conceptual model can be applied to a real-world cybercrime case.

3. Outline the design and architecture of the operational part of the framework that will aid practitioners in the operational aspects of their investigation.

5.2 The Framework

For cybercrime investigators to be able to analyse and investigate cybercrimes, they need to better understand the techniques, processes, and procedures used by cyber criminals. To do so, a robust framework that provides a holistic overview is needed.

In the interviews we conducted with cybercrime investigators (Chapter 4), they described the different factors they need to consider when investigating a crime. These factors consist of three main elements: (1) the offender, (2) the victim, and (3) the incident. Based on a set of properties that describe each of these three elements, they will be able to determine how this particular incident will be investigated. In particular, which cybercrime unit will handle it (national, regional, local), what type of resources will be allocated (e.g., time and man power), and what type of intelligence sources will be considered. In the following sections, we provide a complete description of each of these elements through the CCINT conceptual model. We then show how the model can be used in a real-world cybercrime case.

Next, we present the design and architecture of the operational side of the framework, where we describe the main methods and functionalities that constitute the framework.

5.3 CCINT Conceptual Model

Any given crime consist of three main elements. *The offender*, who commits the crime, *the victim*, who suffers from that crime, and *the incident*, the event that constitute the crime. Previous studies, as well as current work, attempted to understand the different elements of cybercrime from multiple perspectives [161–164]. Some focused on understanding the victims-side, while others focused on the

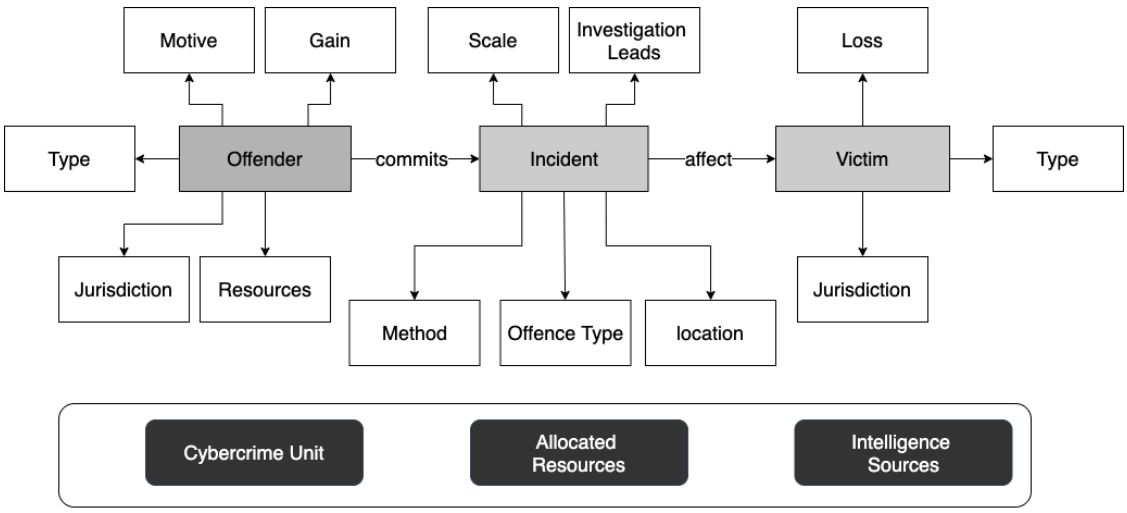


Figure 5.1: CCINT Model

offenders. In addition, some included defenders in their model. For instance, Arief et al. [161] define the cybercrime stakeholders to include the defenders in addition to attackers and victims. In the following, we present our model, which include three main elements, the relationships between them, and the properties describing each of the elements. Figure 5.1 illustrates an overview of CCINT conceptual model.

5.3.1 Offenders

Offenders of cybercrimes can be described by five main properties. (1) *Motive*, (2) *gain*, (3) *type*, (4) *resources*, and (5) *jurisdiction*.

Motive — In cybercrime, similar to traditional crime, there is always a motive behind committing such crime. Typically, one of the most common motives is economic gain. However, there are multiple cases where the motive behind the crime is completely different. Understanding the main driving motive behind committing the crime will allow investigators better investigate these crimes. Prior research investigated the different motives [165] and psychology of online offenders [162] in general, and cyber-terrorist in particular [163]. Offenders might have composite motives behind committing crimes. We break down these motives as follows:

1. **Economic:** It is considered one of the most common motives for cybercrimes. For example, ransomware, one of the most common attacks in recent years,

is primarily motivated by financial factors. A typical way to financial gain would be through the victim himself (e.g., paid ransom) or through the sale of victim data in black markets.

2. **Political:** It includes disrupting another state by attacks targeting a critical national infrastructure (e..g, power grid). It may also involve attacking government websites and taking them down, or defacing them to spread a political message. Additionally, manipulating elections in favour of particular party is another type of attack that we have seen recently that is politically motivated.

3. **Ideology:** This include attacks by violent extremists trying to promote their ideology and spread propaganda. Additionally, hacktivism activities are ideology motivated where the attackers try to deliver their message (social, cultural) by breaking into systems.

4. **Esteem:** This include attacks that are motivated by gaining status, power, or purely for fun. Typically, these are conducted by script-kiddies who are looking to prove their abilities or groups trying to gain reputation in the cyber world.

Gain — While the defining motive behind the crime significantly influences the gain the offender obtains, we separate them to capture the details of the actual advantage gained by committing the crime. This may include money, data, competitive advantage, and fame.

Type — Offender type captures the category to which the cyber criminal belongs. This is divided into (1) individual actor, (2) group, (3) organised group, (4) state-sponsored actors. Knowing the offender type will give the investigator a sense related to what kind of resources they may have.

Resources — This includes the type of resources the offender has access to in order to commit the crime. These include technical knowledge, access to malicious software, computation power, and financial means.

Jurisdiction — This is the jurisdiction that the offender is part of. This will help in coordinating with other cybercrime units, either locally, regionally, nationally, or globally.

5.3.2 Victims

Three main factors need to be defined to describe victims of cybercrimes. These are: (1) type of the victim, (2) the level of loss/harm incurred, and (3) the jurisdiction where the victim resides. Other studies in the literature [164] also included in their definition of victims, the awareness-level and attractiveness level to attackers. We exclude these from our definition, as they do not necessarily contribute to assisting the investigation.

Victim Type — The victim may be an *individual* who has fallen a victim to a cybercrime. An *enterprise* is further broken down into two types: a *small/medium enterprise*, as they would typically need more assistance in dealing with the aftermath of the crime due to their limited resources. A *large enterprise* is usually more capable of handling and overcoming the effects of the incident through proper processes and procedures. Moreover, victims may be government entities, or the *public* which is when the cyber attack has affected the general population. Investigators also need to identify if the victim is considered a "vulnerable victim" such as a minor individual or a charity organisation as these will need further assistance and guidance when handling the incident.

Loss / Harm — The investigator needs to have an understanding of the level of harm/loss incurred by the victim because of the incident, as this will have and effect on the level of resources to put in place. The topic of online harm and the cost of cybercrime has been studied in the literature [166–168]. The studies suggest the concept of cyber harm must consider the wide variety of stakeholders it may affect. A detailed taxonomy of cyber-harm has been presented in [168]. They explain different characteristics of harm and how they can influence its magnitude. We break down the types of cybercrime harm to six types:

1. **Economic:** Amount of financial loss incurred because of the cybercrime. Either directly (e.g., payment of a ransom) or indirectly (loss of availability of a service because of the attack).

2. **Digital and Physical Assets:** Loss of digital assets such as data and IP, or abuse of resources such as computation power. The level of loss relates to the nature of damage caused. For example, if the victim can restore from a recent back up then this would be a small loss. If the victim lost an IP related data then it would be a high loss for them.

3. **Political:** This type of loss includes loss of geo-political power, damaged relations with other nations, disturbance to the electoral process, and loss of political influence on the public.

4. **Psychological:** This type of loss causes emotional and cognitive disturbance such as feelings of depression, anxiety, negative emotions, and self-harm. This can be clearly manifested in cybercrimes that target the human such as cyber-bullying, cyber-stalking, and online radicalisation.

5. **Reputation:** The attack could cause severe reputation damage to the victim. Reputation damage may affect different types of victims, such as individuals, enterprises, and states.

Jurisdiction — The jurisdiction that the victim belongs to. This will help in coordinating with other cybercrime units, either locally or globally.

5.3.3 Incident

The incident represents the main event under investigation caused by the offender and which affects the victim. It is described according to five properties as follows.

Offence Type — As discussed in Section 2.1, there are different classifications of cybercrimes in the literature. Some are based on the target of the crime (crime against the machine, crime using the machine, crime in the machine), while others are based on the fundamental principles of cyber security (i.e., attacks on confidentiality,

integrity, or availability). The most commonly used classification of offences is the breakdown to cyber-enabled and cyber-dependent crimes. As described by the interviewed professionals (Chapter 4), the common classification they use is cyber-enabled (fraud and financial offences) and cyber-dependent crimes (intrusions (e.g., hacking) and disruption (e.g., DDOS) of computers and networks).

Method — This includes a description of the MO used in the offence and how the incident occurred. For example, was it through a certain phishing campaign that lead to a malware infection? What types of malware have been used? What vulnerabilities were exploited and what tools were used? Identifying the method will aid investigators better understand the incident, identify MO patterns, and profile cyber criminals.

Scale — The scale of the incident and level of impact it caused. This is in terms of the frequency and urgency of the situation, if it is an on-going incident, and the number of victims it is affecting.

Location — The domain where the incident occurred and its impact is visible. This can occur purely online, offline, or it may have a combined online/offline impact.

Investigation Leads — The types of available investigation leads and verifiable information related to the incident. Depending on the verifiable information and leads, the crime may be allocated for investigation or transferred or cancelled due to limited investigation leads.

5.3.4 Dependent Variables

Three dependent variables will be influenced by the described elements of the cybercrime model. These are the *cybercrime unit* that will handle the investigation, the *resources allocated* to conduct the investigation, and the type of *intelligence sources* that will be utilised to investigate the incident.

Cybercrime Unit — Based on the properties of the above described elements of cybercrimes, it will determine the type of cybercrime unit (s) that will handle the investigation of the incident. This will be a national agency (e.g., NCA), regional cybercrime unit (ROCU), or a local cybercrime unit within a police force. For

Table 5.1: Examples of mapping different cybercrimes according to offender's intent and victim's loss

Cybercrime	Offender:Intent	Victim:loss
Malicious Software (Use & Manufacturing)	Economic, Political, Esteem	Economic, Reputation, Digital,& Physical assets
Hacktivism	Political, Ideology	Reputation, Political, Economic, Digital & Physical assets
Cyber-terrorism	Political, Ideology	Economic, Political, Psychological, Reputation, Digital & Physical assets
Cyber-stalking, bullying, & trolling	Esteem, Ideology, Political	Psychological, Reputation
Phishing scams	Economic	Economic, Digital assets
Online Romances	Economic	Economic, Psychological, Reputation

example, if the incident has a large scale and high impact on the victim such that it causes disruption to essential services, or leads to severe economic or social loss, or has a political intent, then it will have to be handled and coordinated by a national crime agency. On the other hand, if the incident affects an enterprise with a serious impact on the victim then it will be handled by the regional cybercrime unit with assistance from the national crime agency. Finally, if the incident affects an individual or a small/medium enterprise with limited economic loss then it would be handled by a local cybercrime unit.

Intelligence Sources — The type of intelligence sources to consider while investigating the incident is largely dependent on the incident properties, such as offence type, method, and investigation leads. In addition, the offender/victim types will influence the type of sources to consider during the investigation. Intelligence sources can be open source intelligence (OSINT) from online public sources such as social media, forums, white papers, or closed intelligence sources from local intelligence reports or global intelligence from international agencies.

Allocated Resources — The amount of resources that will be allocated to investigate an incident is affected by the loss, incident scale, and victim/offender types. The resources include man power, financial resources, and investigation time.

5.4 Applications of CCINT Model

In this section, we demonstrate the framework's use by applying it to different examples of cybercrimes. In Table 5.1, we show how different types of cybercrimes (cyber-enabled, cyber-dependent) map to the offenders' intents and victims' losses. For instance, the intent of a creator of a malware may be an economic gain by selling the malware in the dark web. Similarly, it may be for gaining reputation and esteem in their community, or it can be politically motivated where the malware is written to target a foreign state. On the other hand, the victims of such crime may suffer different losses. For instance, this might be economic loss if the malicious software caused disruption to the victim's business services. Reputation loss may occur if being targeted by a cyber attack made the news (e.g., Yahoo attack). Digital loss may occur if the malicious software wiped the data. A physical loss may be incurred, for example, if the malware targets a critical national infrastructure such as a power plant.

To demonstrate the details of how the framework can be used in investigating a given cybercrime, we provide a detailed example by applying it to a real-world case study.

5.4.1 Case: Buying & Selling of Hacking Tools

Case Description — A website is used for selling a hacking tool, which allows remote access to victim machines. The tool is called *Imminent Monitor RAT (IM RAT)*, and was purchased by around 14,500 people from 124 countries for as little as US$25. The tool gave full remote control of tens of thousands of victims' computers. Once installed, the IM RAT allowed the hacker full access to the infected device, enabling the attacker to disable anti-virus software, steal data or passwords, record key strokes, and watch victims via their webcams [169].

Case Analysis — Figure 5.2 shows how the case is mapped against our model. The model allows the investigator to visualise the details of the case and map the different properties associated with each of the three main elements: incident, offenders, and victims. The incident element shows the different offences that have

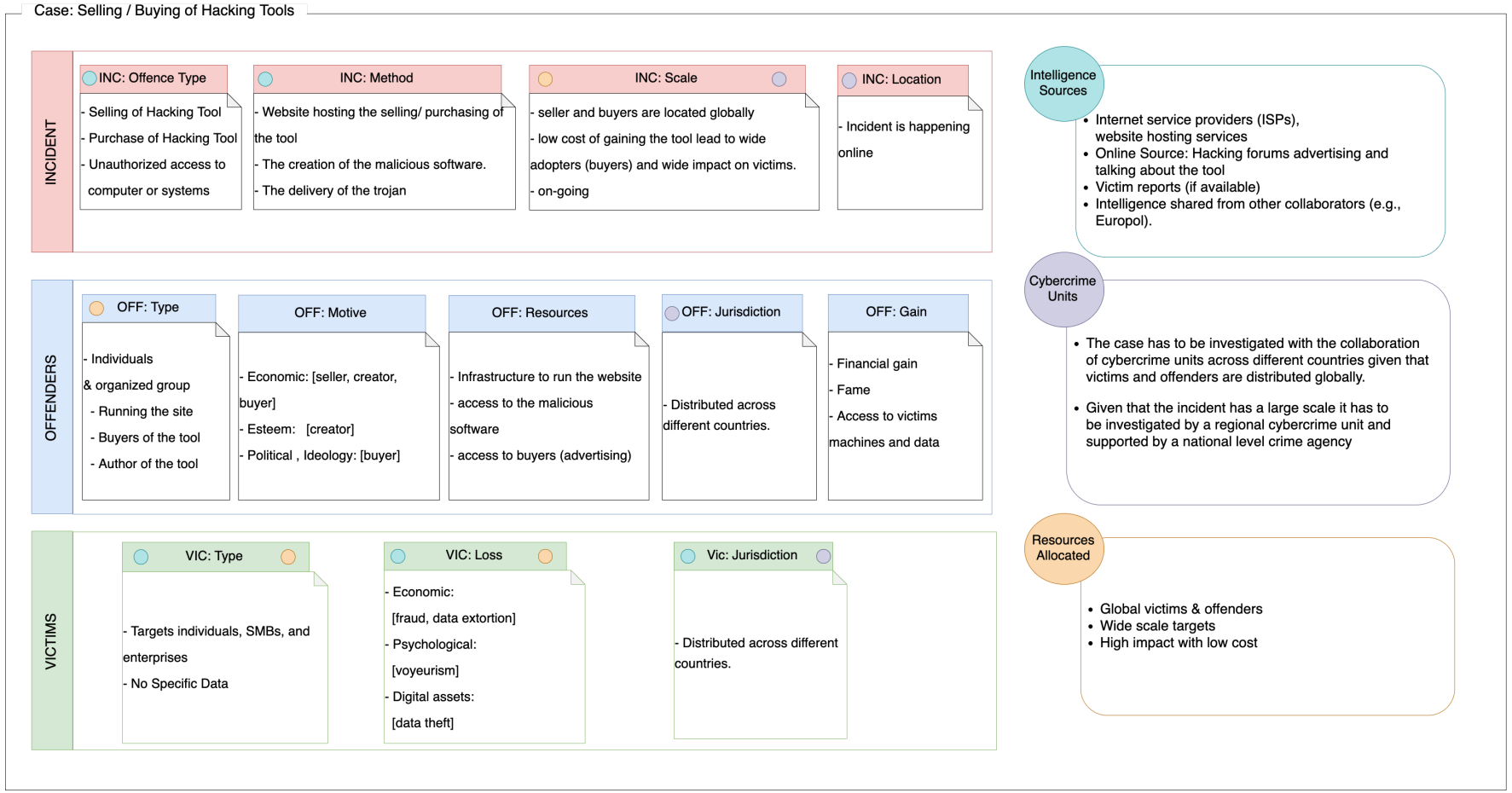


Figure 5.2: Case Example: Selling/buying of hacking tools

been committed, such as selling of the hacking tool, the purchase of the tool, and the utilisation of the tool to gain unauthorised access to victims' machines. We also capture the methods of how the offences have been implemented, which includes, for instance, the manufacturing of the malicious software and the hosting of the website to buy/sell the RAT tool. The scale of this incident is also captured and considered high due to the wide-scale target of victims, the wide features and control it provided to the attacker, the ease of use, and the low cost. We also model the information related to the offenders and victims as they become available during the investigation. For example, we capture that there are different types of offenders: those who sell/create the RAT and those who buy it and utilise it. Moreover, the nature of the cybercrime means the offenders (buyers, sellers) are located in multiple jurisdictions which warrants a global collaboration between authorities in multiple countries for coordinated capture of offenders and the taking down of the selling site. The victims element captures the types of victims who are targeted by the RAT and the types of losses they suffered. For example, psychological and reputation losses are incurred due to voyeurism, and economic loss due to data theft and extortion.

This case demonstrates one example of how our model can be used to map a case of selling and buying of malicious tools. It helps organise the investigation process and inform what kind of resources, collaboration, and intelligence are needed. The model can also be applied to other cases of cybercrimes including phishing, cyber-stalking, and cyber-terrorism just to mention a few.

For instance, let us consider a case of cyber-terrorism that is based on a recent case that took place in the US, where the offender designed a process that uses a computer script to make ISIS propaganda more conveniently accessed and disseminated to users on social media platforms [170]. The case describes how the offender distributed the script with instructions on how to use it to individuals whom he believed to be ISIS supporters and members of pro-ISIS organisations. The offender is an individual actor, who used online chatting services and social media platforms to provide easy access to terrorist material such as online magazines and

videos. The motive is ideological as he wanted to provide easy access to terrorist-created material to radicalise and inspire attacks against the US. Although the distribution scale was limited to specific social media groups, the impact of sharing such material that advocates for carrying out attacks is significant. Therefore, this incident requires the investigation to be conducted by a national level agency. Moreover, the intelligence sources required would include coordination with the social media platforms used by the offender, online information related to the publications of online media by the ISIS group, and evidence found from the offender's devices. In this example, although there are no direct victims, there are indirect victims who will be targeted and recruited by the dissemination of the terrorist media.

5.5 CCINT: Operational Framework

One of the challenges that analysts face when investigating cybercrimes, is making sense of large amounts of data. As such, we designed the operational side of the framework to assist in handling complex data and support multiple analytical techniques. In addition, the framework is designed to account for the six key steps in the analytical process: problem definition, hypotheses generation, information collection, hypotheses evaluation, selecting the most likely hypothesis, and continuous monitoring of new information [85].

An overview of the architecture is shown in Figure 5.3. The initiation of any investigation, as described by the practitioners, is usually triggered by either an internal event (uncovered intelligence information) or external event (reported crime). These are modelled in our framework by providing two modes of operation: bottom-up investigation where the system detects intelligence information from raw data and generates alerts to the user to further investigate, or top-down investigation where the investigator has outside knowledge and wants to generate a hypothesis in order to prove it using the data. Furthermore, typically any cybercrime investigation involves various specialists and investigator roles working together to solve a case, thus we design the framework with support for collaborative investigation sessions, such that several specialists can work together on a shared case scenario.

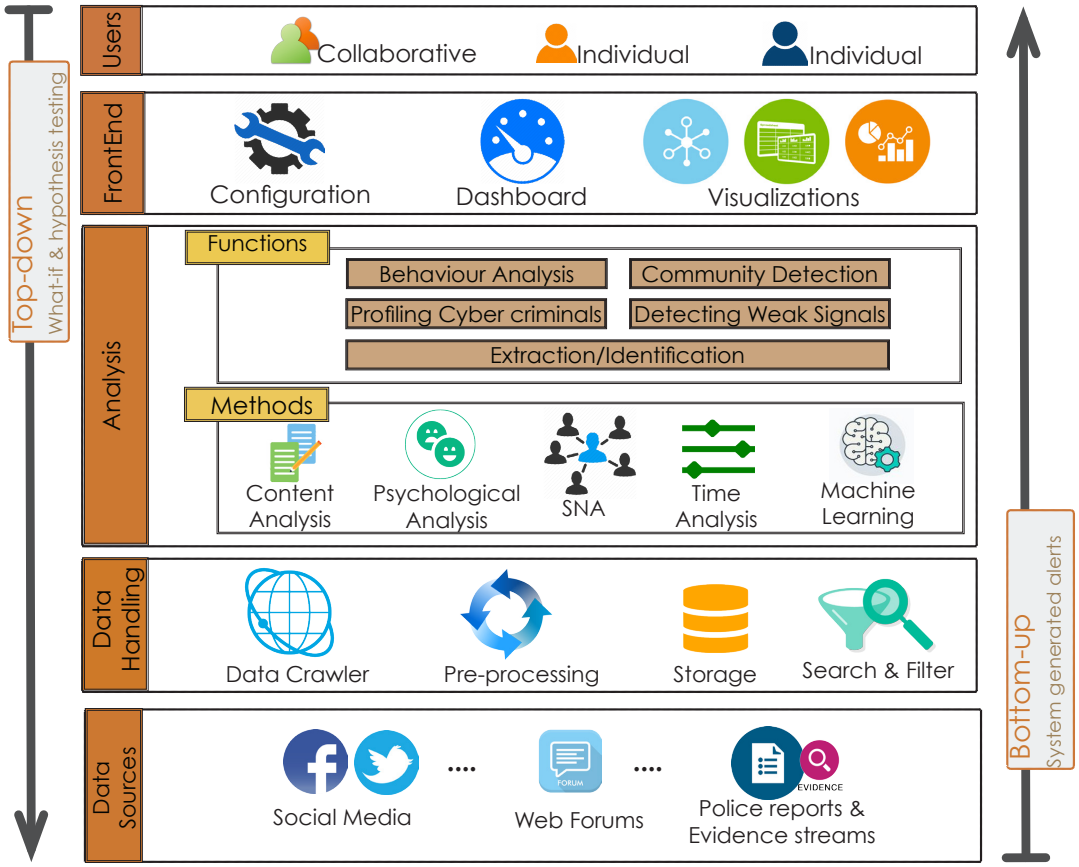


Figure 5.3: CCINT Architecture

5.5.1 Framework Components

The framework consists of three main layers: data-handling, analysis, and front-end, and two external layers: users, and data-sources. The data-sources layer contains any open online or offline data sources that are of interest to the analyst and can be plugged into the framework through an Application Programming Interface (API) to feed data to the data-handling layer.

The data-handling layer is responsible for collecting data from the different sources based on the user's defined configuration. A plug-in for each data-source is created to facilitate the collection of the data. Moreover, this layer consists of data acquisition using a crawler, pre-processing and cleaning of the data, search and filtering, and a database for storage.

The next layer is the analysis layer, which provides a collection of analytical

methods to support the detection and analysis of cyber criminal activities. Methods such as content and sentiment analysis, psychological and personality analysis, time analysis, and social network analysis are used to perform these operations. These methods are utilised to support different functionalities that an investigator may need to perform. We focus here on six main functions that have been found to be important to investigators: behaviour analysis, community detection, profiling cyber criminals, detection of weak signals, and identification of criminal content.

The front-end layer supports user interaction and consists of a dashboard, multiple visualisation views and a configuration panel for the analysts to customise the different processes within the framework. The top layer is the user layer that supports both individual and collaborative sessions.

Furthermore, as described by professionals, during the investigation they may combine two approaches: bottom-up and top-down. Thus, the framework should support the two modes of analysis. The bottom-up mode allows the analyst to start from the data level without having any previous hypothesis, and allows him to observe the data and look for anomalies and abnormalities. In this analysis mode, the system may suggest possible leads for the analyst to investigate in the form of alerts, based on anomalies detected and unusual patterns. For example, the system may alert the investigator if it detects an increase in the number of negative sentiment tweets originating from a given location of interest. Thus, the investigator would then hypothesise that some event may have occurred in that location and perform a more detailed investigation. On the other hand, the top-down approach allows the analyst to start with a hypothesis that they want to test, and examine different what-if scenarios in order to confirm or reject their hypothesis using the data. For example, the investigator may have a lead that two online personas are actually operated by the same criminal. He would then test this hypothesis using stylometric content analysis, and the result would be the probability that this hypothesis is true.

The analysis layer is the core component of the CCINT architecture. As such, in the following sections, we will describe in detail the main elements that constitute the analysis layer.

5.5.2 Analysis Methods

Based on the interviews with cybercrime investigators, they described a variety of methods that they use while investigating different cybercrimes. Some of these methods are still conducted manually (e.g., reading through online content), assisted with simple methods (e.g., keyword-based filtering), or conducted using a diverse, incompatible, set of tools. Investigators highlighted the need for (semi)-automated methods to improve the process of investigation and detection of hidden patterns in the investigation data. As such, in the following, we outline the main methods used in CCINT framework to support the needs of investigators.

Content Analysis — The content analysis method allows cybercrime investigators to analyse the various material posted by cyber-criminals, such as text, images, and videos. Currently, we focus only on textual content however we plan to add support for multimedia such as images and videos. This method utilises techniques from natural language processing, data mining, and machine learning fields. For example, *entity extraction* is used to allow investigators to extract specific patterns such as usernames and URLs. Additionally, this method allows the investigator to analyse the text in order to identify topics, detect most frequently used words, and compare the writing style of different suspects, or criminal groups.

Psychological and Personality Analysis — Previous research in the criminology and psychology fields suggested that criminals may exhibit behaviour and personality changes that may act as indicators to them committing a crime [171, 172]. Based on this premise, we provide investigators with analytical support to explore personality and psychological characteristics of their suspects. This is achieved by proxy of analysing textual content written by these suspects in order to identify features such as *emotions* (e.g. anger, disgust, sadness, joy, and fear), and *personality traits* (e.g., OCEAN Big-five).

Social Network Analysis (SNA) — The SNA method allows cybercrime investigators to capture interactions between different nodes in a criminal network. For example, it can be used to create a network of possible criminals "nodes" and the interactions/relationships "edges" between them. Using SNA techniques, the investigator will be able to detect hidden relations among possible suspects, understand who is the most influential member in an organised criminal group, and what is the best way to disrupt such a network. SNA will also enable us to provide the investigator with a visual representation of the criminal networks. Additionally, this can capture different patterns of the suspect behaviour, for example, in terms of the number of posts they share, discussions they participate in, and friendship connections they create or destroy [173].

Time Analysis — Time analysis methods provide support for detecting patterns in the data when temporal information is available. For example, how frequently the suspect posts, or a sudden increase of suspicious network activity before an attack is launched. Analysing these patterns can sometimes provide interesting insights to investigators.

5.5.3 Analysis Tasks

From our investigation of the literature (Section 2.2), we found that the main research problems tackled in the literature within the cybercrime-intelligence domain generally revolve around the following key areas:

Detection of communities and organisational structure — This includes the identification of key members in criminal networks, the discovery of sub-communities within the network and the different properties they possess, the detection of strong ties between nodes to evaluate relationships between criminals, and finally studying organisational structure of the criminal groups to understand the hierarchy of the criminal network. This aids law enforcement in detecting the leaders and influential members in order to target them instead of wasting resources on low-level, non-influential criminals.

Behaviour analysis and interaction patterns — This includes analysing behaviour and finding patterns in the cyber-criminal networks, establishing interaction patterns between individual actors or between sub-groups within the network. By observing how the network changes over time and studying meta data information, we can possibly predict when the network is conspiring to commit a crime.

Profiling cyber criminals — This includes identifying language stylometric fingerprints and author characteristics such as their age, gender, and ethnicity. As most criminals use screen-names and disguise their information, it is critical to study techniques that can reveal characteristic information about them. Profiling cyber-criminals aids in predicting if two online accounts are operated by the same individual, and to track criminals across different online platforms.

Identifying Weak Signals — This includes identifying weak signals of disruptive events and predicting offline events [174] (i.e., indicators that initially appear insignificant but actually are early indicators of large-scale real-world phenomena). Predicting tipping points, the likelihood of rumour spread, information propagation, and the expected reactions of the public.

Extraction and identification of online criminal content — This includes the development of techniques to automatically identify criminal-related content or individuals. This involves building crawlers that are able to extract content and classify it as criminal-related or non-criminal content.

5.6 Research Focus – The Battle for the Heart and Mind

As we have seen in previous sections, cybercrimes are quite diverse and vary considerably in terms of methods, motives, and victims. Therefore, we need to focus the research on a specific type of cybercrime. A considerable amount of research has focused on studying cybercrimes that target networks and systems, such as malware, DDoS, and cyber-fraud [34–36]. However, in recent years, a different kind of threat has grown and became prevalent, which targets people's hearts and minds [175]. These crimes cause online harm and have ideological, political, or psychological

motives. Such threats include the use of online resources to manipulate the public for political or social gain, hate crimes, cyber-bullying, trolling, spread of fake news, and spread of violent extremist content and behaviour. Recognising the danger of these emerging online threats, the UK government's Department for Digital, Media, Culture and Sport along with the Home Office recently published a joint white paper aiming at limiting online harm and establishing new regulatory frameworks that can properly address these issues [37]. Among the categorise identified in the report, is the topic of *extremist online content and activity*, which is highlighted as one of the areas that have a less clear definition and is still lacking a proper and effective response [37].

Moreover, the Internet organised crime threat assessment report published by the European Cybercrime Center (EC3) [6], emphasise the threat caused by the wide exploitation of online service providers by terrorist groups to distribute their propaganda material. This forms a significant challenge to law enforcement as well as online platforms to disrupt such activities. Therefore, there is a need for collaborative efforts from the research community, the private, and the government sectors to create appropriate methods and policies to counter this threat. Several studies in the literature looked at the role of the Internet in terrorism and whether it is mainly used to instigate or facilitate involvement [176, 177]. Some studies categorise its role to three goals: instrumental such that it facilitate the access to information, identity such that it provides ideological belief, and relational such that it satisfies affiliation and connection goals [178]. Moreover, other studies have showed links between limiting terrorists abilities to carry out attacks and the disruption of their flow of online propaganda [6]. For example, reports suggested that the Boston Marathon bombers in 2010 were inspired by extremist online propaganda [26].

Furthermore, the report from Europol [6] points out the importance of counter-terrorism cooperation and sharing of information between law enforcement authorities as well as with the private sector. This supports our findings from the study in Chapter 4. As such, it has become a priority for law enforcement practitioners to have a better understanding of emerging threats and technologies. This is important

to have an effective measure to counter the extremist online recruitment campaigns and the spread of their online propaganda. Going forward in this research, we will focus our scope on the topic of online extremist content and activities. We will utilise our framework to better understand the properties of the extremist narrative and develop methods to defend against its spread. This will aid law enforcement and online platforms in their fight against online radicalisation and the spread of extremist propaganda.

5.7 Limitations and Implications

The method applied to create the CCINT framework depends on combining previous knowledge from the literature with our understanding of existing processes followed by practitioners and their needs. As such, it is limited to the trends and the processes discovered during the interviews as described by practitioners. Additionally, although we have demonstrated the applicability of the framework to different examples of cybercrimes and presented a case study to validate the proposed framework, further validation with practitioners is needed to test the framework in a real-life investigation environment.

Our work has implications for practitioners working on investigating cybercrimes. Our work can assist investigators by providing a holistic view of the investigation and improve their understanding of the different factors relevant to the three elements of the crime (i.e., offender, victim, and incident). Also, it aids them in managing allocated resources such as time and manpower, identifies the different intelligence sources to consider, and which cybercrime unit to oversee the investigation. Moreover, through the operational side of the framework, we give a road-map that allows for integration between different analytical techniques covering behavioural and linguistic aspects, as well as a combination of multiple data sources. Furthermore, the framework components act as a general definition of models that need to be specialised for each type of cybercrime. As such, going forward, we demonstrate the use of the framework components by focusing on a particular type of cybercrime and show how it can support investigators to combat such crime.

5.8 Summary

In this chapter, we presented the CCINT framework that is inspired by our investigation of the literature and interviews with practitioners from both government and private sectors. We provided a conceptual model to be used during investigations of cybercrimes, which will help investigators visualise the details of the case and map the properties associated with different elements of the crime. Moreover, we presented a working example of how to apply the model on a real-world cybercrime case. Next, we outlined the operational side of the framework, where different analytical methods are adopted to achieve various functionalities related to the investigation of cybercrimes.

To narrow the scope, we decided to focus our research on a particular type of cybercrime, namely extremist online content and activities, which have been highlighted as one of the emerging online threats that require further investigation and more effective responses. As such, the following chapters will focus more closely on the topic of online extremism and radicalisation and, in particular, the *Islamic State of Iraq and Syria (ISIS)* extremist group, aiming to identify measures to automatically detect the spread of radical extremist propaganda and activities online.

6

Identifying Signals of Extremist Propaganda

6.1 Introduction

Several terrorist and extremist groups have benefited tremendously from technology such as Internet and Online Social Network (OSN) platforms as they provide them with opportunities to spread their propaganda, widens their reach for victims, and facilitates potential recruitment opportunities. Research into extremism and online radicalisation revealed that the MO of those groups starts by exploiting popular public platforms to capture attention, maximise impact, and widely spread their ideology and propaganda [6]. They then move potential recruits to more controlled environments such as smaller platforms with end to end encryption for further grooming and radicalisation. A study released by UNESCO [179] discussed how young people and vulnerable individuals are affected by violent extremism on social media. The study looks at existing research in the area and suggests that although the Internet and social media platforms may not directly cause violent behaviour, they can be used as a first step by extremists towards facilitating violent radicalisation.

To limit the reach of cyber-terrorist and extremist groups, several private and governmental organisations are policing online content and utilising big data

technologies to minimise damage and counter the spread of such information. For example, the UK launched a Counter Terrorism Internet Referral Unit in 2010 aiming to remove unlawful Internet content and it supports the police in investigating terrorist or radicalising activities online. The Unit reports that among the most frequently referred links were those coming from several OSNs, such as Facebook and Twitter [24]. Similarly, several OSNs are constantly working on detecting and removing users promoting unlawful and extremist content. Big tech companies such as Facebook, Google and Twitter have established dedicated teams to develop policies and tools to counter terrorism and extremist content and their use of the corresponding platforms. For example, in 2017, Twitter announced that they suspended around 300,000 accounts globally that were linked to terrorism [108]. In addition, the establishment of the Global Internet Forum to Counter Terrorism (GIFCT)[1] by a collection of tech companies with the vision of preventing terrorist from exploiting platforms is a major step towards creating a safe online environment. The GIFCT initiative announced the commitment to invest in technology and improve the existing capabilities to detect and remove terrorist and violent extremist content online.

In this chapter, we utilise computational methods to analyse online extremist propaganda and reveal their ideology and tactics. Our analysis uses data mining techniques to computationally uncover textual properties associated with these groups. Furthermore, we explore new methods to uncover not only the syntactic properties associated with online extremist propaganda, but also look at the contexts and semantics of words used from the perspective of the extremist group. Finally, we adopt psycholinguistic methods to extract a number of psychological properties. These properties correspond to the radicals' interests, personality, and emotions. This allows us to create a general radical psychological profile that we can later utilise as a signal to detect possible radicalisation behaviour. Our hypothesis is that the psychological and linguistic properties of those who produce extremist propaganda are shared by those who may be influenced by it. By having a better

[1] https://www.gifct.org/

understanding of the linguistic cues and psychological properties conveyed in the extremist propaganda, we can identify extremist content and people who might be at risk of radicalisation.

In summary, the main contributions of this chapter are outlined as follows:

1. An analysis of the textual properties of ISIS online extremist magazine, Dabiq, in order to empirically identify their linguistic models and cues. This includes identification of themes, topics, and how they change over time.

2. Generate a radical-language model trained on the ISIS propaganda material. The model captures contextual properties and the semantic relationships between words.

3. An investigation into the psycholinguistic properties conveyed in these articles and how they change over time. This leads us to create a general profile for ISIS inferred from their published propaganda.

6.2 Method

In order to understand how radical propaganda are constructed and used, we analyse all issues of *Dabiq* magazine, an online English magazine published by ISIS terrorist group with the purpose of recruiting people and promoting their propaganda and ideology. A general overview of our approach is presented in Figure 6.1, where articles from Dabiq extremist magazines are used to perform two parallel tasks. (1) Build a language model using (a) Term-Frequency Inverse-Document-Frequency (TF-IDF) scores of uni-, bi-, and tri-grams, and (b) create word embeddings generated from a word2vec model. The output of this task is a radical corpus of top k-grams, and a word embedding model giving a vector representation for each word in the corpus. (2) Create a psychological profile based on the language used in the extremist propaganda articles, consisting of a set of emotional and topical categories using LIWC dictionary-based tool.

Dataset — A summary of the radical corpus dataset (C_{RAD}) is shown in Table 6.1. The corpus consists of 15 magazine issues with total of around $400K$

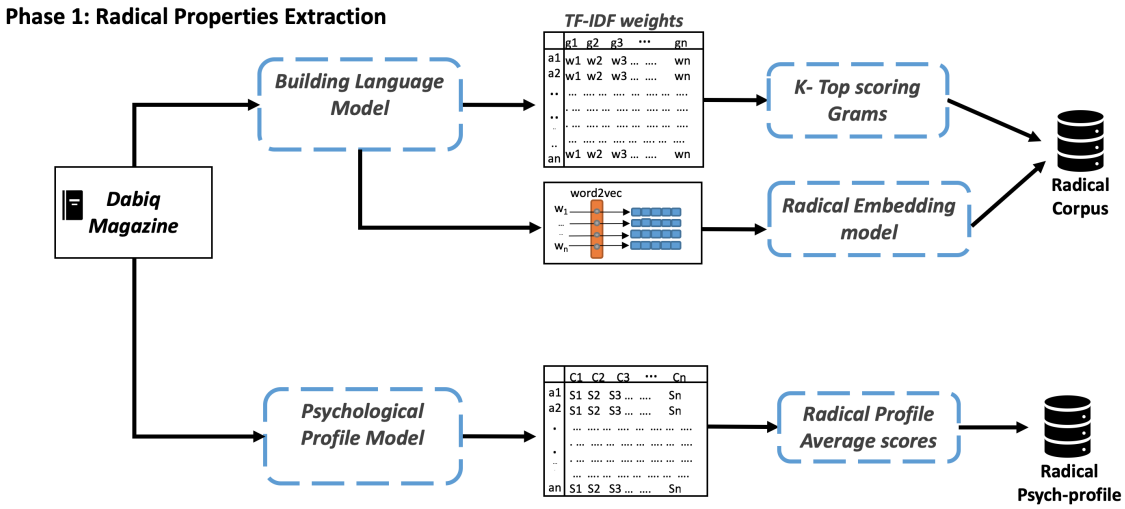


Figure 6.1: Approach Overview

words and $2M$ characters. Each issue typically consists of a collection of articles, editorial snippets, and visual images. The purpose of these magazines is to spread ISIS ideology, motivate, and recruit people. In our analysis, we focus only on the text and discard any graphical content. Using this dataset, our aim is to (1) investigate what topics, textual properties, and linguistic cues exist in these magazines, and (2) uncover psycholinguistic properties including emotions and personality traits using these magazines as proxy.

Approach — In order to be able to analyse the text in an automated way, we need to transform the text into a format that we can computationally process and analyse. To do so, we transform each magazine issue M in our dataset to a vector of words (Bag of Words). We run a set of pre-processing steps and convert all words to lower-case, remove all stop-words (e.g., and, or, the) and punctuation marks (e.g., . , ! ?). We perform lemmatization aiming to remove inflectional endings and return words to their roots in order to avoid duplicates of the same word. For example, words with different endings (e.g., killed, killing) will be mapped to a single word (e.g., kill) [180].

After pre-processing the text, the first step focuses on extracting n-grams from the magazines in order to capture not only the words, but the context in which they were used. Each magazine M is now represented as a vector containing the

Table 6.1: Radical corpus (C_{RAD}) summary

Issue	Date	#Words
1	July 2014	26,498
2	July 2014	35,194
3	September 2014	9,817
4	October 2014	30,637
5	November 2014	19,464
6	December 2014	29,020
7	February 2015	11,725
8	March 2015	32,344
9	May 2015	33,882
10	July 2015	39,164
11	September 2015	12,095
12	November 2015	46,130
13	January 2016	28,601
14	April 2016	12,326
15	July 2016	40,601

n-grams ($n = 1, 2, 3$). We then calculate a composite weight $weight_{i,j}$ for each gram g_i occurring in a given article M_i using term-frequency inverse-document-frequency (tf-idf) as follows:

$$weight_{i,j} = tf(g_i, M_j) \times log(\frac{N}{df_i})$$

where $tf(g_i, M_j)$ is the frequency of gram g_i in magazine M_j, N is the total number of issues in the corpus, and df_i is the total number of magazines that contain the gram g_i.

After applying this process to each magazine issue in our dataset, we get a weighted matrix of size $m \times n$, where m is the number of magazine issues we have in the dataset, and n is the number of unique grams in the corpus. Grams with high weights means that they are more important in our collective corpus.

The second step in our approach is to create a word embedding model for the radical corpus. A popular idea in modern machine learning is to represent words by vectors, as vectors capture hidden information about the language, such as word analogies or semantic aspects. It is also used to improve the performance of text classifiers. Research in NLP has studied the effectiveness of word embedding methods

for encoding semantic meaning and found that semantic relationships between words are best captured using vectors representations within word embedding models [181]. We train word2vec model [134] on our corpus to build the lexical semantic aspects of the text using vector space models. We learned word embeddings using skip-gram Word2Vec model implemented in the gensim package[2] with vector size of 100 and window size of 5.

The final step in the analysis of the extremist content is to investigate the psycholinguistic properties of the text. Research in fields such as linguistics, social science, and psychology suggests that the use of language and the word choices we make in our daily communication, can act as a powerful signal to detect our emotional and psychological states [182]. Several psychological properties are unintentionally transmitted when we communicate. It can be detected in the way we speak or express ourselves in writing. For example, research in the field of psychology has found relations between the choices of words and symptoms like depression, suicide, and anxiety [183]. Moreover, previous research studied the social psychology of online interaction and investigated the extent to which principles of social psychology carry over into the online domain. Indeed, Li and Chignell [184] found that online blog readers are significantly more attracted to blog writers with similar personalities.

Different text analysis tools and dictionaries have been developed and used in the literature to identify psychological properties from written text. One of the very first tools to achieve this is LIWC [102]. The tool focuses on counting the frequencies of words occurring in a given text and mapping these to a set of pre-defined lexical categories, such as emotions (e.g., anger, fear), social interests (e.g., family, friends), morality foundation (e.g., care/harm, authority/subversion), and cognitive processes (e.g., certainty, tentative). These categories together can paint a picture of the author's interests, personality, and emotional states as conveyed in the analysed text. Although LIWC is not a personality measure in itself, previous work has found strong associations between LIWC categories and personality measures [103–105].

[2] `https://radimrehurek.com/gensim/models/word2vec.html`

They demonstrated the ability of these computerised text analysis tools to detect psychological meaning in different experimental settings [105]. By utilising the LIWC dictionary, we analyse the text in our corpus (C_{RAD}) and calculate the frequencies of words that map to the different categories.

6.3 Results

6.3.1 Textual Analysis of Dabiq Magazines

First, we perform a macro level analysis by looking at the top scoring TF-IDF grams in C_{RAD}. Among the top scoring grams in the corpus are words related to Islamic concepts, which is expected as the group focuses on associating its legitimacy with religion. Grams such as *Allah, Islamic state, Muslim* were among the top scoring grams in the corpus together with Arabic words such as *Khilafah, Jihad, Ummah (nation).*

To discover the main topics present in the corpus, we apply topic modelling techniques using Latent Dirichlet Allocation (LDA) algorithm [185]. We find common themes that appear across different issues, and these are Religion, war/violence, and people/places. Given that the group is using religion to legitimise their action, the articles contain a heavy Islamic tone, which is apparent through the excessive use of words such as *Allah*, *Muslim*, and *Islam* which all appear among the top 10 weighted words. Additionally, words such as *Kuffar* (infidels), *Hijrah* (migrate), and *Khalifa* (ruler) are used. Moreover, references to other religious groups such as *Jews, Christians* are becoming more present in later issues of Dabiq, mainly Issue#13 and Issue#15.

In terms of the war/violence theme, it includes grams corresponding to religiously motivated wars such as *crusade* and *Jihad*. Additionally, another interesting finding is that in Dabiq articles words like *fight* had higher weights than *kill*, and *war* had higher weights than *attack*, & *blood*. This may be a conscious strategy to legitimise their activities in the eyes of the reader by giving it such a lawful impression.

As for the people/places theme, we can see, for instance, that places including *Iraq*, *Syria* appear frequently, which aligns with the origins of both groups and

Table 6.2: Examples of top-grams per Dabiq issue

Issue#	Top grams
1	Say, imamah, millah, one, Iraq, khalifah, mujahidin, jihad, army, fight
2	Mubahalah, say, people, nuh, ark, flood, one, Jawlani, wilayat, make, front
3	Say, people, hijrah, one, land, Iraq, American, messenger, Sham, prophet
4	Say, one, war, kill, crusader, fight, take, people, Iraq, enslave
5	Say, remain, land, mujahidin, prophet, arabian peninsula, fight, jihad, expand, khilafah
6	Tandhim, say, make, emirate, Waziristan, brother, one, people, gold, sharia
7	One, grayzone, crusader, people, religion, brother, make, hypocrite, kill, jihad
8	Say, people, faction, one, religion, fight, brother, kufr, know, land
9	Yarmuk, crusader, faction, ribat, plot, fight, regime, camp, region, war
10	Jawlani, fight, rule, coalition, law, sahwah coalition, front, faction, jihad, take
11	Crusader, war, fight, darnah, khilafah, religion, leader, Iraq, American, follow
12	Faction, khilafah, jihad, crusader, Syrian, soldire, mujahidin, domestic council, Sham, war
13	Rafidah, khilafah, kill, Taliban, mourn, rafidi, jihad, muharib, Jew, attack
14	Bengal, Ikhwan, kill, religion, jihad, Muslim brotherhood, take, khilafah, soldier, fight
15	Jesus, prophet, lord, God, religion, Levant, soldier caliphate, Christian, worship, creator

their geopolitical standings. Also, locations related to where they carried attacks and locations of their declared enemies like *Paris* and *America* are among the top weighted words.

To have a better understanding of the topics discussed in each issue and how they changed over time, we extract the top scoring grams per issue. However, first we filter out grams related to generic religious terms such as *Allah, Islam, Muslims*, and terms referencing names such as *Abu, Ibn, Ibrahim, Mohammad* to focus our analysis and extract relevant topics of each magazine issue. Table 6.2 presents a list of the top scoring grams in each issue of the Dabiq magazine.

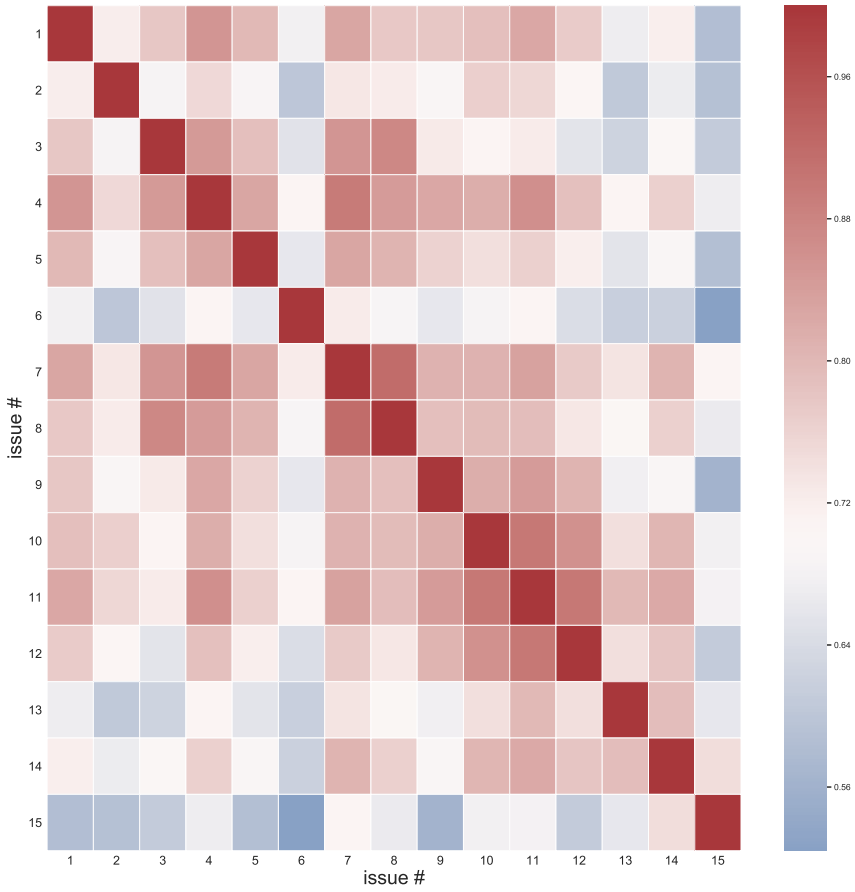


Figure 6.2: Cosine Similarity of the 15 Dabiq issues

Moreover, to identify the overall similarity between each issue of Dabiq magazine, we calculate the cosine similarity between the weighted vector representation for

each issue. Figure 6.2 shows a heat map comparing the different Dabiq issues. We can see from the Figure that there are two odd parts where the similarity scores are decreasing. Mainly, Issues #6 and #15 are quite distinct and seem to have different focus from all other issues. For instance, Issue#15 which is the last published issue of Dabiq to date focuses on other religious groups, mainly Christians, while Issue#6 seem to focus on other Jihadi groups such as organisation (Tandhim) of al-Qaeda in Waziristan (Table 6.2).

6.3.2 Word Embeddings for Radical Propaganda

After exploring the general themes and topics discussed in Dabiq issues and the main keywords used in each issue, we want to investigate the relationship between words used in the context of ISIS propaganda. To do so, we use word embeddings to capture these semantic connections.

We train a word embedding model on the corpus to obtain a vector representation for each word. We then build a semantic network of the top 500 scoring words (TF-IDF), such that we calculate the cosine-similarity score between each of these words. This will result in a graph (571 nodes, 2965 edges), where nodes represent words, edges represent a similarity relationship, and edge weights represent the similarity score. In other words, every node (word) will be connected to other words close to it in the vector-space [186].

Figure 6.3 represents the giant component of the semantic network. We use a modularity algorithm [187] to measure the network modularity and partition the network into communities (groups of semantically related words). Words are scaled based on weighted degree measure.

To validate if the model is able to capture semantics of the Dabiq magazine, we query the model to get the most similar semantic neighbours of some selected words. For example, the word ‘taghut’ is most similar to some political figures such as ‘Erdogan’ and ‘Saddam’ which gives an indication that ISIS view these politicians as tyrant and evil. Moreover, the word ‘Kuffar’ (infidel) is mostly associated with

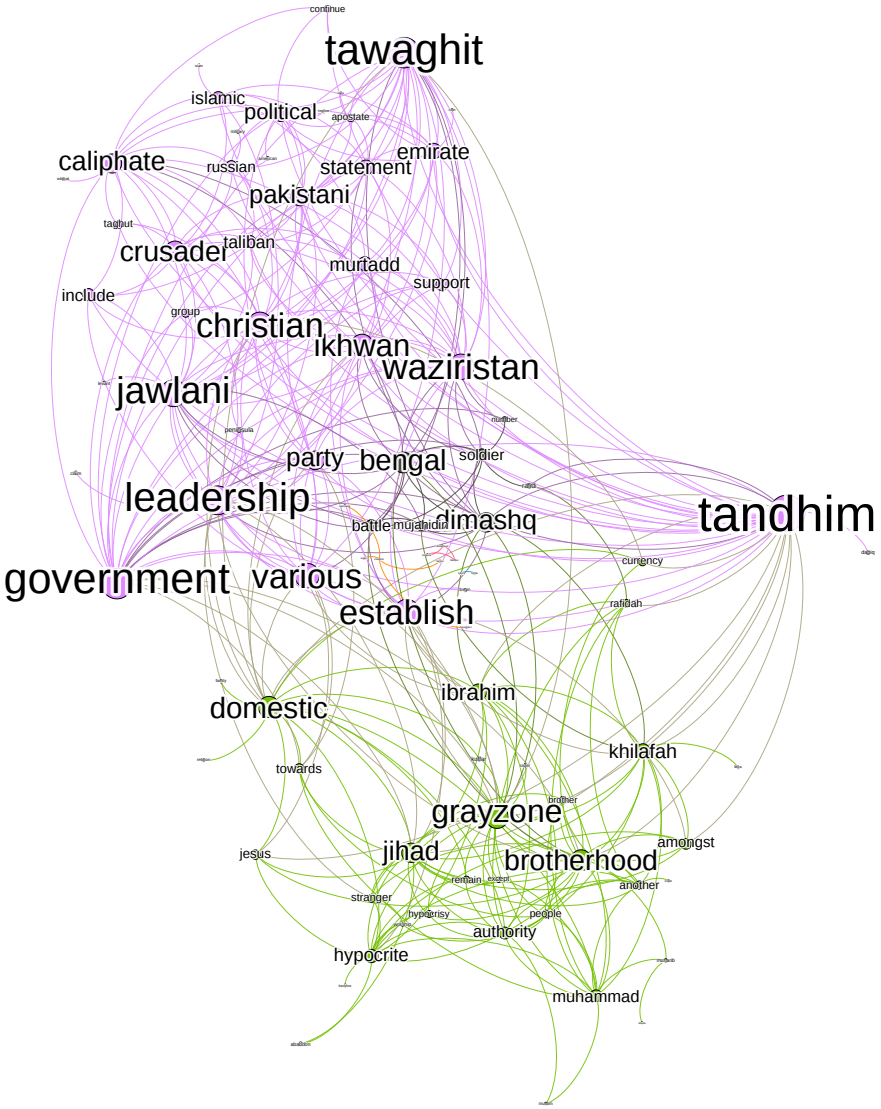


Figure 6.3: Giant component of C_{RAD} semantic network

'weak', 'tawaghit', and 'secular'. Additional examples of semantic relations between words appearing in ISIS propaganda are available in Table 6.3.

6.3.3 Psycholinguistic Properties of Dabiq

Literature from the fields of terrorism and psychology suggests that terrorists may differ from non-terrorists in their psychological profiles [172, 188]. A number of studies looked at the motivating factors surrounding terrorism, radicalisation, and recruitment tactics. Terrorist groups tend to target vulnerable individuals who

Table 6.3: Examples of similar words in the Dabiq context

Word	Most Similar Words
kuffar	weak, tawaghit, secular, hypocritical
rafidi	houthis, safawiyyah, murder, slaughter, mobilization
taghut	Erdogan, Saddam, vote, participation
jihad	fard, defensive, claimant
ummah	reign, mislead, whole, pious

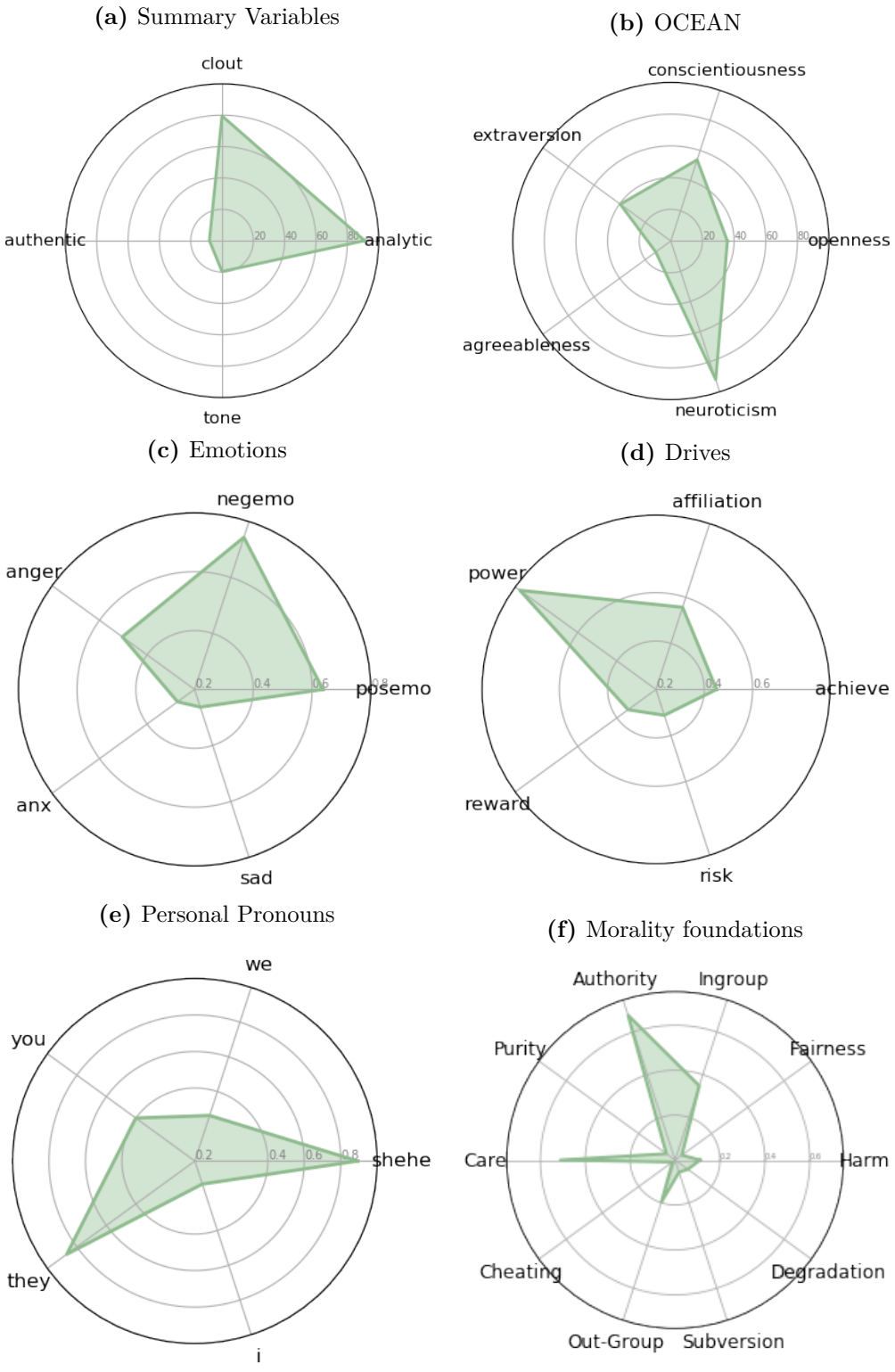
have feelings of desperation and displaced aggression. In particular, research into the recruiting tactics of the ISIS group found that they focus on harnessing the individual's need for significance. They seek out vulnerable people and provide them with constant attention [189]. Similarly, these groups create a dichotomy and promote the mentality of dividing the world into "us" versus "them" [190].

Inspired by previous research, we extract psychological properties from the radical corpus in order to understand the personality, emotions, and the different psychological properties conveyed in these articles. Using existing psycholinguistic lexicons (e.g., LIWC dictionaries), we explore the different categories related to emotions, drivers, morality, and personality and compare how these properties change across issues and over time.

We analyse the text in our corpus (C_{RAD}) and calculate the frequencies of words that map to different categories. In a macro scale, we look at the following properties:

Summary variables – *Analytically thinking* which reflects formal, logical, and hierarchical thinking (high value), versus informal, personal, and narrative thinking (low value). *Clout* which reflects high expertise and confidence levels (high value), versus tentative, humble, and anxious levels (low value). *Tone* which reflects positive emotions (high value) versus more negative emotions such as anxiety, sadness, or anger (low value). *Authentic* which reflects whether the text is conveying honesty and disclosing (high value) versus more guarded, and distanced (low value).

Our analysis shows that the magazines reflect confidence and formal logical thinking style with high scores of analytical thinking and clout. Adopting this style of writing for the propaganda material may reflect why they have been successful in their recruitment campaigns. On the other hand, low scores for tone reveal high



(a) Summary Variables **(b)** OCEAN **(c)** Emotions **(d)** Drives **(e)** Personal Pronouns **(f)** Morality foundations

Figure 6.4: Personality and psychological analysis of C_{RAD}

levels of negative emotions being conveyed in the text. Similarly, low scores for authenticity suggests more guarded and deceptive text (Figure 6.4a).

OCEAN / Big five – Measures the five personality traits, namely Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. Looking at the Big Five personality properties, derived from LIWC dictionaries, we find that Dabiq articles exhibit significant levels of neuroticism followed by conscientiousness and extraversion, while the text showed lower levels for openness and agreeableness. This suggests that the text conveys high levels of anxiety, depression, and anger due to high neruoticism, as well as high conscientiousness which is typically associated with being focused and non impulsive (Figure 6.4b).

Emotional Analysis – Previous research that studied the motivations of terrorism has mostly focused on the role of negative emotions such as anger, hate, and anxiety. For example, anger has been linked in the literature as a mobilising emotion for motivating collective group actions [191]. It is equally important to examine the role of positive emotions such as love and compassion in activating and motivating extremists [192]. Therefore, this property measures the emotions conveyed in the text from both positive and negative sides (including anger, sadness, anxiety). The results show that both emotions are conveyed in the text, although the negative emotion is higher mainly due to the prevalence of the anger emotion (Figure 6.4c).

Personal Drives – Focuses on five personal drives, namely power, reward, risk, achievement, and affiliation [105]. The analysis revealed a significant effect of power. This may be linked to the purpose of these magazines which is to advertise their strength and recruit and attract individuals (Figure 6.4d).

Personal Pronouns — Counts the number of 1^{st}, 2^{nd}, and 3^{rd} personal pronouns used. The analysis for the use of personal pronouns reflects the dichotomy mentality of *us-vs-them*. The articles focus more on the 3^{rd} person pronouns (they, she, he) as opposed to less focus on the use of 1^{st} person pronouns (I, we). This suggests high levels of extremism as found in previous research that reveals high use of 3^{rd} person plural pronouns (e.g., they) is one of the predictors of extremism since this suggests that the group is defining themselves against the existence of an opposite group [193]. In addition, it is interesting to note that the use of the second

person pronoun (you) is higher than first person singular pronouns (I), which shows the strategy of these magazines in making the focus more on the reader (Figure 6.4e).

Morality — The moral foundations theory [194] is a psychological theory that describes how morality differs between cultures and yet shares some fundamental characteristics. The universal moral foundations that hold across cultures as described by the theory are: (1) harm and care, (2) fairness and cheating, (3) authority and subversion, (4) purity and degradation, (5) loyalty/in-group and betrayal/out-group. The theory suggest that cultures differ on the degree to which they build virtues on these five foundations. For instance, if we take political parties as an example, political liberals value virtues based on the harm/care and fairness/cheating foundations, while conservatives value virtues based on all five foundations [195]. We analyse the five morality foundations across all Dabiq issues and found overall very low scores for morality with the two most valued virtues being authority and care (Figure 6.4f).

Micro Level Analysis — We want to investigate how the different properties evolve over time across different Dabiq issues. Such micro level analysis revealed some interesting observations and trends as seen in Figure 6.5 and Figure 6.6.

The morality values associated with each issue of Dabiq magazine are shown in Figure 6.5a and Figure 6.5b. The authority foundation was at its peak in the very first issue published as the group needed to establish respect and legitimate authority. We then notice a drop until Issue#3 when it starts to pick up again with minor fluctuations. Similarly, the care foundation shows some major fluctuations with two main peaks in the 5^{th} and the 13^{th} issues before it drops again to a level similar to that of the first issue.

With regards to the drives and motives of the magazines, power, achievement, and affiliation are the core drives and needs present in each issue of Dabiq. As shown in Figure 6.5c, power fluctuates over time with peaks in the 1^{st} and 4^{th} issues. Similarly, affiliation-related words show minor fluctuations with a tendency to increase over time as opposed to achievement-related words which seem to decrease over time. By comparing the first and last issues of Dabiq, we see how ISIS tended

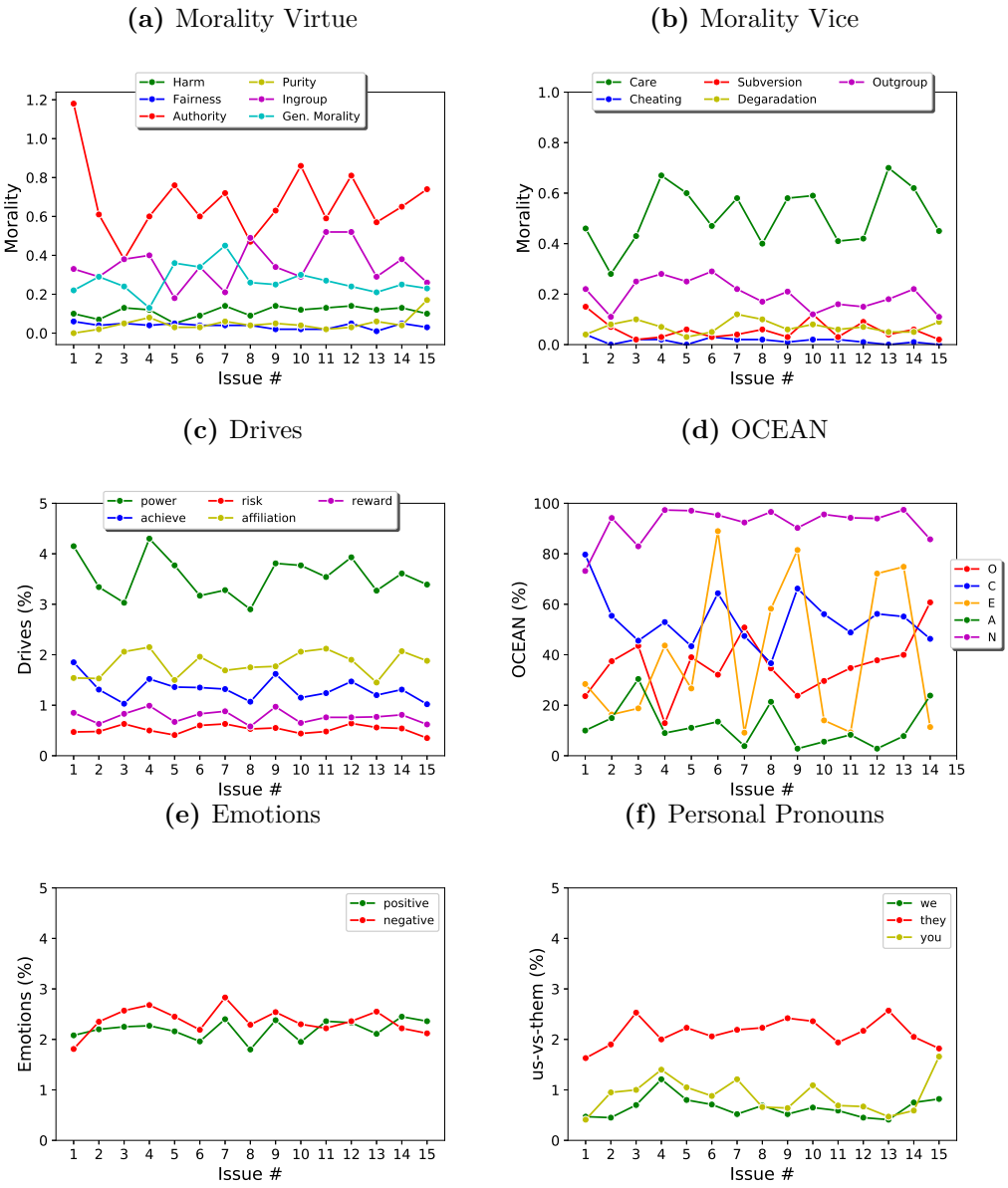


Figure 6.5: Psycholinguistic analysis per Dabiq issue

to convey more power and sense of achievement in its propaganda at the start but then these drives decreased by the time of the last publication. On the other hand, the sense of affiliation had an opposite trend as it grew over time.

Looking at how the OCEAN variables change over time (Figure 6.5d), we see an overall increase for the openness property over time until it peaks in the last issue, while conscientiousness shows an opposite decreasing pattern. This suggests that the group propaganda strategy over time may have been less focused than in their first issue. Similarly, studies suggest that low scores for conscientiousness may be linked

to more antisocial and criminal behaviour [196, 197]. Furthermore, extraversion show a significant fluctuation behaviour across different issues, while agreeableness had significant spikes in the $3^{rd}, 8^{th}$, and last issues. Finally, neuroticism has remained at similar levels following a spike in the 4^{th} issue.

Moreover, as we have seen from the macro analysis both positive and negative emotions are properties conveyed in Dabiq magazines. Comparing the distribution of positive and negative emotions across each issue (Figure 6.5e) reveals fluctuations with negative emotions slightly higher across the majority of the issues with exceptions in the 1^{st}, 11^{th}, and the last two issues). The negative emotion peaks in the 7^{th} issue which, upon examination, shows that the issue's feature articles focus on who ISIS deem to be enemies of Islam including other Islamic groups, leaders, and the west [198].

Looking at the use of personal pronouns over time (Figure 6.5f), we see an overall increase in using more 3^{rd} person plural pronouns such as "they" until about the 13^{th} issue where it starts to drop. Similarly, the emphasis seem to be more on the reader through the use of "you" pronoun which tends to increase until it spikes in the last issue of Dabiq.

Moreover, one of the main themes prevalent in ISIS is religion. Figure 6.6a shows how this theme is mostly steady across all issues with the exception of two drops in issue 6 that focus on rejection of other Jihadi groups and issue 12 that followed the 2016 Paris attack [199]. Moreover, the final issue of Dabiq shows a peak in its focus on the topic of religion even exceeding the initial peak in issue#1. This is in line with our observation while analysing the topics for each issue as the focus in the final issue seems to be on motivating war against other religious groups.

The emotional tone (Figure 6.6b) shows an interesting pattern as it started with very high levels in Issue#1 and dropped after that in the following issues before it peaked again in the last couple of issues. Another inclination in the levels of emotional tone is observed in Issue#11, which focuses on out-group conflict and builds on different conspiracy theories against ISIS [199].

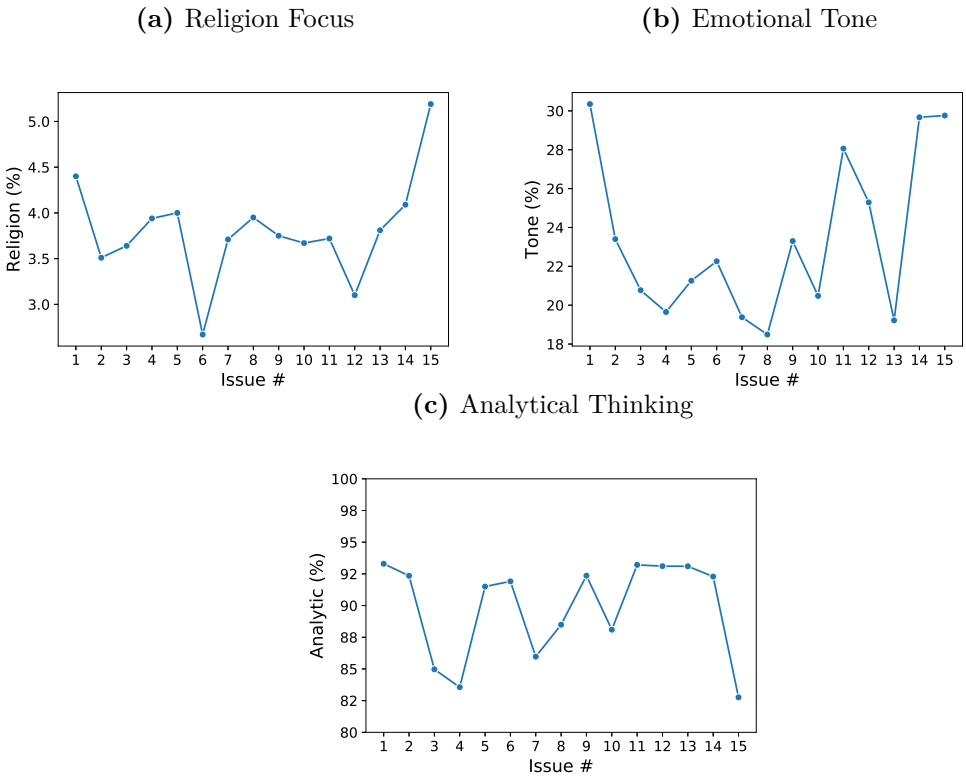


Figure 6.6: Changes over time in Dabiq; focus on religion, emotional tone, and analytical thinking

The last variable we will look at is the analytical thinking variable (Figure 6.6c). Overall, the variable scores are high (over 80%), with its highest levels in the very first issue and its lowest in the final issue with fluctuations in between. This suggests that ISIS' approach to constructing the propaganda has changed over time which may be related to the group's sustained losses in territory.

6.4 Discussion

In this chapter we made progress in understanding how extremist propaganda is constructed by attempting to computationally identify the different hidden markers that are associated with propaganda published by the ISIS group. We analysed the text published in all issues of Dabiq magazines from textual and psycholinguistic perspectives to uncover generic themes, topics, and psycholinguistic properties as depicted in the the ISIS propaganda.

The textual analysis of the articles showed several themes that appear across different issues. As expected, religiously motivated violence was the main overarching theme across all issues. Similarly, the motivation for war and violence against the declared enemies of ISIS included groups such as other Jihadi groups, Islamic politicians, and the West. In addition, the contextual analysis of the relations between different words used by ISIS allowed us to extract semantic meaning of how they describe different concepts. Using word embeddings to represent the different words appearing in the articles, allows us to extract semantic relations between words such that words with similar meanings will have similar representations and thus cluster together. This approach revealed a number of interesting semantic associations regarding the ISIS ideology. For example, the association between 'Taghut' (tyrant) and a number of political figures shows how ISIS portrays those politicians. Similarly, the high similarity between 'Jihad' and words such as 'defensive' and 'Fard' (obligation) illustrates how ISIS portrays jihad as a defensive tactic and as an obligation to all followers.

Moreover, the psycholinguistic analysis revealed that the magazines exhibit strong emotions from both positive and negative sides. The most prevalent negative emotion is the anger emotion, which as suggested in the literature, is typically considered one of the main mobilising factors for groups [191]. Additionally, looking at the Big Five psychological traits, we found that the neuroticism trait is extremely dominant throughout the 15 issues followed by the conscientiousness trait, while the agreeableness trait was the weakest of all five. The analysis of the morality foundations showed that the main valued virtues for ISIS are authority, care, and in-group loyalty.

In order to identify if these properties are capturing radical signals or are common across any normal discourse. We need to compare the results against a control group. To capture the properties related to a normal "non-radical" discourse (i.e., a control group), we used a dataset consisting of different news articles. The data [3] include articles from different publishers, such as the New York Times, CNN, and

[3]https://www.kaggle.com/snapcrack/all-the-news

Table 6.4: News Articles Data Summary "Control Group"

#Article	#Words mean (std)	Publication Year	Key Topics	Publishers
31	18,139 (15,852)	2011, 2015, 2016, 2017	Politics, War, Culture, Debate, Religion	Washington Post, Atlantic, Vox, CNN, New York Times, Buzzfeed News, National Review, Breitbart, Guardian

The Guardian. The articles discuss a range of topics and include a mix of print and digital publications. We randomly sampled 31 articles with varying lengths, such that the lengths are comparable to those in the Dabiq magazine. A summary of the news dataset "control group" is presented in Table 6.4.

To identify if mean differences between properties of both groups are statistically significant or merely due to chance, we perform an unpaired two-sample comparison using Mann–Whitney U test at the 5% level of significance ($\alpha = 0.05$). The null-hypothesis (H_0) is formulated such that there is no difference among the groups (the radical propaganda and the news groups). Alternatively, rejecting the null hypothesis means that the two groups are different. We report the mean scores of all properties for both Dabiq and News groups in Appendix A Table A.1. The results of the test showed that the differences between means of a number of properties are significant. For instance, properties related to the us-vs-them dichotomy (Figure 6.7a), the clout and authentic properties of the summary variables (Figure 6.7b), and the authority and degradation morality foundations (Figure 6.7c) were indeed significantly different from those of the control group.

It is critical that we discuss the topic of "warning behaviours", in the context of extremism, as a method to identify a set of indicators that may precede an act of violence. Meloy et al. [200] have identified the following eight warning behaviours: (1)**Pathway**, any behaviour that is part of research, planning, preparation, or implementation of an attack. (2) **Fixation**, any behaviour that indicates an increasing pathological obsession with a person or a cause. (3) **Identification**, the identification with a group or larger cause and have a warrior-like mentality. (4)

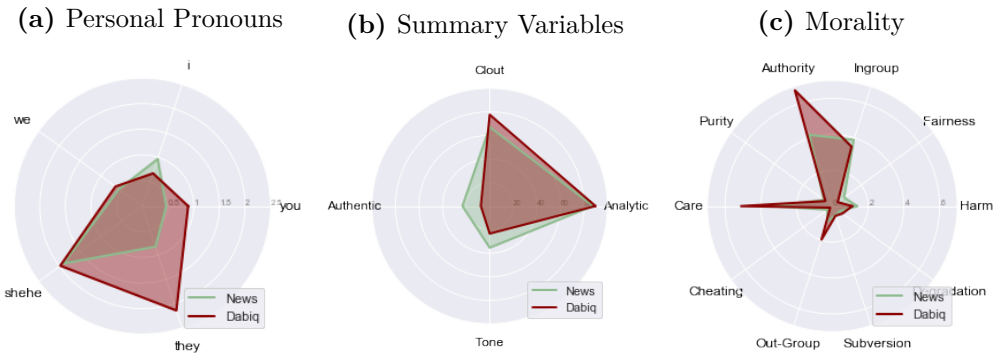


Figure 6.7: Comparison of Dabiq properties v.s. News Articles

Novel aggression, a behaviour change that results in acts of violence committed for the first time. (5) **Energy burst**, an increase in the frequency of activities (related to the target) before the attack. (6) **Leakage**, any leak of the intent through communication to a third party. (7) **Last resort**, where the subject feels increasing desperation forcing them into a position of last resort. (8) **Directly communicated threats**, where a threat is communicated to the target (implicitly or explicitly) explaining an intent to do harm.

Some of these warning behaviours are implicit and relate to the psychological properties of the subject (e.g., fixation and identification) while others can be explicitly identified through actions (e.g., energy burst and novel aggression). In light of our analysis, we reflect on some of these indicators (warning behaviour) in the context of ISIS. For instance, the fixation behaviour of the group is manifested in topics of religiously motivated violence against out-groups. Similarly, the group fixation with properties such as power and authority is prevalent from our analysis. Additionally, identification is established through the dichotomy of in-group and out-group by alienating enemies of ISIS and refusing what they call grey-zone (i.e., if you are not with us you are against us). As demonstrated in Figure 6.7a, this property clearly differs from a normal discourse of news articles, mainly through the focus on the other (i.e., they) and the ratio use of 'you' to 'I'. Similarly, the focus on the language that supports sense of affiliation and care feeds into enhancing the identification level of the reader with the group. The manifestation of last resort mentality was established by utilising emotional language with special focus on

the anger negative emotion, which can be linked to preparing the reader into a desperation state and trying to justify the use of violence as a last resort.

6.5 Limitations and Implications

To understand ISIS extremist narrative and ideology, our method focused on analysing the textual content of Dabiq magazine. We used the text of the magazine as a proxy to identify underlying properties related to psychological and personality traits. These magazines are carefully crafted by ISIS and created with great effort and resources to reflect how they want to appeal to their supporters and the public. However, analysing the personality traits through the text is representative of the authors of the magazine articles (members of ISIS) rather than being representative of the whole group. The identified properties describing the extremist narrative is not meant to be generalised to other extremist groups. Further analysis is required of different extremist narratives to evaluate if the approach can be applied to other groups before attempting to generalise the findings. We will touch on this topic towards the end of this thesis. Finally, as the method applied is quantitative, future validation of the study results through a combination of qualitative analysis of the magazines and expert interviews would complement the findings.

Through this study, we addressed our third research question by identifying properties and signals that describe ISIS strategy in crafting their propaganda material. For example, the use of personal pronouns shows the emphasis on the dichotomy between the in-group and out-group. This is an important strategy to create fixation with group identity and increase the readers association with the group. Similarly, the main group drive being depicted through the propaganda is power, which reflects the image they try to communicate to the readers. Having such an understanding of the ideology and propaganda strategy adopted by ISIS is important to law enforcement and other stakeholders such as social media platforms to improve the narrative of counter-terrorism communication. In the following chapter, we will show how we utilise this understanding in building a system to detect the spread of extremist narrative online.

6.6 Summary

In this chapter, we analysed all issues of Dabiq magazine to reveal characteristics related to ISIS recruitment and propaganda strategy. We adopted computational methods to analyse the propaganda and reveal underlying signals related to the ideology and recruitment strategies of the group. We used data-mining techniques to identify key linguistic and psycholinguistic properties represented in ISIS propaganda. The analysis showed that the main drive for the ISIS group is power, which is expected as they aim to recruit and attract individuals. Additionally, the propaganda published in Dabiq illustrates traits such as confidence and formal logical thinking styles. Adopting this style of writing for the propaganda material may reflect why they have been successful in their recruitment campaigns. The dichotomy mentality of us-vs-them is reflected in the Dabiq articles, and this was exposed based on the distinction of using 3^{rd} person pronouns (they, she, he) and the use of 1^{st} person pronouns (I, we) in the articles. Moreover, the use of the 2^{nd} person pronoun (you) was higher compared with other pronouns. This reveals a strategy used by these magazines to put the focus and emphasis on the reader.

Informed with the results of this analysis, in the next chapter, we will integrate the different models and properties learned in a pipeline to detect pro-ISIS users in OSN. We hypothesise that supporters of ISIS may share similar characteristics to the ones that are conveyed in the ISIS propaganda.

7

Detection of Radicalisation on Microblogging platforms

7.1 Introduction

In this chapter, we aim to study the effects of using textual and psycholinguistic signals to detect online extremist content. These signals are developed based on the insights gathered from analysing propaganda material published by extremist groups (Chapter 6). Mainly, we will focus on detecting pro-ISIS supporters on Twitter, as they are one of the leading terrorist groups that have utilised Twitter to gain a wide audience, disseminate their propaganda, and recruit individuals.

In summary, the main contributions of this chapter are as follows:

- Design and develop a system to detect online radical content using properties inferred from radical magazines.
- Develop a general psychological profile for radicals inferred from the ISIS propaganda material. This profile acts as a baseline that is used to develop a radical distance measure.
- Evaluate our system on Twitter focusing on the detection of pro-ISIS tweets. We show how psychological and radical language models can be used to improve the detection mechanisms of radical content on Twitter.

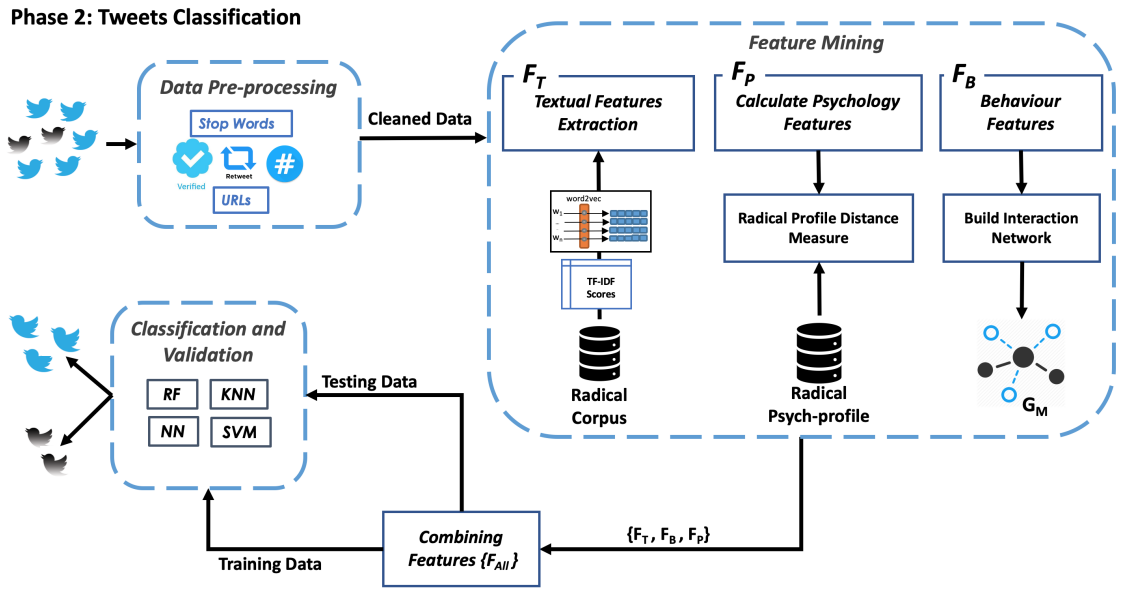


Figure 7.1: Approach Overview

7.2 System Design

We define our task as a binary classification problem where we classify tweets as ISIS supporters or not. An overview of the approach is illustrated in Figure 7.1 with three main steps: (1) Data pre-processing, (2) Feature engineering, (3) Classification and Validation. We utilise the models generated from the previous chapter to engineer textual and psycholinguistic features related to radical activities. In addition, we look at behavioural patterns that are relevant to OSNs and Twitter in particular to capture activities of pro-ISIS users.

We explore three categories of information to identify relevant features to detect radical content. Some features are user-based while others are message-based. The three categories are: (1) Radical language (Textual feature F_T), (2) Psychological signals (Psychological features F_P), and (3) Behavioural features (F_B). In the following, we detail each of these categories.

7.2.1 Radical Language

In Chapter 6, we explored the different topics, textual properties, and linguistic cues related to the propaganda produced by ISIS. We conjecture that these textual properties would be shared by people influenced by it.

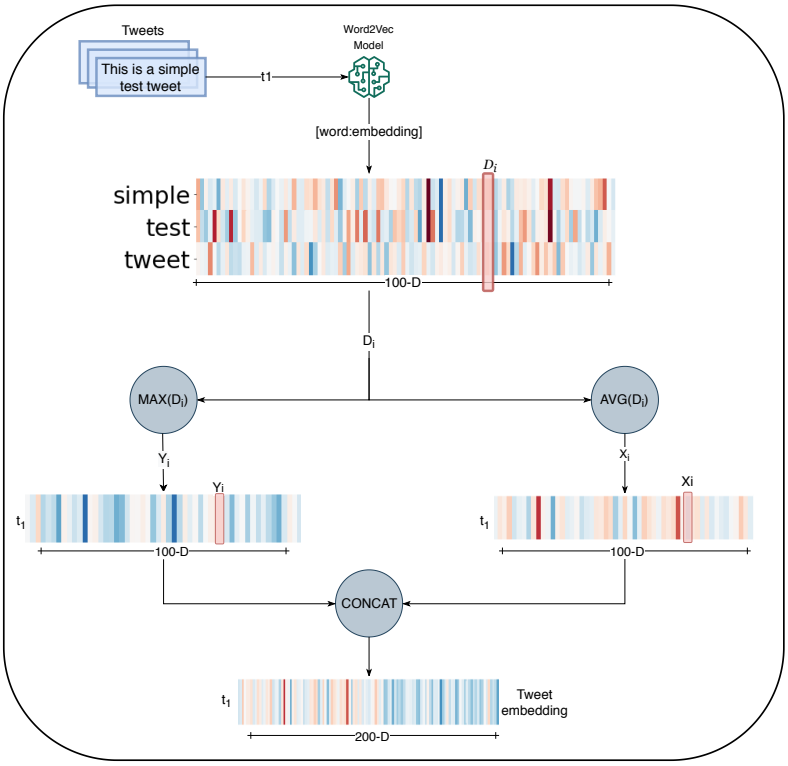


Figure 7.2: Tweet embedding

We select the top scoring TF-IDF grams from the propaganda corpus and use them as features for the language model. N-grams and words frequency have been used in the literature to classify similar problems, such as hate-speech and extremist text and have proven successful [201].

Moreover, in order to get the vector representation for words in tweets. We use the method described in Section 3.4.1 of Chapter 3 to seed the model using available pre-trained Twitter embeddings that has been trained on 1-billion tweets. Then, we update these embeddings based on the specialised corpus of Dabiq. Thus, we obtain the word embedding '100-dimension vector' representation for each word in the tweet. We then aggregate the vectors for a single tweet, and concatenate the maximum and average for each word vector dimension, such that any given tweet is now represented by a 200 dimension sized vector [202]. Figure 7.2 illustrates an example of the tweet embedding process.

Furthermore, since ISIS supporters typically advocate violent behaviour and

Table 7.1: Textual features groups (F_T)

Textual Feature groups	# Features
Top TF-IDF Scores	300 grams
Embeddings (Avg & Max)	200 dim
Offensive language (violent words, curse words, all-capital words)	3 groups
Total	**503**

tend to use offensive language, we use dictionaries of violent words[1] and curse words[2] to record the ratio of such words in the tweet. We also measure frequencies of words featuring all capital letters as they are usually used to convey a shouting behaviour. A summary of the textual-based features (F_T) is presented in Table 7.1.

7.2.2 Psycholinguistic Signals

Previous research found that online personality may be signalled by several linguistic cues [184]. Similarly, studies found that analysis of language used in social media communication can be utilised for gaining psychological insights of the authors [203].

Furthermore, previous research [204–206] looking at the motivating factors surrounding terrorism, radicalisation, and recruitment tactics found that terrorist groups tend to target vulnerable individuals who have feelings of desperation and displaced aggression. In particular, research into recruiting tactics of ISIS group found that they focus on harnessing the individual's need for significance. They seek out vulnerable people and provide them with constant attention [189]. Therefore, we explore the effect of psycholinguistic features on the detection of tweets supporting ISIS.

Using the psycholinguistic properties described in Chapter 6, we calculate the properties' scores and create a psycholinguistic profile (*psyc-profile*) for each Twitter user. Additionally, we calculate a distance measure (*psyc-profile-dist*) using Manhattan distance that measures the distance between each of the user's created psych-profiles and the average scores of the psycholinguistic properties

[1] `https://myvocabulary.com/word-list/terrorism-vocabulary`
[2] `https://www.cs.cmu.edu/~biglou/resources/bad-words.txt`

Table 7.2: Psycholinguistic feature groups (F_P)

Psycholinguistic Features	# Features
Summary Variables	4
OCEAN	5
Emotional Analysis	5
Drives	5
Personal Pronouns	3
Morality	5
Psyc. Profile Dist	1
Total	**28**

of the Dabiq magazine. A summary of the psycholinguistic-based features (F_P) is presented in Table 7.2.

7.2.3 Behavioural Signals

This category consists of measuring behavioural features to capture different properties related to the user and their behaviour. We focus on behaviours that can be used to describe some of the warning behaviours described in Chapter 6. For example, energy burst warning behaviour can be measured using levels of user's activity (i.e., frequency of tweets posted), #followers,#following, and #hashtags and URLs used. Similarly, identification can be measured based on following/followers' behaviour of known pro-ISIS accounts. Additionally, we use features to capture users' engagement and interactions with others through using hashtags, and engagement in discussions using *mention*, *reply*, and *retweet* actions.

Moreover, to capture the user's influence in their network, we construct an interaction graph (G_M) from our dataset, such that $G_M = (U, E)$, where U represents the users' nodes and E represents the set of edges. The graph G_M is a directed graph, where an edge e exists between two user nodes A and B, if user A mentions (replies to) user B. After constructing the graph, we measure the degree of influence each user has using different centrality measures, including degree centrality, betweenness centrality, and HITS-Hub. Such properties have been adopted in the research literature to study characteristics of cyber-criminal

Table 7.3: Behaviour feature groups (F_B)

Behaviour Features	# Features
Energy burst	5
Engagement	4
Influence levels	5
Total	**14**

networks and their behaviour [173], [207]. A summary of the behaviour-based features (F_B) is presented in Table 7.3.

7.3 Experimental Setup

We perform multiple experiments and compare the performance of our approach. In this section, we describe how each experiment is constructed, the data used, and the adopted evaluation metrics.

7.3.1 Dataset

We acquired a publicly available dataset of tweets posted by known pro-ISIS Twitter accounts that was published during the 2015 Paris attacks by Kaggle data science community [3]. The dataset consist of around 17,000 tweets posted by more than 100 users. These tweets were labelled as being pro-ISIS by looking at specific indicators, such as a set of keywords used (in the user's name, description, tweet text), their network of follower/following of other known radical accounts, and the sharing of images of the ISIS flag or some radical leaders. To validate that these accounts are indeed malicious, we checked the current status of the users' accounts in the dataset and found that most of them have been suspended by Twitter. This suggests that they did in fact demonstrate malicious behaviour that opposes the Twitter platform terms of use which caused them to be suspended. We filter out any tweets posted by existing active users and label this dataset as *known-bad*.

To model the normal behaviour, we collected a random sample of tweets from ten trending topics on Twitter using the Twitter streaming API. These topics were

[3] `www.kaggle.com/fifthtribe/how-isis-uses-twitter/data`

related to news events and on-going social events (e.g., sports, music). We filter out any topics and keywords that may be connected to extremist views using a dictionary of violent and abusive terms. This second dataset consists of around 8,000 tweets published by around 1,000 users. A random sample of 200 tweets was manually verified to ascertain it did not contain radical views. We label this dataset as our *random-good* data.

A third dataset is used which was acquired from Kaggle community[4]. This dataset is created to be a counterpoise to the pro-ISIS dataset (our known-bad) as it consist of tweets talking about topics concerning ISIS without being radical. It contains 122,000 tweets from around 95,000 users collected on two separate days. We verify that this dataset is indeed non radical by checking the status of users on Twitter and found that a subset (24,000 users) was suspended. We remove those from the dataset and only keep users that are still active on Twitter. This dataset is labelled as *counterpoise* data.

7.3.2 Data Preprocessing

We perform a series of preprocessing steps to clean the complete dataset and prepare it for feature extraction. These steps are:

1. We remove any duplicates from the dataset in order to reduce noise.

2. We filter out any non-English tweets.

3. We remove tweets that have been authored by verified users' accounts, as they are typically accounts associated with known public figures.

4. All stop words (e.g., and, or, the) and punctuation marks are removed from the text of the tweet.

5. If the tweet text contains a URL, we record the existence of the URL in a new attribute, *hasURL*, and then remove it from the tweet text.

[4] `https://www.kaggle.com/activegalaxy/isis-related-tweets/home`

6. If the tweet text contains emojis (e.g., :-), :), :P), we record the existence of the emoji in a new attribute, *hasEmj*, and then remove it from the tweet text.

7. If the tweet text contains any words with all capital characters, we record its existence in a new attribute, *allCaps*, and then normalise the text to lower-case and filter out any non-alphabetic characters.

8. We tokenize the cleansed tweet text into words, then we perform lemmatization, the process of reducing inflected words to their roots (lemma), and store the result in a vector.

7.3.3 Classifier Design

We conduct two experiments using the datasets described in Section 7.3.1. Our hypothesis is that supporters of groups such as ISIS may exhibit similar textual and psychological properties when communicating on social media, to the properties seen in the propaganda magazines. A tweet is considered radical if it promotes violence, racism, or supports violent behaviour.

In *exp 1* we use the first two datasets, i.e., the *known-bad* and the *random-good* datasets to classify tweets into radical and normal classes. For *exp 2* we examine if our classifier can also distinguish between tweets that are discussing similar topics (ISIS related) by using the *known-bad* and the *counterpoise* datasets.

The classification task is binomial (binary) classification where the output of the model predicts whether the input tweet is considered radical or normal. In order to handle the imbalanced class problem in the dataset, there are multiple techniques in the literature to tackle this issue. Over-sampling or under-sampling of the minority/majority classes are common techniques. Another technique that is more related to the classification algorithm is cost sensitive learning, which penalises the classification model for making a mistake on the minority class. This is achieved by applying a weighted cost on the misclassifying of the minority class [208]. We will use the last approach to avoid down-sampling of our dataset.

Table 7.4: Exp 1: Evaluation metrics across all feature groups.

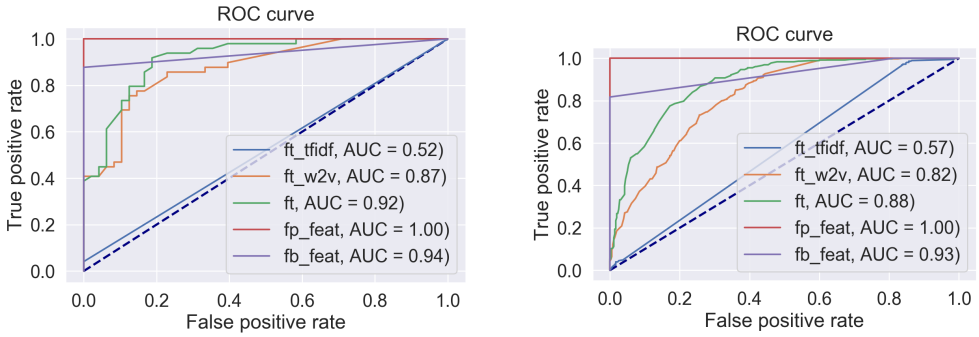
Features	AC	Precision	Recall	F-measure
$F_T(tf-idf)$	0.52	0.76	0.52	0.37
$F_T(w2v)$	0.81	0.82	0.81	0.81
F_T	0.84	0.84	0.84	0.84
F_B	0.94	0.95	0.94	0.94
F_P	1.0	1.0	1.0	1.0
F_{ALL}	1.0	1.0	1.0	1.0

Table 7.5: Exp 2: Evaluation metrics across all feature groups.

Features	AC	Precision	Recall	F-measure
$F_T(tf-idf)$	0.56	0.69	0.56	0.48
$F_T(w2v)$	0.73	0.73	0.73	0.73
F_T	0.80	0.80	0.80	0.80
F_B	0.91	0.92	0.91	0.91
F_P	1.0	1.0	1.0	1.0
F_{ALL}	1.0	1.0	1.0	1.0

For the purpose of this study, we experimented with multiple classification algorithms, including RF, NN, SVM, and KNN and found that RF and NN produced the best performance. We only report results obtained using RF model. We configured the model to use 100 estimators trees with a maximum depth of 50, and we selected gini impurity for the split criteria. We used the out-of-bag samples (oob) score to estimate the generalisation accuracy of the model. Additionally, since RF tends to be biased towards the majority class, we apply the cost sensitive learning method described earlier to make RF more suitable for imbalanced data [208].

We divide the dataset to training set (80%) and testing set (20%), where the testing set is held out for validation. We report validation results using different combinations of the features categories (i.e., F_T, F_B, F_P) and different evaluation metrics: accuracy, recall, precision, f-measure, and area under the ROC curve.



(a) ROC curve for Exp1 **(b)** ROC curve for Exp2

Figure 7.3: Results for Exp1 & Exp2

7.4 Results

7.4.1 Exp 1: Identifying Pro-ISIS Tweets

The classification results using the *known-bad* and *random-good* datasets are reported in Table 7.4. The table shows the average accuracy, precision, recall and f-measure scores obtained from each feature category (F_T, F_P, F_B) and their combination (F_{All}). We also compared the two textual models, and found that results obtained from using word embedding outperforms the use of n-grams TF-IDF scores. This confirms that contextual information is important in detecting radicalisation activities. Furthermore, our model performs best using the F_P features across all metrics. This means that the model is able to distinguish between both radical and non-radical with high confidence using only F_P.

7.4.2 Exp 2: Detecting Radicals from ISIS-related Tweets

In this experiment we test the performance of our classifier in distinguishing between radical and normal tweets that discusses ISIS related topics. Although this task is more challenging given the similarity of the topic discussed in the two classes, yet we found that the model still achieves high performance. Table 7.5 shows the different metrics obtained from each feature category. The F_T feature group obtained 80% accuracy, and 91%, 100% for F_B and F_P feature groups, respectively. The results are consistent with the ones obtained from the first experiment with the features

Table 7.6: Exp 3: Evaluation metrics across all feature groups

Features	AC	Precision	Recall	F-measure
$F_T(tf-idf)$	0.58	0.74	0.07	0.43
$F_T(w2v)$	0.76	0.72	0.73	0.76
F_T	0.78	0.78	0.69	0.77
F_B	0.88	0.94	0.78	0.88
F_P	1.0	1.0	0.99	1.0
F_{ALL}	1.0	1.0	1.0	1.0

from F_P group contributing to the high accuracy of the model. The area under the Receiver Operator Characteristic (ROC) curve, which measures accuracy based on *TP*, and *FP* rates for each experiment is shown in Fig 7.3.

7.4.3 Exp 3: Longitudinal Study for Pro-ISIS Users

One of the main limitations of the previous two experiments is the relatively small number of observations which are collected over a small time span. Thus, we performed a third experiment on a larger dataset collected over a long period of time.

We obtained a dataset of accounts associated with ISIS members and sympathisers, which has been manually validated via crowd-sourcing initiative called Lucky Troll Club. This data has been also verified by Twitter's anti-abuse team manually and all the accounts related to ISIS have been suspended due to the violation of Twitter's Terms of Service policy [209]. The data consist of around $25,000$ Twitter accounts. These users were responsible for posting over 1.9 million tweets from January 2014 to June 2015 [210, 211]. This data is labelled as Class_{Rad}. Furthermore, to model the normal behaviour Class_{Norm} we collected tweets from the top 20 most occurring hashtags in the radical tweets, that were published during the same period and their authors still have active accounts in Twitter.

We tested our model on the new dataset and report the results in Table 7.6. The results are comparable with the previous experiments. Using the feature group F_T we get f-measure of 77%, which is 3% lower from *Exp2*. The same trend is true for feature group F_B. However, for the F_P features, the results are consistent at 100% f-measure.

Table 7.7: Features Importance

Top 10	Features	Category
1	Radical psych-profile distance	F_P
2	Us-Them dichotomy	F_P
3	# of mentions a user make	F_B
4	User rank (hub and authority)	F_B
5	Sad emotion	F_P
6	Risk driver	F_P
7	All-caps count	F_T
8	URL count	F_T
9	Violent-word ratio	F_T
10	Hash count	F_T

7.5 Feature Importance

We investigate which features contribute most to the classification task to distinguish between radical and non-radical tweets. We use the mean decrease impurity method of random forests [212] to identify the most important features in each feature category. The ten most important features are shown in Table 7.7. We found that the most important feature for distinguishing radical tweets is the psychological feature distance measure. This measures how similar the twitter user is to the average psychological profile calculated from the propaganda magazine articles. Following this is the Us-them dichotomy which looks at the total number of pronouns used (I,they, we, you). This finding is in-line with the tactics reported in the radicalisation literature with regards to emphasising the separation between the radical group and the world.

Moreover, among the top contributing features are behavioural features related to the number of mentions a single user makes, and their HITS hub and authority rank among their interaction network. This relates to how active the user is in interacting with other users and how much attention they receive from their community. This links to the objectives of those radical users in spreading their ideologies and reaching out to potential like-minded people. As for the F_T category we found that the use of word2vec embedding improves the performance in comparison with using the TF-IDF features. Additionally, all bi-grams and tri-grams features did not

contribute much to the classification, only uni-grams did. This can be related to the differences in the writing styles when constructing sentences and phrases in articles and in the social media context, particularly given the limitation of the number of words allowed by the Twitter platform. Additionally, the *violent word ratio*, *long-Words*, and *allCaps* features are among the top contributing features from this category. This finding correlates to high extent with observations from the literature dealing with similar problems, where the use of dictionaries of violent words aids with the prediction of violent extremist narrative.

7.6 Comparison with Related Work

There is a range of work that focuses on detecting online radicalisation activities and analysing such behaviour. In Table 7.8 we compare the related work that focuses on detecting online radicalisation activities on Twitter to our proposed approach using a set of criteria. The comparison criteria looks at the size of dataset used to evaluate the approach, the generalisability of the used features and how much it is influenced by the used dataset. This is important as it can influence whether or not the approach can be successful when applied to other datasets. Additionally, we compare the types of features used, mainly sentiment, textual, behavioural, and psychological features. Finally, the best performance measure obtained is reported. From Table 7.8, we can see that the majority of the literature combines different types of features to detect online radicalisation. This is important as each feature type provides an insight and an additional signal that aids in distinguishing this behaviour. The main limitation of previous work is the generalisability aspect, as few articles adopted features that are independent of the testing dataset, or performed an extensive evaluation of multiple datasets.

Moreover, some of the features proposed in the literature may seem similar to our work, however, there are key differences that we would like to highlight. For example, in the work of Ashcroft et al. [28] (achieved 100% accuracy), they used stylometric features that contain a set of words that are most frequent in the dataset. This raises the question of the applicability of the approach to different datasets.

Table 7.8: Comparison between related work and our approach. [FT] results using only textual features. [FP] results using only psychological features. * considers only the five emotions as features.

Authors	Dataset size	Generalizability	Sentiment	Textual	Behavioural	Psycholinguistics	Performance
Kaati et al. [109]	81 K	✓	✓	✓			99%
Ashcroft et al. [28]	7500		✓	✓			100%
Lara-Cabrera et al. [213]	266 K	✓		✓		✓	NA
Berger et al. [214]	300 K				✓		NA
Devyatkin et al. [201]	493			✓		✓	[FT]: 92%,[FP]: 76%
Agarwal et al. [215]	1 M		✓	✓			97%
Saif et al. [216]	2 M	✓	✓	✓			92%
Smedt et al. [217]	90 K			✓			82 %
Hartung et al. [218]	45 K			✓	✓	✓*	93%
Benigni et al. [219]	22 K				✓		96%
Watanabe et al. [220]	25 K		✓	✓			87%
Our Approach	**200 K**	✓	✓	✓	✓	✓	**100%**

Additionally, they used sentiment features to capture the polarity of the text from being very negative to very positive. This method captures part of the psychology of the writer (having negative, or positive emotion), however it reduces it to one variable which does not capture the context of that emotion. Understanding the context of the sentiment is important to detect radicalisation because it paints a picture regarding what the positive/negative sentiment is about. For example, a positive sentiment can be used to express radical ideas.

7.7 Limitations and Implications

In this section, we highlight some of the identified research limitations and the potential implications of the study.

The dataset used in all three experiments were validated through crowd-sourcing techniques. To guarantee the correctness of the validation, we performed a second level of validation by checking the current status of the accounts on Twitter (active/suspended). Thus, our ground-truth is reliant on what Twitter's anti-abuse team identify as abusive. Moreover, online extremism behaviour is constantly evolving. Some radical extremist supporters are changing their tactics to avoid detection by using different terminology and more subtle references to violence. Since our textual and psycholinguistic features depend on the keywords used and emotions expressed, the system might not be as successful in detecting those that try to masquerade their support. As such, future work should investigate if the

model is resilient to different evasion techniques that radical supporters may adopt. Furthermore, the textual and psycholinguistic features are language-dependant. Currently, we focus only on the English language since the embedding model is trained on English text and the LIWC dictionaries are for the English language. A possible future study would be to expand and test the approach on other languages, this can be achieved by adapting the embedding models and the LIWC dictionaries based on the target language.

Our results show that using linguistic and psycholinguistic characteristics we can accurately identify radical content. The radical psych-profile distance measure has been found to have the most contributing effect to the detection task. Similarly, using word embeddings trained on official published extremist propaganda we can detect radical tweets with 76% accuracy. This result agrees with our earlier hypothesis that suggests that supporters of extremist groups are influenced by the published extremist propaganda and share psychological and linguistic properties. Our system can aid law enforcement and OSN platforms to better address the threat of online radical content and help solve a challenging real-world problem.

7.8 Summary

In this chapter, we identified different signals that can be utilised to detect evidence of online radicalisation. We derived linguistic and psychological properties from propaganda published by ISIS for recruitment purposes. We utilised these properties to detect pro-ISIS tweets that are influenced by their ideology. Unlike previous efforts, these properties do not only focus on lexical keyword analysis of the messages but also add a contextual and psychological dimension. We validated our approach using different experiments and the results show that this method is robust across multiple datasets. The psychological features identified had an important role in improving the overall detection accuracy of the model.

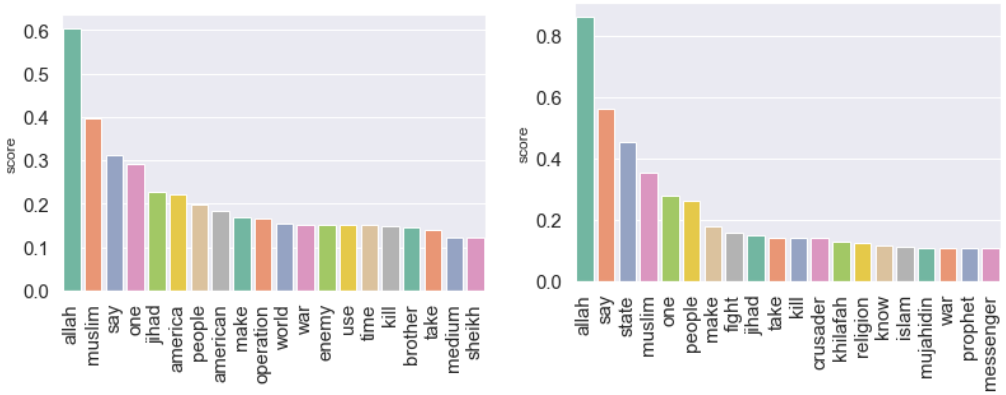
8

Discussion on Other Extremist Groups

8.1 Introduction

As we have seen in the previous chapter, the psycholinguistic profile was identified as a powerful contributor to the successful detection of supporters for extremist groups. Therefore, in this chapter, we wanted to investigate if different extremist groups who originate from both similar and different ideologies share some of these properties. We compare the linguistic properties of online propaganda material published by different extremist groups against the properties of the Dabiq magazine previously described in Chapter 6. Mainly, we study Al-Qaeda and Alternative Right (Alt-Right) groups and compare their linguistic properties against those of Dabiq and the control group. The aim of this chapter is to first highlight the main similarities and differences between these groups in crafting their propaganda material. Second, if similarities are identified, does that mean that our proposed detection approach outlined in Chapter 7 would work on these different groups.

In our analysis we followed a similar methodology to the one described in Chapter 6 to extract textual and psycholinguistic properties from the extremist propaganda. We started by calculating the top-scoring grams based on TF-IDF weights. Then, using LIWC and morality foundation dictionaries, we calculated



(a) Grams for Inspire

(b) Grams for Dabiq

Figure 8.1: Top TF-IDF grams

the frequencies of words that map to the different categories. Finally, we compared the scores against Dabiq and the control group.

8.2 Al-Qaeda: Inspire Magazine

Inspire magazine is an online English magazine published by Al-Qaeda group in the Arabian Peninsula (AQAP). Similar to Dabiq, it aims to radicalise individuals, promote and justify the activities of the group, and improve the group's brand and image. We collect all 17 issues of Inspire that were published between 2010 and 2016. Similar to Dabiq, each issue typically consists of a collection of articles, editorial snippets, and visual images. We focus only on the text and discard any graphical content.

Textual Analysis — Figure 8.1a presents the top-scoring grams ordered by the TF-IDF values. As both groups stem from a similar ideology, we find several similarities related to the general articles' focus. Along with the religion theme that is apparent through the top-scoring grams (e.g., Allah, Muslim), there seems to be a big focus on the United States through the grams America (ranked 6^{th}) and American (ranked 8^{th}). This is linked to the history of AL-Qaeda with the US and how their leader, Osama bin Laden, declared war against the United States [221]. Furthermore, terms such as Jihad (ranked 5^{th}) have a strong presence in the magazine, followed by terms representing violence such as war and kill.

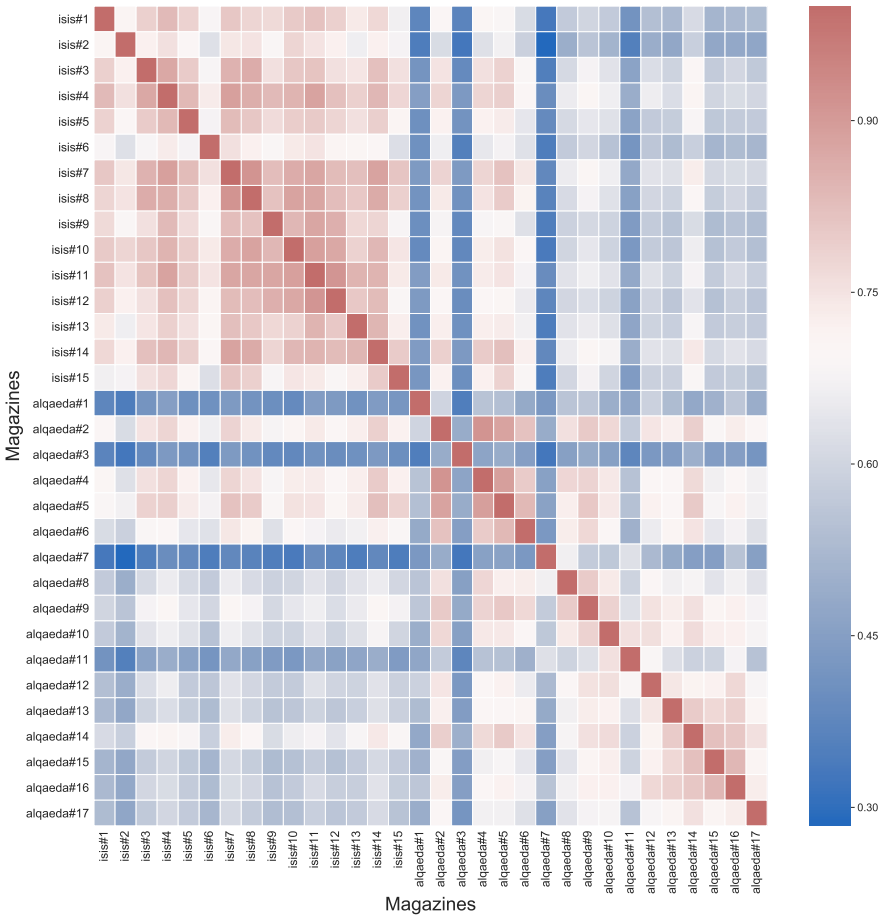


Figure 8.2: Heat-map representing the cosine similarity values between articles published by ISIS and Al-Qaeda.

We can see that there are a lot of similarities between ISIS and AL-Qaeda concerning the topics and words they use, however, there are also some subtle differences. For example, in Inspire the group's enemy is apparent, while in Dabiq, we find a heavier Islamic tone and more focus on the concept of Khilafah that does not seem to be as apparent in the Inspire articles. Similarly, topics of religiously motivated wars (e.g., crusaders) are more prevalent in Dabiq than Inspire.

Furthermore, to measure the similarity between articles appearing in ISIS publications and Al-Qaeda ones, we calculate the cosine similarity between their vector representations. Cosine similarity has been widely used in the literature to compare text documents in the vector space to determine the degree to which they are similar. The values range between $[0, 1]$, where 0 means that the documents have no similar words, and 1 means that they are identical. Figure 8.2 shows a

heat-map of the cosine similarity measure of magazines from both groups. We can see from the figure that ISIS articles have high similarities among themselves (cluster in the upper left corner) and Al-Qaeda articles have more similarities among themselves (cluster in the lower right corner). In summary, although the articles from both groups share some similar themes/topics, they do differ in the choice of words used to express these topics.

Psycholinguistic Analysis — We analyse the Inspire articles to identify the psycholinguistic profile for the magazine. We calculate the same properties described in Chapter 6 for each Inspire magazine issue, and report the mean and standard deviation (std) in Appendix A Table A.1. We then compare it to Dabiq's profile and use the Mann–Whitney U test to identify if the difference between the groups' styles is significant or due to chance. Figure 8.3 shows a comparison between Dabiq and Inspire with regards to some of the key properties.

Overall, both groups share similar trends across most of the properties. For instance, Dabiq shows more analytic and clout signs for summary variables, while Inspire exhibits higher levels of positive and negative emotions for emotion variables. The main drive for both groups is power followed by affiliation. We see different behaviour in the use of personal pronouns (Figure 8.3d). While Dabiq has a clear focus on the "other" (i.e., out-group focus) through the emphasis on the use of 3^{rd} person pronouns, Inspire focuses more on 1^{st} person plural pronouns which emphasise the in-group identity. Finally, looking at the morality foundations, both groups share a similar moral foundation that focuses more on care and authority, with Dabiq devotes significantly more attention to authority.

8.3 Alt-Right: American Renaissance Magazine

American Renaissance (AR) magazine is a monthly white supremacist online publication by the New Century Foundation [222]. The foundation describes itself as a "race-realist", and has been described as alternative right (Alt-Right) by The Guardian [223]. We collect 22 issues of AR [1] that were published between

[1] https://www.amren.com/archives/back-issues

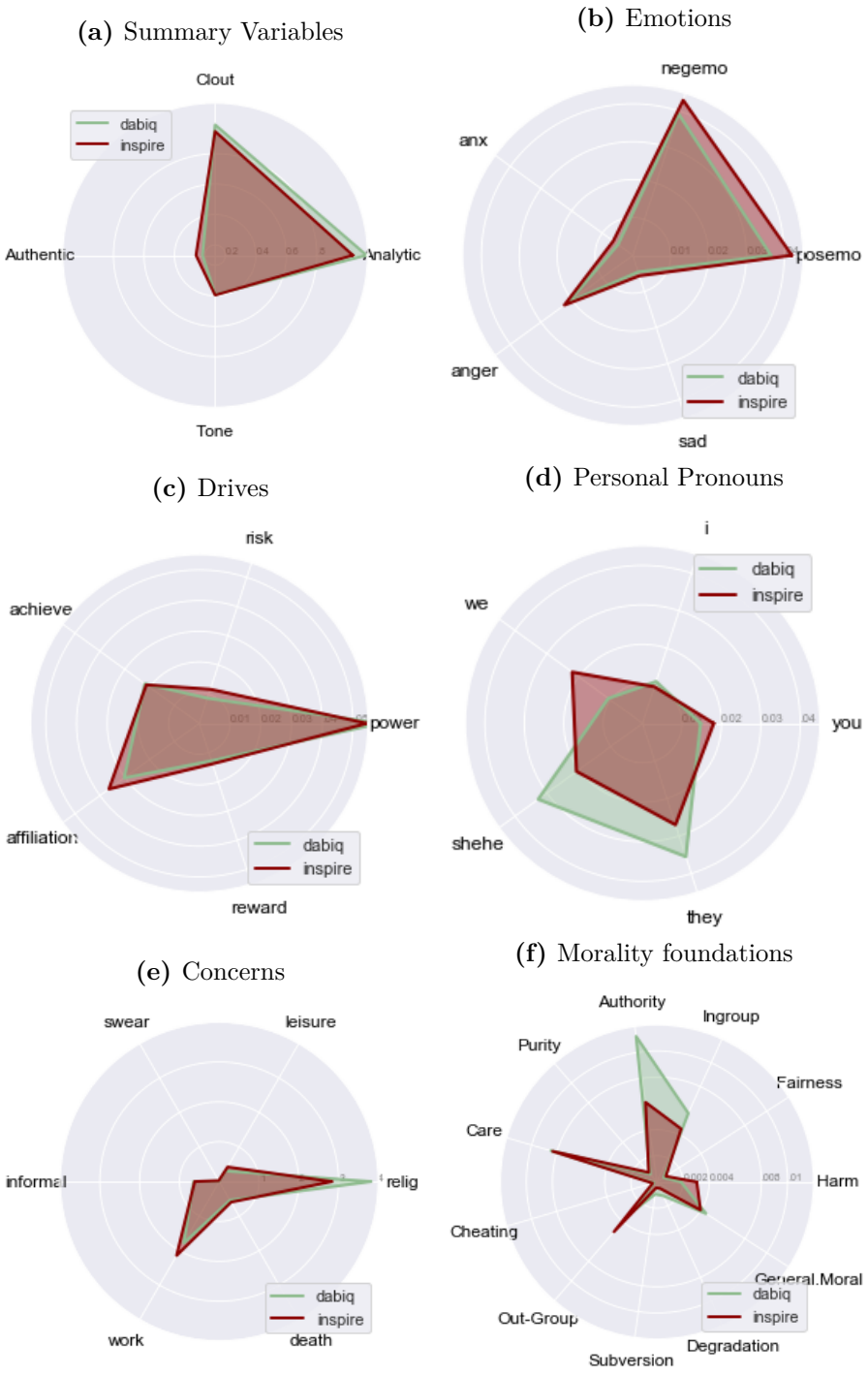


Figure 8.3: Psycholinguistic analysis of Inspire & Dabiq

2009 and 2012. After 2012, the group shifted to publishing online short articles, rather than combined magazines. Similar to Dabiq, each issue typically consists of a collection of articles, editorial snippets, and visual images. We extract the text and discard any graphical content.

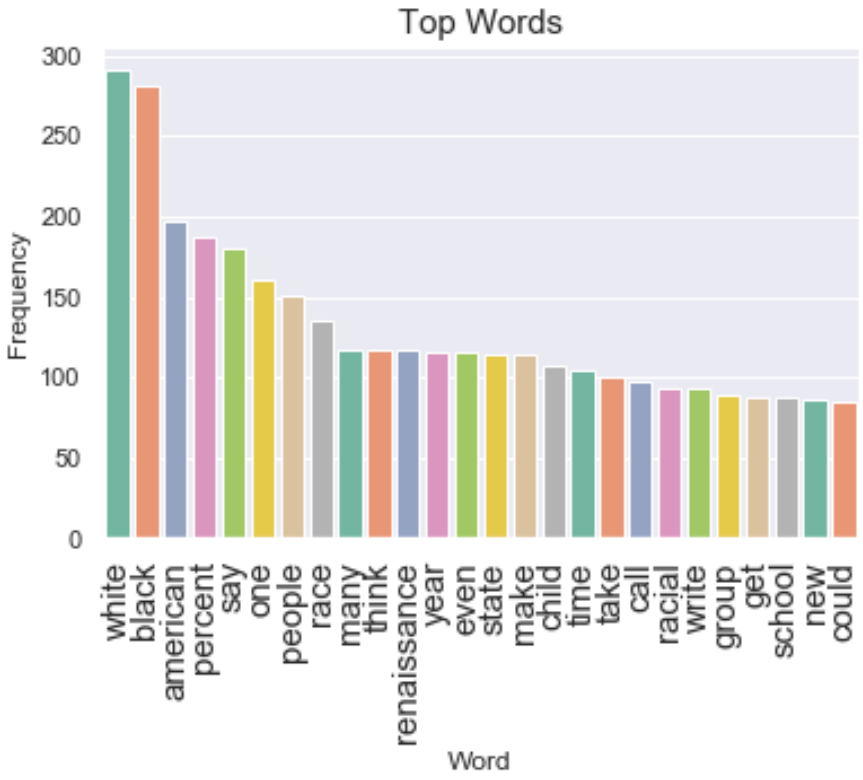


Figure 8.4: Top TF-IDF grams for AR

Textual Analysis — Figure 8.4 presents the top TF-IDF scoring grams for the Alt-Right propaganda. In line with the group focus on race and particularly the supremacy of the white race, the top-ranked gram is indeed *white*. Similarly, the group's ideology places emphasis on the alleged inferiority of other races and particularly people of colour; this explains the high rank of *black* as the 2^{nd} most used gram. Furthermore, the *American* gram (ranked 3^{rd}) is highly relevant since the group is mainly based in the United States and is targeting an American audience. In comparison with the top-ranked grams in Dabiq and Inspire, we notice that AR does not place explicit emphasis on violence, which is clearly due to the absence of violence-related grams.

Psycholinguistic Analysis — We analyse the articles in AR magazine to identify the psycholinguistic profile for the magazine. We calculate the same properties described in Chapter 6 for each AR magazine issue, and report the mean and standard deviation (std) in Appendix A Table A.1. We then compare it

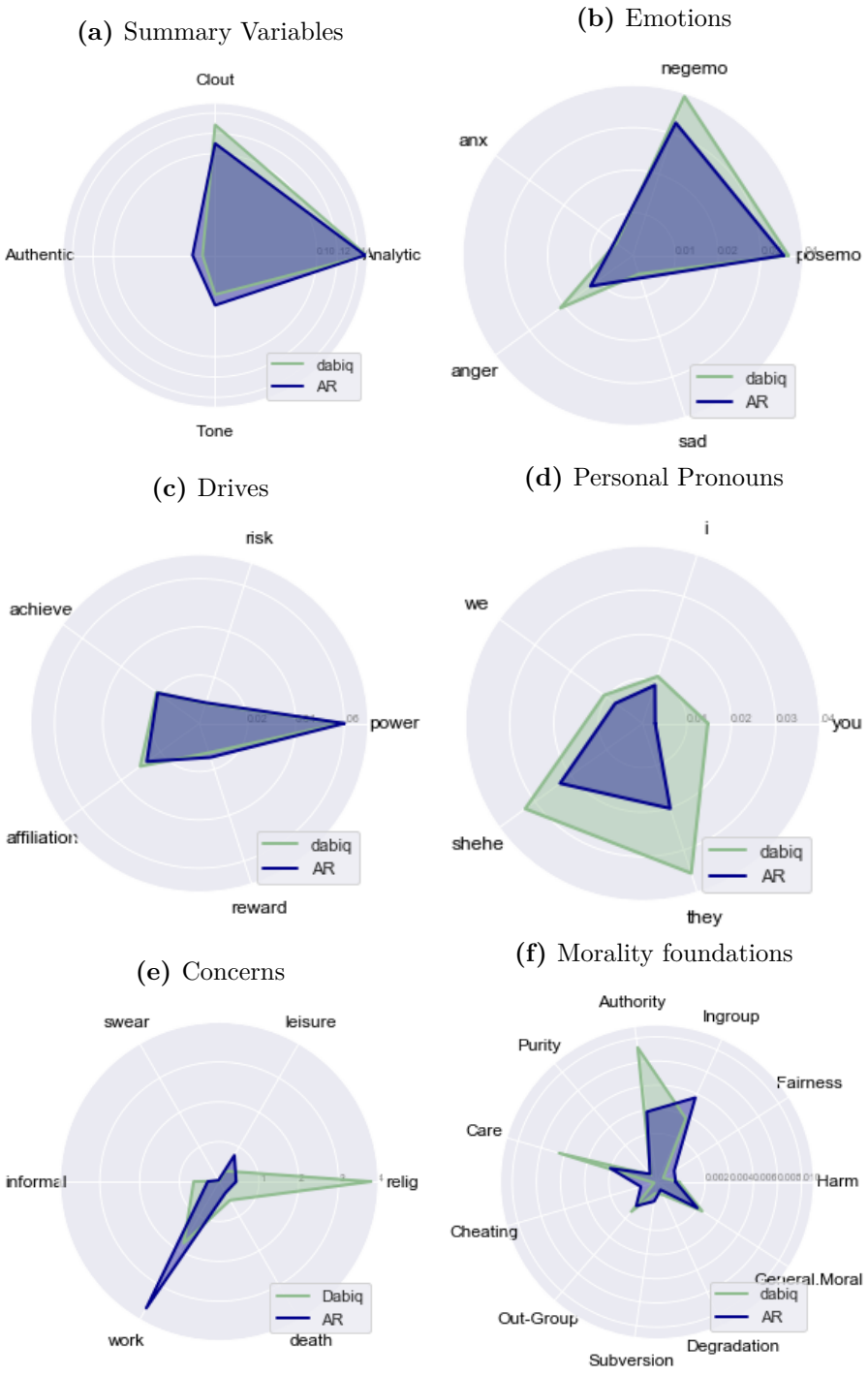


Figure 8.5: Psycholinguistic analysis of AR & Dabiq

to Dabiq's profile and use the Mann–Whitney U test to identify if the difference between the groups' styles is significant or due to chance. Figure 8.5 shows a comparison between Dabiq and AR with regards to some of the key properties. A complete comparison of the two groups' properties is presented in Appendix A.

Both groups share similar trends for the summary variables, emotions, and drives. Dabiq exhibits higher negative emotion, mainly in the anger emotion and AR exhibits a slightly higher sadness emotion. Additionally, both groups share similar levels for the main drive, power.

Furthermore, when comparing the use of pronouns, we find that both groups share the tendency to focus more on 3^{rd} person pronouns, which is associated with distinguishing between in-group/out-group identities. However, AR utilises 1^{st} person pronouns more than 2^{nd} person pronouns, unlike Dabiq which puts the focus more on the reader.

Comparing the groups' attitudes towards different concerns, as expected, we see clear emphases on religion for Dabiq which is significantly higher than AR, while AR focuses more on work-related issues. Finally, looking at the morality foundation, the AR driving moral foundation is the in-group (loyalty) foundation followed by authority. On the other hand, Dabiq is more motivated by authority and care.

8.4 Discussion

Online radical propaganda has been used as a way to promote extremist ideology and gather support for their causes. Studies in the literature suggest that individuals who are exposed to online extremist propaganda are influenced by it and are more likely to use violence in support of their cause [224]. As a result, it is crucial to understand the discourse of such propaganda and how it is created and crafted to deliver its message.

In this chapter, we compared different propaganda material published by three different radical groups; two stem from religious ideology and one stem from racial ideology. We found that there are some common properties and trends that appear in the propaganda material of all three groups despite their ideology, and some properties that are a group-focused. For instance, conveying emotion is a common factor across all groups including both positive and negative emotions, with the anger emotion being the main driving emotion for radical propaganda.

Furthermore, regardless of groups' ideologies or causes, the main driving motives being conveyed in the propaganda text are power followed by affiliation. Such a

Table 8.1: Properties that rejected the H_0 (i.e., statistically significant)

Comparisons (p<= 0.05)	Dabiq v.s. News	Inspire v.s. News	AR v.s. News
Features	clout, authentic, I, you, shehe, they, affect, negative emotion , anger, drives, affiliation, achieve, reward, work, leisure, religion, death, informal, care, out-group, authority, subversion, degradation, morality general	analytic, clout, authentic, I, we, you, they, affect, positive emotion, negative emotion, anxiety, anger, sad, affiliation, work, leisure, religion, death, informal, swear, harm, care, out-group, purity, degradation, morality general	authentic, you, they, drives, achieve, religion, death, swear, fairness, cheating, out-group, subversion, degradation, morality general

projection of power and a sense of affiliation may be the reason why these groups have been successful in gathering supporters and influencing people. Also, strong group identification can be inferred from the text of all three groups. This is achieved through the emphasis being focused on in-group and out-group mentality. In many cases, radicalisation intensifies when people feel strong links to the group's identity combined with feelings of being at risk and feeling threatened by the out-groups [225].

As for the main driving morality foundation for radical groups, we found that each group varies slightly in terms of the main morality foundation, although they are mainly focused around three morality foundations: authority, in-group loyalty, and care. For example, while ISIS tends to show more authority in its propaganda followed by care and in-group loyalty, AL-Qaeda focuses more on care. On the other hand, Alt-Right group is concerned the most with in-group loyalty followed by authority and care.

To identify if mean differences between groups' properties are statistically significant or merely due to chance, we performed an unpaired two-sample comparison of each of the radical groups against a set of news articles "control group" using Mann–Whitney U test at the 5% level of significance ($\alpha = 0.05$). The null-hypothesis (H_0) was formulated such that there was no difference among the groups (the radical propaganda and the news groups). Alternatively, rejecting the null hypothesis means that the two groups are different. In Table 8.1 we list the variables that resulted in rejecting the null-hypothesis ($p <= 0.05$) for each comparison group. In other words, these properties are able to distinguish between these groups. Furthermore, we performed a Kruskal-Wallis test on all three radical groups and report the H-scores and p-values in Appendix A Table A.2.

8.5 Summary

This chapter presents a comparison between properties of propaganda material published by different radical groups. Mainly, we studied properties of Inspire (Al-Qaeda magazine) and American Renaissance (Alt-Right magazine) and compared their linguistic and psycholinguistic properties against those of Dabiq. We found a number of similarities in how these group construct their propaganda material. Similarly, we identified a number of properties that can be used as distinguishing features between these groups. We tested the properties against a control group represented by randomly selected set of news articles, and highlighted the main properties that were indeed significantly different from the control group. Finally, these findings could suggest that there is a potential for a wider application of our proposed approach to the detection of online radicalisation beyond the group of ISIS.

9 Conclusions

This thesis aimed to examine and address challenges faced by cybercrime investigators and to provide novel techniques for analysing and detecting such crimes. In this final chapter, we conclude our research. First, we summarise our work and demonstrate how it addresses our research questions. We then critically reflect on the main research limitations, before we discuss future research directions. Finally, we conclude with final remarks.

9.1 Summary

The threat posed by cybercriminals is constantly increasing as they target new attack vectors and adopt different strategies. To develop an effective framework to respond to cybercrimes, it is essential to approach it with a holistic view. In this thesis, we addressed this by focusing on three main pillars: People, Process, and Technology.

People — We started this research by approaching the people who work on investigating cybercrimes. This allowed us to better understand their environment, background, and capabilities. Many practitioners who work on investigating cybercrimes come with a strong background in investigative tactics and have long expertise in investigating traditional crimes, but may be limited in their technical background. In our study outlined in Chapter 4, we formulated a set of challenges

and recommendations that relate to building the capacity and technical skills (especially at the local-force units) of people working in investigating cybercrimes. Moreover, we identified a set of challenges related to the investigation process. For instance, we found that different cybercrime investigation units operate their own set of processes and workflows. Similarly, a proper process that facilitates the sharing of information across different organizations (e.g., government-government, government-private, and private-private) seemed lacking.

Process — After understanding the process being followed by professionals during investigations, and the main needs and challenges they face, we outline in Chapter 5 *CCINT*, a framework for cybercrime investigations. CCINT constitutes a conceptual model inspired by how professionals described the investigation process and an operational framework that facilitates different analytical capabilities. We provided an example of how to apply the model to a real-world case study, such that we aid investigators to organise the investigation process and inform what type of resources, collaboration, and intelligence are needed when investigating an incident.

Technology — The complexity of cybercrimes is constantly increasing with new methods, tools, and attack vectors being adopted. Similarly, such complexity generates large amounts of data and digital footprints that require advanced tools to handle. As a result, making use of new methods and technological advancements is critical to be able to analyse the data and defend against such crimes. By understanding users' capabilities and needs, we build tools that can seamlessly fit into their workflow and can assist them in conducting their investigations.

Through the three-pillar approach, we answered the first two research questions: *RQ1. What are the main challenges currently being faced by cybercrime investigators?* In Chapter 4, we outlined a set of socio-technical challenges and recommendations. The wide scope of these challenges and recommendations, which can potentially fill many theses, made us decide to focus our research to address the process and technology-related aspects of them. This led us to design the CCINT framework, and answer the second research questions: *RQ2. What are*

the requirements for cybercrime intelligence systems that would be appropriate and helpful for investigators?.

Furthermore, recognising the danger of the increasing spread of extreme online speech "extremist content", which has been described by the UK government report on online harm [37] as still lacking a proper and effective response, we decided to further focus the research into this topic of cybercrime. Tackling such an issue requires coordination between government, academia, and private sectors to formulate policies and supporting technologies that can help limit the spread of extremist narrative online while preserving the societal rights of freedom of speech. As such, we approach this by exploring the properties associated with extremist discourse and focusing on analysing online propaganda published by them. In Chapter 6, we answer the 3^{rd} research question: *RQ3. Can we identify hidden properties and signals that describe extremist strategies in crafting their online propaganda?* This was achieved by utilising computational methods to model the language used in these magazines and extract psycho-linguistic properties that describe the overall personality, morality, and psychology patterns being conveyed in these magazines.

After identifying these properties that stem from the extremist narrative, we test the hypothesis that supporters of extremist groups who are influenced by the ideology will share similar linguistic and psycho-linguistic properties. Therefore, we perform an experiment on Twitter to detect supporters of ISIS, which allow us to answer the 4^{th} research question: *RQ4. Can we detect extremist narrative in social media and therefore help mitigate it's spread?* We were able to achieve this with high accuracy using random forest classifiers trained on different groups of features. We found that using word embeddings, trained on extremist propaganda, to represent tweets demonstrate significant improvement over the use of n-grams and TF-IDF weights. Additionally, throughout the three experiments we performed with different datasets, the psycho-linguistic features contributed the most to the successful detection of pro-ISIS users on Twitter. Using a crafted measure that calculates the distance of the user's psych-profile from a pre-calculated profile of the extremist magazines.

Finally, due to the successful nature of the psych-profile in detecting supporters of extremist groups, we wanted to investigate if different extremist groups who originate from similar and different ideologies share some of these properties. Thus, addressing our final question: *RQ5. To what extent are extremist narrative properties shared across multiple extremist groups?* The results showed that groups that share the same ideology have more properties in common, which is expected. However, we also found that some properties were global across different extremist ideologies. For instance, the us-vs-them dichotomy was affirmed in all extremist propaganda. Similarly, they shared the emphasis on affiliation and achievement drives in their propaganda. As for morality foundations, authority, care, and in-group loyalty were the main moral foundations conveyed in all extremist propaganda.

In summary, the level of ease by which online extremist material is being widely distributed and shared represents a considerable challenge to online platforms and law enforcement. As extremists moved towards exploiting small online platforms, which lack the resources and capabilities to monitor its data, to widely distribute their propaganda they create difficulties for law enforcement in disrupting the distribution. Therefore, we hope that our contribution can enable practitioners in the field to properly disrupt such activities and minimise the spread of harmful content.

9.2 CCINT Framework for Online Radicalisation

The threat of cybercrime is expanding as criminals are migrating to the cyber world to benefit from a low-risk high return environment. Similarly, the convergence of cyber and terrorism introduces new threats with terrorist groups exploiting emerging online platforms to spread their propaganda and radicalise individuals. As these threats continue to evolve, it is important to provide appropriate support for online platforms and law enforcement to effectively respond to these threats.

In this thesis, we achieve this by providing a common framework for practitioners to follow to aid them in extracting effective intelligence and better understand, analyse, and detect such crimes. Most previous frameworks in the literature focus on describing cybercrimes from a single viewpoint, such as the victim, offender, or

the incident. The CCINT framework, on the other hand, aims to provide a holistic view and understanding of the crime by describing it from different factors, and provide practitioners with a combination of analytical techniques and supporting tools that they need to analyse such crimes.

As we apply the framework to the problem of online radicalisation. We define the three elements as follows:

Offenders— are an organised extremist group who are ideologically motivated to promote and spread their propaganda and influence people. Through their propaganda, they aim to recruit supporters and sympathisers, celebrate their victories, and project power and authority. For example, as the military fight against ISIS was successful and resulted in them losing territory, however, this defeat made the digital world an even more important platform for them to promote their propaganda, inspire new attacks, and celebrate their ideology.

Incident — is the spreading of extremist material online to recruit and promote extremist activities. Based on the matrix of cybercrimes presented in Table 2.1, the incident is described as a cyber-dependent, crime in the machine, since it is content-related, and the opportunity here is to spread violence and online grooming. Thus, the scale and the impact of the incident is big since it spreads online, widely targets victims, and can lead to offline real-world attacks.

Victims — The victims are individuals who might get influenced by such propaganda and slowly groomed to join an extremist group. The type of harm caused is psychological. Similarly, when the propaganda entices an attack or an act of violence the general public would be considered as victims and additional types of harm such as economic and loss of digital and physical assets may be incurred.

Since the M.O depends on the spread of propaganda material to recruit individuals and promote the extremist ideology, we utilised them as data sources to study their properties. It is important to note that radicalisation is a process that does not happen overnight. It occurs over time and is linked to multiple online and offline factors. Exposure and interacting with extremist propaganda is one factor that can lead to radicalisation. Through our analysis of extremist propaganda, we

were able to identify factors and signals that can explain their recruitment strategy and why they have been successful in radicalising people. These factors were shared among different radical groups regardless of their ideology.

Several theories in the radicalisation and social psychology literature discussed how the radicalisation process occurs and how to influence group behaviour. Our findings support these theories as it showed how a dichotomy mentality of us-vs-them is apparent in the construction of the propaganda material through the analysis of the use of pronouns. Additionally, the strong manifestation of power and authority was also apparent in the construction of the radical propaganda material. Such focus on strong associations with the in-group ideology and alienating out-groups, radicals can establish strong identification with the group ideology and its superiority. This is in line with the social identity theory that explains how social identity is formulated through positive in-group emotion and affiliation, and hostility towards out-group. Similarly, this dichotomy help groups to create a collective identity and thus influence the behaviour of individuals to commit actions that are of benefit to the group even if they are against the individual's interest. Furthermore, we showed how emotions play a strong role in radical discourse. Utilising emotional language, both positive and negative, allows them to strengthen the individual's sense of affiliation and motivates the collective group action. For example, the anger emotion which is one of the most prevalent emotions in radical propaganda is known in the literature to be a group mobilising factor.

The successful utilisation of these properties in detecting supporters of extremists in social media demonstrates the importance of understanding these factors in detecting radicalisation. Especially when combined with linguistic cues, we can capture extreme ideological discourse that calls for violence.

9.3 Research Limitations

There are a few limitations to our research which should be noted.

In our study that focuses on understanding the investigators' needs and challenges, we opted to collect our data by conducting semi-structured interviews. While

this approach allowed us to gather detailed data from participants, which would not be possible using survey methods, it limited us in the number of participants that we were able to reach. Especially given that the targeted community is considered difficult-to-reach for academics. Additionally, although each interview was guided by the same set of predefined questions, there are unavoidable variations in the level of details gathered from each participant depending on their time availability and their role in investigating cybercrimes. Furthermore, to understand the problem in-depth and capture the views of participants with their own words, we chose to conduct face-to-face interviews which gave us an understanding of how participants perceived their role and the challenges they face. However, relying on methods such as participants' observations as a tool for collecting data would give a better understanding of the people, processes, and cultures of the organisations. Yet, there are many difficulties inherent in observing this kind of activity directly. Therefore, we rely on participants' reports via interviews to collect our data. Similarly, the sampling of participants was chosen from different agencies (government and private) who deal with cybercrime investigations. From the government agencies, we were able to gather the perspective of local cybercrime task forces and regional cybercrime units. However, we could not get access to participants at a national cybercrime agency level. In future work, it would be interesting to include the views and perspectives of experts from national cybercrime agencies, as well as expanding our sample size to allow even more insight into the problems faced.

With regard to researching the area of detection of online radicalisation, access to quality data is a major challenge. The collection of a large dataset that is representative of the community being studied and is not biased is difficult. Using Twitter as a data source involves potential sampling biases enforced by Twitter Sampling API that delivers a random 1% sample of all tweets. Another limitation is related to obtaining "ground-truth". Although the data used in this research has been validated by different means, including crowd-sourcing manual validation. We also validated the data by querying Twitter's API to verify if all users that are labelled as radical in our dataset, currently have suspended accounts. Similarly, we

query all non-radical users accounts to verify that they still have active accounts. Since the data is more than 3-years old, our assumption relies on the ability of Twitter's anti-abuse team to correctly verify and suspend abusive accounts.

9.4 Future Research Directions

The work presented in this thesis offers several potential directions for future work.

We have explored the behaviour of pro-ISIS users on Twitter. One possible extension of this work is to study their behaviour across different online platforms. Since mainstream social media platforms have become more focused on minimising the spread of extremist narrative in their platforms, supporters of extremist groups have migrated to smaller or less open platforms, hiding behind encrypted communication (e.g., telegram), gaming-platforms, and the dark-web. Moreover, the work in this thesis explored the textual content of radical propaganda. Another avenue of future work may focus on multi-media data sources. Large amounts of propaganda being distributed online consist of imagery and video. Creating tools capable of processing multi-media and detecting links to extremism is critical. Furthermore, stemming from our final study on comparing properties of different radical discourse, future work would investigate if we can generalise our approach to detecting supporters of different extremist groups.

We outlined a set of open-research directions in Chapter 4 that relates to formulating appropriate methods to facilitate information sharing across different cybercrime agencies and private organisations. Additionally, the CCINT framework proposed in this thesis consists of a wide set of methods and tools that are useful for cybercrime investigators. In our application of the framework to the topic of online radicalisation, we implemented a select number of the framework components. Mainly, we focused on the data and analysis layers as we designed and built a system that can utilise data sources from online social network platforms and online magazines to detect weak signals (warning behaviour) for adopters of extremist ideologies. We used behaviour analysis to identify patterns of behaviour for ISIS supporters. As such, we leave the remaining components for future work. For

instance, exploration of visual analytics methods and how they can be used to assist investigators in detecting and understanding radical behaviour is a natural expansion to this research. Visual analytics is an important aspect that supports analytical reasoning through visual and interactive user interfaces [226]. Having an appropriately crafted visual representation of complex data plays an important role in achieving effective understanding, reasoning and ultimately results in better decision making. Additionally, work on improving the usability of tools targeted to law-enforcement and ensuring seamless integration and compatibility between existing tools they use and newly designed tools are topics to be investigated further.

9.5 Final Remarks

This work has explored the wide topic of cybercrime investigations before it zoomed-in to the topic of online radicalisation. Through a detailed study including practitioners from both government and private sectors, we identified existing processes, needs, and challenges they face. Informed by our review of the literature and the participants' interviews, we designed CCINT, a cybercrime intelligence framework to aid practitioners better analyse and investigate cybercrimes. We apply framework methods to the problem of online radicalisation, where we investigate properties related to extremist propaganda. Using systematic methodology we design an approach to extract contextual and psycholinguistic properties from extremist propaganda. Subsequently, we demonstrate how these properties can aid in detecting supporters in online social network platforms. Extremists are getting more and more creative in the ways they exploit online platforms. They started by spreading their propaganda online to radicalise and celebrate their attacks, and have now evolved to live-broadcasting attacks as we have seen in the Christchurch mosque attack. We, as the research community, public, and private sectors, need to form collaborations and be one step ahead of them by developing proper methods, tools, and policies to limit the spread of such harmful extremist content online.

Appendices

A

Psycholinguistic Profiles

Table A.1: Comparing Psych-profile Properties of the three radical magazines. N is number of magazines per group. Numbers are mean percentages of total words per text reported as $Mean(std)$

Feature	Dabiq N = 15	Inspire N = 17	AR N = 22	News N = 31
Analytic	89.80 (3.81)	87.21 (2.78)	89.25 (4.31)	87.77 (8.85)
Clout	77.57 (2.85)	78.62 (3.06)	66.55 (4.00)	67.08 (13.04)
Authentic	7.50 (1.39)	12.13 (3.83)	13.59 (3.77)	23.13 (18.30)
Tone	23.42 (4.20)	25.44 (7.40)	29.49 (12.83)	35.39 (22.29)
i	0.67 (0.32)	0.63 (0.33)	0.54 (0.42)	0.97 (1.82)
we	0.65 (0.21)	1.42 (0.60)	0.46 (0.23)	0.56 (0.47)
you	0.89 (0.35)	1.16 (0.41)	0.17 (0.11)	0.46 (0.39)
shehe	1.98 (0.39)	1.34 (0.65)	1.38 (0.53)	1.91 (1.75)
they	2.14 (0.26)	1.74 (0.30)	1.22 (0.19)	0.83 (0.40)
affect	4.59 (0.36)	5.51 (0.85)	4.16 (0.53)	4.33 (1.14)
posemo	2.20 (0.19)	2.71 (0.37)	2.13 (0.49)	2.32 (0.81)
negemo	2.37 (0.25)	2.77 (0.58)	1.99 (0.46)	1.95 (0.99)
anx	0.29 (0.04)	0.41 (0.15)	0.32 (0.11)	0.30 (0.20)
anger	1.27 (0.17)	1.44 (0.35)	0.75 (0.27)	0.73 (0.68)
sad	0.28 (0.07)	0.37 (0.13)	0.34 (0.10)	0.31 (0.21)
drives	7.03 (0.57)	7.93 (0.59)	7.17 (0.94)	8.22 (1.94)
affiliation	1.83 (0.24)	2.34 (0.55)	1.62 (0.41)	1.63 (0.65)
achieve	1.32 (0.23)	1.37 (0.20)	1.29 (0.33)	1.60 (0.58)
power	3.55 (0.40)	3.52 (0.46)	3.65 (0.57)	4.22 (1.64)
reward	0.77 (0.12)	0.89 (0.12)	0.90 (0.19)	1.05 (0.53)
risk	0.52 (0.09)	0.75 (0.19)	0.55 (0.15)	0.66 (0.33)
work	1.78 (0.30)	2.14 (0.35)	3.66 (0.75)	4.16 (1.81)
leisure	0.31 (0.09)	0.42 (0.16)	0.76 (0.22)	1.17 (1.39)
religion	3.81 (0.56)	2.83 (0.68)	0.42 (0.26)	0.34 (0.70)
death	0.54 (0.16)	0.60 (0.18)	0.32 (0.20)	0.32 (0.54)
informal	0.64 (0.14)	0.62 (0.21)	0.29 (0.07)	0.38 (0.27)
swear	0.01 (0.01)	0.02 (0.02)	0.04 (0.02)	0.02 (0.04)
netspeak	0.19 (0.07)	0.28 (0.20)	0.13 (0.05)	0.22 (0.24)
Harm	0.11 (0.03)	0.20 (0.06)	0.10 (0.08)	0.14 (0.16)
Care	0.51 (0.12)	0.54 (0.22)	0.25 (0.19)	0.34 (0.38)
Fairness	0.04 (0.01)	0.05 (0.03)	0.10 (0.03)	0.08 (0.12)
Cheating	0.01 (0.01)	0.02 (0.02)	0.09 (0.07)	0.03 (0.07)
In-group (loyalty)	0.35 (0.10)	0.29 (0.10)	0.46 (0.18)	0.39 (0.23)
Out-group (betrayal)	0.20 (0.06)	0.32 (0.10)	0.16 (0.08)	0.09 (0.17)
Authority	0.68 (0.19)	0.40 (0.09)	0.35 (0.12)	0.42 (0.35)
Subversion	0.06 (0.04)	0.03 (0.02)	0.10 (0.09)	0.05 (0.08)
Purity	0.05 (0.04)	0.06 (0.02)	0.05 (0.03)	0.06 (0.08)
Degradation	0.07 (0.03)	0.03 (0.02)	0.04 (0.03)	0.05 (0.15)
Morality General	0.27 (0.07)	0.26 (0.06)	0.24 (0.09)	0.18 (0.13)

Table A.2: Comparing Dabiq, Inspire, and AR using Kruskal–Wallis test. ($p <= 0.05$)

Category	**Feature**	**H-score**	**p-value**
Summary Variables	Analytic	7.03	0.029716488
	Clout	38.86	3.65E-09
	Authentic	24.08	5.91E-06
Personal Pronouns	we	35.95	1.57E-08
	you	39.45	2.72E-09
	shehe	14.48	0.000717532
	they	38.46	4.45E-09
Emotions	affect	23.37	8.42E-06
	posemo	16.04	0.000329024
	negemo	17.02	0.000201671
	anx	7.02	0.029952208
	anger	29.67	3.61E-07
	sad	6.09	0.047499146
Drives	affiliation	25.05	3.63E-06
	reward	6.24	0.044216285
	risk	14.93	0.000573923
Concerns	work	39.71	2.39E-09
	leisure	33.45	5.45E-08
	religion	43.06	4.46E-10
	death	18.10	0.000117492
	informal	35.93	1.58E-08
	swear	14.17	0.000835645
Morality Foundations	Harm	22.01	1.66E-05
	Care	20.72	3.16E-05
	Fairness	32.38	9.28E-08
	Cheating	25.64	2.70E-06
	In-group (Loyalty)	13.21	0.001351838
	Out-group (Betrayal)	21.10	2.62E-05
	Authority	26.53	1.73E-06
	Subversion	18.30	0.000106345
	Degradation	11.74	0.002828199

References

[1] Sarah Gordon and Richard Ford. "On the definition and classification of cybercrime". In: *Journal in Computer Virology* 2.1 (2006), pp. 13–20.

[2] Canadian Centre for Intelligence and Security Studies. *A Framework for Understanding Terrorist Use of the Internet*. Tech. rep. Volume 2006-2. Integrated Terrorism Assessment Centre (ITAC), 2006.

[3] Department for Digital, Culture, Media and Sport. *Cyber Security Breaches Survey 2019*. `https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/813599/Cyber_Security_Breaches_Survey_2019_-_Main_Report.pdf`. [Online; accessed 16-Jan-2020]. 2019.

[4] Paul Hunton. "The stages of cybercrime investigations: Bridging the gap between technology examination and law enforcement investigation". In: *Computer Law & Security Review* 27.1 (2011), pp. 61–67.

[5] Raj Samani and Francois Paget. *Cybercrime Exposed: Cybercrime-as-a-Service*. `http://www.mcafee.com/us/resources/white-papers/wp-cybercrime-exposed.pdf`. [Online; accessed 10-Nov-2018].

[6] Europol: European Cybercrime Center (EC3). *Internet Organised Crime Threat Assessment*. `https://www.europol.europa.eu/activities-services/main-reports/internet-organised-crime-threat-assessment-iocta-2019`. [Online; accessed 16-Jan-2020]. 2019.

[7] S.W. Brenner. *Cybercrime: Criminal Threats from Cyberspace, 2nd Edition*. Praeger Security International. ABC-CLIO, LLC, 2018.

[8] Lynne Yarbro Williams. "Catch me if you can: a taxonomically structured approach to cybercrime". In: *The Forum on Public Policy*. 2008, pp. 28–30.

[9] Carl Miller. *British police are on the brink of a totally avoidable cybercrime crisis*. `https://www.wired.co.uk/article/british-police-cybercrime-hacking`. [Online; accessed 9-Jan-2019]. 2018.

[10] Ian Murphy. *Cisco to train 120,000 cyber police*. `https://www.enterprisetimes.co.uk/2018/11/30/cisco-to-train-120000-cyber-police/`. [Online; accessed 9-Jan-2019]. 2018.

[11] Jason R.C. Nurse. "Cybercrime and You: How Criminals Attack and the Human Factors That They Seek to Exploit". In: *The Oxford Handbook of Cyberpsychology*. Ed. by Alison Attrill-Smith et al. Oxford: Oxford University Press, 2018.

[12] HMICFRS. *Cyber: Keep the light on. An inspection of the police response to cyber-dependent crime.* `https://www.justiceinspectorates.gov.uk/hmicfrs/wp-content/uploads/cyber-keep-the-light-on-an-inspection-of-the-police-response-to-cyber-dependent-crime.pdf`. [Online; accessed 16-Jan-2020]. 2019.

[13] Hergis Jica. "Cooperation between Cyber Criminals and Terrorist Organizations". In: *Mediterranean Journal of Social Sci.* 4.9 (2013), p. 532.

[14] Kim-Kwang Raymond Choo. "The cyber threat landscape: Challenges and future research directions". In: *Computers and Security* 30.8 (2011), pp. 719–731.

[15] Rob Procter et al. "Reading the riots: what were the police doing on Twitter?" In: *Policing and Society* 23.4 (2013), pp. 413–436.

[16] Yong Lu et al. "Social Network Analysis of a Criminal Hacker Community". In: *Journal of Computer Information Systems* 51.2 (2010), pp. 31–41.

[17] Matthew Edwards, Awais Rashid, and Paul Rayson. "A Systematic Survey of Online Data Mining Technology Intended for Law Enforcement". In: *ACM Comput. Surv.* 48.1 (Sept. 2015), 15:1–15:54.

[18] Mariam Nouh and Jason R.C. Nurse. "Identifying Key-Players in Online Activist Groups on the Facebook Social Network". In: *2015 IEEE Int. Conference on Data Mining Workshops (ICDMW)*. Nov. 2015, pp. 969–978.

[19] John Scott. *Social network analysis*. Sage, 2012.

[20] Raoul Chiesa, Stefania Ducci, and Silvio Ciappi. *Profiling hackers: the science of criminal profiling as applied to the world of hacking*. Auerbach Publications, 2008.

[21] Jason R.C. Nurse and Maria Bada. "The Group Element of Cybercrime: Types, Dynamics, and Criminal Operations". In: *The Oxford Handbook of Cyberpsychology*. Ed. by Alison Attrill-Smith et al. Oxford: Oxford University Press, 2018.

[22] Andranik Tumasjan et al. "Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment". In: *Int. Conference on Web and Social Media*. 2010.

[23] Fred Morstatter et al. "Finding eyewitness tweets during crises". In: *arXiv preprint arXiv:1403.1773* (2014).

[24] Charlie Edwards and Luke Gribbon. "Pathways to Violent Extremism in the Digital Era". In: *The RUSI Journal* 158.5 (2013), pp. 40–47. URL: `%5Curl%7Bhttp://dx.doi.org/10.1080/03071847.2013.847714%7D`.

[25] The Editorial Board. The New York Times. *The New Radicalization of the Internet*. `www.nytimes.com/2018/11/24/opinion/sunday/facebook-twitter-terrorism-extremism.html`. [Online; accessed 09-04-2019]. 2018.

[26] Antonia Ward. *ISIS's Use of Social Media Still Poses a Threat to Stability in the Middle East and Africa*. Tech. rep. Georgetown Security Studies Review, 2018.

[27] Twitter Public Policy. *Expanding and building #TwitterTransparency*. `https://blog.twitter.com/en_us/topics/company/2018/twitter-transparency-report-12.html`. [Online; accessed 10-April-2019]. 2018.

[28] M. Ashcroft et al. "Detecting Jihadist Messages on Twitter". In: *Proceedings of the Intelligence and Security Informatics Conference (EISIC), 2015 European*. Sept. 2015, pp. 161–164.

[29] Matthew Rowe and Hassan Saif. "Mining Pro-ISIS Radicalisation Signals from Social Media Users." In: *ICWSM*. 2016, pp. 329–338.

[30] Adam M Bossler and Thomas J Holt. "Patrol officers' perceived role in responding to cybercrime". In: *Policing: an international journal of police strategies & management* 35.1 (2012), pp. 165–181.

[31] Thomas J Holt and Adam M Bossler. "Predictors of patrol officer interest in cybercrime training and investigation in selected United States police departments". In: *Cyberpsychology, Behavior, and Social Networking* 15.9 (2012), pp. 464–472.

[32] Scott R Senjo. "An analysis of computer-related crime: Comparing police officer perceptions with empirical data". In: *Security Journal* 17.2 (2004), pp. 55–71.

[33] Lee Hadlington et al. "A Qualitative Exploration of Police Officers' Experiences, Challenges, and Perceptions of Cybercrime". In: *Policing: A Journal of Policy and Practice* (2018).

[34] Rodrigo Carvalho, Michael Goldsmith, and Sadie Creese. "Applying semantic technologies to fight online banking fraud". In: *European Intelligence and Security Informatics Conference (EISIC)*. 2015.

[35] Bushra A. AlAhmadi and Ivan Martinovic. "MalClassifier: Malware Family Classification Using Network Flow Sequence Behaviour". In: *2018 APWG Symposium on Electronic Crime Research (eCrime)*. May 2018, pp. 1–13.

[36] Felix C Freiling, Thorsten Holz, and Georg Wicherski. "Botnet tracking: Exploring a root-cause methodology to prevent distributed denial-of-service attacks". In: *European Symposium on Research in Computer Security*. Springer. 2005, pp. 319–335.

[37] HM Government. *Online Harms White Paper*. Tech. rep. Home Office and DMCS, Apr. 2019. URL: `https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/793360/Online_Harms_White_Paper.pdf`.

[38] William Allendorfer and Susan Herring. "ISIS vs. the US government: A war of online video propaganda". In: *AoIR Selected Papers of Internet Research* 5 (2015).

[39] European Commission. *Towards a general policy on the fight against cyber crime*. `http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2007:0267:FIN:EN:PDF`. [Online; accessed 11-May-2016]. 2007.

[40] Gregor Urbas et al. *Resource materials on technology-enabled crime*. Australian Institute of Criminology, 2008.

[41] Claire Hargreaves and Dr Daniel Prince. *Understanding Cyber Criminals and Measuring Their Future Activity Developing cybercrime research*. Tech. rep. Lancaster University, 2013.

[42] David S Wall. "Policing cybercrimes: Situating the public police in networks of security within cyberspace". In: *Police Practice and Research* 8.2 (2007), pp. 183–205.

[43] Harmandeep Singh Brar and Gulshan Kumar. "Cybercrimes: A proposed taxonomy and challenges". In: *Journal of Computer Networks and Communications* 2018 (2018).

[44] Mike McGuire and Samantha Dowling. *Cyber crime: A review of the evidence. Home Office Research Report 75*. Tech. rep. Home Office, UK, Oct. 2013.

[45] David S Wall. "The Internet as a conduit for criminal activity". In: *Information Technology and The Criminal Justice System, Pattavina, A., ed* (2010), pp. 77–98.

[46] H. Sarvari et al. "Constructing and Analyzing Criminal Networks". In: *Security and Privacy Workshops (SPW), 2014 IEEE*. 2014, pp. 84–91.

[47] Jonathan F. Spencer. "Using XML to Map Relationships in Hacker Forums". In: *Proceedings of the 46th Annual Southeast Regional Conference on XX*. ACM-SE 46. New York, NY, USA: ACM, 2008, pp. 487–489.

[48] Chaochang Chiu et al. "Internet auction fraud detection using social network analysis and classification tree approaches". In: *International Journal of Electronic Commerce* 15.3 (2011), pp. 123–147.

[49] Hsinchun Chen et al. "Uncovering the Dark Web: A Case Study of Jihad on the Web". In: *J. Am. Soc. Inf. Sci. Technol.* 59.8 (June 2008), pp. 1347–1359.

[50] Thabit Sabbah et al. "Hybridized term-weighting method for Dark Web classification." In: *Neurocomputing* 173 (2016), pp. 1908–1926.

[51] M. McCord and M. Chuah. "Spam Detection on Twitter Using Traditional Classifiers". In: *Proceedings of the 8th International Conference on Autonomic and Trusted Computing*. ATC'11. Banff, Canada: Springer-Verlag, 2011, pp. 175–186.

[52] Kurt Thomas et al. "Trafficking Fraudulent Accounts: The Role of the Underground Market in Twitter Spam and Abuse". In: *Presented as part of the 22nd USENIX Security Symposium*. Washington, D.C., 2013, pp. 195–210.

[53] A. H. Wang. "Don't follow me: Spam detection in Twitter". In: *Security and Cryptography (SECRYPT), Proceedings of the 2010 International Conference on*. July 2010, pp. 1–10.

[54] Alex Beutel et al. "CopyCatch: Stopping Group Attacks by Spotting Lockstep Behavior in Social Networks". In: *Proceedings of the 22nd International Conference on World Wide Web*. Rio de Janeiro, Brazil: ACM, 2013, pp. 119–130.

[55] Gianluca Stringhini, Christopher Kruegel, and Giovanni Vigna. "Detecting Spammers on Social Networks". In: *Proceedings of the 26th Annual Computer Security Applications Conference*. ACSAC '10. Austin, Texas, USA: ACM, 2010, pp. 1–9.

[56] Emilio Ferrara et al. "The Rise of Social Bots". In: *CoRR* abs/1407.5225 (2014).

[57] Richard M. Everett, Jason R. C. Nurse, and Arnau Erola. "The Anatomy of Online Deception: What Makes Automated Text Convincing?" In: *proceedings of the ACM/SIGAPP Symposium on Applied Computing (SAC)*. Apr. 2016.

[58] Gang Wang et al. "Social Turing Tests: Crowdsourcing Sybil Detection". In: *CoRR* abs/1205.3856 (2012).

[59] Chao Yang et al. “Analyzing Spammers’ Social Networks for Fun and Profit: A Case Study of Cyber Criminal Ecosystem on Twitter”. In: *Proceedings of the 21st International Conference on World Wide Web*. WWW ’12. New York, NY, USA: ACM, 2012, pp. 71–80.

[60] Gianluca Stringhini et al. “The Harvester, the Botmaster, and the Spammer: On the Relations Between the Different Actors in the Spam Landscape”. In: *Proceedings of the 9th ACM Symposium on Information, Computer and Communications Security*. ASIA CCS ’14. Kyoto, Japan: ACM, 2014, pp. 353–364.

[61] Abdullah Almaatouq et al. “If it looks like a spammer and behaves like a spammer, it must be a spammer: analysis and detection of microblogging spam accounts”. In: *International Journal of Information Security* (2016), pp. 1–17.

[62] Vaibhav Garg et al. “Computer-supported cooperative crime”. In: *Financial Cryptography and Data Security*. Springer, 2015, pp. 32–43.

[63] H. Chen et al. “Crime data mining: a general framework and some examples”. In: *Computer* 37.4 (Apr. 2004), pp. 50–56.

[64] Pankaj Choudhary and Upasna Singh. “A Survey on Social Network Analysis for Counter-Terrorism”. In: *International Journal of Computer Applications* 112.9 (2015).

[65] Jialun Qin et al. “Intelligence and Security Informatics: IEEE International Conference on Intelligence and Security Informatics, ISI 2005, Atlanta, GA, USA, May 19-20, 2005. Proceedings”. In: ed. by Paul Kantor et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005. Chap. Analyzing Terrorist Networks: A Case Study of the Global Salafi Jihad Network, pp. 287–304.

[66] Jennifer J. Xu and Hsinchun Chen. “Fighting Organized Crimes: Using Shortest-path Algorithms to Identify Associations in Criminal Networks”. In: *Decis. Support Syst.* 38.3 (2004), pp. 473–487.

[67] Jennifer J. Xu and Hsinchun Chen. “CrimeNet Explorer: A Framework for Criminal Network Knowledge Discovery”. In: *ACM Transactions on Information Systems* 23.2 (Apr. 2005), pp. 201–226.

[68] Malcolm K. Sparrow. “The application of network analysis to criminal intelligence: An assessment of the prospects”. In: *Social Networks* 13.3 (1991), pp. 251–274.

[69] Yulei Zhang et al. “Developing a Dark Web collection and infrastructure for computational and social sciences”. In: *IEEE International Conference on Intelligence and Security Informatics (ISI): Public Safety and Security*. 2010, pp. 59–64.

[70] Joseph Mei and Richard Frank. “Sentiment Crawling: Extremist Content Collection Through a Sentiment Analysis Guided Web-Crawler”. In: *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*. ASONAM ’15. Paris, France: ACM, 2015, pp. 1024–1027.

[71] Martin Bouchard, Kila Joffres, and Richard Frank. “Computational Models of Complex Systems”. In: ed. by Kumar Vijay Mago and Vahid Dabbaghian. Cham: Springer International Publishing, 2014. Chap. Preliminary Analytical Considerations in Designing a Terrorism and Extremism Online Network Extractor, pp. 171–184.

[72] H. Chen et al. “The Dark Web Forum Portal: From multi-lingual to video”. In: *Intelligence and Security Informatics (ISI), 2011 IEEE International Conference on*. 2011, pp. 7–14.

[73] Hsinchun Chen et al. “COPLINK: Managing Law Enforcement Data and Knowledge”. In: *Commun. ACM* 46.1 (Jan. 2003), pp. 28–34.

[74] Jennifer Xu and Hsinchun Chen. “Criminal network analysis and visualization”. In: *Communications of the ACM* 48.6 (2005), pp. 100–107.

[75] Awais Rashid et al. “Who am I? analyzing digital personas in cybercrime investigations”. In: *Computer* 46.4 (2013), pp. 54–61.

[76] John Stasko, Carsten Görg, and Zhicheng Liu. “Jigsaw: supporting investigative analysis through interactive visualization”. In: *Information visualization* 7.2 (2008), pp. 118–132.

[77] Jeremy JD White and Robert E Roth. “TwitterHitter: Geovisual analytics for harvesting insight from volunteered geographic information”. In: *Proceedings of GIScience*. Vol. 2010. 2010.

[78] Gianluca Stringhini et al. “EVILCOHORT: Detecting Communities of Malicious Accounts on Online Services”. In: *24th USENIX Security Symposium*. Washington, D.C.: USENIX Association, Aug. 2015, pp. 563–578.

[79] Paterva Inc. *Maltego*. `http://www.paterva.com/web7/buy.php`. [Online; accessed 11-Nov-2018].

[80] Wynyard. *Wynyard Advanced Crime Analytics: Powerful Software to Prevent and Solve Crime*. `https://www.wynyardgroup.com/media/746922/wynyard-advanced-crime-analytics-overview.pdf`. [Online; accessed 11-May-2018].

[81] Pete Burnap et al. “Tweeting the terror: modelling the social media reaction to the Woolwich terrorist attack”. In: *Social Network Analysis and Mining* 4.1 (2014), pp. 1–14.

[82] Bashar Nuseibeh and Steve Easterbrook. “Requirements engineering: a roadmap”. In: *Proceedings of the Conference on the Future of Software Engineering*. ACM. 2000, pp. 35–46.

[83] William Elm et al. “Finding decision support requirements for effective intelligence analysis tools”. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 49. 3. SAGE Publications. 2005, pp. 297–301.

[84] Eva-Catherine Hillemann, Alexander Nussbaumer, and Dietrich Albert. “The Role of Cognitive Biases in Criminal Intelligence Analysis and Approaches for their Mitigation”. In: *Intelligence and Security Informatics Conference (EISIC), 2015 European*. IEEE. 2015, pp. 125–128.

[85] Richards J Heuer. *Psychology of intelligence analysis*. Center for the Study of Intelligence. Central Intelligence Agency, 1999.

[86] Séamus Ó Ciardhuáin. “An extended model of cybercrime investigations”. In: *International Journal of Digital Evidence* 3.1 (2004), pp. 1–22.

[87] Todd Bacastow and Karen Schuckman. *The Intelligence Process*. `https://courseware.e-education.psu.edu/courses/bootcamp/lo07/09.html`. [Online; accessed 08-Aug-2018]. 2007.

[88] A Tradecraft Primer. "Structured Analytic Techniques for Improving Intelligence Analysis". In: *CIA Center for the Study of Intelligence* (2009).

[89] BL William Wong and Margaret Varga. "Black holes, keyholes and brown worms: Challenges in sense making". In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 56. 1. SAGE Publications. 2012, pp. 287–291.

[90] Community Oriented Policing Services [COPS]. *Online Radicalisation to Violent Extremism: Awareness Brief*. Tech. rep. Washington, DC, 2014.

[91] Haroro J Ingram. "An analysis of Islamic State's Dabiq magazine". In: *Australian Journal of Political Science* 51.3 (2016), pp. 458–477.

[92] Haroro J Ingram. "An analysis of inspire and Dabiq: Lessons from AQAP and Islamic state's propaganda war". In: *Studies in Conflict & Terrorism* 40.5 (2017), pp. 357–375.

[93] Robyn Torok. "Developing an explanatory model for the process of online radicalisation and terrorism". In: *Security Informatics* 2.1 (2013), p. 6.

[94] David G Myers and Helmut Lamm. "The group polarization phenomenon." In: *Psychological Bulletin* 83.4 (1976), p. 602.

[95] Paul Elzinga et al. "Terrorist threat assessment with formal concept analysis". In: *Intelligence and Security Informatics (ISI), 2010 IEEE International Conference on*. IEEE. 2010, pp. 77–82.

[96] Yla R Tausczik and James W Pennebaker. "The psychological meaning of words: LIWC and computerized text analysis methods". In: *Journal of language and social psychology* 29.1 (2010), pp. 24–54.

[97] Robert R McCrae and Paul T Costa Jr. "The five-factor theory of personality". In: *The Guilford Press* (2008).

[98] Sheryl Prentice and Paul J Taylor. "Psychological and behavioral examinations of online terrorism". In: *Violent Extremism: Breakthroughs in Research and Practice*. IGI Global, 2019, pp. 450–470.

[99] Samuel D Gosling et al. "Manifestations of personality in online social networks: Self-reported Facebook-related behaviors and observable profile information". In: *Cyberpsychology, Behavior, and Social Networking* 14.9 (2011), pp. 483–488.

[100] Joanne Hinds and Adam Joinson. "Human and computer personality prediction from digital footprints". In: *Current Directions in Psychological Science* 28.2 (2019), pp. 204–211.

[101] Michele Settanni and Davide Marengo. "Sharing feelings online: studying emotional well-being via automated text analysis of Facebook posts". In: *Frontiers in psychology* 6 (2015), p. 1045.

[102] J.W. Pennebaker et al. "The development and psychometric properties of LIWC 2015". In: University of Texas at Austin. 2015.

[103] Erin O'Carroll Bantum and Jason E Owen. "Evaluating the validity of computerized content analysis programs for identification of emotional expression in cancer narratives." In: *Psychological assessment* 21.1 (2009), p. 79.

[104] Munmun De Choudhury et al. "Predicting depression via social media." In: *ICWSM* 13 (2013), pp. 1–10.

[105] Yla R. Tausczik and James W. Pennebaker. "The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods". In: *Journal of Language and Social Psychology* 29.1 (2010), pp. 24–54.

[106] Matteo Vergani and Ana-Maria Bliuc. "The Language of New Terrorism: Differences in Psychological Dimensions of Communication in Dabiq and Inspire". In: *Journal of Language and Social Psychology* (2018), p. 0261927X17751011.

[107] Ted Grover and Gloria Mark. "Detecting Potential Warning Behaviors of Ideological Radicalization in an Alt-Right Subreddit". In: *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 13. 01. 2019, pp. 193–204.

[108] Adam Satariano. *Twitter Suspends 300,000 Accounts Tied to Terrorism in 2017*. `https://www.bloomberg.com/news/articles/2017-09-19/twitter-suspends-300-000-accounts-in-2017-for-terrorism-content`. [Online; accessed 21-Dec-2019]. 2017.

[109] Lisa Kaati et al. "Detecting Multipliers of Jihadism on Twitter". In: *Data Mining Workshop (ICDMW), 2015 IEEE International Conference on*. IEEE. 2015, pp. 954–960.

[110] G. R. S. Weir et al. "Positing the problem: enhancing classification of extremist web content through textual analysis". In: *2016 IEEE International Conference on Cybercrime and Computer Forensic (ICCCF)*. June 2016, pp. 1–3.

[111] Jytte Klausen, Christopher Marks, and Tauhid Zaman. "Finding Online Extremists in Social Networks". In: *arXiv preprint arXiv:1610.06242* (2016).

[112] Joel Brynielsson et al. "Harvesting and analysis of weak signals for detecting lone wolf terrorists". In: *Security Informatics* 2.1 (2013), p. 11.

[113] Katie Cohen et al. "Detecting linguistic markers for radical violence in social media". In: *Terrorism and Political Violence* 26.1 (2014), pp. 246–256.

[114] Matthew L Williams, Pete Burnap, and Luke Sloan. "Crime sensing with big data: the affordances and limitations of using open source communications to estimate crime patterns". In: *British Journal of Criminology* (2016), azw031.

[115] Nasser Alsaedi, Pete Burnap, and Omer Rana. "Sensing Real-World Events Using Arabic Twitter Posts". In: (2016). URL: `http://www.aaai.org/ocs/index.php/ICWSM/ICWSM16/paper/view/13016`.

[116] Jennifer Fereday and Eimear Muir-Cochrane. "Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development". In: *International journal of qualitative methods* 5.1 (2006), pp. 80–92.

[117] N. King, C. Horrocks, and J. Brooks. *Interviews in Qualitative Research*. London, UK: SAGE Publications, 2010. URL: `https://books.google.co.uk/books?id=FYj2nQEACAAJ`.

[118] Virginia Braun and Victoria Clarke. "Using thematic analysis in psychology". In: *Qualitative research in psychology* 3.2 (2006), pp. 77–101.

[119] Jane Ritchie et al. *Qualitative research practice: A guide for social science students and researchers*. sage, 2013.

[120] Claire Anderson. "Presenting and evaluating qualitative research". In: *American journal of pharmaceutical education* 74.8 (2010), p. 141.

[121] Mira Crouch and Heather McKenzie. "The logic of small samples in interview-based qualitative research". In: *Social science information* 45.4 (2006), pp. 483–499.

[122] Rensis Likert. "A technique for the measurement of attitudes." In: *Archives of psychology* (1932).

[123] Maurits Clemens Kaptein, Clifford Nass, and Panos Markopoulos. "Powerful and Consistent Analysis of Likert-Type Ratingscales". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '10. Atlanta, Georgia, USA: Association for Computing Machinery, 2010, pp. 2391–2394.

[124] Anna Huang. "Similarity measures for text document clustering". In: *Proceedings of the sixth new zealand computer science research student conference (NZCSRSC2008), Christchurch, New Zealand*. Vol. 4. 2008, pp. 9–56.

[125] *On the surprising behavior of distance metrics in high dimensional space*. Springer. 2001, pp. 420–434.

[126] Student. "The probable error of a mean". In: *Biometrika* (1908), pp. 1–25.

[127] Henry B Mann and Donald R Whitney. "On a test of whether one of two random variables is stochastically larger than the other". In: *The annals of mathematical statistics* (1947), pp. 50–60.

[128] William H Kruskal and W Allen Wallis. "Use of ranks in one-criterion variance analysis". In: *Journal of the American statistical Association* 47.260 (1952), pp. 583–621.

[129] Robert Ackland. *Web social science: Concepts, data and tools for social scientists in the digital age*. Sage, 2013.

[130] Gobinda G Chowdhury. "Natural language processing". In: *Annual review of information science and technology* 37.1 (2003), pp. 51–89.

[131] Yoav Goldberg. "Neural network methods for natural language processing". In: *Synthesis Lectures on Human Language Technologies* 10.1 (2017), pp. 1–309.

[132] Juan Ramos et al. "Using tf-idf to determine word relevance in document queries". In: *Proceedings of the first instructional conference on machine learning*. Vol. 242. Piscataway, NJ. 2003, pp. 133–142.

[133] Omer Levy and Yoav Goldberg. "Dependency-Based Word Embeddings". In: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Baltimore, Maryland: Association for Computational Linguistics, June 2014, pp. 302–308. URL: https://www.aclweb.org/anthology/P14-2050.

[134] Tomas Mikolov et al. "Distributed representations of words and phrases and their compositionality". In: *Advances in neural information processing systems*. 2013, pp. 3111–3119.

[135] Tomas Mikolov et al. "Efficient estimation of word representations in vector space". In: *arXiv preprint arXiv:1301.3781* (2013).

[136] Piotr Bojanowski et al. "Enriching word vectors with subword information". In: *Transactions of the Association for Computational Linguistics* 5 (2017), pp. 135–146.

[137] Shai Shalev-Shwartz and Shai Ben-David. *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.

[138] Ronan Collobert and Jason Weston. "A unified architecture for natural language processing: Deep neural networks with multitask learning". In: *Proceedings of the 25th international conference on Machine learning*. ACM. 2008, pp. 160–167.

[139] Majid Alfifi and James Caverlee. "Badly Evolved? Exploring Long-Surviving Suspicious Users on Twitter". In: *Social Informatics: 9th International Conference, SocInfo 2017, Oxford, UK, Proceedings, Part I*. Springer International Publishing, 2017, pp. 218–233.

[140] Leo Breiman. "Random forests". In: *Machine learning* 45.1 (2001), pp. 5–32.

[141] B Krishna Murthy. "Neural networks for natural language processing". PhD thesis. 1996.

[142] Michael Buckland and Fredric Gey. "The relationship between recall and precision". In: *Journal of the American society for information science* 45.1 (1994), pp. 12–19.

[143] Stephen P Borgatti. "Centrality and network flow". In: *Social networks* 27.1 (2005), pp. 55–71.

[144] Marc Barthelemy. "Betweenness centrality in large complex networks". In: *The European physical journal B* 38.2 (2004), pp. 163–168.

[145] Gabriele Lohmann et al. "Eigenvector centrality mapping for analyzing connectivity patterns in fMRI data of the human brain". In: *PloS one* 5.4 (2010), e10232.

[146] Amy N Langville and Carl D Meyer. *Google's PageRank and beyond: The science of search engine rankings*. Princeton University Press, 2011.

[147] Ramesh Prajapati. "A survey paper on hyperlink-induced topic search (HITS) algorithms for web mining". In: *International Journal of Engineering* 1.2 (2012).

[148] Darren Quick and Kim-Kwang Raymond Choo. "Impacts of increasing volume of digital forensic data: A survey and future research challenges". In: *Digital Investigation* 11.4 (2014), pp. 273–294.

[149] Quick, Darren and Choo, Kim-Kwang Raymond. "Pervasive social networking forensics. Intelligence and evidence from mobile device extracts". In: *Journal of Network and Computer Applications* 86 (2017), pp. 24–33.

[150] David Lillis et al. "Current Challenges and Future Research Areas for Digital Forensic Investigation". In: *11th Annual ADFSL Conference on Digital Forensics, Security and Law (CDFSL 2016), Daytona Beach, Florida, USA, 24-26 May 2016*. 2016.

[151] Daniel W Turner III. "Qualitative interview design: A practical guide for novice investigators". In: *The qualitative report* 15.3 (2010), pp. 754–760.

[152] Peter H Carstensen and Kjeld Schmidt. "Computer supported cooperative work: New challenges to systems design". In: *In K. Itoh (Ed.), Handbook of Human Factors*. Citeseer. 1999.

[153] Patrick Biernacki and Dan Waldorf. "Snowball sampling: Problems and techniques of chain referral sampling". In: *Sociological methods & research* 10.2 (1981), pp. 141–163.

[154] Action Fraud. *The National Fraud Intelligence Bureau*. `https://www.actionfraud.police.uk/`. [Online; accessed 9-Jan-2019].

[155] Oliver Hutt et al. "Data and evidence challenges facing place-based policing". In: *Policing: An International Journal of Police Strategies & Management* 41.3 (2018), pp. 339–351.

[156] Jerry H Ratcliffe. "Knowledge management challenges in the development of intelligence-led policing". In: *The Handbook of Knowledge-Based Policing: Current Conceptions and Future Directions. Chichester: John Wiley and Sons* (2008), pp. 205–220.

[157] Lord Stevens. *Policing for a better Britain: Report of the independent police commission*. 2013.

[158] UK Government. *National Cyber Security Strategy 2016 to 2021*. `https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/567242/national_cyber_security_strategy_2016.pdf`. [Online; accessed 04-Jan-2020]. 2016.

[159] UK Government. *UK Serious and Organised Crime Strategy*. `https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/752850/SOC-2018-web.pdf`. [Online; accessed 04-Jan-2020]. 2018.

[160] Benjamin Saunders et al. "Saturation in qualitative research: exploring its conceptualization and operationalization". In: *Quality & quantity* 52.4 (2018), pp. 1893–1907.

[161] Budi Arief, Mohd Azeem Bin Adzmi, and Thomas Gross. "Understanding cybercrime from its stakeholders' perspectives: Part 1 attackers". In: *IEEE Security & Privacy* 13.1 (2015), pp. 71–76.

[162] Grainne Kirwan and Andrew Power. *Cybercrime: The psychology of online offenders*. Cambridge University Press, 2013.

[163] Rabiah Ahmad et al. "Perception on cyber-terrorism: A focus group discussion approach". In: *J. Information Security* 3.3 (2012), pp. 231–237.

[164] Budi Arief and Mohd Azeem Bin Adzmi. "Understanding cybercrime from its stakeholders' perspectives: Part 2 defenders and victims". In: *IEEE Security & Privacy* 13.2 (2015), pp. 84–88.

[165] Kai Koong and Manal Yunis. "A Conceptual Model for the Development of a National Cybersecurity Index: An Integrated Framework". In: *INFORMATION SYSTEMS SECURITY, ASSURANCE AND PRIVACY* (2015).

[166] Ross Anderson et al. "Measuring the cost of cybercrime". In: *The economics of information security and privacy*. Springer, 2013, pp. 265–300.

[167] Massimo Felici et al. "What's New in the Economics of Cybersecurity?" In: *IEEE Security & Privacy* 14.3 (2016), pp. 11–13.

[168] Ioannis Agrafiotis et al. "A taxonomy of cyber-harms: Defining the impacts of cyber-attacks and understanding how they propagate". In: *Journal of Cybersecurity* 4.1 (2018), tyy006.

[169] South East Regional Organised Crime Unit. *Cyber-crime site selling hacking tool taken down following international operation.* `https://serocu.police.uk/2019/11/cyber-crime-site-selling-hacking-tool-taken-down-following-international-operation/`. [Online; accessed 19-Dec-2019]. Nov. 2019.

[170] Department of Justice. *Chicago Man Charged with Attempting to Provide Material Support to ISIS.* `https://www.justice.gov/usao-ndil/pr/chicago-man-charged-attempting-provide-material-support-isis`. [Online; accessed 19-Dec-2019]. Nov. 2019.

[171] Robert F Krueger. "Personality from a realist's perspective: Personality traits, criminal behaviors, and the externalizing spectrum". In: *Journal of Research in Personality* 36.6 (2002), pp. 564–572.

[172] Arie W Kruglanski and Shira Fishman. "The psychology of terrorism:"Syndrome" versus "tool" perspectives". In: *Terrorism and Political Violence* 18.2 (2006), pp. 193–215.

[173] Mariam Nouh and Jason R. C. Nurse. "Identifying key-players in online activist groups on the Facebook social network". In: *IEEE International Conference on Data Mining Workshop (ICDMW)*. 2015, pp. 969–978.

[174] Christos Charitonidis, Awais Rashid, and Paul J. Taylor. "Weak Signals As Predictors of Real-World Phenomena in Social Media". In: *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*. ASONAM '15. Paris, France: ACM, 2015, pp. 864–871.

[175] Tom Holt et al. "Political radicalization on the Internet: Extremist content, government control, and the power of victim and jihad videos". In: *Dynamics of Asymmetric Conflict* 8.2 (2015), pp. 107–120.

[176] Paul Gill et al. "Terrorist use of the Internet by the numbers: Quantifying behaviors, patterns, and processes". In: *Criminology & Public Policy* 16.1 (2017), pp. 99–117.

[177] Paul Jonathon Taylor, Donald Holbrook, and Adam Joinson. "A same kind of different: affordances, terrorism and the internet". In: *Criminology and Public Policy* 16.1 (2017), pp. 127–133.

[178] Paul J Taylor. "The role of language in conflict and conflict resolution". In: *Handbook of language and social psychology* (2014), pp. 459–470.

[179] Alava, Séraphin and Frau-Meigs, Divina and Hassan, Ghayda. *Youth and Violent Extremism on Social Media.* `http://unesdoc.unesco.org/images/0026/002603/260382e.pdf`. [Online; accessed 21-Dec-2019]. 2017.

[180] S Vijayarani, Ms J Ilamathi, and Ms Nithya. "Preprocessing techniques for text mining-an overview". In: *International Journal of Computer Science & Communication Networks* 5.1 (2015), pp. 7–16.

[181] Tobias Schnabel et al. "Evaluation methods for unsupervised word embeddings". In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Lisbon, Portugal: Association for Computational Linguistics, Sept. 2015, pp. 298–307.

[182] Ethan Fast, Binbin Chen, and Michael S Bernstein. "Empath: Understanding topic signals in large-scale text". In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM. 2016, pp. 4647–4657.

[183] Mohammed Al-Mosaiwi and Tom Johnstone. "In an Absolute State: Elevated Use of Absolutist Words Is a Marker Specific to Anxiety, Depression, and Suicidal Ideation". In: *Clinical Psychological Science* 6.4 (2018), pp. 529–542.

[184] Jamy Li and Mark Chignell. "Birds of a feather: How personality influences blog writing and reading"". In: *International Journal of Human-Computer Studies* 68.9 (2010), pp. 589–602.

[185] David M Blei, Andrew Y Ng, and Michael I Jordan. "Latent dirichletiliev2015automated allocation". In: *Journal of machine Learning research* 3.Jan (2003), pp. 993–1022.

[186] William L. Hamilton, Jure Leskovec, and Dan Jurafsky. "Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change". In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Berlin, Germany: Association for Computational Linguistics, 2016, pp. 1489–1501.

[187] Vincent D Blondel et al. "Fast unfolding of communities in large networks". In: *Journal of statistical mechanics: theory and experiment* 2008.10 (2008), P10008.

[188] S Robert Lichter and Stanley Rothman. "The radical personality: Social psychological correlates of new left ideology". In: *Political Behavior* 4.3 (1982), pp. 207–235.

[189] Vern Pierson. *Western radicalization: rethinking the psychology of terrorism*. Tech. rep. Naval Postgraduate School Monterey United States, 2017.

[190] Monica Lopez et al. "A Forensic Psychological Assessment of Terrorists: An Anti-Terrorism Approach for Radicalized Westerners". In: *Journal of Aggression, Maltreatment & Trauma* 9.1-2 (2015), pp. 33–43.

[191] Martijn Van Zomeren, Tom Postmes, and Russell Spears. "Toward an integrative social identity model of collective action: A quantitative research synthesis of three socio-psychological perspectives." In: *Psychological bulletin* 134.4 (2008), p. 504.

[192] Simon Cottee and Keith Hayward. "Terrorist (e) motives: The existential attractions of terrorism". In: *Studies in Conflict & Terrorism* 34.12 (2011), pp. 963–986.

[193] James W Pennebaker et al. "Computerized text analysis of Al-Qaeda transcripts". In: *A content analysis reader* 453465 (2008).

[194] Jesse Graham et al. "Chapter Two - Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism". In: ed. by Patricia Devine and Ashby Plant. Vol. 47. Advances in Experimental Social Psychology. Academic Press, 2013, pp. 55–130.

[195] Jonathan Haidt and Jesse Graham. "When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize". In: *Social Justice Research* 20.1 (2007), pp. 98–116.

[196] Wiebe R. "Delinquent behaviour and the five factor model: Hiding in the adaptive landscape?" In: *Individual Differences Research* 29.2 (2004), pp. 38–62.

[197] Mohammad Rahim Kamaluddin et al. "Linking psychological traits with criminal behaviour: A review". In: *ASEAN Journal of Psychiatry* 16.2 (2015), pp. 13–25.

[198] Julian Droogan and Shane Peattie. "Mapping the thematic landscape of Dabiq magazine". In: *Australian Journal of International Affairs* 71.6 (2017), pp. 591–620.

[199] Julian Droogan and Shane Peattie. "Mapping the thematic landscape of Dabiq magazine". In: *Australian Journal of International Affairs* 71.6 (2017), pp. 591–620.

[200] J Reid Meloy. "Identifying warning behaviors of the individual terrorist". In: *FBI Law Enforcement Bulletin* 85 (2016), pp. 1–9.

[201] D. Devyatkin et al. "Exploring linguistic features for extremist texts detection (on the material of Russian-speaking illegal texts)". In: *2017 IEEE International Conference on Intelligence and Security Informatics (ISI)*. July 2017, pp. 188–190.

[202] Ping Liu et al. "Forecasting the presence and intensity of hostility on Instagram using linguistic and social features". In: *ICWSM*. 2018.

[203] H Andrew Schwartz et al. "Toward Personality Insights from Language Exploration in Social Media." In: *AAAI Spring Symposium: Analyzing Microtext*. 2013, pp. 72–79.

[204] Tony Munton et al. "Understanding vulnerability and resilience in individuals to the influence of al Qaida violent extremism". In: *Croydon: Great Britain Home Office* (2011).

[205] Bader Araj. "Harsh state repression as a cause of suicide bombing: the case of the Palestinian–Israeli conflict". In: *Studies in Conflict & Terrorism* 31.4 (2008), pp. 284–303.

[206] Shazadi Beg and Laila Bokhari. "Pakistan: In search of a disengagement strategy". In: *Leaving Terrorism Behind*. Routledge, 2008, pp. 242–260.

[207] Peter J Carrington. "Crime and social network analysis". In: *The SAGE handbook of social network analysis* (2011), pp. 236–255.

[208] C Chen, A Liaw, and L Breiman. *Using random forest to learn imbalanced data*. Tech. rep. Dept. Statistics, University of California, Berkeley, 2004.

[209] Emilio Ferrara et al. "Predicting online extremism, content adopters, and interaction reciprocity". In: *International conference on social informatics*. Springer. 2016, pp. 22–39.

[210] Adam Badawy and Emilio Ferrara. "The rise of Jihadist propaganda on social networks". In: *Journal of Computational Social Science* 1.2 (Sept. 2018), pp. 453–470.

[211] Clayton A Davis et al. "OSoMe: the IUNI observatory on social media". In: *PeerJ Computer Science* 2 (2016), e87.

[212] Gilles Louppe et al. "Understanding variable importances in forests of randomized trees". In: *Advances in neural information processing systems*. 2013, pp. 431–439.

[213] Raúl Lara-Cabrera, Antonio Gonzalez-Pardo, and David Camacho. "Statistical analysis of risk assessment factors and metrics to evaluate radicalisation in Twitter". In: *Future Generation Computer Systems* (2017).

[214] JM Berger and Bill Strathearn. "Who Matters Online: Measuring influence, evaluating content and countering violent extremism in online social networks". In: *International Centre for the Study of Radicalisation and Political Violence* (2013).

[215] Swati Agarwal and Ashish Sureka. "Using knn and svm based one-class classifier for detecting online radicalization on twitter". In: *International Conference on Distributed Computing and Internet Technology*. Springer. 2015, pp. 431–442.

[216] Hassan Saif et al. "A Semantic Graph-Based Approach for Radicalisation Detection on Social Media". In: *The Semantic Web*. Ed. by Eva Blomqvist et al. Cham: Springer International Publishing, 2017, pp. 571–587.

[217] Tom De Smedt, Guy De Pauw, and Pieter Van Ostaeyen. "Automatic Detection of Online Jihadist Hate Speech". In: *CoRR* abs/1803.04596 (2018). arXiv: 1803.04596. URL: %5Curl%7Bhttp://arxiv.org/abs/1803.04596%7D.

[218] Matthias Hartung et al. "Identifying Right-Wing Extremism in German Twitter Profiles: A Classification Approach". In: *Natural Language Processing and Information Systems*. Ed. by Flavius Frasincar et al. Cham: Springer International Publishing, 2017, pp. 320–325.

[219] Matthew C. Benigni, Kenneth Joseph, and Kathleen M. Carley. "Online extremism and the communities that sustain it: Detecting the ISIS supporting community on Twitter". In: *PLOS ONE* 12.12 (Dec. 2017), pp. 1–23.

[220] H. Watanabe, M. Bouazizi, and T. Ohtsuki. "Hate Speech on Twitter: A Pragmatic Approach to Collect Hateful and Offensive Expressions and Perform Hate Speech Detection". In: *IEEE Access* 6 (2018), pp. 13825–13835.

[221] DOMINIC TIERNEY. *The Twenty Years' War*. https://www.theatlantic.com/international/archive/2016/08/twenty-years-war/496736/. [Online; accessed 12-Jan-2020]. 2016.

[222] Carol M Swain. *The new white nationalism in America: Its challenge to integration*. Cambridge University Press, 2002.

[223] Jason Wilson. *The races are not equal': meet the alt-right leader in Clinton's campaign ad*. https://www.theguardian.com/us-news/2016/aug/26/jared-taylor-alt-right-clinton-trump. [Online; accessed 12-Jan-2020]. 2016.

[224] Philip Baugut and Katharina Neumann. "Online news media and propaganda influence on radicalized individuals: Findings from interviews with Islamist prisoners and former Islamists". In: *new media & society* (2019), p. 1461444819879423.

[225] Akemi Takeoka Chatfield, Christopher G Reddick, and Uuf Brajawidagda. "Tweeting propaganda, radicalization and recruitment: Islamic state supporters multi-sided twitter networks". In: *Proceedings of the 16th Annual International Conference on Digital Government Research*. ACM. 2015, pp. 239–249.

[226] Daniel Keim et al. "Visual analytics: Definition, process, and challenges". In: *Information visualization*. Springer, 2008, pp. 154–175.