2022

Exploring Cyberterrorism, Topic Models and Social Networks of Jihadists Dark Web Forums: A Computational Social Science Approach

Vivian Fiona Guetler
West Virginia University, vfg0002@mix.wvu.edu

Recommended Citation
Guetler, Vivian Fiona, "Exploring Cyberterrorism, Topic Models and Social Networks of Jihadists Dark Web Forums: A Computational Social Science Approach" (2022). *Graduate Theses, Dissertations, and Problem Reports*. 11253.
https://researchrepository.wvu.edu/etd/11253

Exploring Cyberterrorism, Topic Models and Social Networks of Jihadists Dark Web Forums: A Computational Social Science Approach

by

Vivian Fiona Guetler

Dissertation submitted to the
Eberly College of Arts and Sciences
at West Virginia University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Sociology

Jesse Wozniak, Ph.D., Chair
Katie Corcoran, Ph.D.
Abhik Roy, Ph.D.
Joshua Woods, Ph.D

Department of Sociology & Anthropology

Morgantown, West Virginia
2022

Keywords: cyberterrorism, terrorism, Internet, Dark Web, machine learning, natural language processing, topics models, sentiment analysis, social networks, community detection

Abstract

Exploring Cyberterrorism, Topic Models and Social Networks of Jihadists Dark Web Forums: A Computational Social Science Approach

by Vivian Fiona Guetler

This three-article dissertation focuses on cyber-related topics on terrorist groups, specifically Jihadists' use of technology, the application of natural language processing, and social networks in the analysis of text data derived from terrorists' Dark Web forums. The first article explores cybercrime and cyberterrorism. As technology progresses, it facilitates new forms of behavior, including tech-related crimes known as cybercrime and cyberterrorism. In this article, I provide an analysis of the problems of cybercrime and cyberterrorism within the field of criminology by reviewing existing literature focusing on (a) the issues in defining terrorism, cybercrime, and cyberterrorism, (b) ways that cybercriminals commit a crime in cyberspace, and (c) ways that cyberterrorists attack critical infrastructure, including computer systems, data, websites, and servers.

The second article is a methodological study examining the application of natural language processing computational techniques, specifically latent Dirichlet allocation (LDA) topic models and topic network analysis of text data. I demonstrate the potential of topic models by inductively analyzing large-scale textual data of Jihadist groups and supporters from three Dark Web forums to uncover underlying topics. The Dark Web forums are dedicated to Islam and the Islamic world discussions. Some members of these forums sympathize with and support terrorist organizations. Results indicate that topic modeling can be applied to analyze text data automatically; the most prevalent topic in all forums was religion. Forum members also discussed terrorism and terrorist attacks, supporting the Mujahideen fighters. A few of the discussions were related to relationships and marriages, advice, seeking help, health, food, selling electronics, and identity cards. LDA topic modeling is significant for finding topics from larger corpora such as the Dark Web forums. Implications for counterterrorism include the use of topic modeling in real-time classification and removal of online terrorist content and the monitoring of religious forums, as terrorist groups use religion to justify their goals and recruit in such forums for supporters.

The third article builds on the second article, exploring the network structures of terrorist groups on the Dark Web forums. The two Dark Web forums' interaction networks were created, and network properties were measured using social network analysis. A member is considered connected and interacting with other forum members when they post in the same threads forming an interaction network. Results reveal that the network structure is decentralized, sparse, and divided based on topics (religion, terrorism, current events, and relationships) and the members' interests in participating in the threads. As participation in forums is an active process, users tend to select platforms most compatible with their views, forming a subgroup or community. However, some members are essential and influential in the information and resources flow within the networks. The key members frequently posted about religion, terrorism, and relationships in multiple threads. Identifying key members is significant for counterterrorism, as mapping network structures and key users are essential for removing and destabilizing terrorist networks. Taken together, this dissertation applies a computational social science approach to the analysis of cyberterrorism and the use of Dark Web forums by jihadists.

This dissertation is dedicated to my dear mama, Rose Atieno, and my entire family: my grandparents, aunts and uncles, siblings, cousins and nephews, nieces, and extended family.

"It takes a village to raise and educate a child"

– Author Unknown

"I am because we are, and since we are, therefore I am"

– J.S Mbiti

Acknowledgments

I would like to thank my dissertation advisor, Dr. Jesse Wozniak, for the support, encouragement, guidance, advice, and patience. You have always supported my ideas, and I am grateful for that. I inspire to be the same advisor to my future students.

I am grateful for my dissertation committee members Dr. Abhik Roy, Dr. Katie Corcoran, and Dr. Joshua Woods. Thank you for the advice, feedback, help with data analysis, and the countless recommendation letters you wrote on my behalf.

I am thankful for the entire WVU sociology and anthropology department faculty and staff, cohorts, and friends. Without your support, encouragement, and feedback, I would not have been where I am today.

I am thankful to the WVU Research Computing for the access of supercomputing systems (Thorny Flat) and their help in navigating the system.

I am thankful to my family for their prayers, support, and encouragement.

Lastly, I am thankful for Almighty God for making this possible.

Asanteni Sana, Thank You!

Contents

List of Figures

List of Tables

Acronyms

ML	Machine Learning
NLP	Natural Language Processing
LDA	Latent Dirichlet Allocation
SA	Sentiment Analysis
SNA	Social Network Analysis
ANTMN	Analysis of Topic Model Networks

Chapter 1

Introduction

1.1 Background

The digital revolution and development of new technologies, including the Internet, Web 2.0, artificial intelligence, and social media networks, have rapidly transformed all aspects of social, political, economic, communications, and cultural life. As of January 2022, more than 50% of the world's population is online (Forum, 2020; Kemp, 2019). Criminals and terrorists have embraced the development of the digital landscape to further their goals. Terrorists are attacking computer systems and networks, stealing data and information, thus increasingly posing serious threats to national security, economy, and public policy (Carlin & Graff, 2018; K.-s. Choi et al., 2018; Hardy & Williams, 2014; Holt, 2012; Holt et al., 2017a). While not attacking cyberspace, terrorists predominantly use the Internet to spread propaganda/publicity, psychological warfare, data mining, fundraising, recruitment and mobilization, networking, and secure communication (Behr et al., 2013; Bloom et al., 2019; Conway, 2017; Edwards & Gribbon, 2013; Gill et al., 2017; Sageman, 2011). Hence, facilitating the promotion, popularization, and operation of radical ideologies and ideologically motivated violence. The availability of large-scale information generated by terrorists and their supporters online can be collected and analyzed. However, the traditional manual methods for collecting and detecting terrorist content can be challenging, costly, and time-consuming (Brynielsson et al., 2013; Scrivens et al., 2021; Scrivens et al., 2018). This dissertation argues that novel computational methods such as machine learning and natural processing language techniques can advance criminological research on terrorism online content, cybercrime, and cyberterrorism.

Research into terrorists and technology use has gained pace in recent years, driven by the growth of domestic and international terrorist groups. S. Furnell et al. (2015) define cybercrime as cyber-dependent and cyber-enabled. Cyber-dependent crimes are offenses that can only be conducted using a computer, computer networks, or technologies such as spreading viruses, malware, hacking, and distributed denial of service (DDoS) attacks directed against computers or networks. In contrast, cyber-enabled crimes are traditional crimes that use computers or computer networks to widen their reach or scale. Cyber-enabled crimes include fraud, theft, and sexual offending. Hacking is the starting point from which cyber-dependent crimes begin by deliberately accessing an unauthorized computer system, networks, and data (S. Furnell, 2003; S. Furnell et al., 2015; Maimon & Louderback, 2019; Schell & Dodge, 2002). Online offenses include hacking, website defacement, Distributed Denial of Service (DDoS), malware and ransomware, piracy, cyberbullying, and online fraud and scams. It is crucial to note cybercrime has become more professional, automated, complex, and stealthy. The technology used to commit cybercrime is commercialized, and experienced hackers can be hired to commit the crime (Grabosky, 2016; Holt, 2012; Wall, 2001b). Cybercrime is predicted to inflict damages totaling $ 6 trillion by 2025 (Morgan, 2021).

Terrorists are known to engage in cybercrime offenses, referred to as cyberterrorism, which is the convergence of cyberspace and terrorism (K.-s. Choi et al., 2018; Conway, 2014; Denning, 2000, 2001). The term cyberterrorism has sparked an intense debate over whether terrorist groups are capable of attacking cyberspace and critical infrastructures to further their goals (Conway, 2003, 2014; Hardy & Williams, 2014; Jarvis & Macdonald, 2014;

Weimann, 2004, 2005). Although terrorist groups have yet to cause massive loss of life or economic chaos by hacking into critical infrastructure systems, they have engaged in cyber-attacks against websites and other non-essential infrastructure through hacking (Carlin & Graff, 2018; K.-s. Choi et al., 2018; Hardy & Williams, 2014; Holt, 2012). Additionally, the ability to attack cyberspace has a far-reaching impact on terrorist groups' enemies, unlike bombings restricted to a specific physical location and communities (Carlin & Graff, 2018; S. M. Furnell & Warren, 1999). Cybercriminals and cyberterrorists are attacking computer systems and networks, critical infrastructures, stealing data and information, thus increasingly posing serious threats to national security, economy, and public policy.

Most of the studies on terrorists' use and/or misuse of the Internet have identified different ways in which terrorists use the internet. The fundamental use is to spread propaganda/publicity, psychological warfare, data mining, fundraising, recruitment and mobilization, networking, information dissemination, and secure communication (Conway, 2017; Hoffman, 2006b; Weimann, 2006). The groups use both the indexed and accessible Surface Web and the non-indexed and hidden Dark Web to further their goals. The radical right, ISIS, and other Jihadist groups use multiple platforms; for instance, they would use Twitter for audience development and YouTube or Instagram to share videos with the hope of recruiting (Weimann, 2016b). Interactive platforms such as chat rooms, message boards, forums, and social media platforms are beneficial for the interactions amongst like-minded individuals. Hence bringing together individuals, fostering a sense of belonging and self, loyalty, emotional attachment, and meaning amongst participants (Bloom & Daymon, 2018; Gaudette et al., 2020; Piazza & Guler, 2019). Hence, active online discussions about ideologies and opinions with friends or strangers have fostered like-minded individuals' belongingness and virtual communities of like-minded individuals (Bowman-Grieve, 2009; Scrivens et al., 2018). Once in the virtual community, users' language and social interactions change over time, mirroring those within their online community.

Terrorists use the internet predominantly to spread propaganda and psychological warfare. Islamist groups, for instance, Al-Qaeda and ISIS, have dedicated media councils overseeing their communication strategy (Hoffman, 2006b; Ligon et al., 2017). The role of the media jihad (Erez et al., 2011) is to produce, distribute, and collect jihadist content for TV, Radio, websites, and social media platforms. ISIS is the most prolific group on the internet, the most sophisticated in media and technology usage (Bisgin et al., 2019; Ligon et al., 2017). To reach their global audience, the groups publish and share information about attacks, ideologies, strategies, and justifications in Arabic, English, and other local languages. They publish online magazines such as Al-Qaeda's Inspire and ISIS' Dabiq and Rumiyah, in Arabic, English, French, Russian and Turkish (Bisgin et al., 2019; Welch, 2018; Wozniak et al., 2020) to encourage their supporters to commit terrorist attacks, provide instructions on how to make a bomb, assemble weapons, and become a homegrown terrorist. In addition, they create films known as 'Mujahideen films,' which have been posted on the internet or sold as DVDs. The films depict groups attacking U.S military forces on patrol, transmitting the last words of kidnapped Iraqis and foreigners about to be executed, and appealing for financial contributions.

The internet has become a better platform for terrorist groups to spread their ideologies through meta-narratives. A meta-narrative is a "theory that tries to give a totalizing, comprehensive account to various historical events, experiences, social, and cultural phenomena based upon the appeal to universal truth or universal values" (Ali & Bwana, 2015, p. 22). The metanarratives of Islamist terrorist groups such as Al-Shabaab, al-Qaeda, and ISIS are based on the misinterpretation of Islam's religious traditions and historical events for ideological purposes. Al Qaeda and ISIS base their political ideology on Islam, "on the religious tradition of Islam, appropriating and transforming key elements from the Quran and the Hadith, from Mohammed's life story and from the early history of Islam for its own ideological purposes" (Schmid, 2014, p. 4). Their messaging is strategic; for instance, Al-Shabaab's metanarratives for its targeted audience in Kenya include: there is a war against Islam in

Kenya; Muslims in Kenya have a duty to wage ‘holy war’ against non-Muslims, and the government of Kenya; and the only way to address and achieve victory against the ‘war on Islam’ is through armed struggle, self-sacrifice and/or active support to their version of ‘Jihadist’ cause (Ali & Bwana, 2015). They use the meta-narratives of victimization to appeal to the recruits, inform, educate and solicit support (Anzalone, 2016; Hoffman, 2006a).

Jihadist’s ideologies ascribed by Al Qaeda, ISIS, and their affiliates follow three themes; the West is hostile to Islam and the Muslim world; the caliphate, a political entity will replace the corrupt rules under the Western influence; the only way to address the West is in the language of violence, by waging jihad (holy war) against the non-believers, the West (Hoffman, 2006b; Pelletier et al., 2016; Schmid, 2014; Weimann, 2011). The ideology is meant to inspire believers to join the global jihad to create an Islamic State that enforces strict Islamic law. As Weimann (2006) notes, a terrorist campaign has four target audiences for its messages, namely the supporters, the population being served, the enemy, and international public opinion. Messages vary depending on the audience; the message amplifies pride, success, commitment, and vision to the supporters, potential recruits, and community. To the enemy, the message is meant to cause fear and threat with the hope of pressuring the government to accept the terrorists’ demands. To the rest of the world, the message is to justify their cause explain why and what they want to achieve, i.e., freedom, equal rights, a free state, etc. Therefore, the messaging seeks to express outrage, call for jihad, violence against the West, and to gain support for their Islamic cause. The groups have gained sympathizers and insurgents through their propaganda tactics and narratives.

There has been a growing effort in preventing and countering violent extremism (P/CVE) (Greenberg, 2016; Romaniuk, 2015; Saltman et al., 2021; Tuck & Silverman, 2016). As a policy and practice, countering violent extremism (CVE) has rapidly emerged as the most significant development for counterterrorism efforts. The common strategies include disruption, diversion, and counter-messaging (Greenberg, 2016). Officials and practitioners are responding to online extremism by implementing counter-narrative measures as the ‘soft’ solution to the problem of terrorism. The measures include using social media, newspapers, radio, and texting messages as tools for countering online extremism. The counternarratives provide discussions of Islam that correct the misinterpretation of the Quran and religion, refuting the "idealization of life inside the Caliphate. Instead of a perfect, protected, peaceful life, viewers see images of rape, death, and general suffering for members of the Caliphate" (Greenberg, 2016, p. 172). Other CVE programs include community engagement and outreach through town halls, roundtables, capacity building of youth, and community development (Gartenstein-Ross & Barr, 2016; Hussain & Saltman, 2016; Romaniuk, 2015).

Diversion and disruption occur through technical interventions by removing and disrupting terrorist platforms and online content. For instance, social media platforms used by terrorist groups and their supporters such as Facebook and Twitter, and forums such as Gab and Stormfront are being monitored and suspended (Behr et al., 2013; J. M. Berger, 2016; Caló & Hartley, 2019; Zelin & Fellow, 2013). The monitoring and disruption from Surface Web have led to the migration to encrypted platforms such as Telegram and the Dark Web (Malik, 2018; Weimann, 2016a). Most Internet users only access the surface web, also known as the Internet, easily accessible through browsers such as Google and Yahoo. The Dark Web, a subsection of the Deep Web, is a hidden part of the Internet that can only be accessed via special browsers such as TOR (Malik, 2018; Mirea et al., 2019; Monk et al., 2018; Weimann, 2016b). The Dark Web or Dark Net can only be accessed through encrypted Internet browsers; it offers anonymity, making it attractive for individuals interested in privacy and avoiding censorship and criminals, including terrorist groups evading law enforcement. Terrorists use the Dark Web for secure communication, discussion, and networking with their supporters through forums, as a repository of their propaganda materials, and financing using cryptocurrencies (Malik, 2018; Weimann, n.d.).

While previous research has highlighted the problem of terrorist use of the Internet, specifically social media platforms and websites, there has been little research exploring cyberterrorism and the use of Dark Web forums by terrorist groups and their supporters.

Additionally, although criminologists have applied classical research methods in the analysis of crime and terrorism and contributed to the development of novel research methods, crime theories, findings, and policies, few have yet to embrace computational approaches to the study of crime and crime-related problems, specifically terrorism and extremism studies (Bisgin et al., 2019; Greene & Lucas, 2020; Hernandez-Suarez et al., 2018; Scrivens et al., 2018). Social scientists can benefit from the new interdisciplinary field of computational social science, which is the development and application of computational methods, such as machine learning and natural language processing, to analyze complex, large-scale behavioral data (Hofman et al., 2021; Keuschnigg et al., 2018; Lazer et al., 2020). The digitization and availability of large-scale information generated by internet users (Hofman et al., 2021; Kemp, 2022) have rendered traditional manual methods for collecting and analyzing terrorist content challenging, time-consuming, and costly (Brynielsson et al., 2013; Scrivens et al., 2018). As such, social scientists working with text data can benefit from several statistical and computational techniques such as topic modeling algorithms latent Dirichlet allocation (LDA) that automatically uncovers topics and the hidden thematic structure of texts (Blei et al., 2003; DiMaggio et al., 2013; Maier et al., 2018; Walter & Ophir, 2019). Additionally, apply social network analysis (Borgatti et al., 2018; Scott & Carrington, 2014) to expose and map covert networks and structures (Koschade, 2006; Mullins, 2013; Perliger & Pedahzur, 2011).

The availability of digital data and the advancement of computational processing power and techniques has not only reduced the human effort in analyzing text data but allowed social science researchers to analyze large-scale text data, ask new questions, observe human interactions and behavior, thus advancing criminological research on terrorism, cyberterrorism, extremist online content, and counterterrorism. This dissertation aims to address this gap by exploring cyberterrorism and applying computational techniques and social networks to analyze Dark Web forums of international jihadists and their supporters.

1.2 Research Objectives

This three-article dissertation address three problems from an interdisciplinary, computational social science perspective, the first is an exploration of existing research on cybercrime and cyberterrorism. The second is methodological; providing insight into how researchers can analyze the large-scale text data generated by terrorists and their supporters. The article introduces and demonstrates topic modeling, specifically, Latent Dirichlet Allocation (LDA), in highlighting topics and themes from jihadists Dark Web forums. The third article explores Jihadists' network structure and behavior in Dark Web forums using social network analysis techniques.

The purpose of this three-article dissertation composed of three distinct but connected studies is threefold a) to explore the issue of cybercrime and cyberterrorism from a criminological perspective, b) highlight Jihadist topics and discussion on the Dark Web using innovative computational techniques, specifically topic modeling, c) illuminate and map Jihadist Dark Web forums' social structure and behavior using social network analysis. Additionally, the dissertation aims to contribute to the field of criminology innovative computational social science approaches for detecting terrorists' online contents and technology use and adds to the existing literature knowledge on cyberterrorism and online behavioral patterns of Jihadists and their supporters.

1.3 Research Organization

The dissertation is organized as follows, the first article, chapter two, explores the literature on terrorists' use of technology, cybercrime, and cyberterrorism. The nature of cybercrime and cyberterrorism are technological, criminal, and social, requiring analysis from a disciplinary and interdisciplinary lens. In this chapter, I provide an analysis of the problems of cybercrime and cyberterrorism within the field of criminology. The article focuses on (a) the issues in defining terrorism, cybercrime, and cyberterrorism, (b) ways that cybercriminals commit a crime in cyberspace, and (c) ways that cyberterrorists attack critical infrastructure, including computer systems, data, websites, and servers. The article ends with a discussion of cybercrime and cyberterrorism's future directions.

The second article, chapter three, is a methodological study introducing the application of topic modeling in analyzing large-scale text data. I make a case for applying natural language processing computational techniques, specifically LDA topic models and topic network analysis of text data. I begin by discussing machine learning, natural language processing, its challenges, and the various topic models algorithms. I then demonstrate the potential of topic models by inductively analyzing large-scale textual data of terrorist groups and supporters from three Dark Web forums to uncover underlying topics. Topic models' inductive approach and measurements are a powerful tool for the exploratory and descriptive text analysis for social scientists and data. I conclude the chapter by discussing the benefits and limitations of topic models in uncovering topics.

Chapter four, the third article, is an exploratory study of the social networks of Dark Web forums used by international Jihadist groups and their supporters. Using social network analysis, I construct undirected networks to analyze the network's structure, the interactions and participation patterns between members within the forums and identify the important members and virtual communities. The research questions guiding this study are: what is the network structure of jihadist Dark Web forums, and how does the structure affect interactions? What communities can be detected from the network? What are the posting activities of key members in the forums? The paper begins by describing the technological structure of the Internet, specifically the Dark Web. Next, the related literature in terrorists' use of the Internet, virtual communities and terrorists' online content, terror on the Dark Web, and social networks of terrorist groups. The last sections explore the research design, methods, data, and results. Finally, the article concludes with a discussion about Dark Web forums and their network structures. The dissertation concludes with a summary of the major findings, limitations, and future research on cyberterrorism and use of the Internet.

Chapter 2

Technology, Terrorism, and Crime: Exploring the Intersections of Cybercrime and Cyberterrorism

2.1 Introduction

The digital revolution and the development of new technologies, including the Internet, Web 2.0, artificial intelligence, Internet of Things, the metaverse, and social media networks, have rapidly transformed all aspects of social, political, economy, communications, and cultural life. More than 50% of the world's population is online, with about one million additional people joining the Internet each day (Forum, 2020; Kemp, 2019). As a result, cybercrimes and cyberterrorism are the unintended consequences of the digital revolution. Cyberspace has become a powerful platform for criminals. Cybercriminals and cyberterrorists are attacking computer systems and networks, critical infrastructures, stealing data and information, thus increasingly posing serious threats to national security, economy, and public policy. Progressing in number, scale, and consequences, cybercrimes are a challenge for intelligence and security agencies tasked with combating online crimes. According to the FBI Internet Crime

Complaint Center (IC3, 2020), in 2020, they received 791,790 complaints from Americans with a reported loss exceeding $4.2 billion, representing an increase of nearly 69% in total Internet crime complaints between 2019 and 2020. The most common types of cybercrime reported by victims include phishing scams, non-payment/non-delivery, extortion, personal data breach, and identity theft. Some victims also complained of ransomware, denial of service, malware, terrorism, and hacktivism (IC3, 2020). Cybercrime is predicted to inflict damages totaling $6 trillion by 2025 (Morgan, 2021).

Research examining the advancement of technology and terrorist groups has increased dramatically over the past decade, the focus has been chiefly on the use of the Internet, including social media platforms and websites for communication, financing, radicalization, and recruitment (Aly et al., 2014; Bermingham et al., 2009; Bloom et al., 2019; Gill et al., 2017; Hoffman, 2006b; Macdonald et al., 2019; Sageman, 2011; Scrivens et al., 2021; Weimann, 2011). However, there is a limited but growing body of criminological scholarship examining cybercrimes (Burruss et al., 2012; Holt, 2013; Holt & Bossler, 2009; Holt, Cale, et al., 2021; Howell et al., 2019; Lee & Holt, 2020; Maimon & Louderback, 2019). Few researchers have examined terrorist groups attacking digital infrastructures, stealing information, sensitive data, and finances to further their ideologies (K.-s. Choi et al., 2018; Holt, Cale, et al., 2021; Holt et al., 2019; Holt & Steinmetz, 2020). Some scholars consider such cyberattacks by terrorist groups an act of cyberterrorism (T. M. Chen et al., 2014; S. M. Furnell & Warren, 1999; Holt et al., 2019; Holt, Navarro, et al., 2020). Others debate whether cyberterrorism is a real threat (Conway, 2003, 2014; Denning, 2000, 2001; Pollitt, 1998; Weimann, 2004). Nonetheless, cyberspace remains a valuable borderless environment for criminal and terror activities, where the "likelihood of detection and prosecution is estimated to be as low as 0.05% in the United States" (Forum, 2020, p. 63).

Prior to the 9/11 terror attacks, criminological and sociological studies on terrorism were nonexistent, overlooked, and neglected (Armborst, 2010; Black, 2004; Deflem, 2004; Silke, 2004). The field was dominated by political scientists, international relations, and terrorism studies scholars, conducting research on terrorism from a political and psychological perspective. Furthermore, "criminology has been remiss in its research into the phenomena of cybercrime and has been slow to recognize the importance of cyberspace in changing the nature and scope of offending and victimization" (Jaishankar, 2007, p. 1). Despite the empirical gap, much has since changed within criminology and the research on terrorism. Criminologists are now researching terrorism, and a limited number of criminological studies examining cybercrime and cyberterrorism are published in top-tier criminological journals (A. M. Bossler & Berenblum, 2019; Bossler A.M., 2017; Maimon & Louderback, 2019). In addition, criminology scholars are advancing a new and developing field of criminological study known as cyber criminology, defined as "the study of causation of crimes that occur in the cyberspace and its impact in the physical space" (Jaishankar, 2007, p. 1). Thus, a multidisciplinary field relying upon criminology, sociology, and computer sciences (Diamond & Bachmann, 2015). However, as Diamond and Bachmann (2015) note, the field of cyber criminology is "largely ignored or marginalized by mainstream criminology, and that many criminologists refrain from examining this important, future-oriented issue" (Diamond & Bachmann, 2015, p. 25). Furthermore, the shortage of cybercrime and cyberterrorism research has been attributed to data limitations, the lack of official statistics on cybercrime and cyberterrorism incidents (A. M. Bossler & Berenblum, 2019; Bossler A.M., 2017; Diamond & Bachmann, 2015; Holt, Navarro, et al., 2020) and criminologists' unfamiliarity with the technology and tools for collecting online data (Diamond & Bachmann, 2015; Maimon & Louderback, 2019). Research on cybercrime and cyberterrorism is increasing despite these challenges, albeit still limited.

The nature of cybercrime and cyberterrorism are technological, criminal, social, and economical, requiring analysis from a disciplinary and interdisciplinary lens (Payne & Hadzhidimova, 2020). Drawing from previous research that highlighted the state of the field and gap in cybercrime (Maimon & Louderback, 2019; Payne & Hadzhidimova, 2020) and ideologically-motivated cyberattacks (Holt, Lee, et al., 2020; Holt et al., 2012; Holt, Turner, et al., 2021), this study aims to provide an analysis of the problems of cybercrime and cyberterrorism within the field of criminology. Focusing on (a) the issues in defining terrorism, cybercrime, and cyberterrorism, (b) ways that cybercriminals commit a crime in cyberspace, and (c) ways that cyberterrorists attack critical infrastructure, including computer systems, data, websites, and servers. Finally, the paper ends with discussing cybercrime and cyberterrorism's future directions.

2.2 Understanding terrorism, cybercrime, and cyberterrorism: Definitions

To understand the relationship between cybercrime and cyberterrorism, it is critical to understand terrorism and extremism. Unfortunately, there is no single consensus on terrorism, cybercrime, and cyberterrorism (Crenshaw, 2011; Hoffman, 2006a; LaFree & Ackerman, 2009; Schmid, 2004; Silke, 2004). Most books and studies on terrorism begin with discussing the definition of terrorism and its varieties (Silke, 2004). Consequently, studying terrorism is complicated, as it poses conceptual and methodological challenges as opposed to most types of criminal violence (Armborst, 2010; LaFree & Ackerman, 2009; LaFree & Dugan, 2015; Schuurman, 2019). Initially, criminologists shied away from studying terrorism, citing a) terrorism is different from other forms of violence, terrorism is not 'common' violence, despite terrorists being arrested and tried on murder, assault, weapon charges, b) terrorism cannot be explained by theories of violence, c) terrorism is political violence, criminological theories are intended to explain non-political violence (Rosenfeld, 2002). Yet, "while it seems reasonable that criminological theory alone cannot sufficiently explain terrorist violence, it is not convincing to exclude it from criminological research for this reason alone" (Armborst, 2010, p. 417). Neither does seeking first to find a "general theory" of terrorism necessary (Laqueur, 2004). Terrorism is a criminal matter (Hamm, 2007; Hamm & Van de Voorde, 2005), and terrorist groups engage in other types of crime, including bank robbery, international drug trade, drug trafficking, smuggling humans, and money laundering.

The FBI defines international terrorism as the "violent, criminal acts committed by individuals and/or groups who are inspired by, or associated with, designated foreign terrorist organization or nations" they also define domestic terrorism as "violent, criminal acts committed by individuals and/or groups to further ideological goals stemming from domestic influences, such as those of a political, religious, social, racial, or environmental nature" (FBI, 2021). In this study, I adopt the criminologist's operational definition of terrorism, which is "the threatened or actual use of illegal force directed against civilian targets by nonstate actors in order to attain a political goal through fear, coercion, or intimidation" (LaFree & Ackerman, 2009, p. 348). Hence, the varying definitions of terrorism are framed in the context of socio-political, religious, cultural identification and the use of violence, fear, and threats to accomplish the ideological goals (Borum, 2011; LaFree & Dugan, 2009; Silke, 2004).

In addition, there are varying definitions of the term extremism. The definitions are based in the context of violent extremism and are frequently used interchangeably with the term terrorism (J. M. Berger, 2018; J. Berger, 2017). However, extremism is not always violent and associated with non-state actors. J. Berger (2017) defines extremism in terms of in-group/out-group relations, perceptions, and beliefs. Thus, extremism is "a spectrum of beliefs in which an in-group's success is inseparable from negative acts against an out-group. Negative acts can include verbal attacks and diminishment, discriminatory behavior, or violence" and violent extremism as "the belief that an in-group's success is inseparable from violence against an out-group. A violent extremist ideology may subjectively characterize this violence as defensive, offensive, or pre-emptive" (J. Berger, 2017, p. 6). Ideologies are the central element of violent and non-violent extremism (McCauley & Moskalenko, 2008). Categorized as international terrorism, individuals and groups espousing Islamist ideologies practiced by jihadists seek a violent military strategy to accomplish their goal of establishing an Islamist society. Often categorized as domestic terrorism, the far-right ideology is espoused by White supremacists, militia, and militant gun rights adhering to racial supremacy and anti-government beliefs. In contrast, the far-left groups advocate for workers and animal rights, environment, and race-based issues.

There is no agreement on what constitutes cybercrime. As such, the debate on defining cybercrime centers around whether cybercrime should be conceptualized as a new crime type or traditional crime conducted through a new medium, in this case, technology (Diamond & Bachmann, 2015; Holt, 2016; Holt et al., 2012). For example, Furnell and colleagues (2015) define cybercrime as cyber-dependent and cyber-enabled. Cyber-dependent crimes are offenses that can only be conducted using a computer, computer networks, or technologies such as spreading viruses, malware, hacking, and distributed denial of service (DDoS) attacks directed against computers or networks. In contrast, cyber-enabled crimes are traditional crimes that use computers or computer networks to widen their reach or scale. Cyber-enabled crimes include fraud, theft, and sexual offending (S. Furnell et al., 2015).

Some scholars define cybercrime based on four types of activities and behaviors (Holt & Bossler, 2014; Wall, 2001a). The four categories include cybertrespass or hacking, which is the unauthorized access of a networked system or computers. It is considered an intrusion when an individual accesses a computer system, network, or data without the owner's permission. Hackers are known to access the networks of individuals, businesses, and governments without authorization.

Cyber-deceptions or cyber-theft include identity theft, online fraud, and digital piracy. The use of the Internet to steal information, credit and debit cards, money from bank accounts belonging to individuals or businesses. Cyber-deception also includes crimes simplified by the internet but do not require it. For instance, the Nigerian or 419 messages and work-at-home schemes using spam emails have adapted the offline fraud schemes to the online environments. Additionally, digital piracy, which is the stealing of intellectual properties, has been simplified by the Internet and technology.

Cyber-pornography/obscenity crimes include the distribution of sexually explicit materials online, child sexual exploitation materials, and the virtual sex trade. Cyber-violence causes psychological harm to victims through cyberstalking, cyberbullying, harassment, and hate speech. Extremists and terrorist groups also engage in acts of cyberviolence (A. M. Bossler & Berenblum, 2019; Holt, 2016; Holt et al., 2012; Wall, 2001b). Based on these definitions and categorizations, the range of activities viewed as cybercrimes are challenging for any definition of cyberterrorism, and terrorist groups may also engage in the same acts as non-ideologically motived cybercriminals (Holt et al., 2012).

The term cyberterrorism has sparked an intense debate over whether terrorist groups are capable of attacking cyberspace and critical infrastructures to further their goals (Conway, 2003, 2014; Hardy & Williams, 2014; Jarvis & Macdonald, 2014; Weimann, 2004, 2005). Similar to the term terrorism, various definitions of cyberterrorism exist. Cyberterrorism referred to the convergence of cyberspace and terrorism and was coined by Barry Collin in the 1980s (K.-s. Choi et al., 2018; Conway, 2014; Denning, 2000, 2001). Subsequent definitions have focused on the previous definitions and challenges in defining terrorism. Denning (2000, 2001) defines cyberterrorism as the

> "Unlawful attacks and threats of attacks against computers, networks and the information stored therein when done to intimidate or coerce a government or its people in furtherance of political or social objectives. Further, to qualify as cyberterrorism, an attack should result in violence against persons or property or at least cause enough harm to generate fear. Attacks that lead to death or bodily injury, explosions, or severe economic loss would be examples. Serious attacks against critical infrastructures could be acts of cyberterrorism, depending on their impact. Attacks that disrupt nonessential services or that are mainly a costly nuisance would not" (Denning, 2000, p. 29).

It is implied that for cyberterrorism to occur, terrorist groups must attack computers or networks, for instance, attack a power grid or access air traffic control leading to death, injury, or economic loss. In addition, (Conway, 2014) argues contrary to what the media reports or government authorities, per this definition, terrorist groups sending pornographic emails to minors, defacing web pages, posting offensive content on the Internet, stealing credit card information does not constitute cyberterrorism. Thus, an attack only qualifies as cyberterrorism if it fits the definition of terrorism. However, some scholars argue terrorist groups can commit cyberterrorism by targeting government or businesses' online critical infrastructures (S. M. Furnell & Warren, 1999; Holt et al., 2019; Holt, Lee, et al., 2020). Furthermore, law enforcement and government agencies acknowledge the capabilities and potential cyberattacks by terrorist groups (Holt, Lee, et al., 2020). In this case, cyberterrorism can be understood as "the use of computer network tools to shut down critical national infrastructures (such as energy, transportation, government operations) or to coerce or intimidate a government or civilian population" (Yannakogeorgos, 2014, p. 44). This definition does not include violence as the end goal; thus, cyberterrorism should be distinctly defined from other forms of terrorism. Therefore, "while there is no single agreed upon definition for cyberterror, it is clear that this term must encapsulate a greater range of behavior than physical terror due to the dichotomous nature of cyberspace as a vehicle for communications as well as a medium for attacks" (Holt et al., 2012, p. 341). An act of cyberterrorism should not necessarily result in physical harm or intended to generate fear if its intention is the interference of the political, social, or economy of a nation or lead to physical violence.

Weimann (2004, 2005) claims at present, terrorists are using the internet and benefit from it rather than attacking it. Yet, cyberterrorism is an attractive option for terrorists who value its potential to inflict massive damage, its psychological impact, and the media appeal. No one has yet been physically hurt or killed by cyberterrorism. However, this does not mean terrorist groups are not interested in cyberterrorism. Nevertheless, whether terrorist groups weigh the benefits and costs of a cyberattack or possess the technical skills to attack critical infrastructure instead of bombing or suicide should not be the focus of a debate. The fact that the groups have access to the Internet, can learn technical skills, or hire hackers deserves serious attention from scholars, governments, and national security policymakers. The lack of consensus on the definitions or elements comprising cyberterrorism should not be the focus of disciplinary and interdisciplinary debate but rather on the threat and changing dynamics of cyberspace and technology and the motivations and capabilities of terrorist groups to engage in cyberterrorism.

2.3 Cybercrime and Cybercriminals

Criminological research examining cybercrimes has increased over the past decade and addresses a variety of online offenses, including hacking, website defacement, malware, piracy, and cybercrime offenders (Holt & Bossler, 2009; Holt, Turner, et al., 2021; Maimon & Louderback, 2019; Wall, 2001b), advance-fee fraud and romance scams (Dion, 2010; Holt & Graves, 2007; Kopp et al., 2016; Whittaker & Button, 2020), identity theft (Archer, 2012; J. Choi et al., 2021; Gies et al., 2021; Rudner, 2008) and cyberbullying (Kerstens & Veenstra, 2016; Marcum & Higgins, 2012; Su & Holt, 2010; Zhang et al., 2021). Some scholars have analyzed the application and testing of criminological theories in explaining cybercrime, such as the explanatory power of self-control theory and routine activities theory (Back et al., 2018; A. Bossler & Burruss, 2011; A. Bossler & Holt, 2009; Howell et al., 2019; Kigerl, 2011; Leukfeldt & Yar, 2016; Louderback & Antonaccio, 2020; Reyns et al., 2011), social learning theory (Holt, 2010; Li et al., 2016; Miller & Morris, 2016; Morris & Higgins, 2010; Skinner & Fream, 1997), neutralization theory (Higgins et al., 2008; Morris, 2011) and strain theory (Hay et al., 2010; Jang et al., 2014). Cohen and Felson (1979) routine activities theory has been "applied in most of the victim-based studies of cyber-dependent crime" (Maimon & Louderback, 2019, p. 201), finding moderate support for the theory. As a result, "while support for different theories exists, no apparent general theory of cybercrime explains all of the offenses. Instead, some of them appear to be better suited for certain types of crimes" (Payne & Hadzhidimova, 2020, p. 82). Indeed, finding the general theories of crimes inadequate in explaining cybercrime, cyber criminologist Jaishankar (Jaishankar, 2007) developed the space transition theory to explain the causation of crimes within cyberspace. The theory claims people behave differently (conform or non-conform) when moving from one space to another, i.e., from physical space to cyberspace and vice versa.

Hacking is the starting point from which cyber-dependent crimes begin, whereby criminal hackers deliberately access a computer system, networks, and data unauthorized (S. Furnell, 2003; S. Furnell et al., 2015; Maimon & Louderback, 2019; Schell & Dodge, 2002). Once hackers gain access to a computer system, they may develop or spread viruses and malicious software, launch website defacement or distributed denial of service attacks, destroy or alter files, redesign or configure hardware and software systems (Marcum et al., 2014). In addition, criminal hackers are behind major ransomware attacks in businesses and governments, posing infrastructure and human risk.

Cybercriminals' predominant hacking tools and tactics include phishing or spamming, a social engineering attack where an attacker uses social skills to obtain and compromise information from an individual or organization's computer system. For instance, a user may click a malicious link or file in an email from a 'reputable' credit card company or bank, allowing hackers access to a personal account or financial information. A typical cyberattack is the use of malicious software or malware; this occurs when a virus, worm, or Trojan horse is spread to gain access into a computer system. Ransomware is malware that extracts and encrypts data making it inaccessible and unusable. The victim must first pay a ransom, often using cryptocurrency before the information or system is released and available. It is essential to note the malware software is automated. It infects the computer, takes the victim's money, and deposits it to the attacker's account without the attacker being present or near the victim's computer. Malware can be spread via email attachments, advertisements, fake software installations, phishing emails, text messages, infected USB drives, and apps.

Web defacements or vandalism occurs when an attacker changes the "front face" of the website with their messages, images, or text. Such defacement attacks are meant to cause the website to be offline and unavailable, damage the reputation or brand of the individual, organization, or government. Distributed Denial of Service (DDoS) attacks intentionally overwhelms a website, usually a government or business website, by disrupting the regular traffic of the server, service, or network, thus becoming unavailable to users (Hayward & Maas, 2021; Holt, 2010; Holt & Steinmetz, 2020; Singer & Friedman, 2013).Cybercrime has become more professional, automated, complex, and stealthy (Wall, 2015). Furthermore, the technology used to commit cybercrime is commercialized, and experienced hackers can be hired to commit the crime. While some hacker tools may be free to download, the better ones are available for sale or rent (Grabosky, 2016). For instance, the Cybercrime-as-a-Service (CaaS) kits sold on the Dark Web enable cybercriminals to attack cyberspace without the technical expertise of the computer or systems. The commercialization of technology tools for cybercrime may be due to the low technical capabilities and skills of most hackers but also provides cybercriminals with access to vulnerable and compromised systems with the possibility of evading detection or arrest.

Hackers are motivated by curiosity, technical achievement, fun, addiction, power and prestige, peer associations, ideology, and, most importantly, monetary gain (Grabosky, 2016; Weulen Kranenbarg et al., 2018; Weulen Kranenbarg et al., 2019). Within the hacker subculture, norms such as privacy and secrecy, meritocratic value, knowledge distribution networks exist (Holt & Graves, 2007; T. Jordan & Taylor, 1998; Steinmetz et al., 2020). In addition, hacking is dominated by males; often a male world, young, technologically oriented, and unfriendly to women (Adam, 2005; Bachmann, 2010; Holt, Navarro, et al., 2020; Hutchings & Chua, 2016; T. Jordan & Taylor, 1998; Steinmetz et al., 2020; Turkle, 2005). Hacking behaviors begin during early adolescence (Holt, 2007; T. Jordan & Taylor, 1998; Lee & Holt, 2020), and many cybercriminals may work alone, but "a great deal of cybercrime is the work of organizations" (Grabosky, 2016, p. 37). Furthermore, cybercrime organizations have structures and roles including "disorganized groups of cyber-dependent offenders that mainly operate in online environments yet have no clear chain of command (i.e., a swarm structure) and organized groups with clear leadership and command structure (i.e., a hub structure)" (Maimon & Louderback, 2019, p. 195). For instance, the hacktivist group Anonymous, which hacks for a political cause, is decentralized and coordinated (Singer & Friedman, 2013). Anonymous have launched distributed denial-of-service (DDoS) attacks against the government and prominent organizations' websites. These include attacking the Australian Parliament, the US Federal Bureau of Investigation, the US Department of Justice, PayPal, MasterCard, Visa (Hardy & Williams, 2014). Hackers can also be classified as white hat hackers, employed to test system security thus authorized, versus black hat hackers who access systems unauthorized. Since cybercrime is a global phenomenon and a transnational business, the offenses can originate globally, and offenders can easily cross borders without detection (Payne & Hadzhidimova, 2020).

2.4 The cybercrime-cyberterrorism nexus: Relationship between cybercrime and cyberterrorism

The ability to attack cyberspace has a far-reaching impact on terrorist groups' enemies, unlike bombings restricted to a specific physical location and communities (Carlin & Graff, 2018; S. M. Furnell & Warren, 1999). Terrorists use the Internet to further their goals and advance and adapt to technological changes. As a result, cyberspace is indeed "the new great equalizer, the Internet of Things is the new battlefield, anonymizing tools are the new weapons for stealth, malware is the new air strike, vulnerabilities are the new focus of attack, and robust zero day arsenals are the new symbols of supremacy as the world order is technologically toppled and supplanted by the new" (Scott & Spaniel, 2016, p. 2). Although terrorist groups have yet to cause massive loss of life or economic chaos by hacking into critical infrastructure systems, they have engaged in cyber-attacks against websites and other non-essential infrastructure through hacking (Carlin & Graff, 2018; K.-s. Choi et al., 2018; Hardy & Williams, 2014; Holt et al., 2017a; Holt et al., 2012). Indeed, the first recorded cyberattack by terrorist groups was by the Tamil Tigers, who carried out a denial-of-service attack against

Sri Lankan embassies worldwide (S. M. Furnell & Warren, 1999). Recently, in 2020, the US Justice Department, the Department of Homeland Security, and the Department of Treasury dismantled three terrorist cyber-enabled finance campaigns. The Al-Qassam Brigade terror group ran an online cryptocurrency fundraising effort soliciting bitcoins from its supporters. Al-Qaeda and its affiliated groups also operated a bitcoin money-laundering network using Telegram channels and other social media platforms and a scheme by an ISIS facilitator to sell fake masks online for funding purposes (ICE, 2020).

However, less research has explored cybercrimes and similar methods conducted by terrorist groups and ideologically motivated cyber attackers (T. M. Chen et al., 2014; Holt et al., 2019; Jarvis & Macdonald, 2014; Yannakogeorgos, 2014). As such, the dearth of cyberterrorism research poses challenges in developing counterterrorism policy and prevention strategies related to ideologically motivated cyberattacks (Holt et al., 2017b). There is no difference in skills and techniques between cyberterrorists and cybercriminals. Both require and utilize the same techniques for attacking cyberspace. However, cyberterrorists are motivated by political or ideological issues (S. M. Furnell & Warren, 1999; Holt et al., 2012). Although cyberterrorists are also known to hire and fund available hackers to carry out attacks on their behalf (Carlin & Graff, 2018), some hackers may not believe in the terrorist's cause but only work for financial gain, signifying collaborations between criminals and terrorist groups. In addition, some of the terrorist supporters and group members possess less technologically sophisticated skills and are taught by the skilled hackers "operating ISIS's Cyber Help Desk" (Scott & Spaniel, 2016, p. 2). As a result, cyberattacks may require only a few skilled hackers to train the rest and fewer financial resources and workforce.

Furthermore, terrorist groups such as ISIS recruit technology-savvy hackers known as cyber-jihads, members of the Cyber Caliphate, a hacker division of ISIS (Scott & Spaniel, 2016). The group mostly attacked websites by replacing the content with the ISIS flag and phrases. The digital jihadis are believed to have begun with the British terrorist of Pakistan descent, Junaid Hussain (Carlin & Graff, 2018). With a group of seven friends, Junaid formed a hacker group called TeaMpOisoN. The group became famous in 2011 after they defaced/vandalized websites with pro-Palestine messages. In addition, they attacked websites belonging to NASA, NATO, UN, and figures such as former Prime Minister Tony Blair. Hussain was a prominent recruiter within the Islamic State, and it is believed he is responsible for developing much of their cyber and social media strategies. He may also have been developing al-Qaeda's Dabiq and Kybernetiq publications on information technology, communication, and security. Another hacker, Ardit Ferizi, accessed U.S computer systems to obtain personal information on 1,300 military and federal employees and then provided this information to ISIS members (Holt et al., 2017a). More cyber jihadis groups have been formed, including the Cyber Caliphate Army and United Cyber Caliphate.

As mentioned elsewhere, known cyber-attacks by terrorist groups include web defacements and information release, killing lists of military personnel with the hope that lone wolf recruits will conduct attacks on the targets, and credit card fraud to finance their cause. The defacements are meant to scare Western businesses and organizations than recruit new followers (Scott & Spaniel, 2016). Distributed denial of service (DDoS) is another method terrorist groups may use to attack websites and servers. This attack involves flooding requests to the website and server, rendering it 'busy' and unusable to others, causing financial harm to the victim (Holt, Turner, et al., 2021). The economic costs for the victims include damage and destruction of data, deletion of hacked data and systems, loss of revenue from site downtime, productivity costs, and reputational damages (Holt, Lee, et al., 2020).

Additionally, supporters of Al-Qaeda also operate web forums to distribute hacker tools and coordinate attacks. For instance, the hacker Younis Tsoulis promoted hacking tools against various targets in support of global jihad. Using the handle Irhabi 007, or Terrorist 007, he published a manual entitled "The Encyclopedia of Hacking the Zionist and Crusader Websites," detailing attack methods and a list of vulnerable targets online (Holt et al., 2012, p. 344). As a result, Tsoulis and his compatriots became the first people in Britain to be convicted of conspiring to commit terrorism on the Internet (Carlin & Graff, 2018).

Indeed, as Verton and Brownlow (2003) claim, “the cyber-threat from terrorist organizations such as al-Qaeda does not hinge on the willingness of the core leadership or even the most dedicated and radicalized jihadists within the movement to adopt cyber-tactics as a main attack method. There are already a vast number of cyber-based alliances of hackers and internet activists (aka “hacktivists”) cyber-jihad movement, or electronic holy war, against the West” (Verton & Brownlow, 2003, p. 101). Like the cyber jihadis, far-left extremists have also engaged in website defacement campaigns against industrial and governmental targets to shame their victims for harming animals and the environment. The far-right groups have also defaced websites serving Holocaust memorial sites (Holt, Lee, et al., 2020). Indeed, “such actions would correspond in part to physical terror and extremist violence, which target infrastructure and persons who symbolically represent their ideological beliefs” (Holt, Lee, et al., 2020, p. 4). As such, high-profile websites such as those belonging to governments, businesses, and organizations are likely targeted by terrorist groups and extremists.

Terrorist groups are developing their cyberattacks strategies. For instance, the “lower-level actors such as Boko Haram are upgrading their 419 scams with ransomware, RaaS, and MaaS” (Scott & Spaniel, 2016, p. 2). Primarily by using Ransomware as a Service (RaaS) and Malware as a Service (MaaS) which ransomware developers lease to people without much technical knowledge and skills enabling them to launch ransomware attacks just by paying for the service (Europol, 2021; Ledesma, 2021). Ransomware remains a threat as cybercriminals tend to increase pressure by threatening the publication of data if victims do not pay (Europol, 2021), affecting not only the organizations or governments but those whose data is compromised. The prepackaged exploitation of vulnerabilities such as malicious software or malware (Singer & Friedman, 2013) is significant for terrorist groups. Mainly since “the use of known vulnerabilities may be associated with semi-skilled attackers as they may be able to utilize existing, pre-written attack programs to compromise vulnerable systems. In fact, a number of attack tools available for download include exploitable code for known vulnerabilities so as to facilitate attacks by lower-skilled cyberattackers” (Holt, Lee, et al., 2020, p. 5). The dark web is the marketplace for malicious software and hacker tools, where cybercriminals and terrorists buy and sell the tools and services needed for a cyberattack (Singer & Friedman, 2013; Weimann, 2016a).

The technical sophistication and capabilities of terrorist groups to engage in cyberterrorism is often the central contention in the cyberterrorism debate (Evan et al., 2017; Singer & Friedman, 2013). The extent to which terrorist groups may engage in cyberterrorism is based on their technical capabilities and sophistication. For instance, according to Scott and Spaniel (2016), Al-Qaeda relied on more innovative partner terrorist organizations such as the Tunisian Cyber Army for cyber-attacks. In 2015, the group developed its cyber-arm that performs electronic jihad operations. The group was active on Twitter and video distribution sites but did not possess the social media presence, recruiting capabilities, and technical sophistication such as ISIS. The less technically sophisticated Al-Shabab shares many strategies with Al-Qaeda. Al-Shabaab predominantly uses the Internet to disseminate propaganda and recruitment. However, Al-Shabaab's "disorganized internet strategy and lack of technical proficiency prevent it from posing a real threat in cyberspace" (Scott & Spaniel, 2016, p. 8). Since Boko Haram switched allegiance to ISIS, Boko Haram's social media presence and the distribution of propaganda have become more sophisticated. However, the group still raises funds through fee fraud or 419 scams. ISIS is the most sophisticated group. It has recruited a following of technology-savvy kiddies and wannabe hackers (Scott & Spaniel, 2016). ISIS has dedicated its offensive cyber capabilities to compromise social media accounts and websites belonging to individuals, businesses, and government organizations. They have also used malware, preconfigured tools, and insider threats (Scott & Spaniel, 2016). Thus, the groups are also innovating and adapting to technological changes.

On the other hand, Evan et al. (2017) developed three levels of capability scales, enabling, disruptive and destructive, to classify the cyber-capabilities of these groups. The groups use the Internet for publicity, propaganda, recruitment, communication, and disseminating manuals/magazines online within the enabling level. The disruptive level is online activities disrupting the computer systems, e.g., cyber breaches, malware, financial theft and fraud, denial of service attacks, phishing, and web defacements. Within the destructive level, the groups engage in cyber-attacks that trigger physical damage or injury through sensor spoofing technology, digital control systems, disabling control, and safety systems. However, Evan et al. (2017) claim most terrorist organizations such as ISIS, Al-Qaeda, Al-Shabaab, Boko Haram possess the 'enabling' cyber capability as they have websites, social media accounts, and media. They have some 'Disruptive' abilities, as seen by the cyber breaches, web defacement, and financial fraud. However, none of the groups have shown that they possess the higher destructive capability yet. An organization would have to have supporters and group members possessing advanced hacking and engineering skills for higher cyber capability. So far, none of the terrorist groups have this capability. However, "the imminent acquisition of cyber capabilities by terrorist groups has been long expected but has so far failed to materialize, and there have been no known terrorist attacks using cyber means to trigger physical damage and destruction" (Evan et al., 2017, p. 4). Overall, the benefits of cyberterrorism include reduced risk of capture, inflicting financial damage without loss of life, hackers and tools are available for hire, and the groups may widely publicize the successful attack while failure would go unnoticed. As such, researchers should not perceive the threat of cyberattacks by extremists and terrorist groups as unrealistic or non-existent. Evidence suggests that terrorist groups have the skills and means to engage in cyberattacks and use the Internet to further their goals.

2.5 Conclusions: Future Directions for Cybercrime and Cyberterrorism Research

Due to the rapid technological advancement and emerging technologies, including artificial intelligence, and the unintended consequences of these technologies, it is challenging to predict cybercriminals and cyberterrorists' technical capabilities and cyberattacks. Indeed, while there have been few incidents of cyberterrorism globally, mostly attacks against government, organizations, and businesses networks, we cannot ignore the threat of severe cyberattacks and critical infrastructures posed by terrorist groups. Countries face a growing risk of cybercrime and cyberterrorism as we rely more on the Internet and technologies for communication, business, health, economy, governing, and critical infrastructures. While terrorism scholars have been hesitant in acknowledging that terrorists engage in cyberterrorism, law enforcement and government agencies do recognize terrorists and extremist groups may target government and civilian managed online critical infrastructure (Holt, Lee, et al., 2020; Holt, Turner, et al., 2021).

The lack of access to official data on cyberattack incidents is one of the reasons for the dearth of cyberterrorism and cybercrime research (A. M. Bossler & Berenblum, 2019; Bossler A.M., 2017; Diamond & Bachmann, 2015; Holt, Lee, et al., 2020). A few official data sources provide data on computer crimes such as hacking. Industry or open-source government data "are rarely available due to underreporting on the part of victims, particularly within the private industry over fears of economic loss and a decrease in consumer confidence" (Holt, Lee, et al., 2020; Holt, Turner, et al., 2021, p. 6). Businesses and organizations tend not to report cyberattacks to the police or the public. Furthermore, cyberattack incidents are often omitted from terrorism databases "due to their failure to qualify as a "violent" incident" (Freilich et al., 2014; Holt, Lee, et al., 2020; Holt, Turner, et al., 2021). As such, most scholars have resorted to examining a sample of college and school students, mainly surveying their hacking and digital piracy experiences, analyzing bulletin boards and blogs, and field experiments (A. M. Bossler & Berenblum, 2019; Bossler A.M., 2017; Burruss et al., 2012; Holt & Steinmetz, 2020). Perhaps the innovative data collection from honeypots, simulated computer systems created to be attacked, researchers then analyze the data on system intrusions (Maimon et al., 2013; Testa et al., 2017; Wilson et al., 2015) could solve the lack of data challenges. However, while honeypots collect data on behavior (Bossler A.M., 2017), this type of data collection has limitations, including the inability to collect demographic data from the actors.

There are currently no existing statutes for cyberterrorism within US federal criminal code. Terrorists engaging in cyberattacks are often prosecuted under existing computer hacking criminal laws (Holt, Lee, et al., 2020; Holt, Turner, et al., 2021). Thus, to defeat cyberterrorism globally, creating a "legal framework for prosecution with a strong foundation in international law" (Stockton & Golabek-Goldman, 2014, p. 214) would be significant for pursuing and prosecuting cybercriminals. Perhaps, cybercrime and cyberterrorism researchers may consider a collaborative research analysis with legal scholars to analyze the existing laws and contribute to the implementation of relevant policies on cybercrime and cyberterrorism from the perspectives of criminology.

Crime is one of the unintended consequences of old and new technologies. The field is changing, criminological research on cybercrime is progressing, researchers and colleges are advancing and offering a new subfield of cyber criminology. Research clusters are also developing, for instance, digital criminology (Powell et al., 2018), computational and Big Data criminology (Chan & Bennett Moses, 2016; Smith et al., 2017), and artificial intelligence crime (AI-Crime) (Hayward & Maas, 2021; King et al., 2020). Research groups are formed in educational institutions such as the Evidence-Based Cybersecurity Research Group at Georgia State University, which researches human factors in cybersecurity. Additionally, a significant and growing number of cybercrime and cyberterrorist presentations at national and international conferences exist (A. M. Bossler & Berenblum, 2019). The American Society of Criminology (ASC) Division of Cybercrime was formed in 2019. Undergraduate, graduate, and certificate offerings in cybercrime are increasing. Nodeland et al. (2018) argue, "students pursuing careers in criminal justice must be prepared to respond to modern threats including, but not limited to, a range of cyber-related offenses (e.g. cyberbullying, cyberstalking, sexting and cybersex crimes) as well as the traditional hacker, identity theft, terrorists, and nations wishing harm on the United States" (Nodeland et al., 2018, p. 71). Yet, most universities are lagging in their educational and training opportunities in the field of cybersecurity for their criminal justice students (Nodeland et al., 2018).

Social scientists researching cybercrime tend to focus on the offenders and victims while placing less emphasis on the technical aspects of cybercrime (Holt, 2016). As a result, less criminological research in cybersecurity focuses on securing cyber systems and preventing cybercrimes. Technical analysis has been left to computer scientists with less input from the social sciences. Thus, "it is vital to identify potential methods for proactive actor motivation detection by cybersecurity professionals based on actor behavior during and after the incident is complete to better defend against future attacks and minimize harms" (Holt, Turner, et al., 2021, p. 544). Indeed, the "behavioral side of cybersecurity needs more research and can improve faster if it is integrated with human factors, and benefit from sophisticated modeling and simulation techniques" (Maalem Lahcen et al., 2020, p. 2). As Hayward and Maas (2021) note, "if many computer scientists are understandably guilty of concentrating their efforts on the technology side of what we might call here the 'tech-crime nexus', then the same is true in reverse of criminologists. While our discipline continues to advance knowledge about crime and punishment, it does so largely oblivious of the many social challenges posed by technological disruption" (Hayward & Maas, 2021, p. 210). The study of cybercrime and cyberterrorism requires a multidisciplinary focus and collaboration. Criminologists must collaborate with computer scientists and industry to develop knowledge on cybercrime, cyberterrorism, and cybersecurity to better anticipate cybercriminal activities and prevent, respond to, and mitigate cyberattacks.

With the advancement of artificial intelligence technologies and algorithms, new crimes have been developed, and new types of policing, punishment, and legal decision-making (Brayne, 2020; Brayne & Christin, 2021; Hayward & Maas, 2021). Therefore, criminology research should also focus on the policing and security practices that use technologies and algorithms, such as predictive policing and surveillance technologies, and how these tools are used to investigate criminals of all types, including cyberterrorists. The U.S criminal justice system uses predictive technologies. For instance, police departments use predictive software programs to predict where and when future crimes will occur and target potential victims and offenders (Brayne, 2020; Brayne & Christin, 2021; Ferguson, 2017). Indeed, as Ferguson (2017) argues, "new technologies threaten to impact all aspects of policing, and studying the resulting distortions provides a framework to evaluate all future surveillance technologies" (Ferguson, 2017, p. 4). Hence, criminological research should also focus on the potential harms of predictive and surveillance technologies used by law enforcement.

Finally, social scientists, including criminologists, can also benefit from the adoption and application of a computational social science (Y. Chen et al., 2021; Hofman et al., 2021; Keuschnigg et al., 2018; Lazer et al., 2020; Molina & Garip, 2019; Salganik, 2019) approach for the analysis of crimes in cyberspace. The technological revolution has led to the digitization of social, economic, political, and cultural activities of billions of Internet users, thus generating "vast repositories of digital data as a byproduct" (Hofman et al., 2021, p. 181). While existing research methodologies, theories, and data have proven utility in examining crime, new innovative interdisciplinary and borrowed techniques from computer science and data science are necessary to analyze big data, including text and image data. Future research applying computational and network techniques to explore the online content and messaging of terrorist groups and their supporters on the Dark Web is imperative for counterterrorism. As terrorists have migrated to encrypted platforms such as the Dark Web, which is also a cyberthreat, it has become challenging for law enforcement and security agencies to monitor, detect, and analyze their use of encrypted platforms. The groups use the Dark Web marketplaces to hire hackers and purchase hacking tools to attack cyberspace. The Dark Web forums also provide secure communication and discussion platforms for the groups and their supporters. The following studies examine the topics and networks of Jihadists Dark Web forums from a computational social science approach.

Chapter 3

Text as Data: Topic Modeling of Dark Web Forums

3.1 Introduction

As of January 2022, over half of the global population is online, and Internet users spend more than 40 percent of their day online. As such, billions of Internet users' social, economic, political, and cultural activities have been digitized, generating vast amounts of data (Hofman et al., 2021; Kemp, 2022) that can also be used for descriptive and causal inferences. Traditional research methods and techniques may not analyze the influx of big data. The text and image big data generated from search and social media platforms are often noisier and unstructured, not fitting tidily into the row and column structure of traditional social science data such as surveys and experiments (Hofman et al., 2021). To analyze big data, social scientists can benefit from the new field of computational social science, which is the development and application of computational methods, such as topic modeling, to analyze complex, large-scale behavioral data (Hofman et al., 2021; Keuschnigg et al., 2018; Lazer et al., 2020).

Natural language processing (NLP), a subfield of computer science, uses computational techniques to learn, understand and produce human language. The computational analysis of text data has grown over the past 20 years mainly due to increased computing power, the availability of large amounts of text data, and the development of successful machine learning (ML) methods (Hirschberg & Manning, 2015). NLP combines concepts from linguistics, statistics, machine learning, and deep learning enabling computers to understand spoken and written human language. Some NLP applications on text data include sentiment analysis that measures subjective content such as attitudes and emotions; and named entity recognition that identifies valuable words or phrases such as names and locations. As one of the NLP approaches, topic modeling is a computational technique that uncovers topics and the hidden thematic structure of a collection of texts. As a result, social scientists working with text data can benefit from several statistical topic modeling algorithms, such as the latent Dirichlet allocation (LDA) (Blei et al., 2003; DiMaggio et al., 2013; Maier et al., 2018; Walter & Ophir, 2019). Although there are several topic modeling algorithms, latent Dirichlet allocation (LDA) (Blei et al., 2003) is one of the widely used statistical algorithms for analyzing large-scale text data to find the topics and thematic structure. Most importantly, LDA's inductive approach and quantitative measures make it suitable for the exploratory and descriptive analyses of text data (Maier et al., 2018). Additionally, LDA models are useful for automatically detecting and removing terrorist content from online platforms.

The traditional approach of manual hand-coding text data is limited in terms of resources, the analysis is time-consuming, and concerns arise about biases of researchers or challenges in achieving acceptable intercoder reliability scores, especially when there is a large corpus to analyze (DiMaggio et al., 2013; Roy & Rambo-Hernandez, 2021). It is important to note, however, manual coding of text can be applied to analyze smaller samples of text data and effectively discover themes. Furthermore, the qualitative approach to text data using grounded theory (Glaser Strauss, 1967) and the quantitative approach from natural language processing (Blei et al., 2003) involve similar processes and produce similar results

(Baumer et al., 2017). However, the availability of digital data and the advancement of computational processing power and techniques have reduced the human effort in analyzing text data and allowed social science researchers to analyze large-scale text data, ask new questions, and observe human interactions and behavior.

In this methodological study, I make a case for applying natural language processing computational techniques, specifically LDA topic models and topic network analysis of text data. I demonstrate the potential of topic models by inductively analyzing large-scale textual data of terrorist groups and supporters from three Dark Web forums to uncover underlying topics and sentiments. I argue that topic models' inductive approach and measurements are a powerful tool for the exploratory and descriptive text analysis for social scientists and data. The organization of this article is as follows. I begin by discussing machine learning, natural language processing, its challenges, and the various topic models algorithms. In the next section, to highlight the utility of topic models, I provide an example of topic models by analyzing Dark Web forums of jihadist terrorist groups. I conclude by discussing the benefits and limitations of topic models in uncovering topics.

3.2 Machine Learning and Natural Language Processing

Within the field of artificial intelligence, machine learning (ML) focuses on developing algorithms and computer systems that improve automatically through experience and the use of data (M. I. Jordan & Mitchell, 2015). In ML, the algorithm learns from training data and improves as it learns more from processing data. Machine learning is an interdisciplinary field comprising computer science and statistics. It is used to develop software for computer vision, speech recognition, natural language processing, other applications, and predictions and inferences. Machine learning approaches in text analysis can be categorized into supervised, unsupervised, and semi-supervised learning methods with inputs from both humans and machines. In supervised machine learning, the algorithm 'learns' the patterns from input (training) data to make predictions. For instance, a researcher can use available labeled text data to train the model, then the model learns it and classifies the rest of the unseen (test) data. In unsupervised learning, algorithms learn from unlabeled data to find hidden patterns. Here, the researcher's role is to interpret the model's results. Finally, the algorithms learn from both smaller labeled and larger unlabeled datasets in semi-supervised learning to find patterns. Here the researchers' label/annotate a smaller sample first to train a model (Lucas et al., 2015; Vajjala et al., 2020).

Another subfield of artificial intelligence is natural language processing (NLP), which provides the tools and techniques for the learning, understanding, production, and analysis of human language content (Hirschberg & Manning, 2015). As such, the goal of NLP is to make the human language accessible to computers so that the machines can understand and respond to human languages. Since the "statistical or corpus ("body of words")-based NLP was one of the first notable successes of the use of big data, long before the power of ML was more generally recognized or the term "big data" even introduced" (Hirschberg & Manning, 2015, p. 261), NLP is a suitable approach for the analysis and understanding of language. NLP combines concepts and techniques from linguistics, computer science, statistics, and machine learning models. Methodological applications of NLP include machine translation, speech-to-text dictation, chatbots, spell check, spam detection, spoken-dialog devices (e.g., Siri, Alexa), information extraction, text classification, and automatic summarization, semantic and sentiment analysis.

NLP computing is complex and challenging mainly because natural language is ambiguous, complicated, evolving, and diverse and requires context to convey the intended meaning, domain-specific, social, and world knowledge (Hirschberg & Manning, 2015; Manning & Schutze, 1999). Without context and world knowledge, it is difficult for humans or machines to discern whether an utterance is sarcastic. For instance, the sentence "Oh, it is snowing again" may be sarcastic if you understand the context; it has been snowing for a long time, and the speaker has had enough of the cold. This task may be easy for a human being to understand based on the context. However, while the NLP models may have learned all the definitions and synonyms, understanding the sentence in context is challenging.

Language is diverse and evolving, with different vocabularies, accents, and regional dialects, including slang. While it may be less challenging for some humans to understand the various languages and slang/colloquial based on one's culture, machines may find it challenging to understand the evolving language, including informal expressions and slang, mainly because these words do not exist in the dictionary or have a formal definition. As a result, models trained to detect hate speech are likely to falsely identify a dialect, for instance, the African American Vernacular English, as hate speech (Bender, 2019; Hvitfeldt & Silge, 2021; Sap et al., 2019). Furthermore, NLP "still cannot perform common-sense reasoning or draw on world knowledge in a general and robust manner" (Bird et al., 2009, p. 33). The models must be constantly updated with the language changes to reflect the new changes. The update would require more data for the algorithm to learn. Despite these challenges, statistical NLP approaches solve some of these problems by "automatically learning lexical and structural preferences from corpora [...] the use of statistical models offers a good solution to the ambiguity problem: statistical models are robust, generalize well, and behave gracefully in the presence of errors and new data" (Manning & Schutze, 1999). The statistical NLP approaches include topic modeling, which I describe below.

3.2.1 Topic Modeling

Topic models are generative probabilistic models used for text mining and uncovering latent topics from documents. In generative probabilistic modeling, data is treated "as arising from a generative process that includes hidden variables. This process defines a joint probability distribution over both the observed and hidden random variables" (Blei, 2012, p. 79). Topic model algorithms automatically discover themes and how the themes are connected in a large and unstructured collection of documents (Blei, 2012). As such, "this method scales to billion-word datasets and can arguably provide an analysis driven more by the documents than by human preconceptions" (Baumer et al., 2017, p. 2). In analyzing terrorists' online outlets, topic models are more beneficial in the automatic and quick detection of large-scale terroristic content than manual annotation. Topic modeling can also find patterns in genetic data, images, surveys, and social networks (Blei, 2012; L. Liu et al., 2016).

There are several topic models; in this study, I am interested in the unsupervised latent Dirichlet allocation (LDA) topic model, which does not require prior annotations or labeling of the data (Blei, 2012; Blei et al., 2003). LDA automatically finds topics from documents based on words co-occurrence and frequency distribution. The documents and words are observed, while the topic structure, which includes topics, per-document topic distributions, and the per-document per-word topic assignments, is hidden (Blei, 2012). The benefit of topic models is that the hidden topic structure resembles a document's writing process, making it easier to interpret and label the topics. The high probability words in each topic distribution can be easily interpreted as a theme, therefore a topic (Baumer et al., 2017). Additionally, topics are "inferred from a given collection without input from any prior knowledge. Since topics are hidden in the first place, no information about them is directly observable in the data" (Maier et al., 2018, p. 94). The assumptions of LDA models include a) a topic is a distribution over vocabulary, i.e., a set of words, and b) a document is a mixture over topics. The researcher must specify the number of topics to be modeled.

Extending LDA, other topic models incorporate meta-data into the topic models, such as author, title, location, links, and year (Blei, 2012). For example, both topics and authors are generated in the author-topic model (Rosen-Zvi et al., 2004). In addition, the topics are attached to authors, providing the analysis of which authors are likely to write similar work and topics. Significant for social science research, the structural topic modeling (STM) (Roberts et al., 2013) includes metadata to analyze the topics and topic structure. For instance, the author, political views, and ideologies contribute to the words used in any given topic and document. Both these topic models can be applied to analyze text data; the choice of an algorithm depends on the research questions and objective of the analysis.

Other topic models have extended the function of LDA to combine topic and network analysis. For example, the relational topic model (RTM) (Chang & Blei, 2009) models the documents and assumes the links between the documents depend on the distance between their topic proportion. "Unlike traditional statistical models of networks, the relational topic model takes into account node attributes (here, the words of the documents) in modeling the links" (Blei, 2012, p. 83), taking into account links such as citation and attributes such as text (Chang & Blei, 2009). Words or concepts are modeled as nodes and relationships, such as co-occurrence within documents, as edges (e.g., Walter & Ophir, 2019). Similarly, Walter and Ophir (2019) have developed the analysis of topic model networks (ANTMN) to include topics models and network analysis. In ANTMN, the topics are nodes, and edges are the relationships between the nodes calculated by their co-occurrence over documents and clustered using community detection techniques into topics. Community detection algorithms identify communities in networks based on the structural position of nodes (Molina & Garip, 2019).

3.2.2 Sentiment Analysis

Sentiment analysis, also referred to as opinion mining, is another NLP method for analyzing text data and is significant for sociological research (B. Liu, 2012; Pozzi et al., 2016; Puschmann & Powell, 2018; Roy & Rambo-Hernandez, 2021). Sentiment analysis (SA) extracts and analyzes subjective information from text data, such as opinions, sentiments, evaluations, attitudes, and emotions towards topics, events, organizations, individuals, or things (B. Liu, 2012). Using the words written by a user, SA measures how a person might feel about a topic (positive, negative, neutral, emotions – anger, happy, frustrated etc.); hence textual data expressing opinions, emotions, and sentiments are imperative for SA rather than factual data (Puschmann & Powell, 2018). Opinions influence our behaviors and attitudes, and the analysis of Jihadists forums and their supporters highlights their sentiments towards terrorism (Abbasi & Chen, 2007; Ahmad et al., 2019; Greene & Lucas, 2020).

The approaches to SA are mainly supervised and unsupervised machine learning. In supervised learning, the models learn from labeled data, while in unsupervised learning, the models are dictionary-based, trained using lexicons and sentiment scoring. Lexicons are dictionaries or lists of terms classified with a polarity score as negative or positive and used to analyze the sentiment of the text (Puschmann & Powell, 2018; Roy & Rambo-Hernandez, 2021; Schweinberger, 2022). There are several lexicons; the most common include AFINN (Nielsen, 2011), NRC word-emotion association (Mohammad & Turney, 2013), Bing lexicon (Hu & Liu, 2004) based on customer reviews, and VADER (Valence Aware Dictionary and sEntiment Reasoner) (Hutto & Gilbert, 2014) which is specifically adapted to sentiments from social media data.

Although SA is conducive to measuring opinions from text data that affect our behaviors, the method has limitations. The lexicons used to train the data are domain-specific; while they can accurately detect sentiment in one context/data, for instance, marketing research, the lexicons may not be applicable in other contexts, such as terrorism. Nevertheless, both topic models and sentiment analysis can be applied to analyze and understand the topics and opinions from online platforms. However, the goal of this research was to examine and find the hidden topics discussed on Jihadists' Dark Web forums. Therefore, most of the analysis and discussion will focus on topic models and the results.

3.3 Applications

In the following section, I introduce three applications of LDA models using text data from Dark Web forums as case studies. First, I explore the background and then describe the data and the text analysis process.

3.3.1 Detecting Terror on the Dark Web: A Case Study

Although most Internet users are using it for legitimate reasons and benefiting from the technologies, terrorist groups and their supporters use the Internet to further their ideologies and goals posing new national security and counterterrorism challenges. Terrorism is a form of communicative violence (Aly et al., 2017), as such terrorist groups use the Internet, specifically social media platforms and messaging apps, to spread propaganda, recruitment, radicalization, financing, and communication, establish their appeal and popularity, and interaction with low risk of being captured (Behr et al., 2013; Bloom et al., 2019; Conway, 2017; Edwards & Gribbon, 2013; Gill et al., 2017; Sageman, 2011). One way of examining the online content and narratives of terrorist groups is by applying real-time computational techniques to highlight the topics discussed.

Social media accounts and websites of terrorist groups and their supporters are removed or suspended from the Internet by social media companies and counterterrorism agencies. As a result, the groups have moved to the Dark Web, the part of the Internet that is hidden, not indexed, unregulated, and therefore not accessible via search engines. Within the Dark Web, "the decentralized and anonymous networks enable evading arrest and the closure of the terrorist platforms" (Weimann, 2016a, p. 41). The groups moved to the encrypted platforms to protect their information from being removed and further disseminate news and propaganda, finance, and recruitment. As such, individuals seeking terrorist content can do so on the Dark Web without being monitored or detected (Malik, 2018). Additionally, "terrorist forums on the Darknet not only encourage individual radicalization but also promote a "self-starter" type of terrorism. This is a strategy endorsed by IS both online and offline and motivates vulnerable individuals to commit violence in the organization's name in an attempt to "crowdsource" terrorism" (Malik, 2018, p. 20).

Before the growth of the Internet, most terrorist networks resulted from face-to-face interactions among friends and family (Sageman, 2011). The Internet has allowed like-minded individuals to interact and form virtual communities; however, passive absorption of information does not change someone's mind. Communication through forums or chat rooms contributes to interactions paving the way to changing minds and radicalization. As such, active online discussions about ideologies and opinions with friends or strangers and the sense of belonging within the virtual community (Bowman-Grieve, 2009; Scrivens et al., 2018) or the virtual Muslim community (Ummah) with jihadist ideologies (Erez et al., 2011) has the potential of changing someone's mind rather than just reading impersonal stories. Participants post and share their views and knowledge within the virtual communities and offer support and justification for using violence and validation. The virtual networks comprise unaffiliated sympathizers, propagandists, fighters, recruiters, and rivals. The forums, not passive websites, are crucial in radicalization. Thus, creating opportunities for self-radicalization, connection, and interaction of like-minded people globally while facilitating the promotion, popularization, and operation of radical ideologies and ideologically motivated violence.

Once in the virtual community, users' language and social interactions change over time, mirroring those within their online community. As time goes on, the users in these platforms have roles including producing and aggregating content, translating and curating content across platforms, or passively consuming the information (Bloom et al., 2019; Shehabat & Mitew, 2018). As participants begin to feel comfortable and welcomed on these platforms, skeptics and lurkers will be encouraged to share their beliefs and become more vocal. For instance, ISIS uses an army of 'media mujahideen' to amplify their online messaging campaign, "these IS-supporting internet users meticulously share and repost official content across several social media platforms, as part of a coordinated effort to maximize the organization's impact and relevance, and to assist in the recruitment to the group" (Malik, 2018, p. 21). In addition, since participation in these forums is an active process, users tend to select platforms most compatible with their views (Piazza & Guler, 2019; Sageman, 2011). Therefore, the analysis of topics is significant for understanding terrorists' communication patterns and narratives within online platforms.

The more time individuals spend online interacting with terrorists and their supporters, the more likely they may be radicalized into violent or radical social networks. A user can be defined as radicalized based on the content they post or share, i.e., anti-Western, pro-terrorist rhetoric, if they post very negative and inciting comments to the forum. For example, a long-time member of a discussion forum who posts harmful material over time suggests a dedication to an extremist movement. Or an author who participates for a short time but posts very radical views implies the user has become radicalized (Scrivens et al., 2021; Scrivens et al., 2018). However, researchers attempting to classify users and their online posts as a signifier of radicalization (Ahmad et al., 2019; J. R. Scanlon & Gerber, 2015; Scrivens et al., 2021; Scrivens et al., 2018) have found that classifying a text or user as extremist is challenging due to the different kinds of extremism and ideologies, various targets, different ways of expression and language. For example, a pro-terrorist material and discussion forum would include terms and topics such as a call for jihad, Islam, mujahideen (members of the Taliban), apostate, caliphate, Ummah (Muslim community with jihadists worldview, nation), crusaders, promote violence towards the West, Christians, Jews, call for martyrdom, suicide operations, infidels (enemies, or non-believers), and offensive or abusive language such as kuffar (non-believers, non-Muslims). Forum posts of jihadists can be classified into themes based on information dissemination, religious preaching, instruction or training, and social interactions (Erez et al., 2011; Rowe & Saif, 2016). As such, topic models are imperative in classifying topics and themes within forums and websites.

Criminological/sociological theories and scholarship emphasis social interactions (Goffman, 1959) and criminal behavior (Akers et al., 2014; Gottfredson & Hirschi, 1990) within a social context, providing frameworks for understanding online social ties and terrorism. In the presentation of the self, Goffman (1959) argues social interactions are performances where individuals manage how they want others/ the audience to perceive them. As such, one would expect individuals interacting in an online forum to conform, hide their true identities and beliefs to fit in. Indeed, individuals interacting in online platforms imagine their audiences and act accordingly (Marwick & boyd danah, 2011; Németh & Koltai, 2021). Social control theorists argue we are all capable of committing a crime (Hirschi, 1969, p. 31). The critical question is not what causes crime but why individuals conform. Control perspectives maintain that deviation from conforming behaviors is likely to occur when social bonds to family, peers and conventional society are weak or nonexistent. The Internet's role in radicalization and violent extremism has become a growing concern. Online content and platforms are conducive to online radicalization, as users seeking justification for joining or intending to join terrorist groups are likely to be indoctrinated by these messages and radicalized others.

The social bond theory (Hirschi, 1969) remains one of the most cited theoretical frameworks in sociological and criminological research. The theory holds that individuals are less likely to commit a crime if they have stronger social bonds to family, peers, and legitimate institutions. The theory has four central elements: attachment, commitment, involvement, and belief. The first element is attachment to others and sensitivity to their opinion. Attachment contains a moral component, as an attached individual considers others' reactions, such as parents, peers, or teachers, before engaging in deviant acts. Commitment refers to conventional lines of action; it is the rational component of conformity. An individual invests time in getting an education, starting a business, or establishing a reputation. Thus, before engaging in deviant behavior, they consider the act's cost. Those who have invested time and energy in getting an education, building a business, or establishing a reputation are less likely to engage in crime for fear of losing their investment. Involvement in conventional activities prevents individuals from engaging in deviant behavior. A person busy doing conventional activities has less time to engage in crime. Finally, belief refers to the commitment and devotion to society's standard value system; individuals who believe in society's laws are likely not to break them. The four elements are interrelated; the more individuals are affectingly attached to conventional others, the more they are committed to conventional systems, the more involved in those conventional systems, and the more they believe in conventional values and norms. As a result, they are likely to be more conforming and less delinquent.

Social learning theory (SLT) (Akers et al., 1979; Akers et al., 2014) presumes that individuals' differences in criminal behaviors stem from learning through association or interactions with others. The premise of SLT is that all social behavior is learned, including crime, and has four key elements: 1) differential association, which refers to direct social interaction with members of a primary group. 2) imitation, which occurs when individual copies the behavior of others, 3) definitions, which serve as behavioral guideposts for how we think about certain behaviors as good or bad, rewarding or punishing and 3) differential reinforcement, which exists in both social and nonsocial forms, are anticipatory or prospective (Akers et al., 2014; Akins & Winfree, 2016). Social learning theory utilizes general principles of social behavior to offer a social psychological explanation of criminal and deviant behavior, the acquisition, maintenance, and change in criminal or conforming behavior through a learning process (Akers et al., 2014). Therefore, SLT can explain why individuals support and sympathize with terrorist groups. Exposure to deviant peers and family (differential association) predicts an individual's likelihood of engaging in crime. Individuals associating with others online who define violent extremism as favorable are likely to adopt those beliefs and eventually engage in violent extremism. As such, social bond and social learning theories have great potential in explaining the role of online bonds and interactions in violent extremism. Although these theories were formulated to explain offline interactions and behavior, few studies have yet to apply the theories and topic models in explaining whether online interactions and posting behavior are related to violent extremism.

3.3.2 Dark Web Forum Data

Data used for analysis is accessed from the University of Arizona Artificial Intelligence Lab's Dark Web Forums (https://www.azsecure-data.org/dark-web-forums.html). The forums were collected as part of the lab's study of the international Jihadi social media and movement (H. Chen, 2011; H. Chen et al., 2008). The open-source dataset has provided researchers with data that would typically not be collected without computational knowledge, tools, or access to the forum. The Dark Web Forum repository includes twenty-eight jihadist dark web forums in Arabic, English, Spanish, French, German, and Russian. Forums are organized into threads and posts; the metadata includes members and dates collected between 2004 and 2012.

I analyzed three forums; the first is Ansar1, an English-language forum with 29,492 posts and 11,244 threads made by 382 members between 12/8/2008 and 1/20/2010. The second forum is Gawaher, an English language Islamic forum dedicated to discussing the Islamic world and Islam issues. Some of the forum members sympathize with radical Islamic groups. This forum entails 372,499 posts and 53,235 threads made by 9,269 members between 10/24/2004 and 6/7/2012. Finally, the Islamic Network forum is dedicated to various topics of interest to Muslims, ranging from theology and world events. Some members sympathize with and support terrorist organizations. This forum has 91,874 posts and 13,995 threads created by 2,082 members between 6/9/2004 and 11/10/2010.

Forums	**Year**	**Posts**	**Threads**	**Users**
Ansar1	2008-2010	29,492	11,244	382
Gawaher	2004-2012	372,499	53,235	9,269
Islamic Network	2004-2010	91,874	13,995	2,082

Table 3.1: Dark Web Forums Data

3.4 Text Analysis

Text analysis is conducted using R and Python statistical analysis tools and LDA packages. R packages for LDA include topicmodels (Gruen & Hornik, 2011), LDAtuning (Nikita & Chaney, 2020), tm, quanteda (Benoit et al., 2021), tidyverse packages (Wickham, 2021), tidytext (Queiroz et al., 2021) and Python package langdetect (Danilak, n.d.). In addition,

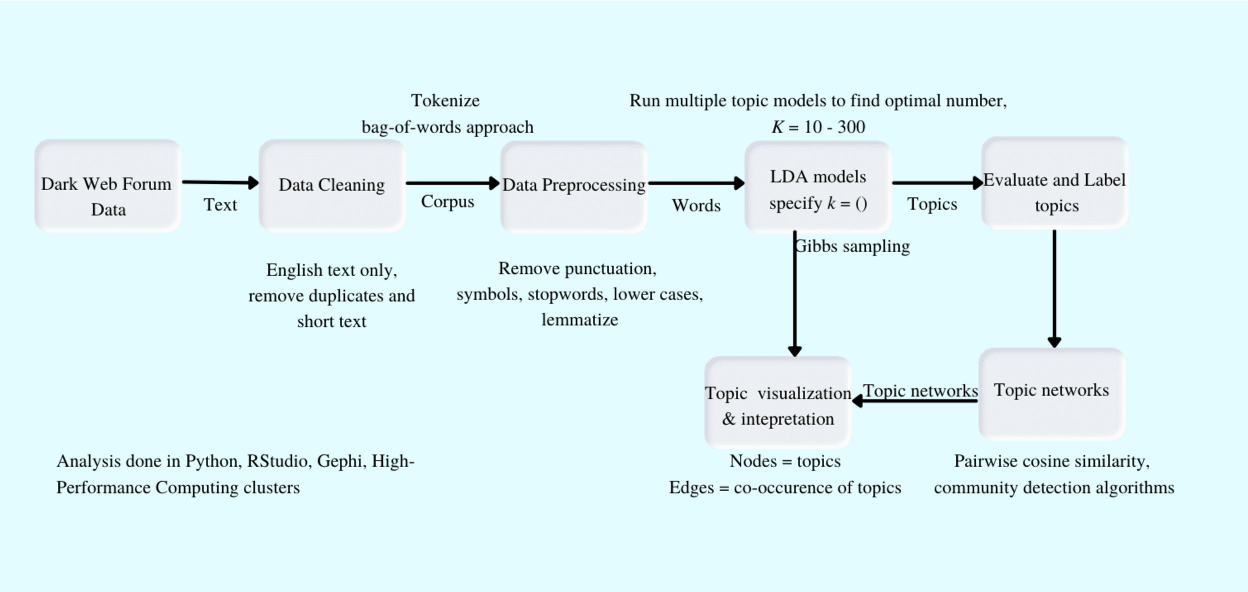
Tokenize
bag-of-words approach
Run multiple topic models to find optimal number,
K = 10 - 300
Dark Web Forum Data
Text
Data Cleaning
Corpus
Data Preprocessing
Words
LDA models specify k = ()
Topics
Evaluate and Label topics
Gibbs sampling
English text only, remove duplicates and short text
Remove punctuation, symbols, stopwords, lower cases, lemmatize
Topic visualization & intepretation
Topic networks
Topic networks
Analysis done in Python, RStudio, Gephi, High-Performance Computing clusters
Nodes = topics
Edges = co-occurence of topics
Pairwise cosine similarity, community detection algorithms

Figure 3.1: Text Analysis Workflow

there are multiple ways for preprocessing text data for analysis (Baumer et al., 2017; Roy & Rambo-Hernandez, 2021), and the order of the preprocessing matters (Maier et al., 2018). Below I outline the steps I took in my analysis.

3.4.1 Data Pre-processing

In LDA, the text collection to be analyzed is referred to as corpus (body of text), for instance, the Ansar Dark Web forum. An item within the corpus is a document, i.e., the message/post, and words within the document are called terms. To analyze text, one must first process it into a structured, machine-readable format. While text analysis tools include translation packages that English content by filtering and removing non-English text from the data by using a language-detection algorithm in Python, mainly because the data is predominantly in English. I then removed short texts and duplicates to analyze a sample of the raw data. The next step was to tokenize the text and break down the texts/sentences into words or tokens using the bag-of-words approach, which analyzes groups of words together, representing words in a phrase disregarding the order in which they appear (Lucas et al., 2015; Silge & Robinson, 2017), resulting in a matrix with the unique terms and words frequencies.

I removed punctuations, numbers, and symbols and converted capital letters to lowercase letters. Common stop-words such as "the," "and" "a" were also removed as they are not helpful for the detection of relevant topics. Additionally, words that appeared frequently and infrequently were filtered and removed. Similar to Walter and Ophir (2019) and Maier et al. (2018), words that occurred in more than 99% of the documents or less than .5% were removed. This process reduces the size of the corpus vocabulary, enhancing the algorithm's performance (Lucas et al., 2015; Maier et al., 2018). Stemming and lemmatization are NLP processes used to reduce inflection forms of a word to a common base or dictionary form, i.e., its stem or root, lemma. However, due to outputs of the stemming process often not making sense, it is difficult to interpret the words, thus impacting the topic quality (Maier et al., 2018; Walter & Ophir, 2019) by degrading topic stability (Schofield & Mimno, 2016). Therefore, I skipped the stemming process as "interpreting word stems correctly can be tough, or even impossible" (Maier et al., 2018, p. 101). Finally, I lemmatized some of the words to their lemma forms, i.e., 'study,' 'studies,' 'studying,' are collapsed to 'study.'

3.4.2 LDA Models

The next step is to train the LDA models using Gibbs sampling iteratively. This algorithm uses "thousands of iterations through the documents, considering the topic proportion of each word token in turn. Words that frequently occur together in documents are likely to be placed into the same topic so that each document contains relatively few topics" (Baumer et al., 2017, p. 7). The number of topics, K, is a parameter that must be defined for an LDA model. The number of topics can be anything from 2 to 100 based on the research questions, the overall size of the corpus, and evaluation metrics. The more topics, the more specific the topic outputs will be. On the other hand, too many topics may lead to a similar outcome that is not meaningful, or too few topics may lead to broad outputs that combine different terms that should have been separated (Maier et al., 2018). Walter and Ophir (2019) suggest specifying many topics rather than a few as duplicate topics can then be grouped using the community detection algorithms.

To assess and evaluate the ideal number of topics for each Dark Web forum, a range of topic numbers from 10 - 300 were estimated and evaluated using four different model fit indicators from the LDAtuning package. The model fit metrics are based on the accuracy (Arun et al., 2010), density (Cao et al., 2009), latent concept modeling (Deveaud et al., 2014), and Markov chain Monte Carlo algorithm (Griffiths & Steyvers, 2004). Then I used the output from the indicators to select a final number of topics for each data set. The next step was to evaluate the relevant topic numbers and then rerun the model. Extract the theta and beta statistics from the output and plot, interpret, and label the topics. Similar to Walter and Ophir (2019), I analyzed the results of the top words in the topics, words that occur frequently and are exclusive to each topic, and the original documents representing each topic. I then interpreted and labeled the topics by examining the words of each topic and the documents. Not all topics in topic models are relevant or interpretable; some of the topics contained words that did not make sense and were, therefore, challenging to interpret. After labeling the topics, a topics network was created using pairwise cosine similarity and community detection algorithms (see Walter & Ophir, 2019). Nodes represent topics, and the edges are the co-occurrence of topics in the documents.

3.5 Results

The unit of analysis in topic models is the key terms or words present within the text data collection. For example, textual data visualizations (figure 3.2) show that the most frequent words in the Ansar1 forum include the terms ‘said,’ which occurred 29611 times, ‘kill’ occurred 18424 times, and ‘attack’ occurred 15,047 in the whole forum. For the Islamic Network forum, the terms ‘quot’ (quote) appeared 92,309, ‘will’ 40,196 times, and ‘Allah’ 39,474 times. The most frequent terms occurring within the Gawaher forum include ‘link’ 84,610 times and ‘post’ 60,468 times.

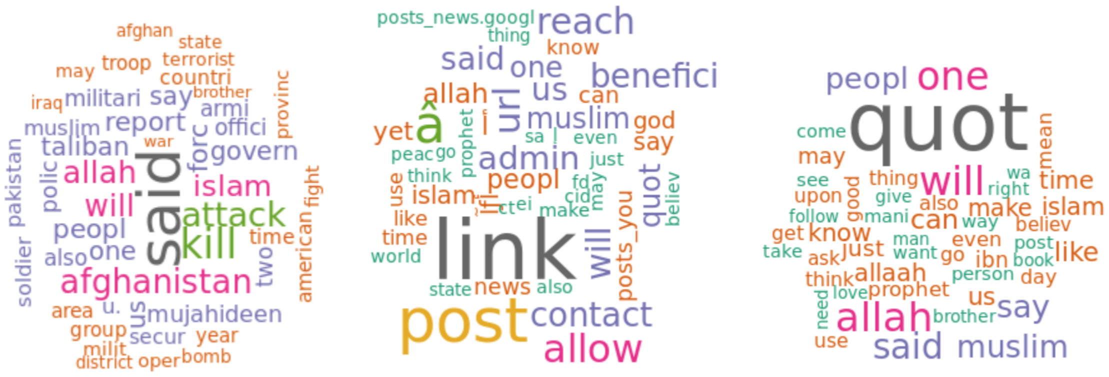

Figure 3.2: Ansar1, Gawaher and Islamic Network Top Frequent Terms

Ansar1 Dark Web Forum

The first topic model was conducted on the Ansar1 Dark Web Forum. The raw sample was N = 29,492 messages and 382 users. After removing duplicates and short text, the final sample was 12,744 messages used to analyze topic models and networks. The topic model for the Ansar Network consisted of 82 topics; 2 of the topics were not coherent enough and to be labeled and included in the network analysis (figure 3.3). Two communities, subgroups connected based on language co-occurrence than the rest of the group, were detected using the Walktrap clustering algorithm (Pons & Latapy, 2005). Walktrap clustering algorithm was selected because it captures the community structures of a network and computes efficiently and faster than other community detection algorithms (Pons & Latapy, 2005). The blue/on the left community focuses on religion and prayers, support for the mujahideen fighters, jihad, Al-Qaeda and Taliban, Muslims vs. the crusaders, the arrest of terrorists, and counterterrorism measures. The other community on the right/in red, topics ranged from mujahideen media, the death of U.S soldiers killed by the terrorists, death of terrorists and their leaders, Somalia government and Al-Shabaab, militants attacks in Afghanistan and Pakistan, suicide attacks and bombings, insurgency in Afghanistan.

A topic is outlined by its probability distribution over words, and the results also include the representative texts. Below I show the ten most probable words for a given topic and the full representative text of these topics. For instance, topic 23, which I labeled ‘Mujahideen

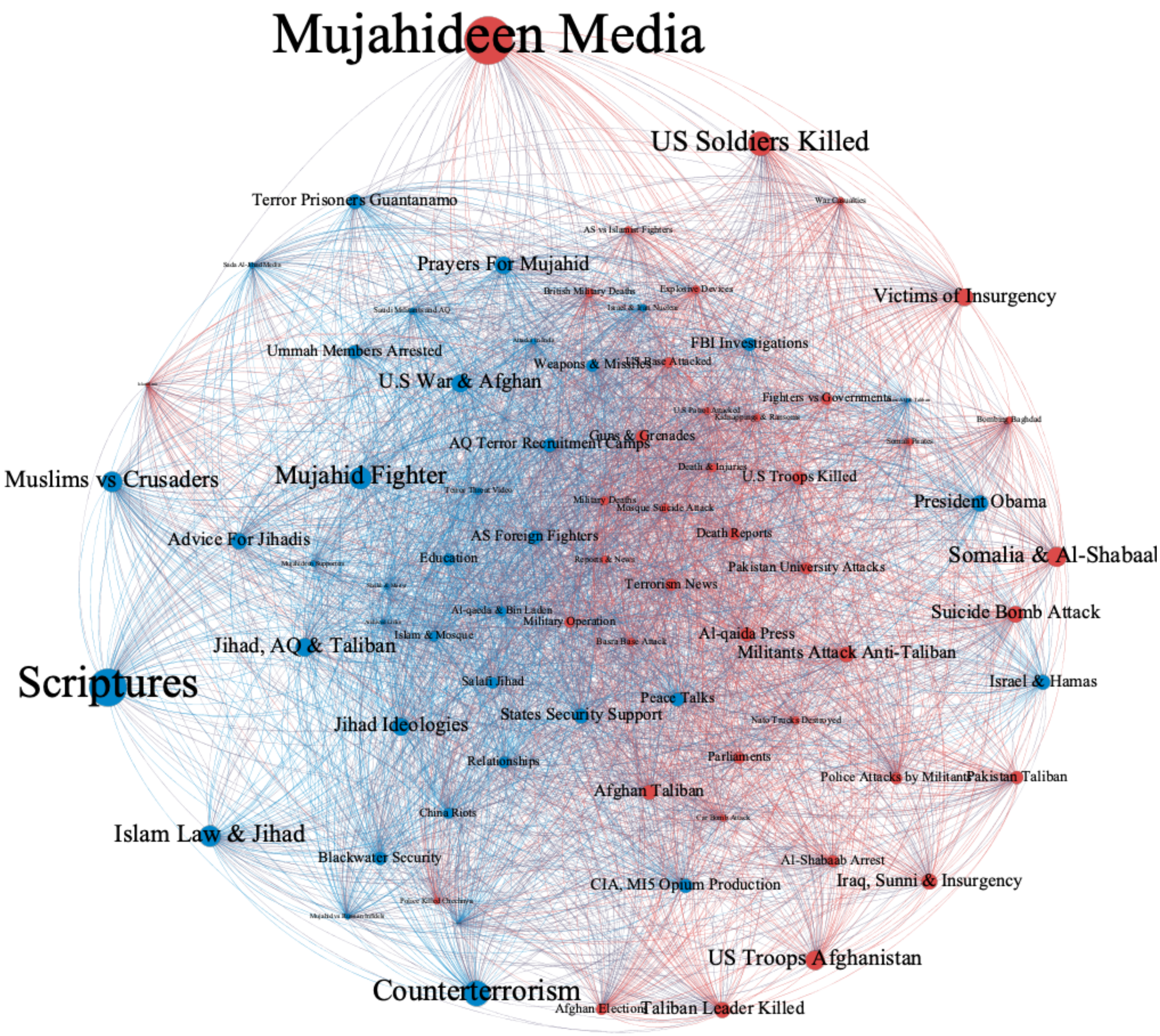

Figure 3.3: Ansar1 Dark Web Forum Topic Network. 80 Nodes represent topics, 3160 Edges represent the co-occurrence of topics within the documents. The node size means the importance of each topic in the network. Color represents the community membership, weighted & undirected network.

Media' after reading the representative texts and words, has the following top 10 terms: Mujahideen, terrorist, kill, report, destroy, Islam, Afghanistan, Allah, Emir, puppet. The first complete text example was:

> "[1] Mujahideen of Islamic Emirate of Afghanistan military operations against the kafirs, munafiqs and the worshippers of Idols. This page is updated throughout the day as new operations are reported [. . .] the Mujahideen operations against the enemies of Islam terrorists in Afghanistan are reported to Theunjustmedia.com by the official Mujahideen of Islamic Emirate of Afghanistan spokesmen Qari Muhammad Yousuf and Zabeehullah Mujahid by e-mails." (Ansar1 Dark Web Forum).

Further reading of the rest representative texts suggests the forum members posted and discussed news from the Mujahideen media. Another example is topic 29, which I labeled 'Advice for Jihadis' has the following top 10 words: God, people, Muslim, religion, said, fight, one, victory, great, believe, faith, infidel. The representative text was:

> "[1] Al-Sahab is pleased to present you with a sound file titled "Advice for the People of Jihad" for Sheikh Mansur al-Shami. (Rajab, 1430) To our Mujahidin brothers greetings and prayers, It is mentioned in the Qur'an, and according to al-Sa'diy that if people don't fight for each other, land will be lost. The infidels will prevent the Muslims from worshipping God, spreading his religion, and they will establish their apostate idols everywhere. Because God is kind, he has allowed Muslims to go to Jihad to fight for the sake of their happiness. He has made this available to them for reasons they do know and reasons they don't know" (Ansar1 Dark Web Forum).

This topic incites and encourages supporters to join the fight for jihad, the Islamic religion, and their nation to prevent the West or non-Muslims from taking over their faith and country.

Gawaher Dark Web Forum

The following topic models were conducted on the Gawaher forum. First, I analyzed a sample (n = 52,588) based on the forum's most active year, 2007, for optimal analysis. Then, after removing duplicates and short text (less than 600), I remained with 19,236 texts for final analysis. Eighty-eight topics were analyzed and plotted; figure 3.4 illustrates the results of the 67 topics. Gawaher forum has two communities; the largest community in green topics are primarily about religion, Islamic scripture, the Bible vs. Quran, converting to Islam, and non-Muslims. When reading the full documents, there were some philosophical discussions about the existence of God and evolution, and relationships. Topics within the smaller community in red were about the U.S military and the U.S in Iraq, terrorists, Israel, Palestine, and the Hamas, the Taliban, and the imprisonment and torture of terrorists by the U.S. A few of the topics not included in the network mainly were posts from the forum admin reminding the members about posting rules and that posting links was forbidden until one had submitted 50 posts.

The most prevalent topic in the Gawaher topic network is topic 80, labeled "U.S Military and Admin." The top ten words include official, new, president, unit, report, American, Washington, military, said, intelligence. The first representative text was about the U.S admin:

> [1] "The Zionists are controlling the US regime and manipulating the USA and they are using the US regime to terrorize and invade Muslim countries and terrorize and massacre innocent and defenseless Muslims for the brutal and barbaric Zionist cause: Zionists in the Bush Administration: Paul Dundes Wolfowitz - Deputy Secretary, Department of Defense Richard Perle - Assistant Secretary of Defense for International Security Policy. Ari Fleischer - White House Press Secretary Josh Bolten - Deputy Chief of Staff . . . ." (Gawaher Dark Web Forum).

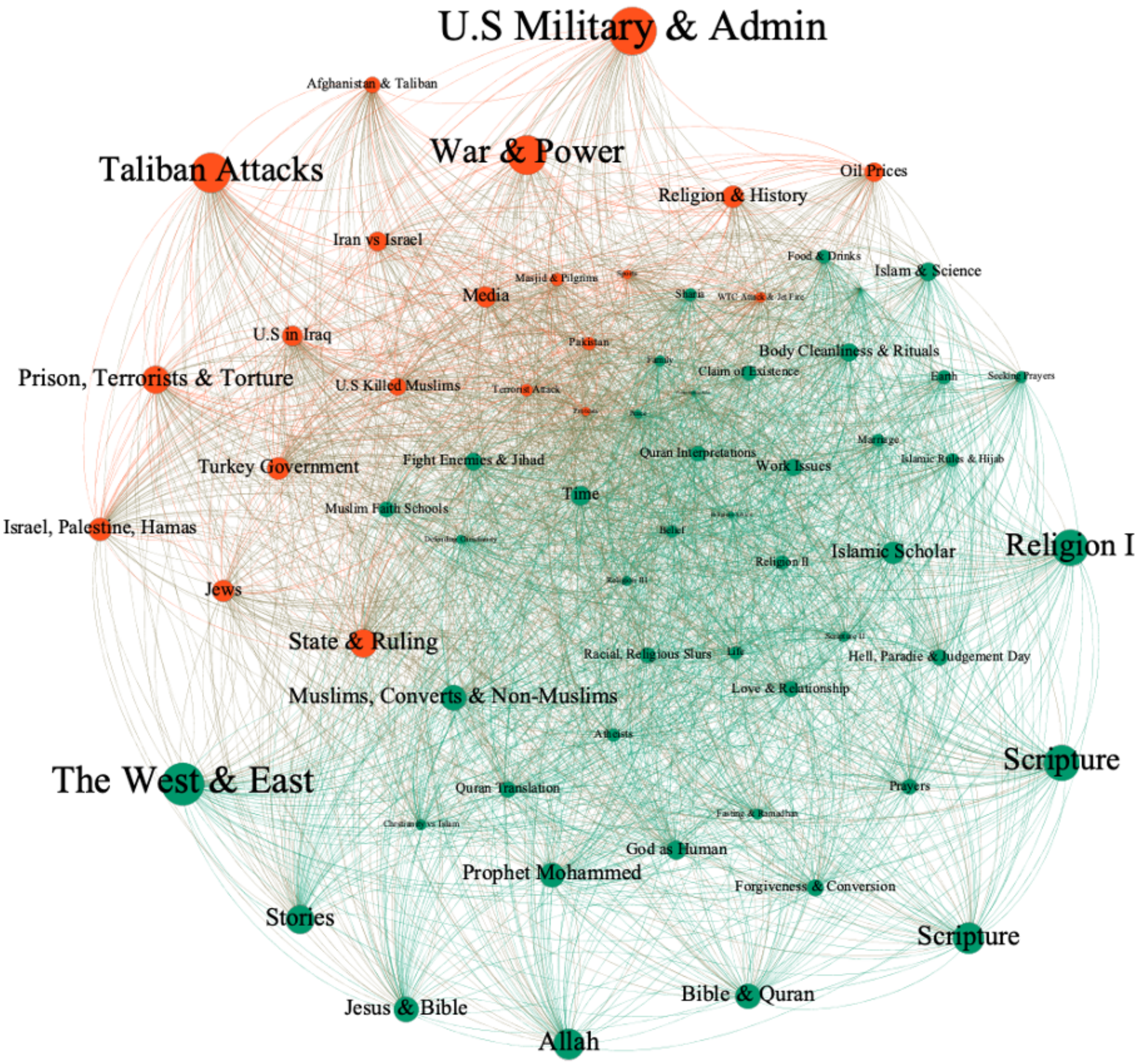

Figure 3.4: Gawaher Dark Web Forum Topic Network. Topic network has 67 Nodes representing topics, 2211 Edges represent the co-occurrence of topics within the documents. The node size means the importance of each topic in the network. Color represents the community membership, weighted and undirected network.

In topic 87, ‘Taliban Attacks,’ the top words include: said, kill, soldier, force, report, police, attack, bomb, military, Baghdad. The first representative text is a news report from the Associated Press:

> [1] “Taliban Torture, Kill 5 Afghan Police Associated Press — November 19, 2007 KANDAHAR, Afghanistan - Taliban militants tortured five abducted policemen in southern Afghanistan and then hung their mutilated bodies from trees in a warning to villagers against working with the government, officials said Nov. 18. The discovery of the bodies came as officials said that recent violence and clashes had left at least 63 other people dead across Afghanistan. The officers had been abducted two months ago from their checkpoint in southern Uruzgan province, said Juma Gul Himat, the provincial police chief. [. . .] Elsewhere, 23 Taliban militants were killed during a U.S.-led coalition operation on Thursday aimed at disrupting a weapons transfer in southern Afghanistan, the coalition said. Copyright 2007 Associated Press. All rights reserved. This material may not be published, broadcast, rewritten or redistributed” (Gawaher Dark Web Forum).

Another interesting topic was Topic 9, “Forum Posting Rules” the probable top words include: forum, discussion, video, topic, can, site, read, please, may, website, thread, member. The first representative text:

> [1] “When starting a topic, please choose a title that describes it as best as you can. Non-descriptive titles confuse others, while good descriptive titles invite more members to come into your topic and share by reading or replying to you. Descriptive titles are beneficial and important in many ways: it makes searching and finding info much easier it invites more members to click your topic to read and reply choosing good descriptive titles allows members to experience ’informative clicking’ titles help search engines index our topics and return more accurate results good titles bring more web searchers and so our daa’wa insha’Allah can

> expand to more horizons Choosing a non-descriptive title for your topic is a violation of our forum rule 7. Kindly read our (you are not allowed to post links yet)"you can't post links until you reach 50 posts you are not allowed to post links yetgawaher(contact admin if its a beneficial link) to make sure you don't break them." (Gawaher Dark Web Forum)

This text exhibits how the forums are organized and how members use descriptive titles to attract volume and interactions.

Islamic Network Dark Web Forum

The Islamic Network forum had 91,874 messages. After removing duplicates and short text, the final sample was 22,787 messages resulting in 92 topic models. After labeling, 72 topics were applied to construct the topic network, which resulted in three topic clusters (figure 3.5). The 72 topics for the Islamic Network contributed to three major topic clusters. The topics in green were on the Guantanamo Bay and how prisoners are tortured and interrogated in that prison. Other topics were terrorism, US troops in Iraq, the death of soldiers from terrorist groups, the Taliban in Pakistan, and the relationship between Israel and Palestine. The topics in red/orange were a mix of electronics sold on the Dark Web, jihad, religion, the discussion on nationalism, people sharing stories, advertisements about the London talk by an Islamic scholar, investments and money, advertisements about selling fake documents, and IDs. Finally, the cluster in blue had topics on religion, Quran recitations and verses, Allah and God, Imam, and Islamic schools/scholars.

For instance, the most prevalent topic 14 is labeled 'Guantanamo,' top words include prison, Guantanamo, torture, detained, court, case, release, charge, terror, British. The representative text was:

> [10] "The crimes at Abu Ghraib are part of a larger pattern of abuses against Muslim detainees around the world, Human Rights Watch said on the eve of the April 28 anniversary of the first pictures of U.S. soldiers brutalizing prisoners

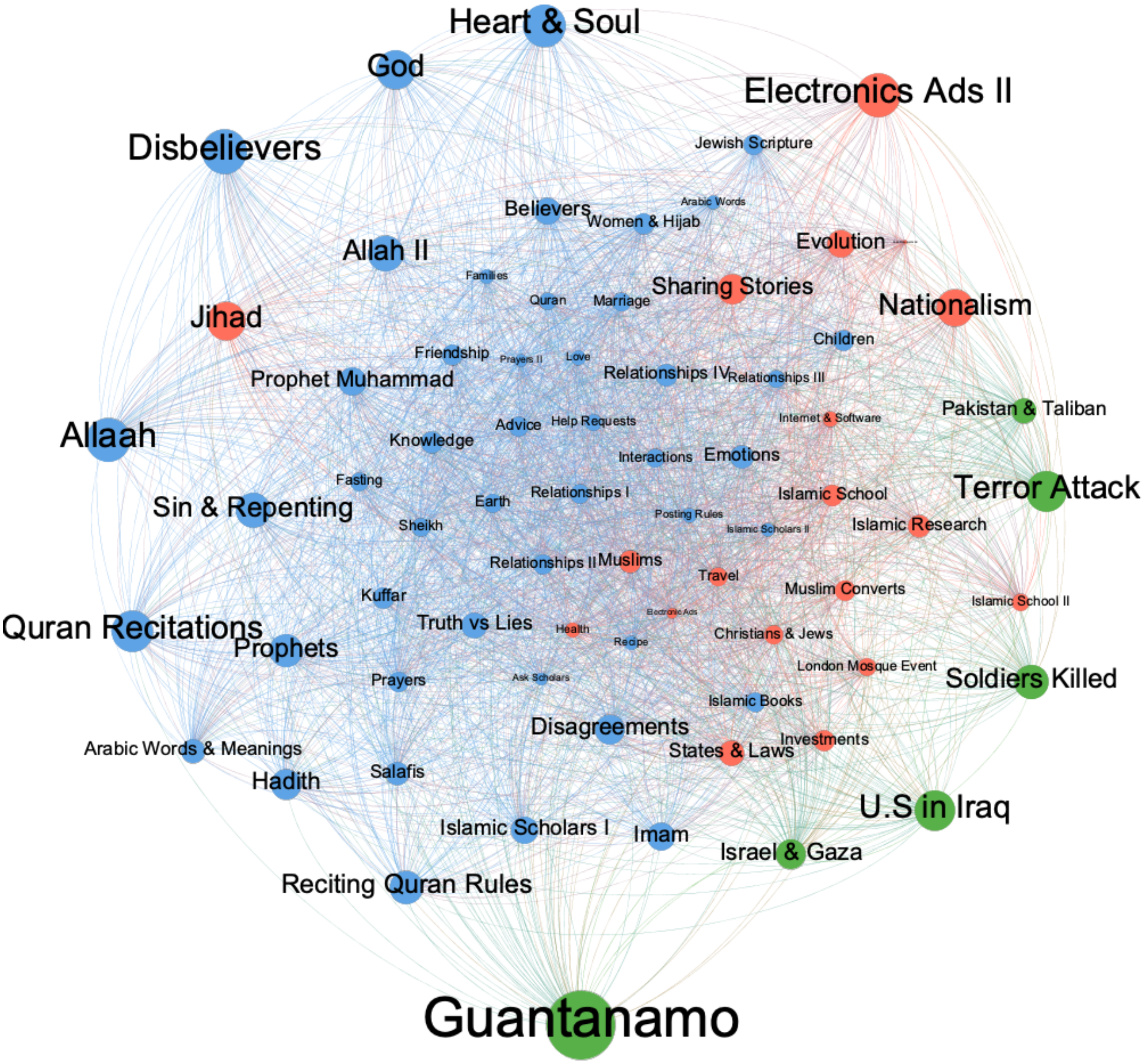

Figure 3.5: Islamic Network Dark Web Forum Topic Network. Topic network with 72 Nodes representing topics, 2556 Edges representing the co-occurrence of topics within the documents. The node size means the importance of each topic in the network. Color represents the community membership, weighted and undirected network.

> at the Iraqi jail. Human Rights Watch released a summary (below) of evidence of U.S. abuse of detainees in Iraq, Afghanistan, and Guantanamo Bay, Cuba, as well as of the programs of secret CIA detention [. . .] Six Algerians held in Bosnia were transferred to U.S. officials in January 2002 (despite a Bosnian high court order to release them) and were sent to Guantanamo. SOURCE: Human Rights Watch" (Islamic Network Dark Web forum).

Another prevalent topic is topic 49, labeled 'Electronics Ads,' the most probable words for this topic include: soni, nokia, apple, htc, Samsung, Ericson, phone, notebook, pioneer, pc, tv, ipod, new, digit. The representative documents include:

> [6] "We deal in the sales of consumer electronic. Main products include portable DVD player, TFT LCD monitor, Mobile phones and Our company have big quantity stock for mobile phone, they are 100% brand new set, wholesale price. . . . PRICE LIST'S BELLOW: NOKIA PHONES N95 8GB Latest/Black Edition.$350usd N73 Cost.....$120usd N75 Cost.....$150usd N76 Cost.....$170usd . . . ." (Islamic Network Dark Web forum).

Although not prevalent, topic 64, labeled 'Fake Documents Ad,' is interesting. The top words include passport, fake, false, sell, new, id, buy, identity, driver, sale, private. The representative text was:

> [1] "Our team is a unique producer of quality fake documents. We offer only original high-quality fake passports, driver's licenses, stamps and other products for following countries: USA, Australia, Belgium, Brazil, Canada, Finland, France, Germany, Israel, Mexico, Netherlands, South Africa, Spain, United Kingdom. To get the additional information and place the order just visit our website: . . . (If in some technical reasons you are unable to visit our website we are always happy to answer your questions on email addresses mentioned below)" (Islamic Network Dark Web forum).

The topics were primarily about religion and a few on relationships, terror attacks, and electronics sales in three clusters. Most interesting is the illegal activities conducted on the Dark Web forums, including forums used by terrorist groups and their supporters.

In addition to the topic models, sentiments displayed by the words/topics were measured using the NRC lexicon (Mohammad & Turney, 2013). Sentiment analysis adds to topic models the sentiments of forum members based on the words and topics discussed. The NRC lexicon categorizes words into eight emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgusts) and polarity (negative and positive). Figure 3.6 shows the result of the forum sentiments, Ansar1 forum posts tend to display sentiments such as anger, fear, sadness, and trust compared to Gawaher and Islamic Network forums. The emotions highly displayed by Gawaher forum include disgust, joy, and trust, while Islamic Network sentiments are anticipation and trust. Overall, the sentiments of fear and trust was prevalent in the forums. Polarity scores (figure 3.7) indicate that the posts in the forums were mostly positive, words used in the forums were scored as positive for the Gawaher forum by (10.9%), Islamic network (10.3%) and Ansar1 (8.6%). Words associated with negative were fewer within the three forum Ansar1 by 8.3%, Gawaher (6.9%), and Islamic Network (6.5%). Overall, Gawaher is the most positive forum, has the highest average polarity ratio followed by Islamic Network while Ansar1 has the most negative sentiments.

Finally, results of sentiment analysis also display the top words contributing to the anger, anticipation, fear, and trust emotions (figure 3.8). For instance, in the Ansar1 forum, the top words associated with anger include attack, fighting, destroyed, enemy, killing. While the terms, words, attack, bad, death and evil were used differently in the Gawaher forum to portray anger. Words associated with fear are used differently in the forums, for instance, the words police, attack, government, military, and war conveyed different polarity within the forums. The words government, military, death, war appear in all three forums and are associated with the sentiment fear. The term, God, appears frequently within the Islamic Network and Gawaher forums and associated with the anticipation, fear, and trust emotions.

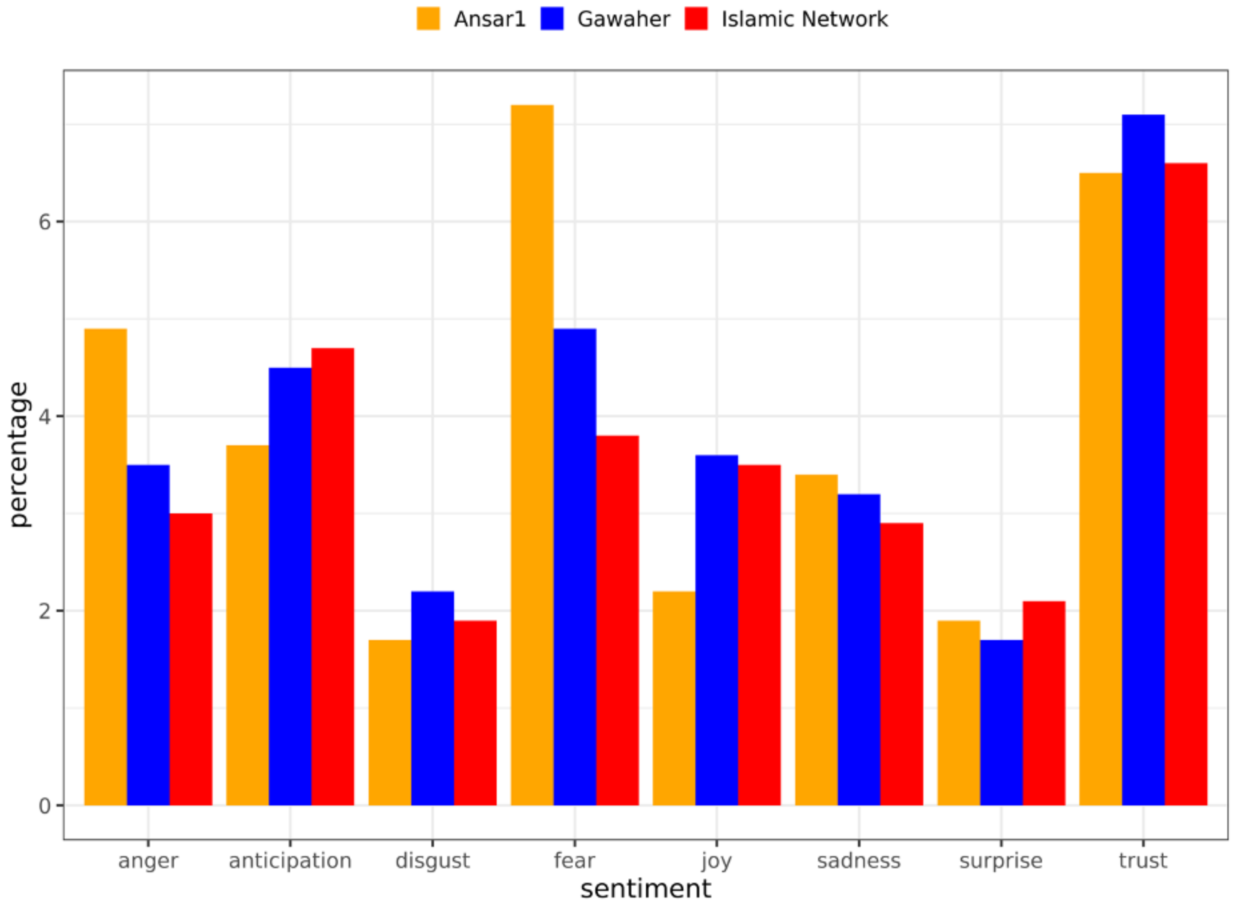

Figure 3.6: Word Sentiments of Dark Web Forums

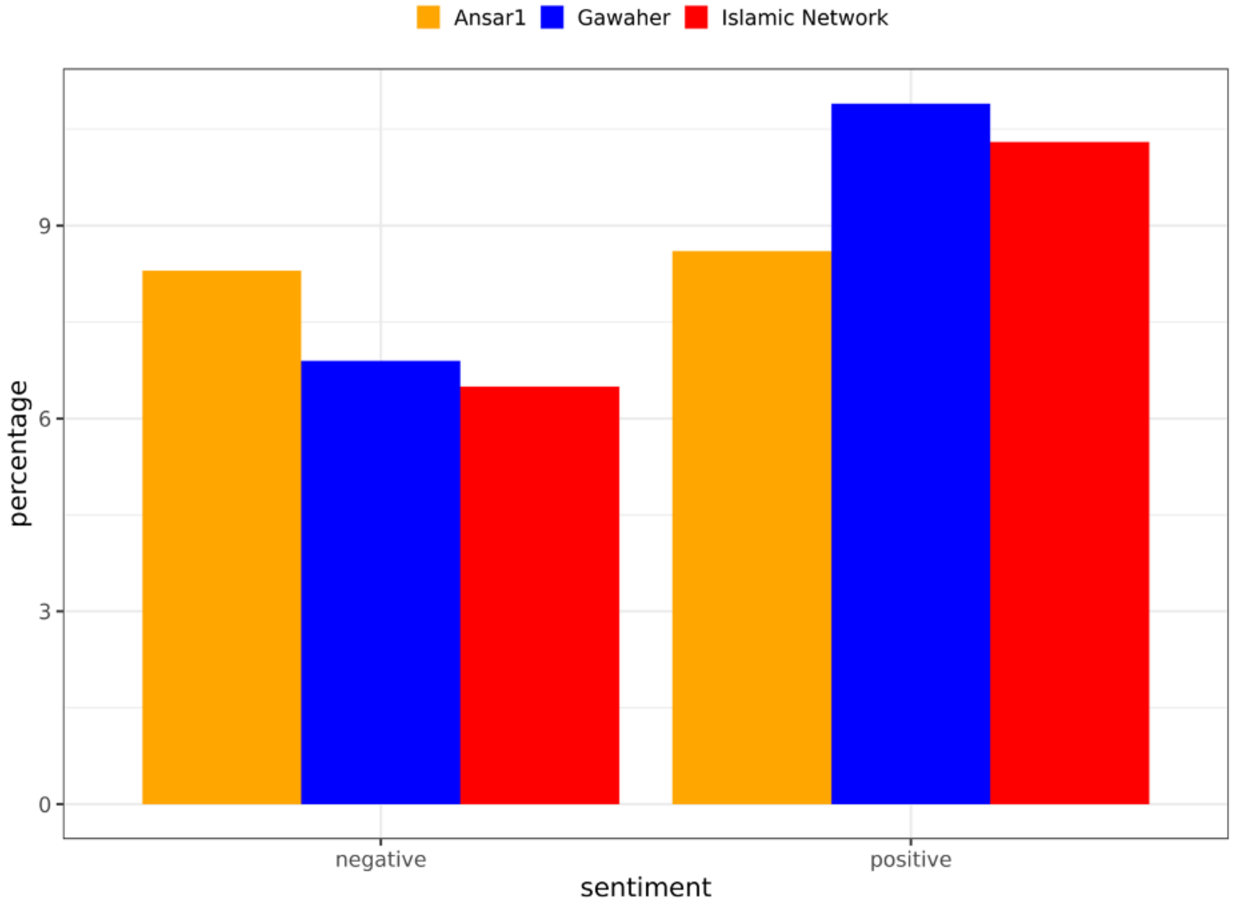

Figure 3.7: Polarity Scores of Dark Web Forums

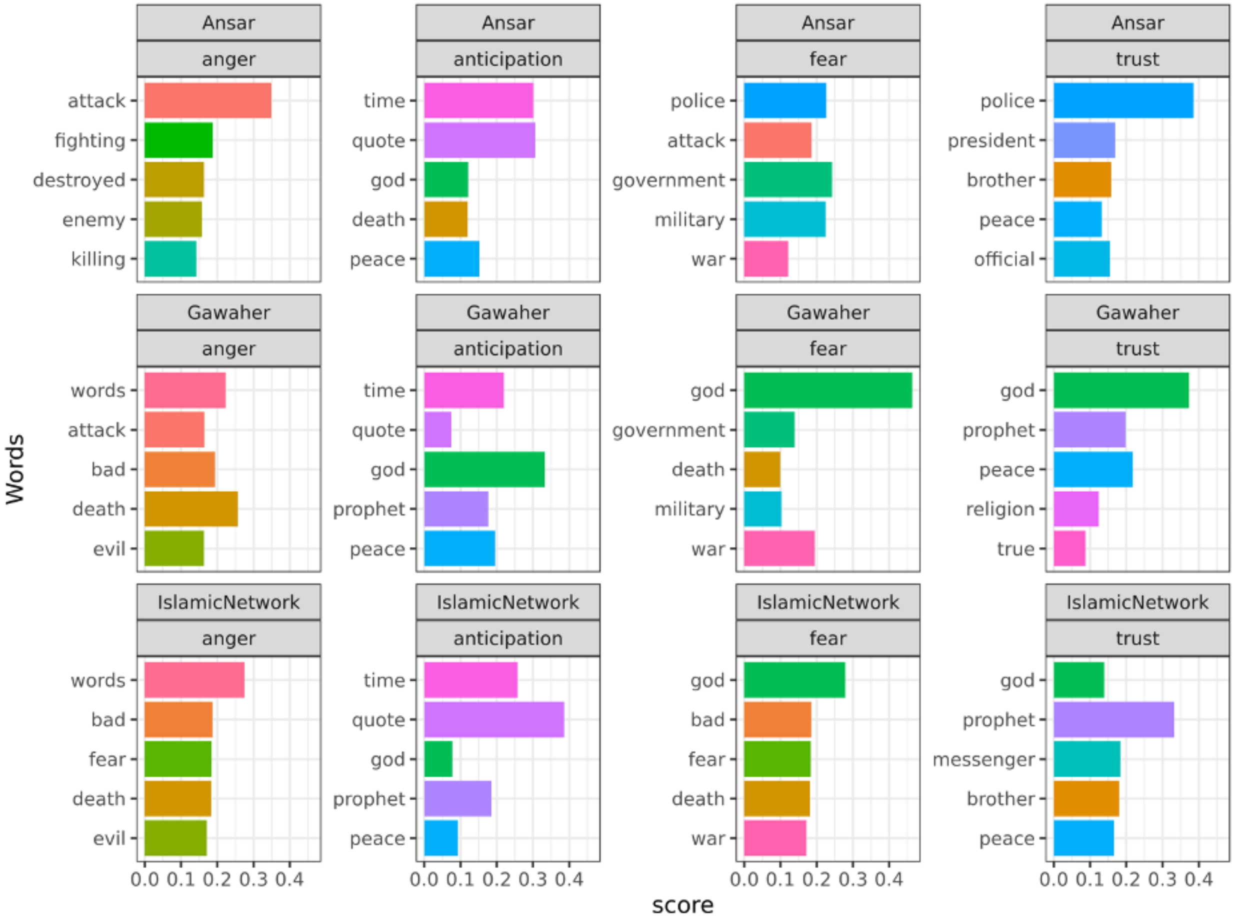

Figure 3.8: Dark Web Forums Top 5 Words and Emotions

By comparing the words and sentiments provides a further understanding of the topics, attitudes and opinions on religion and terrorism and how they differed between the forums.

3.6 Discussion and Conclusion

As Jihadists continue to use the Internet for communication and recruitment, topic models are valuable techniques that automatically detect and classify their online content. Topic model algorithms automatically discover themes and topics and how the themes are connected in a large and unstructured collection of documents based on the co-occurrence and frequency distributions of words (Blei, 2012). While the topic structure is hidden, the documents and words are observed (Blei, 2012; Blei et al., 2003). The high probability words in each topic distribution can be easily interpreted as a topic (Baumer et al., 2017). The unsupervised latent Dirichlet allocation (LDA) topic model does not require prior annotations or data labeling (Blei, 2012; Blei et al., 2003). Therefore, there is no need to manually label a data sample before identifying the topics. The researcher only has to specify the optimal topic numbers and label the output. LDA models assume that a topic is a distribution over vocabulary, i.e., a set of words, and a document is a mixture of topics. Topics can be inductively observed through topics modeling, network analysis, and community detection (Walter & Ophir, 2019). Community detection identifies groups of vertices/nodes, particularly topics, that are more densely connected than the rest of the network. Thus, topic models enable the analysis of text data that is too large to code by hand. Furthermore, topic models might highlight topics that a researcher using manual coding techniques may otherwise not have found (DiMaggio et al., 2013).

Terrorist groups and their supporters use the Internet to further their ideologies and goals posing new national security and counterterrorism challenges. To evade detection or disruption, the groups and their supporters have moved to the Dark Web (Malik, 2018; Weimann, 2016a). To show the utility of LDA topic models, I used three cases from Dark

Web forums of jihadist groups and their supporters, the Ansar1 (Ansar AlJihad) forum, the Gawaher forum, which is dedicated to discussions of Islam and the Islamic world, and the Islamic Network forum, which is also devoted to the theology and world events. Although the forum discussions pertaining to the Islamic world and Islam, some members of these forums sympathize with and support terrorist organizations. Therefore, participating in forum discussions is considered a form of jihad (Erez et al., 2011).

Using LDA, I identified coherent topics and clusters of the forums. As expected, the most prevalent topic in all forums was religion. Topics about religion included Islam, Christianity, and Judaism, discussion about Allah and God, believers and disbelievers, Quran recitations, religious rituals, and philosophical discussions about religion and science. These topics are expected, as the forums were dedicated to the Islamic faith and issues concerning Muslims. The topic of terrorism was prevalent in all forums, mainly within the Ansar1 forum. Here the discussion was about terrorists and attacks, support for the Mujahideen fighters, jihad, Al-Qaeda, Taliban, the U.S in Iraq, Afghanistan, and the arrests and torture of terrorists and foreign fighters. One could argue the discussion or posts about terror groups were created by the terrorist groups and their media centers and supporters, also known as 'media mujahideen' (Malik, 2018). Most of the jihadist discourse online involves religious/ideological indoctrination framed by references to the Quran, Hadith, and the early history of Islam (Erez et al., 2011). A few of the discussions were on relationships and marriages, advice, seeking help, health, food, selling electronics, and identity cards. Topics on selling fake IDs and documents indicate the presence of criminals in terrorist networks and possibly a partnership between criminals and terrorists, also known as a crime-terror nexus. Not all these topics would have been easily identified via manual hand-coding, especially since a predefined set of themes would have been set as guidance for coding.

By clustering the topics into communities, one can highlight the frequency of language use from the corpus used by each community, in this case, the extent of discussions based on terrorism and/or Jihadism. For instance, in the Ansar1 forum, 51% of the corpus was associated with the community on the left, in blue. In contrast, 49% of language used in the forum corpus is likely to be used by the community on the right, in red. Similarly, for the Gawaher forum, the topics in the green community were most likely to be associated with 69% of the language in the corpus and 31% for the topics in the red community. In the Islamic Network forum, the topics in blue are most likely to be associated with 65.3% of the words in the corpus and 26.4% for the topics in red, and 8.33% for the topics in the green community. In addition to identifying the topics, and the prevalence of language use, one can also classify extremist content from the Dark Web. Results suggest that from the 82 topics identified in the Ansar1 forums, there is an equal mixture of religious and terrorism content. Most of the Gawaher forum's 67 topics were about religion (61%) and terrorism content (31%). The Islamic Network forum's 72 topics resulted in three classifications. The prevalent topic is religion (65.3%), a mixture of religion, legal and illegal activities (26.4%), and terrorism (8.33%). The Ansar1 forum was mainly dedicated to discussing or sharing terrorist content as opposed to the other two forums, which were especially about Islam and Muslims. It should be noted that the Ansar1 forum is the English forum of the Arabic Ansar AlJihad Network, which is an invitation-only Jihadist forum popular with western Jihadists (J. Scanlon & Gerber, 2014). Additionally, as mentioned elsewhere, topics on religion can also be classified as Jihadist discourse, using religion to justify their ideology.

LDA topic modeling has several advantages over the manual annotation of text data, especially when analyzing larger corpora. The analysis of text data has predominantly been based on manual content analysis, where researchers create a set of themes based on research questions and objectives, create a coding sheet, and read the text and code by hand. As DiMaggio et al. (2013) note, manual coding is limited when the text to be analyzed is large, and "the more analytically interesting are the research questions, the harder it is to achieve acceptable levels of intercoder reliability; and c) the approach presumes that the researcher knows what is worth finding in the texts before having analyzed them" (DiMaggio et al., 2013, p. 577). However, a manual approach can help analyze smaller samples. When analyzing big data generated by Internet users, automatic computational techniques are more effective, faster, and reliable. Topic models quickly discover hidden topics, thus providing real-time text analysis. The topics are inductively uncovered and require no previous knowledge of the content from the large-scale text data. Additionally, topic models offer insights into extensive unstructured text data instead of the traditional structured data such as surveys and interviews. The keywords discovered from the topic models can further be utilized to organize and search for content in other text data. For instance, the keywords derived from the three Dark Web forums can be used to analyze or search other Dark Web forums or websites for similar topics.

Compared to manual annotation, topic models require less time and money and reduce researcher biases. Topic modeling is data-driven and thus requires very minimal human supervision. The only work for the researcher is to input a topic number and interpret and label the output. Hence, topic models illustrate that textual data can also be analyzed quantitatively and mitigate researcher bias. Additionally, compared to other text classification algorithms of unlabeled data, such as k-means clustering analysis, where documents can be classified into only one cluster, topic models are mixture models where each document is assigned a probability of belonging to a topic; therefore, a document can belong in different topics.

However, some limitations abound when using topic modeling. First, depending on the corpus size to be analyzed, topic modeling can be computing-intensive. Depending on the computer's power, an analysis can run for two days or weeks. Secondly, language is ambiguous, and not all topics can be labeled and interpreted into a recognizable theme. For instance, despite considering the full text and context, some of the topics in the Gawaher and the Islamic Network forums were challenging to interpret and label, were not meaningful, and thus omitted from the analysis. In addition, unrelated topics may be attributed to the stylistic conventions of text data and the algorithm (Walter & Ophir, 2019). Finally, it is important to note while topic models are data-driven, they assist but do not replace human interpretation (Baumer et al., 2017).

Since the focus of this example case study was to exemplify how topic modeling, specifically LDA, can be used to discover topics from the Dark Web forums, I only analyzed the messages posted by members and did not include the members or time, the year they posted or were active on the forums. Future directions may include the analysis of the members and their contributions to the topics. For instance, an analysis of the topic models would consist of the members and using structural topic modeling, author-topic models, or other algorithms to classify the posts and members as radical or non-radical, based on the posts, time spent on the forum, engagement with others, i.e., responding to all threads, etc. Another future study may analyze the social network structure, including the key members, and detect the communities to highlight how the members behave and interact in the Dark Web forums. It is important to note that this work applied topic models in classifying and detecting Jihadist online content; I did not seek to imply that the forum members are terrorists. Instead, I was interested in detecting topics posted on the forums. Lastly, while the application of sentiment analysis to measure polarity was not the scope of this study, sentiment analysis of the Dark Web forums highlights the opinions and attitudes of forum users toward the major topics of religion and terrorism. Overall, Ansar1 sentiments tend to be negative (i.e., in support of terrorism) as compared to Gawaher and Islamic Network where the sentiments were positive i.e., in favor or support of the Islamic religion. As such, analyzing the sentiments in addition to the topics provides an overall idea about the perceptions on religion and terrorism. A future study may analyze the sentiments using the key words identified by the topics, to create a lexicon that is terrorism specific.

From the topic classification of jihadist Dark Web forums, we can conclude that topic modeling and sentiment analysis can be applied automatically to analyze large-scale text from extremist content online. The detection of extremist content online using topic models is significant for analyzing users' beliefs and countering and/or disrupting their online narrative. Additionally, sociological study of social interactions and theories such as social learning and social control theories can be applied to explain online radical content and interactions of like-minded individuals based on their topics. Social bond theory (Hirschi, 1969) holds that individuals are less likely to commit a crime if they have stronger social bonds to family, peers, and legitimate institutions. On the other hand, social learning theory (Akers et al., 1979; Akers et al., 2014) presumes that individuals' differences in criminal behaviors stem from learning through association or interactions with others. Therefore, as social behavior is learned, one would expect individuals participating in the forums and interacting with those who post about terrorism are more likely to change their beliefs and become radicalized into violent extremism. Additionally, individuals posting on the forums may seek to manage their impressions (Goffman, 1959) by selecting what they post in order to be perceived as either in support or against terrorism. Although these theories were formulated to explain offline interactions and behavior, a future study would apply the theories in explaining whether online interactions and posting behavior are related to violent extremism.

Chapter 4

Detecting Virtual Communities and Networks of Jihadists Dark Web Forums: A Social Network Analysis

4.1 Introduction

Counterterrorism experts and social media organizations are disrupting the online presence of terrorist groups. Social media platforms such as Facebook and Twitter and forums such as Gab and Stormfront used by extremists and terrorists are monitored and suspended (Behr et al., 2013; J. M. Berger, 2016; Caló & Hartley, 2019; Zelin & Fellow, 2013). As a result, to avoid detection and disruption, the groups have migrated to encrypted platforms such as Telegram and the Dark Web (Malik, 2018; Weimann, 2016a) that are difficult to regulate and disrupt. Most Internet users only access the surface web, also known as the Internet, easily accessible through browsers such as Google and Yahoo. A subsection of the Deep Web, the Dark Web is a hidden part of the Internet that can only be accessed via special browsers such as TOR (Malik, 2018; Mirea et al., 2019; Monk et al., 2018; Weimann, 2016b). The Dark Web or Dark Net offers anonymity, making it attractive for individuals interested in privacy and avoiding censorship and criminals, including terrorist groups. Terrorists use the surface web for communication, spreading propaganda, recruitment, radicalization, and financing. However, with the Dark Web, they can do the same while hiding from law enforcement, avoiding disruption, and storing their propaganda materials (Malik, 2018; Weimann, 2016a).

The Internet's propensity for anonymity, minimal censorship, expanded reach, and influence on global audiences, has transformed the structure and dynamics of violent extremism by changing how terrorists operate and interact. The Internet has created opportunities for self-radicalization, connection, and interaction of like-minded people globally while facilitating the promotion, popularization, and operation of radical ideologies and ideologically motivated violence (Behr et al., 2013; Bloom et al., 2019; Conway, 2017; Edwards & Gribbon, 2013; Sageman, 2011). Prior to the growth of the Internet, most of the terrorist networks resulted from face-to-face interactions and strong bonds amongst friends and family (Sageman, 2004, 2011). However, the Internet, especially forums or chat rooms and not passive websites, is crucial in creating social bonds, trust, indoctrination, and eventually radicalization into violent extremism. Active online discussions about ideologies and opinions with friends or strangers have led to the sense of belonging within the 'virtual community' (Bowman-Grieve, 2009; Scrivens et al., 2018). As such, virtual communities have the potential of exposing someone to radical content, thus changing someone's mind rather than just reading impersonal stories. Once in the virtual community, users' language and social interactions change over time, mirroring those within their online community.

Generally, research on terrorists and the Internet has focused on the role of the internet in radicalization and recruitment, specifically through social media accounts such as Twitter, encrypted platforms such as Telegram, websites, and forums such as Stormfront and Gab (Bloom et al., 2019; Bodine-Baron et al., 2016; Bowman-Grieve, 2009; Caló & Hartley, 2019; Gaudette et al., 2020; Koehler, 2014; Piazza & Guler, 2019; Rudner, 2017; Winter et al., 2020). Yet, little is known about terrorist use of the Dark Web and the network structure of the forums used by terrorist groups and their supporters. In this exploratory study, I use social network analysis to examine the social networks of two international jihadist groups Dark Web forums. I construct undirected networks to analyze the network structure, the interactions between members within the forums, the essential members, and virtual communities. The research questions guiding this study are: what is the network structure of jihadist Dark Web forums, and how does the structure affect interactions? What communities can be detected from the network? And what are the posting activities of key members in the forums?

The paper is organized as follows; the first section describes the technological structure of the Internet, specifically the Dark Web. The second section discusses related work in explaining virtual communities and terrorists' online content, terror on the Dark Web, and social networks of terrorist groups. The last sections explore the research design, methods, data, and results. Finally, the paper concludes with a discussion about the Dark Web forums and their network structures.

4.2 Dark Web's Technological Structure

The Internet is divided into layers, surface, deep and dark. Most people use the surface web or the Internet, which is reachable, indexed, and accessed via browsers such a Google, Bing, and Yahoo. Information from the surface web is visible to all those who want to access it, and there are no restrictions. The Deep Web is the part of the Internet that is hidden, not indexed, and therefore not accessible via search engines. Access to Deep Web content requires sign-in passwords and is behind paywalls. For instance, medical records, membership websites, social media platforms, online banking sites are part of the Deep Web. The Deep Web is approximately 400 to 500 times larger than the surface web, holding 400 times more content and 7,500 terabytes of information than 19 terabytes on the surface web (Chertoff, 2017; Malik, 2018). The Dark Web, or Darknet hereby referred to as Dark Web, exists within the Deep Web. The Dark Web is even harder to access, is unregulated, and contains approximately 0.01% of the content as opposed to the Deep Web's estimated 90% of content hosted on the internet (Malik, 2018; Weimann, 2016a).

The Deep Web is accessed for anonymity, usually for legal purposes, and access is granted via passwords, encryption, or specific software. The Dark Web is a layer deeper in the Deep Web, and it is intentionally hidden, accessed for anonymity, and sometimes illegal activities. Access to the Dark Web is granted through free and unique software programs supporting encrypted channels. Connections and networks can only be accessed by specific software, configurations, or authorizations such as the Tor (The Onion Router) or I2P (Invisible Internet Project). Dark Web sites provide anonymity and prevent monitoring for Internet users through encryption, which is "the process of encoding information using mathematical algorithms, ensuring that communication is obfuscated to protect against unauthorized access" (Malik, 2018, p. 13). For instance, the commonly used Dark Web browser, the Tor browser (The Onion Router), uses multilayered encryption that relays traffic and encrypts it three times. It is also used to block trackers and ads, defend against surveillance and fingerprinting. The idea of onion routing is to route traffic through multiple servers and encrypt all the way (https://www.torproject.org/). Based on multi-layer encryption and routing, users access the Dark Web sites securely, without revealing their location or their identity and anonymizing visited websites and their users (Chertoff, 2017; Grabosky, 2016; Malik, 2018). Ensuring the identity of users and their computers are anonymous in case they are being monitored.

The Dark Web is accessed and used for both legitimate and illegitimate reasons. Legitimate users of the Dark Web include journalists, whistle blowers, and activists, web users seeking to protect their privacy, or users who want to access hidden sites and services that may have been restricted or censored (Chertoff, 2017; Finklea, 2017; Mirea et al., 2019; Monk et al., 2018). Aside from the anonymity users benefit from while on the Dark Web, the site is useful for cybercriminals. However, "the darknet might not be intrinsically criminogenic – it does not naturally increase criminal activities; rather it might be just another tool that is used by some individuals to carry out illicit activities" (Mirea et al., 2019, p. 104). Dark Web is used by criminals to sell and buy illegal goods. Marketplaces such as the Silk Road, AlphaBay, DreamMarket are used for advertisement and trading of illicit goods such as drugs, weapons, stolen identities, and credit card details, child pornography, hackers, and hacking software for hire (Finklea, 2017; Gupta et al., 2021; Mirea et al., 2019). While law enforcement has shut down the marketplaces, new versions are recreated or moved to forums within the Dark Web where they communicate and perform their illegal transactions (Berton, 2015; Gupta et al., 2021), posing challenges for the detection and swift removal of criminals on the Dark Web.

4.3 Related Work

4.3.1 Virtual Communities and Terrorists Online Content

Most of the studies on terrorists' use and/or misuse of the Internet have identified how terrorists use the internet. The fundamental benefit is to spread propaganda/publicity, psychological warfare, data mining, fundraising, recruitment and mobilization, networking, information dissemination, and secure communication (Conway, 2017; Gill et al., 2017; Hoffman, 2006a). As a result, there has been a legitimate concern about the role of the Internet in radicalization. According to McCauley and Moskalenko, radicalization "means change in beliefs, feelings, and behaviors in directions that increasingly justify intergroup violence and demand sacrifice in defense of the ingroup" (McCauley & Moskalenko, 2008, p. 416). Bartlett and Miller (2012) distinguish between violent and non-violent radicalization. In violent radicalization, the end goal is violence where individuals undertake, aid, or abet terrorist activity. In non-violent radicalization, individuals hold radical views but do not undertake, aid, or abet terrorist activity. Thus, radicalization is a social process in which extremist ideologies and beliefs are developed, stemming from exposure to radical content via social ties such as family and peer relationships offline and online via websites, forums, and social media. While radicalization requires face-to-face interaction and trust, social interactions and networks remain influential in the process of radicalization and violent extremism. However, in the absence of broader social ties to groups or terrorist networks, online communications expose radical content and facilitate violent or non-violent radicalization (Bloom & Daymon, 2018; Holt & Steinmetz, 2020). Nearly all lone-wolf terrorists have been radicalized online (Brynielsson et al., 2013).

Forced out of operating and communicating via public spaces such as the mosques and prisons, terrorists and violent extremists have moved their activities to private areas and personal computers for security, safety, accessibility, and anonymity (Edwards & Gribbon, 2013; Sageman, 2011). Interactive platforms such as chat rooms, message boards, and social media allow interaction amongst like-minded individuals to develop, hence binding, bringing together individuals, fostering a sense of belonging and self, loyalty, emotional attachment, and meaning amongst participants (Bloom & Daymon, 2018; Gaudette et al., 2020; Piazza & Guler, 2019). Interactive platforms have the same influence in the process of radicalization as face-to-face interactions. As such, active online discussions about ideologies and opinions with friends or strangers and the sense of belonging results in a 'virtual community' (Bowman-Grieve, 2009; Scrivens et al., 2018), in the case of jihadists, a 'virtual Muslim community (ummah)' (Erez et al., 2011). Within the virtual communities, participants post and share their views and knowledge, offer support and justification of the use of violence, and validate. Recruitment through the websites and forums is facilitated through community and propaganda efforts (Burris et al., 2000). As a result, a virtual community network is created as extremists and terrorist groups bond and connect with their audiences using these online communication systems, forming a sense of community and belongingness. The virtual networks comprise unaffiliated sympathizers, propagandists, fighters, recruiters, and rivals.

Once in the virtual community, users' language and social interactions change over time, mirroring those within their online community. As time goes by, users in these platforms have roles including producing and aggregating content, translating and curating content across platforms, or passively consuming the information (Shehabat & Mitew, 2018). For instance, in Telegram, ISIS's outlets, there are three kinds of active Telegram users (i) those seeking information, (ii) those who want to engage more fully with the terrorist group, and (iii) propagandists in search of both (Bloom et al., 2019). As participants feel comfortable and welcomed in these platforms, skeptics and lurkers are encouraged to share their beliefs and become more vocal. In addition, since participation in these forums is an active process, users select platforms most compatible with their views (Piazza & Guler, 2019; Sageman, 2011). The more time individuals spend online, the more likely they may be radicalized into violent or radical social networks.

To make their platforms attractive, persuasive, and addictive, terrorist groups engage in "soft power" to influence their target audience using persuasive language and arguments without coercion (Bloom et al., 2019). For instance, ISIS' telegram platform has a different selection of materials such as emojis, stickers, and memes, ISIS's pop culture, which are shared and spread in the forum hence sustaining user engagement and mitigating boredom (Bloom & Daymon, 2018; Bloom et al., 2019). The vulnerable, loners, individuals without positive social ties and bonds, are susceptible to these recruitment tactics and become targets of online recruiters.

The online content and strategy are conducive to radicalization, as users seeking justification for joining or intending to join terrorist groups are likely to be indoctrinated by these messages. Additionally, terrorists such as ISIS online recruitment process follows a specific trajectory: (i) discovery of a potential recruit, or a recruit discovers ISIS, (ii) supporters form a community around potential recruits, (iii) potential recruits are encouraged to isolate and cut ties from their social bonds (families, friends), (iv) move to private or encrypted messaging spaces for conversations, and lastly, identify and encourage action to be taken by the recruit, whether to travel and join IS or carry out an attack at home (J. M. Berger, 2016; J. M. Berger & Strathearn, 2013). While the content attracts sympathizers and would-be extremists to frequent the platforms, recruiters monitor these virtual communities to find potential recruits.

Although the Internet has provided an eager and ready-made audience and acts as a "larger megaphone" (Edwards & Gribbon, 2013) for violent extremists than traditional modes of communication, some scholars are skeptical about the influence of online platforms and communities. Holt and colleagues argue "close to 99%" of individuals exposed to radical messages and content never engage in violent extremism (Holt et al., 2017a). Similarly, Benson (2014) notes the Internet appears to be one influence among many other factors in relation to violent extremism. He claims the Internet is more advantageous to state security forces than terrorist groups. Individuals participating or frequenting these sites are only interested in conversations and not in political violence (Awan, 2017). On the contrary, the Internet is used by violent extremists groups to reinforce the radicalization process as individuals in these online communities experience an echo chamber, where only violent extremists views are expressed and reinforced (Behr et al., 2013; Piazza & Guler, 2019). For instance, as Gaudette et al. (2020) note about the Far-Right, due to the access to extreme right-wing content, exposure, and networks of like-minded individuals online, the internet played a role in facilitating their process of radicalization. Yet, as extremists and terrorists use the Internet and form virtual cliques and communities, terror networks are still a combination of online and offline elements which mutually influence each other (Sageman, 2011).

4.3.2 Terror on the Dark Web

As the terrorist platforms and materials on the surface web are monitored and removed, the groups migrate to the Dark Web, where anonymity is provided if not guaranteed. Private virtual spaces are now encrypted. Administrators must verify members, membership is by referral only, and links and instructions for accessing the Dark Web forums are shared on the social media accounts or encrypted platforms (Berton, 2015; Malik, 2018). The Dark Web sites provide a ‘safe space’ for the groups and their supporters to interact, communicate and access the information without being monitored. Media jihad is the production, distribution, promotion, collection, and redistribution of jihadist content as an obligation to incite violence against the enemies (Erez et al., 2011). As such, the ‘media mujahideen’ (Malik, 2018) amplify terrorists’ online messaging campaign by sharing and reposting official content of the groups across social media platforms and the Dark Web forums. At times, “different terrorist groups compete for control of these forums: Shumukh al-Islam, for example, is a forum which oscillates between ISIL and al-Qaeda supporters” (Berton, 2015, p. 2).

Terrorists and their supporters use the Dark Web for communication via forums and emails, where they discuss religion and jihadist ideologies (Abbasi & Chen, 2007; J. R. Scanlon & Gerber, 2015). Forums are the center of online jihadist activism; terrorists control them, and members can ask questions or contribute to discussion topics and private communications (Erez et al., 2011). The forums encourage radicalization as well as promote “a “self-starter” type of terrorism. This is a strategy endorsed by IS both online and offline and motivates vulnerable individuals to commit violence in the organization’s name in an attempt to “crowdsource” terrorism” (Malik, 2018, p. 20). The jihadist discourse online focuses on religious and ideological indoctrination, framed by references to the Quran and the early history of Islam (Erez et al., 2011). One can find texts such as the ‘Terrorists’ Handbook’ and ‘Explosives Guide’ that can be purchased via the Dark Web marketplaces specializing in selling weapons (Berton, 2015). Yet, with the inaccessible nature of the Dark Web and with most people only accessing the surface web, mass recruitment rarely takes place on the Dark Web. Instead, the groups tend to draw interested supporters from the surface web to the more secure Dark Web platform for further interaction, indoctrination, planning, and launching terror attacks (Malik, 2018).

Although terrorists use the surface web to store and share their information, internet companies and social media platforms often remove the materials. Hence, the Dark Web is used as a repository to store its material. For instance, following the November 2015 attacks in Paris, ISIS used the Dark Web to create a new propaganda hub where they shared news and propaganda to their supporters and also to protect its information from being hacked/removed from the surface web by the Anonymous hacktivists (Malik, 2018; Weimann, 2016a). Furthermore, the groups also use Dark Web for financing through cryptocurrencies, where donations are paid via bitcoins that cannot be traced by law enforcement (Berton, 2015). Therefore, the Dark Web is an ideal ecosystem for terrorist groups; compared to the surface web, they can avoid censorship, discuss, communicate with their supporters, protect their identity and supporters, store their content without removal and fundraise anonymously.

4.3.3 Covert Social Networks

The application of social network analysis (SNA) in the study of terrorism is limited (Koschade, 2006; Perliger & Pedahzur, 2011). Social network analysis (SNA) takes relations as the fundamental unit of social analysis rather than individuals, groups, attributes, or categories (Borgatti et al., 2013; Borgatti & Halgin, 2014). Relations measured are the ties and interactions amongst individuals such as family, friendships, religion, politics, finance, in the case of terrorists' ideological links that form a social network. A social network is represented by nodes (vertices, actors, individuals, or events) connected by complex relationships creating a network. Edges are ties, links, or connections between the nodes, and attributes (attitude, opinions, behavior, demographics) are also fundamental for social network analysis (Borgatti & Halgin, 2014; Scott, 2000).

SNA is suitable for analyzing terrorist networks based on the relationships as the attributes (e.g., real names, gender, age) are often difficult to obtain for covert networks (Basu, 2014). Raab and Milward (2003) introduced the concept of 'dark networks' to describe networks of actors and organizations that engage in covert and illegal activities. Several network studies have been applied to these dark networks, including, organized crime and gangs (Bienenstock & Salwen, 2015; Bright et al., 2012; Duijn et al., 2014), terrorism (Almquist

& Bagozzi, 2019; Asal & Rethemeyer, 2008; Bienenstock & Salwen, 2015; Campedelli et al., 2019; Carley et al., 2001; Koschade, 2006; Krebs, 2002; Milla et al., 2020; Perliger & Pedahzur, 2011; Sageman, 2004). Yet only a few have focused on Dark Web forums of terrorist and hacker groups (Monk et al., 2018; Pete et al., 2020; Phillips et al., 2015).

Research has shown that the covert networks of terrorist groups tend to be decentralized (Raab & Milward, 2003; Sageman, 2011) and connected by actors/nodes who function as 'brokers' (Everton, 2008; Raab & Milward, 2003). For instance, Sageman (2004) study of the global Salafi jihad found the network relatively centralized to decentralized with four clusters. The network had more connected nodes (hubs) significant for the terrorist network. It was a small-world network, where influential nodes connected with more people who then influenced others to join the network. Unlike hierarchical networks that can be destroyed by removing their leadership, the small-world network resists fragmentation because of its dense interconnectivity. Thus, the network may be inefficient by identifying and taking out central actors (nodes). However, only small-world networks with targeted attacks at their hub (actors with a high number of ties) are efficient in disrupting the network (Carley et al., 2001; Sageman, 2004).

Network disruption occurs when a network "cannot efficiently diffuse information, goods and knowledge" (Duijn et al., 2014, p. 2). Network destabilization is often based on the critical player approach that seeks to identify the actors with the most considerable impact on the network and remove them. Some actors are central to the network, key influencers, brokers, leaders (Koschade, 2006; Perliger & Pedahzur, 2011). Detecting actors in the network who are critical to the continuous operation of the group is essential as it highlights the power division and the motives within the network. Such actors have roles imperative for the success of the network. For example, Koschade (2006) social network study of the Jemaah Islamiyah cell operating in Bali found that actors most central and active within the network had the ability to access others and had the greatest control over the flow of information. In contrast, actors with the lowest centrality scores were those in the periphery, isolated, and only called upon for assistance.

Similarly,Milla et al. (2020) study of Indonesian terrorist groups found that operational leaders – those who manage and organize groups, recruit members - are more central than ideological leaders – the preachers. Thus, ideological leaders in operational networks may have less influence than operational leaders. Key members in the network are responsible for coordinating the group's activities, recruitment, and manipulating information flow. In such networks, the removal of operational leaders may disrupt the network. Yet this 'whack-a-mole' strategy (Everton, 2008) may lead to new leaders, and new networks will emerge. Information will continue to flow as covert networks are quick to adapt. Since "covert networks have the ability to heal themselves in the event of loss of a node" (Koschade, 2006, p. 25). Criminal networks become stronger after targeted attacks (Duijn et al., 2014). Similarly, covert networks of Dark Web forums are often decentralized, and the identification and removal of key players usually lead to the networks becoming more decentralized, thus harder to disrupt (Pete et al., 2020). Although removing critical individuals disrupts a network, one of the latent consequences of network destabilization is that decentralized networks tend to reemerge, thus posing challenges in an ultimate and successful removal of covert networks.

Uncovering cohesive subgroups, cliques, groups, or communities is significant for understanding members' network structure and functions. Cliques or communities are formed when each actor in the subgroup is connected to all the other actors (tightly connected) and few connections to the actors outside the subgroup (Newman, 2006; Papadopoulos et al., 2012). Meaning that some actors interact more strongly with members of their communities than they would with members of other subgroups (Fortunato & Hric, 2016). Identifying the subgroups, as Perliger and Pedahzur (2011) note:

> "Allows us to detect different functions of the network (founders, collaborators, passersby), network recruitments paths, operational characteristics (for example, in some networks there is a clear distinction between a suicide bombers subgroup

> and other subgroups, while in others, the suicide bombers are isolated actors in the periphery of the network) and patterns of flow of information. . . . By looking at the attributes of the subgroups, we can evaluate ideological homogeneity and level of solidarity within the network and how this influences the activities and development of the terrorist networks" (Perliger & Pedahzur, 2011, p. 13).

Subgroups or network divisions within the Dark Web forums can be identified and classified based on the related topics and interests of the forum members (Pete et al., 2020). Thus, removing key members in these forums or members who act as brokers between subgroups may disrupt the information flow in that subgroup and the whole network.

Identification of ties, weak ties (relationships with acquaintances), and strong ties (family and friends (Granovetter, 1973) is a significant component for network analyses. The tie strength may be based on the amount of time spent together between a pair of members (Krebs, 2002). Weak ties within terrorist cells are significant for survivability when one of the cells is exposed. On the other hand, strong ties facilitate solidarity and commitment (Perliger & Pedahzur, 2011; Sageman, 2011). Trust is significant within a terrorist network (Krebs, 2002; Milla et al., 2020). Covert or criminal networks are often larger, incomplete, and have fuzzy boundaries as clandestine network borders are unclear; individuals may belong to different groups and networks. The covert networks are dynamic, as the network is constantly changing, with interactions between individuals strengthening or weakening over time (Koschade, 2006). Indeed, as Krebs (2002) noted when collecting data to map the 9/11 airline hijackers, he encountered the problems of incompleteness due to missing nodes and links, difficulty in deciding who to include and who to exclude in the network, and the networks are not static; they are constantly changing.

In sum, research on terrorist use of the Internet, both surface and Dark web, suggest that the groups utilize the platforms to spread propaganda, communicate, recruit, radicalize and finance. Although recruitment often takes place on the surface web, the groups lure potential recruits to the Dark Web for further indoctrination, safer communication, and planning of attacks. The forums provide a safe space, a sense of belongingness for a virtual community of like-minded individuals to interact. Members discuss, share, support, and justify the jihadist ideology within the virtual communities. Terrorists face the threat of disruption from law enforcement and counterterrorism by removing and disrupting social media platforms, websites and detecting and eliminating critical leaders in their online social networks. However, little is known about the Dark Web forums used by terrorist groups and their supporters. There still exists less research applying social network analysis in the study of terrorist groups, especially the groups and their supporters in virtual covert communities. In this study, I aim to add to the literature knowledge about the network structure, behaviors, and significant members in the Dark Web forums of jihadist groups. The research questions guiding this study are: what is the network structure of jihadist Dark Web forums, and how does the structure affect interactions? What communities can be detected from the network? What are the posting activities of key members in the forums?

4.4 Research Design

This study applies social network analysis (SNA) to explore the network of jihadists and their supporters on the Dark Web. SNA uncovers patterns and behaviors of people's interactions. The aim is to understand the social network structure of the forum users, their posting behavior, and communities. SNA requires relational data to measure the relationships, links, or ties (edges) between individuals or groups (nodes or vertices) (Borgatti et al., 2013; Borgatti & Halgin, 2014). Thus, a social network consists of individuals or groups (nodes) connected (ties) by kinship, friendship, common interest, financial interest, and beliefs. Network analysis is significant for this study as it complements conventional studies on terrorism and provides information about the characteristics of the group structure (influencing motives, behaviors, and outcomes), the recruitment processes, and the division of power among its members (Perliger & Pedahzur, 2011).

4.4.1 Data

Data used for this study is from the Dark Web Forum Portal by the Artificial Intelligence Lab, University of Arizona. The Dark Web international jihadist forums are dedicated to Islamic ideology and theology; however, some members sympathize with jihadist groups (H. Chen, 2011). Most of the jihadist discourse online involves religious/ideological indoctrination framed by references to the Quran, Hadith, and the early history of Islam (Erez et al., 2011). The Dark Web Forum Portal contains twenty-eight forums from 2004 to 2012; forums are in English, Arabic, French, German and Russian. I analyze two English forums, the Ansar1, is the English version of the Arabic Ansar AlJihad Network, an invitation-only jihadist forum popular with western jihadists (J. Scanlon & Gerber, 2014; Skillicorn, 2010). The forum has 382 members, 29,492 posts, and 11,244 threads. Previous research (see chapter 3) has identified the topics discussed on this forum include jihad, terrorism, and religion. Myiwc (Islamic Web-Community) forum has 756 members, 25,016 posts, and 6,310 threads. Within the English-language forums, these two are the smallest in terms of members.

Forums	**Year**	**Posts**	**Threads**	**Users**
Ansar1	2008-2010	29,492	11,244	382
Islamic Web-Community	2000-2010	25,016	6,310	756

Table 4.1: Dark Web Forums Data

4.4.2 Analysis

Network Model

Networks are conceptualized as graphs. A graph $G(V, E)$ consists of a set of vertices V (nodes) and a set of edges E (links). G is the whole network. It should be noted that nodes are interchangeably addressed as vertices and ties as edges. Due to different types of vertices and links in networks, it is common to measure sets of vertices and edges (Papadopoulos et al., 2012; Pete et al., 2020). The Dark Web forums are organized around specific topics within threads. The threads are an ordered collection of posts by members. The first post begins a thread and others respond by replying to the author/member of the first post or others. In this case, one could measure the set of vertices V to include the members, the posts, and threads, thus $V = (M, P, T)$. The set of edges E would consist of the member-post, post-threads, member-threads, $E = (MP, PT, MT)$.

In this study, the ties among all pairs of nodes in the network and one-mode and two-mode networks (Borgatti et al., 2013; Borgatti & Halgin, 2014; Wasserman & Faust, 1994) are analyzed. In two-mode, affiliation network models, individuals are connected by attending an event, class, etc. The individuals or events are tied to the extent they share affiliations (Borgatti et al., 2018). In this case, members of the forums are connected when they post on a particular thread. The assumption is that attending an event or participating in the same forum and thread indicates an underlying social relationship between members or a potential opportunity for one to develop (Borgatti et al., 2018; Scott & Carrington, 2014). For this study, I create a bipartite network (two-mode) with two nodes $V = (M, T)$ where $M =$ members, and $T =$ thread and the edges, $E \ = (MT)$. Two-mode data can be converted to one-mode data or analyzed directly as a bipartite graph (Borgatti et al., 2018). I analyzed the one-mode data (unipartite) consisting of members-members interaction, which measures the interaction when two members post in the same thread and threads-threads, where thread nodes will be connected if they share an affiliation to a member. Interactions occur when two members post in the same thread, thus participating or co-occurring within a particular thread. Members who post frequently or together in the same thread will have multiple ties, which can be used as weights associated with the edges (ties).

Thus, the analysis begins with the raw dataset of the forums being pre-processed to produce the interaction network of members for analysis. First, observations with missing values and duplicates were removed. Next, the network was constructed in RStudio using iGraph packages (Csardi & Nepusz, 2022) and imported to Gephi, an open-source network graph and analysis tool (Bastian et al., 2009) for visualization. Finally, the undirected interactions network between members of the forums is measured using network statistics outlined below.

Network Measures

I measure and report the network statistics, structure, and node-level characteristics. Measures of network statistics include size (number of nodes and edges), density, average degrees, degree assortativity, network diameter, and average path lengths. Network density measures the extent to which members are connected by direct relations (Scott, 2000), the more nodes are connected, the more dense the graph will be. Density can indicate how covert a network is (Koschade, 2006). However, most covert networks balance between efficiency (low number of unconnected ties) and survivability (high density and high number of redundant ties) (Basu, 2014; Perliger & Pedahzur, 2011). Covert networks and Dark Web forum users are interested in privacy; thus, high density levels and group centrality increase the chance of being exposed and monitored. On the other hand, high density facilitates indoctrination as information spreads, and diffusion is faster in densely interconnected networks (Koschade, 2006; Perliger & Pedahzur, 2011; Pete et al., 2020). Thus, analyzing the network density of the Dark Web forums of jihadists is useful in understanding the information spread and user engagement within the forums.

Similarly, the average degree measures the number of edges attached to a node. The network diameter measures the length of the longest path between any two vertices in a network. The average path length measures the steps taken on average to reach any node in the graph—the average distance between all pairs of nodes. Degree assortativity measures the similarity of nodes based on their degree, how nodes tend to connect to similar nodes (homophily). When high degree nodes are connected to high degree nodes, and low degree nodes are connected to low degree nodes, the network is considered assortative. Some nodes with a high degree may be connected to low degree nodes; these are disassortative networks

(Barrenas et al., 2009; Foster et al., 2010). Assortativity provides information on the pattern of connections between nodes with a similar degree and whether they are likely to interact (Pete et al., 2020). Assortative networks are likely to remain connected with the removal of a node. In contrast, disassortative networks limit the effects of node removal as important nodes tend to be isolated from each other (Foster et al., 2010).

The following network properties are measured to uncover the large-scale network structure of the Dark Web forums. Hubs and small-world property, and community detection. Hubs in a network are nodes with a larger number of connections. A network with hubs is considered decentralized as opposed to a hierarchical network. Such graphs are likely to have the small-world property (Kleinberg, 2001; Pete et al., 2020; Watts & Strogatz, 1998) where two individuals can reach each other in the network through a short path length, a short sequence of acquaintances. The presence of hubs makes it likely to connect many nodes in the network.

In community detection, vertices or nodes are organized into communities or clusters (Fortunato & Hric, 2016; Papadopoulos et al., 2012). Within a community, a set of vertices or nodes are tightly connected in a set and loosely connected to vertices outside of the community. Thus, highly interconnected but with few ties or links to other nodes (Pons & Latapy, 2006). Community detection algorithms are valuable for analyzing large complex networks by identifying the groups of more densely connected nodes than the rest of the network. Detecting the communities or subgroups in the Dark Web forums is significant for analyzing the interaction network's structure, organization, and function. Thus, identifying the groups that tend to participate or interact in forums and threads about a specific topic, e.g., religion, ideologies, or other forum topics. To detect densely connected communities, I utilize a variety of community detection algorithms from the iGraph package and report the results of the Louvain algorithm (Blondel et al., 2008), one of the most popular, efficient, and faster algorithms for uncovering community structure in larger networks. In addition, the algorithm reports the modularity score, which measures how effective a given clustering algorithm is; higher modularity means the algorithm has identified distinct groups and the extent to which the network can be divided into groups.

At the node level, I next measure network centrality to identify the most influential forum members using eigenvector/degree centrality, betweenness, and closeness (Borgatti et al., 2018; Freeman, 1978). Degree centrality measures the number of connections a node has, eigenvector centrality measures the importance of a node if its neighbors are influential as well. Betweenness centrality identifies the 'bridge' nodes, which are significant in connecting subgroups and fall along the shortest path between two nodes. Closeness centrality measures a node's distance to others how close a node is to others in the network. A node with high closeness score is a short distance to others, meaning that information flows quickly to other nodes and the central nodes.

Below I present the results of the interaction networks of the two Dark Web forums, Islamic Web-Community (Mywic) and Ansar1.

4.5 Results

The networks examined from the Dark Web forums are small; the Islamic Web-Community has 638 nodes and 4438 edges, Ansar1 interaction network has 353 nodes and 8765 edges (see Table 4.2). For both networks, to facilitate further analysis and graph visualization, nodes were pruned by removing nodes with zero connections (isolates) and parallel edges; several edges between one pair of nodes were merged and averaged. The Islamic Web-Community network has a smaller density compared to the Ansar1 network. The density score is an indicator of the connectedness of a network, in the case of the Dark Web forums, the covertness and the ability of members to interact by creating a thread or posting in a thread. Smaller density scores indicate a highly covert network, and higher density scores indicate efficiency (Koschade, 2006). Therefore as ideas spread faster in a denser connected network (Pete et al., 2020), a sparsely connected network may not be as effective. Members

Network Measures	**Islamic Web-Community**	**Ansar1**
Number of nodes (members)	638	353
Number of edges (interaction)	4438	8765
Density	0.022	0.141
Average degree	13.912	49.66
Network diameter	6	5
Assortativity (homophily)	0.022	-0.026
Average path length	2.437	1.939
Average clustering coefficient	0.831	0.73

Table 4.2: Interaction Network Statistics

with the highest degree scores are connected to other nodes on average by 14 nodes for the Islamic Web-Community, and 50 nodes for the Ansar1 network. Therefore, the Ansar1 forum is highly active, and members participate in many threads and interact with others. Assortativity measures homophily, the tendency of nodes with the same degree to connect, thus forming an assortative network. Some nodes with a high degree may be connected to low degree nodes; these are disassortative networks (Barrenas et al., 2009; Foster et al., 2010). The Islamic Web-Community network is assortative, meaning that members with high degree nodes tend to connect with other high degree nodes in the network and low degree nodes to low degree nodes (r = 0.022). The Ansar1 forum has a negative disassortative network structure, meaning nodes with a high degree tend to connect with lower degree nodes (r = -0.026). Hence, low degree nodes tend to participate in the same thread as nodes with a higher degree.

The clustering coefficient and average path length analyze the small world effect. A higher clustering coefficient indicates that many connections of a node are also connected. Both networks have a high clustering coefficient, Islamic Web-Community (0.831) and Ansar1 (0.73). The average path length measures the steps to get from one node to another in the network. On average, members of the Islamic Web-Community and Ansar1 forums are about two steps from another member interacting in any other thread. Meaning that any two random members participating in thread1 and another participating in thread2 are connected to at least one member who participates in both thread1 and thread2 (Pete et al., 2020). The network diameter of the Islamic Web-Community is 6, and for Ansar1, the longest path is 5. Diameter indicates how far information must spread from one node from one end to the other end of the network. Therefore, lower average paths facilitate the faster spread of information and ideas.

Community detection algorithms are applied to find tightly connected nodes in a subgroup/community and loosely connected to nodes outside the community. The modularity algorithm (Newman, 2006) produces an overall modularity score which measures the network division into distinct modules or groups. A higher modularity score indicates the presence of densely connected nodes within a community. Below, I present the network graph showing the two forums' communities (Figures 4.1 and 4.2).

The Islamic Web-Community (Myiwc) resulted in eleven communities with a modularity score of 0.361. 37% of the nodes are part of the largest community in pink, 31% are part of the green cluster, while 12 % and 10% of the nodes were part of the blue and black clusters, respectively. Between 5% and 0.31%, the rest of the nodes were part of the smaller communities. The percentage of nodes in each group suggests that nodes within each forum were classified in the densely connected communities. Since the forums are based on threads with specific topics, the network division seen here could be due to members participating in threads they are interested in discussing.

Network centrality measures highlight the influential members and their positions (see Tables 5.1 and 5.2 in the Appendix). The top 3 influential members of the Islamic Web-Community network are Tayeb, Lulua, and OmMohammed. They all have a higher score in degree (more connections), eigenvector (influential neighbors), betweenness (bridge nodes), and closeness centrality. Further analysis of the influential members posting behavior shows that they tended to post frequently. For instance, in this network, Tayeb, the most central node in the pink community (see network graph), had the most posts (2848 out of 25,016) in the forum, participating in 1258 threads (out of 6,310) by discussing issues related to religion (Islam vs. Christianity), terrorists (Taliban, Al-Qaeda), and relationships. This

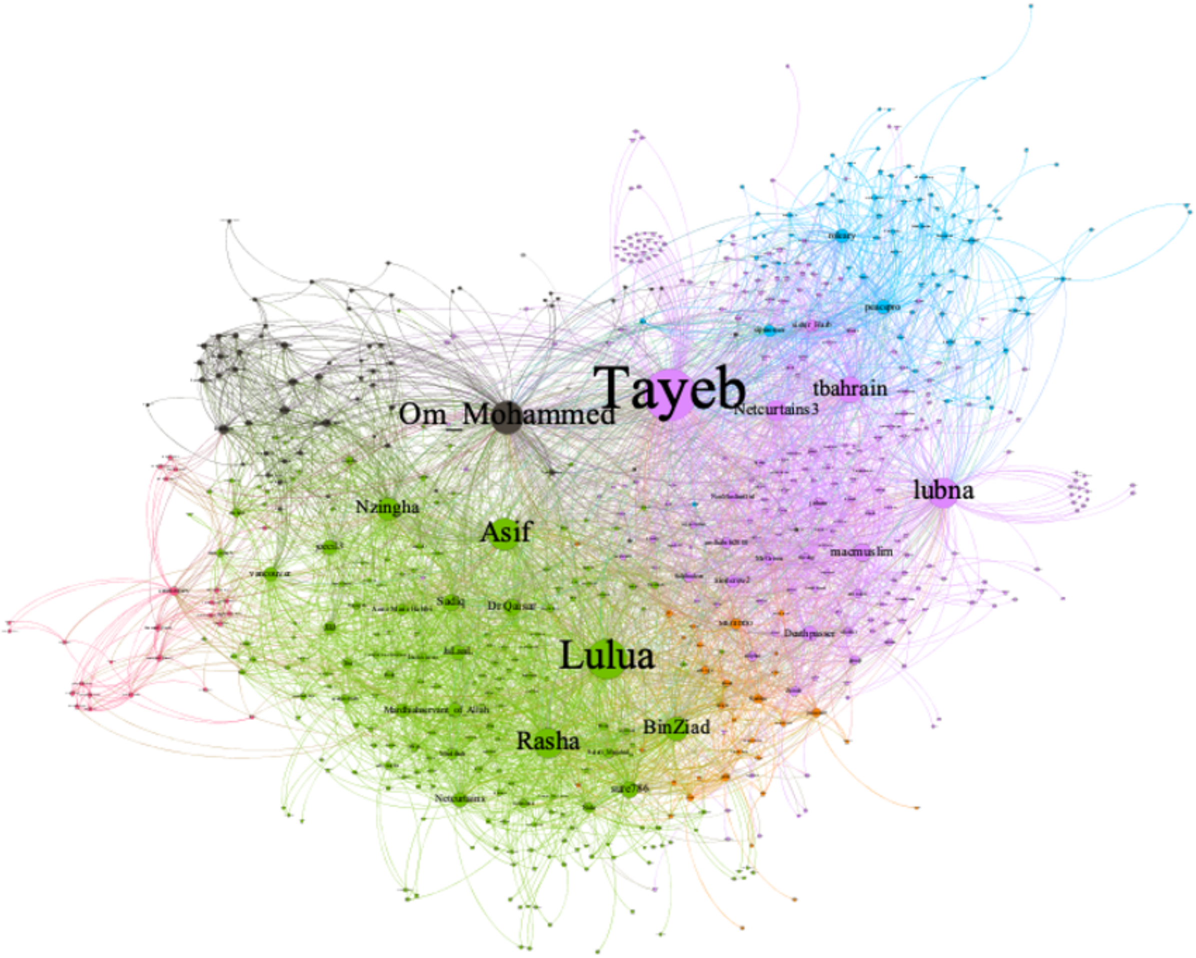

Figure 4.1: Islamic Web-Community (Myiwc) undirected network with eleven communities. 638 nodes, 4438 edges, node size represents the node degree, colors represent communities.

node mainly participated in the thread titled ‘Logic of some Christians. . . not all!’ (62 posts) and ‘Historical Jesus vs historical Muhammad’ (24 posts) discussing Christianity and Islam. The second node with high centrality score is Lulua belonging to the community in green. Lulua had 1659 posts in 1114 threads and responded to threads related to religion, poverty, relationships, politics, beauty and mental illness, news. The node frequently posted in the threads ‘Comparative Study between Quran and Bible. . . ” (35 times), ‘News!’ (25 times) and ‘Islamic Quest’ (14 times). The node OmMohammed in the black community posted less frequently than the other two. This node has 530 posts in 355 threads. Topics discussed include jihad, terrorists and extremism, religion, women's rights, relationships. The node frequently posted in threads about the 'Can Salvation be reached through Christ?' (9 times) and 'A message for People of the Book...' (7 times) and posted a few times in the rest of the threads. One could note the nodes participating in the same threads or mentioning other members with high centrality scores. For instance, OmMohammed responded to a threat titled 'Welcome back bro Tayeb.' Both OmMohammed and Tayeb participated in similar threads. While the node Tayeb frequently posted in this forum, other nodes with lower centrality scores had the most posts (see Appendix). The nodes, Lubna had 1868 posts, netcurtain3 had 1776 posts, while the node Lulua with 1659 posts, is ranked second in all the centrality measures.

The communities in this network are based on topics discussed. The division stems from the specific discussions surrounding religion – Islam vs. Christianity. While the central nodes in the pink and green communities tended to discuss issues related to terrorism, the central node in the green subgroup tended to focus primarily on Islamic topics and discussions related to relationships. Other prominent members in the network tended to discuss similar topics, and some participated in forums about computer viruses and software, religion and poverty, homosexuality, food, and marriage.

The Ansar1 network (see figure 4.2) resulted in four communities with a modularity score of 0.296. In addition, 64% of the nodes are part of the largest community in pink, 17% are part of the green cluster, while 9.35 % are in the orange cluster, and 8.22% of the nodes were part of the blue community. Similarly, the network centrality measures identified the influential members of this network; the top 3 are Asadullah A. (pink community), Insurgent (green community), and Asad'Allah (pink community).

When looking at their communities and the threads/topics, one could see similarities among influential members. For example, the node Asadullah posted 2402 times in 1674 threads. Topics included terrorism - Taliban, terrorist media center Al-Ansar media center, Pakistan under Taliban, Mujahideen, Al-Qaeda, US soldiers' death in Iraq, and religion. The

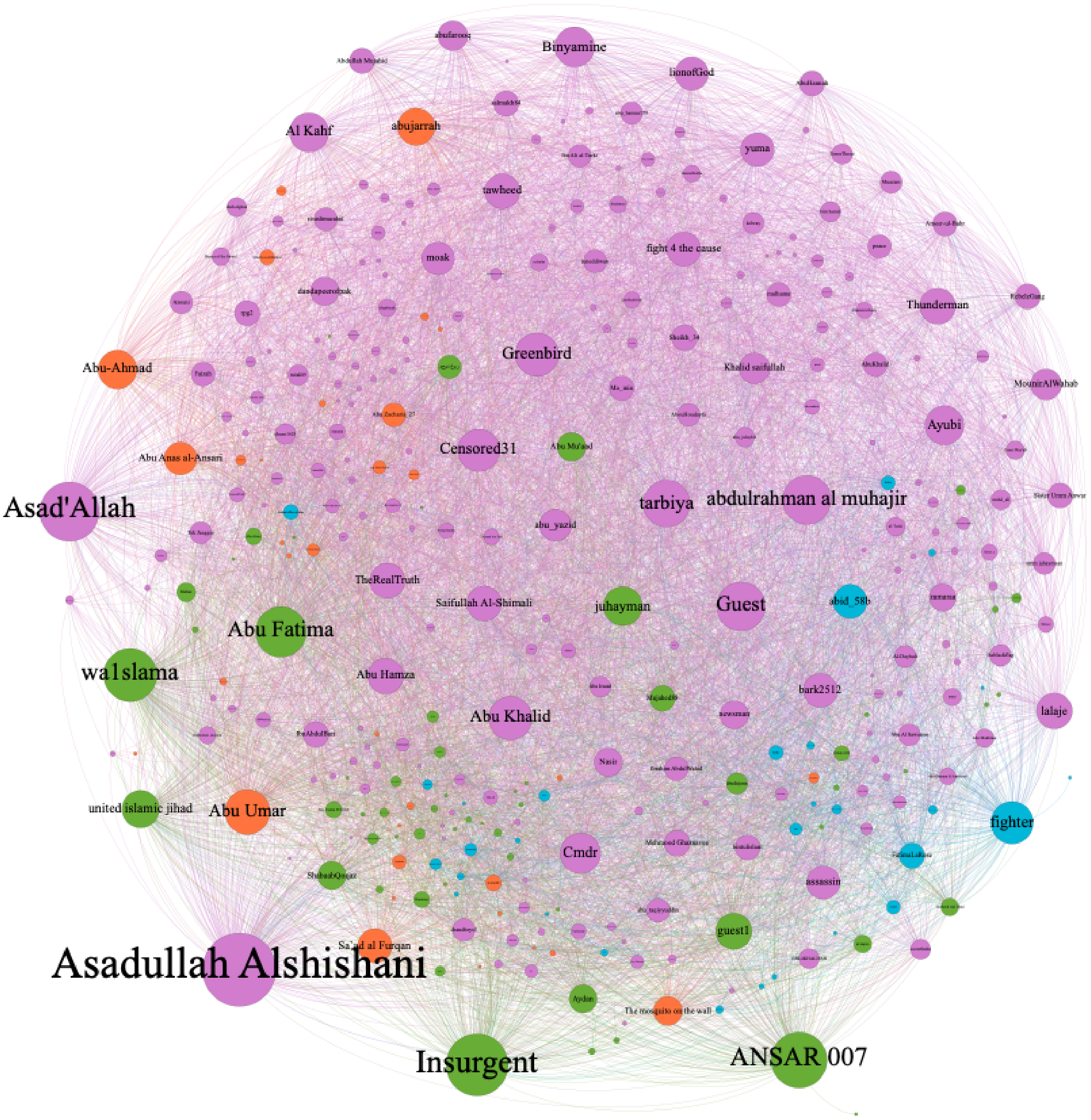

Figure 4.2: Ansar1 undirected network. 353 nodes, 8765 edges with 4 communities. Node size represents the node degree, colors represent communities.

top thread this node participated in was the '13 dead, 30 wounded in dual attacks at Fort Hood, Texas' (38 times), 'Clashes between HAMAS & Jund Ansar Allah kill 24' (28 times), and 'Feds: Leader of radical Islam group killed in raid' (22 times) in this thread, one could see prayers and support for the killed and arrested radical leaders. On the other hand, the node Insurgent posted 2101 times in 1795 threads. Topics ranged from terrorists – Islamic

Emirate of Afghanistan, Al-Ansar media, death of Iraqi police, Taliban, jihad, Al-Qaida, and religion. The thread Insurgent posted mostly is titled ‘Glad Tiding for the Believers and shaking for the disbelievers. . . ” (13 times) and ‘As-Sahab: how to prevent a repeat of Gaza holocaust’ (10 times) but participated less frequently in the rest of the threads.

The node Asad’Allah posted 521 posts in 324 threads, topics include terrorism – Taliban, Afghan attacks, martyrdom bomber, wills of 9/11 martyrs, youth mujahideen movement, jihad, ansar media center, Tamil Tiger rebels, Al-Qaida, Al Shabab in Somalia. The top thread the node participated in was titled “As-Sahab: An address to the American people by the Lion of Islam Sheikh Usama bin Laden (May Allah protect him)’ posted eight times and ‘As-Sahab: 9/11 Surprise, The West. . . and the Gloomy Tunnel’ posted eight times as well. The node posted less in the rest of the threads. It should be noted the important nodes identified here do not necessarily mean they post the most. For instance, while the node ANSAR 007 is ranked fourth in all the centrality measures, the node posted more frequently (2213 posts) than Insurgent and Asad’Allah.

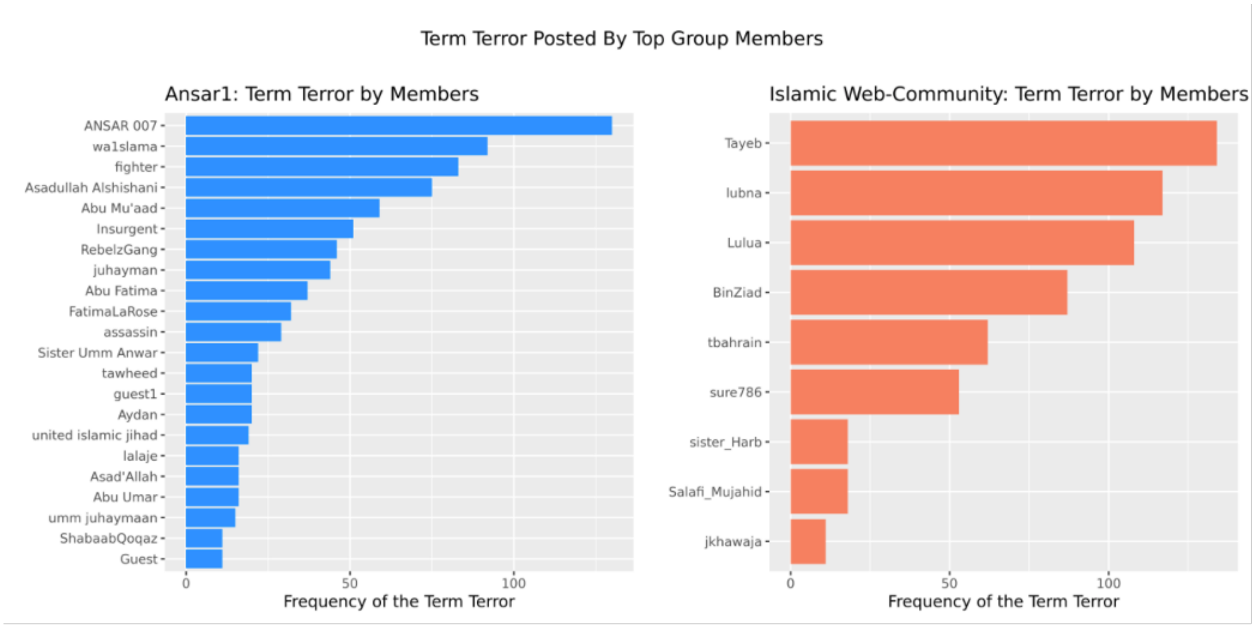

Figure 4.3: Frequency of the term terror posted by key members in the network

The division of the nodes in this network stems from the topics discussed. All three central nodes addressed issues related to terrorism, offering support and sharing news about terrorists and the attacks. Two of the central nodes are part of the large community in pink, while one was in the green community. Other central members in the larger and small subgroups also posted about terrorism. The nodes posting behavior indicate some level of specialization, interests, or roles of the key members in the communities. For example, the central members of the Ansar1 network tend to frequently post about terrorism (figure 4.3), whereas some members of the Islamic Web-Community posted about terrorism. Findings suggest members with high centrality scores may be forum moderators, terrorist recruiters, supporters, or just frequent users with information to share.

4.6 Discussion and Conclusion

Terrorist groups use the Internet for communication, spreading propaganda, recruitment, radicalization, financing. The Internet has created opportunities for self-radicalization, connection, and interaction of like-minded people while facilitating the promotion, popularization, and operation of radical ideologies and ideologically motivated violence. Online interaction occurs in active forums such as the Dark Web forums or chat rooms, creating social bonds, trust, indoctrination, and eventually radicalization into violent extremism. Active online discussions about ideologies and opinions with friends or strangers have led to the sense of belonging within the 'virtual community' (Bowman-Grieve, 2009; Scrivens et al., 2018) and in the case of jihadists, a 'virtual Muslim community (ummah)' (Erez et al., 2011). The virtual communities change users' behavior, mirroring those within the online community.

The network structure of the two jihadist Dark Web forums, Islamic Web-Community and Ansar1, are sparsely connected, smaller density. The density score is an indicator of the connectedness of a network. Members of the Dark Web forums are less connected thus have fewer interactions. However, most covert networks balance efficiency, security, and survivability (Basu, 2014; Perliger & Pedahzur, 2011). Dark Web forum users, in general, are interested in privacy. Members join the forums via invitation or shared links; thus, high density levels increase the chance of being exposed and monitored. Forum users supporting terrorist groups are interested in secrecy as well. On the other hand, high density facilitates indoctrination and information spread; diffusion is faster in densely interconnected networks (Koschade, 2006; Perliger & Pedahzur, 2011; Pete et al., 2020).

Despite the sparse density in these forums, members with the same interests and characteristics, i.e., those who frequently post in various threads, tend to interact with similar members. For example, members in the Islamic Web-Community tended to interact with similar members, i.e., experienced or old members only interacted with experienced members. Thus, this network's information and resource flow were between similar users; one could argue trusted users. However, in the Ansar1 network, some forum members who frequently posted in almost all the threads also interacted with those who posted less or in a few threads suggesting information and resource flow between dissimilar users, the interaction between experienced and new users. Interactions between new and experienced members can shape opinions and influence new members' behavior (Pete et al., 2020). This participating pattern suggests that the Ansar1 network welcomed new members more. On the other hand, some members join the forums to seek information passively, engage with others, or share terrorist propaganda (Bloom et al., 2019; Shehabat & Mitew, 2018). As participants feel comfortable and welcomed in these platforms, skeptics and lurkers are emboldened to share their beliefs and become more vocal. In addition, since participation in these forums is an active process, users select platforms most compatible with their views (Piazza & Guler, 2019; Sageman, 2011).

Detecting the communities or subgroups in the Dark Web forums is significant for analyzing the interaction network's structure, organization, and function. Community detection identifies the groups and members who tend to participate in forums and threads about a specific topic, e.g., religion, than they would with members discussing other issues like terrorism or relationships. Therefore, there is a division of members within the forum; communities can overlap, members can participate in different threads and topics depending on their interests. Participation in forums is an active process; users select platforms most compatible with their views (Piazza & Guler, 2019; Sageman, 2011). For example, the Islamic Web-Community (Myiwc) resulted in eleven communities, and the Ansar1 network resulted in four communities.

Further analysis of the key members in these groups shows the network division is due to the members participating in threads and topics they are interested in discussing. Since the forums are based on threads with specific topics, the topics discussed by the key members in the Islamic Web-Community were mainly about the Islamic religion and a few discussions on terrorism, marriage and relationships, computer viruses, and software. The Ansar1 topics tended to be mostly about terrorism, similar to the results found in a previous paper (see chapter 3). Most of the jihadist discourse online involves religious/ideological indoctrination framed by references to the Quran, Hadith, and the early history of Islam (Erez et al., 2011). The posting activities suggest that some members tended to post in various threads. Some of the central nodes participated in the same forums or mentioned other major nodes in their discussions. Identifying the communities allows us to detect different functions of the network, such as the collaborators, key posters, recruitment patterns (Perliger & Pedahzur, 2011), and in this case, topics of interest and supporters of terrorist groups. The key members identified in these communities influence the topic discussion and information flow within the subgroups and the forums. These could be 'media mujahideen' (Malik, 2018), who amplify terrorists' online messaging campaigns by sharing and reposting official content of the groups across social media platforms and the Dark Web forums. In the posts, it was interesting to note explicit and implicit support for jihad, terrorist groups, and news shared from terrorist media such as Al-Qaeda's as-Sahab, and the Global Islamic Media Front led by supporters of jihad but unaffiliated with any terrorist organization (Erez et al., 2011).

Network destabilization, a common approach in counterterrorism, begins with identifying and removing key members in a network. Some of the actors are central to the network, influencers, brokers, leaders, therefore crucial for the flow of information and resources (Koschade, 2006; Perliger & Pedahzur, 2011). Detecting actors in the network who are key to the continuous operation of the group is essential as it highlights the power division and the motives within the network. Such actors have roles imperative for the network's success; in the case of Jihadists Dark Web forums, they could be unaffiliated sympathizers, propagandists, fighters, recruiters, and rivals. While the content attracts sympathizers and would-be extremists to frequent the platforms, recruiters monitor these virtual communities to find potential recruits.

The effectiveness of disruption strategies is based on network topology and network resilience (Duijn et al., 2014). Disruption strategies target the central members, the hubs. Degree centrality identifies the actor with most direct contacts; actors with high betweennesses centrality scores are the bridge; they control the flow of information and resources within the network. For example, the removal of the top 3 influential members of the Islamic Web-Community network, Tayeb, Lulua, and OmMohammed and the Ansar1 members Asadullah A., Insurgent, and Asad'Allah may disrupt the flow of information in these forums. The key members are responsible for coordinating the group's activities, sharing the content, and starting and responding to a thread. However, the disruption approach may not be effective in the long run as new leaders, and new networks will emerge, information will continue to flow, as covert networks are quick to adapt (Everton, 2008; Koschade, 2006). Similarly, covert networks of Dark Web forums are often decentralized, and the identification and removal of key players often lead the networks to become more decentralized, thus harder to disrupt (Duijn et al., 2014; Pete et al., 2020). It should be noted while an individuals' centrality indicates importance, this does not mean that they hold power in the network (Koschade, 2006). Additionally, some members deliberately operate from the peripheries, thus not easily detected as key members for disruption. Yet, disruption strategies render covert networks more exposed, leading to the practice of network disruption a long-term effort (Duijn et al., 2014). As a result, although network disruption has some challenges as new leaders and networks reemerge, network destabilization is still beneficial. It may temporarily destabilize the network, providing the opportunity to monitor and await new leaders or those in the peripheries to surface.

In summary, this study analyzed two Dark Web forums of jihadist groups and their supporters using social network analysis techniques. The study investigates the Dark Web forum's network structure, behavior, and patterns. The network structure is decentralized, sparse, and divided based on topics (religion, terrorism, current events, and relationships) and the members' interests in participating in these threads. Some members are essential and influential in the information and resources flow within the networks. Identifying key members is significant for counterterrorism, as mapping network structures and key users are important for destabilizing terrorist networks. Ethical implications such privacy violations, lack of informed consent and confidentiality arise from applying SNA to identify and detect key members especially in a forum used by terrorist sympathizers and supporters. While the data analyzed is considered public information therefore requiring no consent from the forum users or IRB approval, the results of this study do not imply the key members identified to be terrorists or supporters of terrorist groups.

Nevertheless, SNA is conducive for predictive and explanatory studies on terrorism and counterterrorism. Future studies perhaps may analyze current Dark Web forums for comparison as one of the limitations of this study was the analysis of forums collected from 2008 to 2012. An investigation that includes member attributes such as time posted, and events may highlight the changes over time to further understand the networks of Dark Web forums of covert networks.

Chapter 5

Conclusions

5.1 Summary

This three-article dissertation sought to explore three distinct but related topics. First was the exploration of cybercrime and cyberterrorism from a criminological perspective. Secondly, to highlight the utility of innovative computational techniques, specifically topic modeling in detecting Jihadist topics and discussion on the Dark Web forums. Thirdly, illuminate and map Jihadist Dark Web forums' social structure and behavior using social network analysis.

In the first paper, I explore existing research on cybercrime and cyberterrorism. It is known that the development of the Internet has provided terrorists with access to a global audience and target. However, criminology researchers examining cybercrime and cyberterrorism are often impeded with the challenge of defining cybercrime, terrorism, cyberterrorism, and access to data that can be used to analyze and understand cybercriminals. Similarly, terrorism and violent extremism scholars are divided as the term cyberterrorism has sparked an intense debate over whether terrorist groups can attack cyberspace and critical infrastructures to further their goals. For some researchers, terrorists have to engage in cyberattacks that result in violence against individuals (death, injury) or property (explosions, economic loss) to be considered cyberterrorism, the convergence of cyberspace and terrorism. Furthermore, terrorists do not possess the technical capability required to attack cyberspace. However, some scholars argue terrorist groups engage in cyberterrorism by targeting government or businesses' online critical infrastructures. Cyberterrorism does not have to result in violence to be taken seriously.

Indeed, while the groups have not yet attacked a critical infrastructure such as power plants, they are attacking computer systems and networks, stealing data and information, thus increasingly posing serious threats to national security, the economy, and public policy. Research shows that the ability to attack cyberspace has a far-reaching impact on terrorist groups' enemies, unlike bombings restricted to a specific physical location and communities. To address their lack of technical sophistication, some group members are tech-savvy, or they recruit tech-savvy supporters to attack cyberspace through website defacement, distributed denial-of-service (DDoS), cyber-enabled financing, credit card fraud, stealing data, and selling on the Dark Web.

The existing research reveals that due to the rapid technological advancement and emerging technologies, including artificial intelligence, and the unintended consequences of these technologies, it is challenging to predict cybercriminals and cyberterrorists' technical capabilities and cyberattacks. Indeed, while there have been few incidents of cyberterrorism globally, mostly attacks against governments, organizations, and business networks, we cannot ignore the threat of severe cyberattacks and critical infrastructures posed by terrorist groups. Countries face a growing risk of cybercrime and cyberterrorism as we rely more on the Internet and technologies for communication, business, health, economy, governing, and critical infrastructures.

The second article is methodological and aims to address the issue of terrorism content on the Dark Web forums and automatic techniques, such as topic modeling that can be applied to classify and detect the topics. Terrorists use the Internet to spread propaganda/publicity, psychological warfare, data mining, fundraising, recruitment and mobilization, networking, and secure communication facilitating the promotion, popularization, and operation of radical ideologies and violence. Interactive platforms such as chat rooms, forums, message boards, and social media allow interaction and communication amongst like-minded individuals to develop, thus fostering a sense of belonging and self, loyalty, emotional attachment, and meaning amongst the virtual community. The Jihadist content on these platforms focuses on religious and ideological indoctrination, inciting violence against their enemies, share information about attacks, strategies, and justifications.

Although criminologists have applied classical research methods to analyze terrorism and contributed to the development of novel research methods, crime theories, findings, and policies, few have yet to embrace computational approaches to the study of crime and crime-related problems, specifically terrorism and extremism studies. In the second paper, I show how natural language processing computational techniques, specifically LDA topic models, can be applied to inductively analyze and uncover topics from the large-scale information generated by terrorists and their supporters. I applied topic models and sentiment analysis on three Dark Web forums of Jihadist groups and their supporters, the Ansar1 (Ansar AlJihad), the Gawaher forum, which is dedicated to discussions of Islam and the Islamic world, and the Islamic Network forum, which is also devoted to the theology and world events. Some members of these forums sympathize with and support terrorist organizations.

Results show that the most prevalent topic in all forums was religion. Topics about religion included Islam, Christianity, and Judaism, discussion about Allah and God, believers and disbelievers, Quran recitations, religious rituals, and philosophical discussions about religion and science. These topics are expected, as the forums were dedicated to the Islamic faith and issues concerning Muslims. The forums also discussed terrorism, mainly within the Ansar1 forum. The discussion was about terrorists and attacks, support for the Mujahideen fighters, jihad, Al-Qaeda, Taliban, the U.S in Iraq, Afghanistan, the arrests and torture of terrorists and foreign fighters. The support of terrorists and sharing news about the attacks signify the media strategy of the groups and their supporters to communicate and justify their ideological goals. Indeed, one could argue the discussion or posts about terror groups were created by the terrorist group's media centers and supporters, also known as media mujahideen. A few of the discussions were related to relationships and marriages, advice, seeking help, health, food, selling electronics, and identity cards.

It is known that Jihadists' ideologies and narratives are based on the misinterpretation of the Islamic religious traditions and historical events framed by references to the Quran, Hadith, and the early history of Islam. Therefore, participating in forum discussions is also considered a form of jihad. LDA topic modeling has several advantages over the manual annotation of text data, especially when analyzing a large body of text such as the Dark Web forums. Topic models quickly discover hidden topics, providing real-time analysis of terrorist online content. The topics are inductively uncovered, thus requiring little or no previous knowledge of the content from the text. Additionally, the keywords discovered from the topic models can further be utilized to organize and search for content in other text data. For instance, the keywords derived from the three Dark Web forums can be used to analyze or search other Dark Web forums or websites for similar topics.

The topic models suggest that terrorist content and discussions occur in Dark Web forums that are created for the discussion of religion. Therefore, counterterrorism measures should also focus on data-driven applications such as topic modeling in detecting and analyzing terroristic content from large-scale textual data. Additionally, monitor Dark Web forums and especially forums that discuss religion, as terrorist groups are known to use religion to justify their goals and recruit in such forums for supporters. Topic modeling for counterterrorism is significant as it is quicker to detect and classify terrorist content for disruption than manual analysis. In addition, the detection of extremist content online using topic models is significant for analyzing users' beliefs and countering/disrupting their online narrative. The application of sentiment analysis provides insight into the opinions and attitudes of forum users toward the prevalent topics of religion and terrorism. For instance, the forum Ansar1 tend to portray negative attitudes towards the war, military, police, and government, while the Gawaher and Islamic Network forums were mostly positive about topics surrounding the Islamic religion.

The third article builds on article two, and it is an exploratory study providing an insight into the network structure and behavior of Dark Web forum users. In this paper, I use social network analysis to examine the networks of two international Jihadist groups' Dark Web forums. I construct interaction networks amongst members of the forums to map, visualize and measure the interactions between members within the forums, illuminating the network structure, how members interact with each other, the key members, virtual communities, and how the network behaves. The research questions guiding this study are: what is the network structure of jihadist Dark Web forums, and how does the structure affect interactions? What communities can be detected from the network? And what are the posting activities of key members in the forums?

Results suggest that the network structure of the two Jihadist Dark Web forums, Islamic Web-Community and Ansar1, are sparsely connected have smaller densities. Thus, implying a decentralized network, a 'leaderless' (Sageman, 2011) network where all members communicate equally, with not one leader but several important members spread across the forums. In such networks, individuals are likely to act independently or in smaller groups. Members of the Dark Web forums are less connected thus have fewer interactions; this is expected in a covert network. Dark Web users, in general, are interested in privacy and anonymity. Members join the forums via invitation or shared links. A well-connected network would increase the chance of being exposed and monitored but also successfully facilitate indoctrination and the spread of information as diffusion is faster in densely interconnected networks. Despite the sparse density in these forums, members in the Islamic Web-Community tended to interact with similar members, i.e., experienced, or old members only interacted with experienced members. Thus, this network's information and resource flow were between similar users; one could argue trusted users. However, in the Ansar1 network, some forum members who frequently posted in almost all the threads also interacted with those who posted less or in a few threads suggesting information and resource flow between dissimilar users, the interaction between experienced and new users. Interactions between new and experienced members can shape opinions and influence new members' behavior.

Furthermore, as participation in forums is an active process, users tend to select platforms most compatible with their views, forming a subgroup or community. For example, the Islamic Web-Community (Myiwc) resulted in eleven communities, and the Ansar1 network resulted in four communities. Further analysis of the key members in these groups shows the network division is due to the members participating in threads and topics they are interested in discussing. Since the forums are based on threads with specific topics, the topics discussed by the key members in the Islamic Web-Community were mainly about the Islamic religion and a few discussions on terrorism, marriage and relationships, computer viruses, and software. The Ansar1 topics tended to be mostly about terrorism, similar to the results found in article two. Some members are essential and influential in the information and resources flow within the networks. Identifying key members is significant for counterterrorism, as mapping network structures and key users are important for destabilizing terrorist networks. However, the results of this study do not imply the key members identified are terrorists or supporters of terrorist groups. The analysis of the interaction networks provides an understanding of terrorist behavior, thus leading to more effective counterterrorism measures such as disruption and quicker intelligence gathering of online terrorist networks and operations.

5.2 Limitations and Future Research

This dissertation has explored the problem of cyberterrorism and the applicability of automatic computational methods in analyzing Dark Web forums and networks. However, this dissertation is not without limitations. First, as noted elsewhere, the lack of access to official data on terrorism and cyberterrorism is one of the limitations for any research on terrorism and cyberterrorism (A. M. Bossler & Berenblum, 2019; Bossler A.M., 2017; Diamond & Bachmann, 2015; Holt, Lee, et al., 2020). Official data on terrorism is limited; if available, it is censored or influenced by political considerations. Some countries do not collect terrorism data, and those who keep terrorism databases are unwilling to share it with the research community (Hodwitz, 2019; LaFree & Ackerman, 2009). Additionally, in some countries such as the U.S, terrorists are often not prosecuted for terrorism acts but other related offenses; therefore, gathering data of those arrested, convicted, or prosecuted for terrorism acts is challenging (LaFree & Dugan, 2015; Rosenfeld, 2002). Industry or open-source government data are rarely available due to victims underreporting over fears of economic loss (Holt, Cale, et al., 2021; Holt & Steinmetz, 2020). Furthermore, cyberattack incidents are often omitted from terrorism databases "due to their failure to qualify as a "violent" incident" (Freilich et al., 2014; Holt, Cale, et al., 2021; Holt, Navarro, et al., 2020). As such, most scholars have resorted to examining a sample of college and school students, mainly surveying their hacking and digital piracy experiences, analyzing bulletin boards and blogs, and field experiments (A. M. Bossler & Berenblum, 2019; Bossler A.M., 2017; Burruss et al., 2012; Holt, Lee, et al., 2020). Others have collected data from simulated computer systems created to be attacked to analyze system intrusions (Maimon et al., 2013; Testa et al., 2017).

Due to the challenge of accessing terrorist online content, in this study, I am analyzing a secondary dataset, the Dark Web Portal, collected by the University of Arizona, Artificial Intelligence Lab. Therefore, the analysis is limited to the data and variables available in the collection. As a result, I may not have been able to analyze other essential variables not available in the database that would be conducive to exploring topics and users of Dark Web forums, such as demographics and locations of the Dark Web forum members. It is important to note while it is easier and faster to collect behavioral data of Internet users online, it is challenging to collect and assess demographic data from Internet users. The Dark Web forums include username, posts, threads, date, and year of the posts. The names are expected to be aliases of the users interested in anonymity. However, including the username and dates they participated in the forums in future research would provide insight into the members' topics and time spent interacting with others. The analysis of time spent on the forums, the topics discussed, and interactions may classify radical, non-radical, or neutral users.

Secondly, although the dataset analyzed of forums was from 2004 – 2012, it still provided insight into the Dark Web forums topics and network behavior. A comparison with a current dataset would provide insight into the changes in topics and posting behaviors of the forum users. A current dataset may also be used to compare the language use of forum users, as language is known to evolve. Thirdly, this dissertation has taken a computational social science approach that integrates social sciences with statistics and computer science techniques focusing on topic modeling, specifically LDA. However, other topic modeling algorithms such as structural topic modeling and author-topic models can be applied to identify the authors' topics. Additionally, different computational approaches can be used to analyze text data, such as semi-supervised machine learning algorithms. For instance, a semi-supervised study of Dark Web forums would entail manually annotating a percentage of the forum posts as radical, non-radical, or neutral and using those annotations as dependent variables that can be applied to classify the rest of the Dark Web forums and users. Another application would be, instead of manually classifying the Dark Web forums, to use the results of topic models and machine learning classifiers to analyze the rest of the Dark Web forums. Additionally, topic models can be combined with sentiment analysis, topic models highlight the topics discussed in the Dark Web forums, and sentiment analysis highlights the attitudes of forum users based on their posts and topics i.e., negative, positive, or neutral. A further analysis of the crime-terror nexus on the Dark Web would shed light into the partnership between cybercriminals and cyberterrorists and their behaviors.

Thirdly, social networks analysis of the Dark Web forums was exploratory. Future directions may include a deeper analysis of the topics identified by topic modeling and key members to highlight what they post and how they behave and interact in the Dark Web forums. Another future direction will employ time-series analysis to measure the network changes, identifying key members and groups and how they change over time. Another approach would be to apply Exponential Random Graph Models (ERGM) to analyze and predict the relationship of forum members based on their posts and time spent on the forums.

Lastly, although this research applied computational social science methods such as topic models and social network analysis, the ethical concerns arising from the computational techniques should be noted. Applying topic models to classify text as radical or non-radical and identifying key forum users based on their network position and posting behavior may lead to unnecessary profiling and/or mislabeling Internet users who may have accessed the Dark Web forums for legitimate purposes.

References

Abbasi, A., & Chen, H. (2007). Affect Intensity Analysis of Dark Web Forums. *2007 IEEE Intelligence and Security Informatics*, 282–288. https://doi.org/10.1109/ISI.2007.379486

Adam, A. (2005). Hacking into Hacking: Gender and the Hacker Phenomenon. In A. Adam (Ed.), *Gender, Ethics and Information Technology* (pp. 128–146). Palgrave Macmillan UK. https://doi.org/10.1057/9780230000520_7

Ahmad, S., Asghar, M. Z., Alotaibi, F. M., & Awan, I. (2019). Detection and classification of social media-based extremist affiliations using sentiment analysis techniques. *Human-centric Computing and Information Sciences*, *9*(1), 24. https://doi.org/10.1186/s13673-019-0185-6

Akers, R. L., Krohn, M. D., Lanza-Kaduce, L., & Radosevich, M. (1979). Social Learning and Deviant Behavior: A Specific Test of a General Theory. *American Sociological Review*, *44*(4), 636–655. https://doi.org/10.2307/2094592

Akers, R. L., Silverman, A. L., & Silverman, A. L. (2014). Toward a Social Learning Model of Violence and Terrorism. Retrieved September 24, 2020, from http://www.taylorfrancis.com/

Akins, J., & Winfree, J., Latham. (2016). Social Learning Theory and Becoming a Terrorist: New Challenges for a General Theory. *The Handbook of the Criminology of Terrorism* (pp. 133–149).

Ali, M. Y., & Bwana, O. M. (2015). Peace-building and Conflict Prevention Training Manual and Resource Guide for Building Resilience Against Violent Extremism. https://cscrcenter.org/

Almquist, Z. W., & Bagozzi, B. E. (2019). Using Radical Environmentalist Texts to Uncover Network Structure and Network Features. *Sociological Methods & Research*, *48*(4), 905–960. https://doi.org/10.1177/0049124117729696

Aly, A., Macdonald, S., Jarvis, L., & Chen, T. M. (2017). Introduction to the Special Issue: Terrorist Online Propaganda and Radicalization. *Studies in Conflict & Terrorism*, *40*(1), 1–9. https://doi.org/10.1080/1057610X.2016.1157402

Aly, A., Weimann-Saks, D., & Weimann, G. (2014). Making 'Noise' Online:An Analysis of the Say No to Terror Online Campaign. *Perspectives on Terrorism*, *8*(5). Retrieved September 24, 2020, from http://www.terrorismanalysts.com/pt/index.php/pot/article/view/376

Anzalone, C. (2016). Continuity and Change: The Evolution and Resilience of Al-Shabab's Media Insurgency, 2006–2016. Retrieved March 11, 2022, from https://www.belfercenter.org/publication/continuity-and-change-evolution-and-resilience-al-shababs-media-insurgency-2006-2016

Archer, N. (2012). Consumer identity theft prevention and identity fraud detection behaviours. *Journal of Financial Crime*, *19*(1), 20–36. https://doi.org/10.1108/13590791211190704

Armborst, A. (2010). Modelling Terrorism and Political Violence [Publisher: SAGE Publications Ltd]. *International Relations*, *24*(4), 414–432. https://doi.org/10.1177/0047117810385779

Arun, R., Suresh, V., Veni Madhavan, C. E., & Narasimha Murthy, M. N. (2010). On Finding the Natural Number of Topics with Latent Dirichlet Allocation: Some Observations. In M. J. Zaki, J. X. Yu, B. Ravindran, & V. Pudi (Eds.), *Advances in Knowledge*

Discovery and Data Mining (pp. 391–402). Springer. https://doi.org/10.1007/978-3-642-13657-3_43

Asal, V., & Rethemeyer, R. K. (2008). The Nature of the Beast: Organizational Structures and the Lethality of Terrorist Attacks. *The Journal of Politics*, *70*(2), 437–449. https://doi.org/10.1017/S0022381608080419

Awan, I. (2017). Cyber-Extremism: Isis and the Power of Social Media. *Society*, *54*(2), 138–149. https://doi.org/10.1007/s12115-017-0114-0

Bachmann, M. (2010). The Risk Propensity and Rationality of Computer Hackers. *International Journal of Cyber Criminology*, *4*(1), 14.

Back, S., Soor, S., & LaPrade, J. (2018). Juvenile Hackers: An Empirical Test of Self-Control Theory and Social Bonding Theory. *International Journal of Cybersecurity Intelligence & Cybercrime*, *1*, 17.

Barrenas, F., Chavali, S., Holme, P., Mobini, R., & Benson, M. (2009). Network Properties of Complex Human Disease Genes Identified through Genome-Wide Association Studies. *PLOS ONE*, *4*(11), e8090. https://doi.org/10.1371/journal.pone.0008090

Bartlett, J., & Miller, C. (2012). The Edge of Violence: Towards Telling the Difference Between Violent and Non-Violent Radicalization. *Terrorism and Political Violence*, *24*(1), 1–21. https://doi.org/10.1080/09546553.2011.594923

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An Open Source Software for Exploring and Manipulating Networks. *International AAAI Conference on Weblogs and Social Media*.

Basu, A. (2014). Social Network Analysis: A Methodology for Studying Terrorism. In M. Panda, S. Dehuri, & G.-N. Wang (Eds.), *Social Networking: Mining, Visualization, and Security* (pp. 215–242). Springer International Publishing. https://doi.org/10.1007/978-3-319-05164-2_9

Baumer, E. P. S., Mimno, D., Guha, S., Quan, E., & Gay, G. K. (2017). Comparing grounded theory and topic modeling: Extreme divergence or unlikely convergence? *Journal of* *the Association for Information Science and Technology*, *68*(6), 1397–1410. https://doi.org/10.1002/asi.23786

Behr, I. V., Reding, A., Edwards, C., & Gribbon, L. (2013). Radicalisation in the digital era: The use of the internet in 15 cases of terrorism and extremism. *RAND*, 76.

Bender, E. (2019). The #BenderRule: On Naming the Languages We Study and Why It Matters. Retrieved February 17, 2022, from https://thegradient.pub/the-benderrule-on-naming-the-languages-we-study-and-why-it-matters/

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., Matsuo, A., Lowe, W., & Müller, C. (2021). Quanteda: Quantitative Analysis of Textual Data. Retrieved February 24, 2022, from https://CRAN.R-project.org/package=quanteda

Benson, D. C. (2014). Why the Internet Is Not Increasing Terrorism. *Security Studies*, *23*(2), 293–328. https://doi.org/10.1080/09636412.2014.905353

Berger, J. M. (2016). Nazis vs. ISIS on Twitter: A Comparative Study of White Nationalist and ISIS Online Social Media Networks, 32.

Berger, J. M. (2018). *Extremism*. The MIT Press.

Berger, J. M., & Strathearn, B. (2013). Measuring influence, evaluating content and countering violent extremism in online social networks. *The International Centre for the Study of Radicalisation and Political Violence*, 56.

Berger, J. (2017). Extremist Construction of Identity: How Escalating Demands for Legitimacy Shape and Define In-Group and Out-Group Dynamics. *Terrorism and Counter-Terrorism Studies*. https://doi.org/10.19165/2017.1.07

Bermingham, A., Conway, M., McInerney, L., O'Hare, N., & Smeaton, A. (2009). Combining Social Network Analysis and Sentiment Analysis to Explore the Potential for Online Radicalisation. https://doi.org/10.1109/ASONAM.2009.31

Berton, B. (2015). The dark side of the web: ISIL's one-stop shop? — European Union Institute for Security Studies. Retrieved March 1, 2022, from https://www.iss.europa.eu/content/dark-side-web-isil%E2%80%99s-one-stop-shop

Bienenstock, E. J., & Salwen, M. (2015). Covert Network Analysis: An Exchange Network Theory Perspective. In L. M. Gerdes (Ed.), *Illuminating Dark Networks: The Study of Clandestine Groups and Organizations* (pp. 8–18). Cambridge University Press. https://doi.org/10.1017/CBO9781316212639.002

Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit.* "O'Reilly Media, Inc."

Bisgin, H., Arslan, H., & Korkmaz, Y. (2019). Analyzing the Dabiq Magazine: The Language and the Propaganda Structure of ISIS. In R. Thomson, H. Bisgin, C. Dancy, & A. Hyder (Eds.), *Social, Cultural, and Behavioral Modeling* (pp. 1–11). Springer International Publishing. https://doi.org/10.1007/978-3-030-21741-9_1

Black, D. (2004). Terrorism as Social Control. In M. Deflem (Ed.), *Terrorism and Counter-Terrorism* (pp. 9–18). Emerald Group Publishing Limited. https://doi.org/10.1108/S1521-6136(2004)0000005003

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77–84. https://doi.org/10.1145/2133806.2133826

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation [Publisher: Microtome Publishing, Brookline, MA]. *Journal of Machine Learning Research (JMLR)*, *3*(4-5), 993–1022. https://doi.org/10.1162/jmlr.2003.3.4-5.993

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

Bloom, M., & Daymon, C. (2018). Assessing the Future Threat: ISIS's Virtual Caliphate. *Orbis*, *62*(3), 372–388. https://doi.org/10.1016/j.orbis.2018.05.007

Bloom, M., Tiflati, H., & Horgan, J. (2019). Navigating ISIS's Preferred Platform: Telegram1. *Terrorism and Political Violence*, *31*(6), 1242–1254. https://doi.org/10.1080/09546553.2017.1339695

Bodine-Baron, E., Helmus, T., Magnuson, M., & Winkelman, Z. (2016). *Examining ISIS Support and Opposition Networks on Twitter*. RAND Corporation. https://doi.org/10.7249/RR1328

Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2013). *Analyzing Social Networks* (1st edition). SAGE Publications Ltd.

Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2018). *Analyzing Social Networks*. SAGE.

Borgatti, S. P., & Halgin, D. S. (2014). Analyzing Affiliation Networks. *The SAGE Handbook of Social Network Analysis* (pp. 417–433). SAGE Publications Ltd. https://doi.org/10.4135/9781446294413.n28

Borum, R. (2011). Radicalization into Violent Extremism II: A Review of Conceptual Models and Empirical Research. *Journal of Strategic Security*, *4*(4). https://doi.org/⟨p⟩http://dx.doi.org/10.5038/1944-0472.4.4.2⟨/p⟩

Bossler, A., & Burruss, G. (2011). The General Theory of Crime and Computer Hacking: Low Self-control Hackers? *Corporate Hacking and Technology-driven Crime: Social Dynamics and Implications*, 38–67. https://digitalcommons.georgiasouthern.edu/crimjust-criminology-facpubs/247

Bossler, A., & Holt, T. (2009). On-line Activities, Guardianship, and Malware Infection: An Examination of Routine Activities Theory. *International Journal of Cyber Criminology*, *3*(1), 400–420. https://digitalcommons.georgiasouthern.edu/crimjust-criminology-facpubs/259

Bossler, A. M., & Berenblum, T. (2019). Introduction: New directions in cybercrime research [Publisher: Routledge]. *Journal of Crime and Justice*, *42*(5), 495–499. https://doi.org/10.1080/0735648X.2019.1692426

Bossler A.M. (2017). Need for Debate on the Implications of Honeypot Data for Restrictive Deterrence Policies in Cyberspace. *Criminology and Public Policy*, *16*(3), 681–688. https://doi.org/10.1111/1745-9133.12322

Bowman-Grieve, L. (2009). Exploring “Stormfront”: A Virtual Community of the Radical Right. *Studies in Conflict & Terrorism*, *32*(11), 989–1007. https://doi.org/10.1080/10576100903259951

Brayne, S. (2020). *Predict and Surveil: Data, Discretion, and the Future of Policing*. Oxford University Press. https://doi.org/10.1093/oso/9780190684099.001.0001

Brayne, S., & Christin, A. (2021). Technologies of Crime Prediction: The Reception of Algorithms in Policing and Criminal Courts. *Social Problems*, *68*(3), 608–624. https://doi.org/10.1093/socpro/spaa004

Bright, D. A., Hughes, C. E., & Chalmers, J. (2012). Illuminating dark networks: A social network analysis of an Australian drug trafficking syndicate. *Crime, Law and Social Change*, *57*(2), 151–176. https://doi.org/10.1007/s10611-011-9336-z

Brynielsson, J., Horndahl, A., Johansson, F., Kaati, L., Mårtenson, C., & Svenson, P. (2013). Harvesting and analysis of weak signals for detecting lone wolf terrorists. *Security Informatics*, *2*(1), 11. https://doi.org/10.1186/2190-8532-2-11

Burris, V., Smith, E., & Strahm, A. (2000). White Supremacist Networks on the Internet. *Sociological Focus*, *33*(2), 215–235. https://doi.org/10.1080/00380237.2000.10571166

Burruss, G. W., Bossler, A. M., & Holt, T. J. (2012). Assessing the Mediation of a Fuller Social Learning Model on Low Self-Control’s Influence on Software Piracy: *Crime & Delinquency*. https://doi.org/10.1177/0011128712437915

Caló, B., & Hartley, E. (2019). ISIL recruiters as social media influencers: Mechanisms of legitimation by Australian Muslim men. *Contemporary Voices: St Andrews Journal of International Relations*, *1*, 2. https://doi.org/10.15664/jtr.1497

Campedelli, G. M., Bartulovic, M., & Carley, K. M. (2019). Pairwise similarity of jihadist groups in target and weapon transitions. *Journal of Computational Social Science*, *2*(2), 245–270. https://doi.org/10.1007/s42001-019-00046-8

Cao, J., Xia, T., Li, J., Zhang, Y., & Tang, S. (2009). A density-based method for adaptive LDA model selection. *Neurocomputing*, *72*(7), 1775–1781. https://doi.org/10.1016/j.neucom.2008.06.011

Carley, K., Lee, J.-S., & Krackhardt, D. (2001). Destabilizing Networks. *Connections*, *24*, 79–92.

Carlin, J. P., & Graff, G. M. (2018). *Dawn of the Code War: America's Battle Against Russia, China, and the Rising Global Cyber Threat*. PublicAffairs.

Chan, J., & Bennett Moses, L. (2016). Is Big Data challenging criminology? *Theoretical Criminology*, *20*(1), 21–39. https://doi.org/10.1177/1362480615586614

Chang, J., & Blei, D. (2009). Relational Topic Models for Document Networks. *Proceedings of the Twelth International Conference on Artificial Intelligence and Statistics*, 81–88. Retrieved February 18, 2022, from https://proceedings.mlr.press/v5/chang09a.html

Chen, H. (2011). *Dark Web: Exploring and Data Mining the Dark Side of the Web*. Springer Science & Business Media.

Chen, H., Chung, W., Qin, J., Reid, E., Sageman, M., & Weimann, G. (2008). Uncovering the dark Web: A case study of Jihad on the Web. *Journal of the American Society for Information Science and Technology*, *59*(8), 1347–1359. https://doi.org/10.1002/asi.20838

Chen, T. M., Jarvis, L., & Macdonald, S. (Eds.). (2014). *Cyberterrorism*. Springer New York. https://doi.org/10.1007/978-1-4939-0962-9

Chen, Y., Wu, X., Hu, A., He, G., & Ju, G. (2021). Social prediction: A new research paradigm based on machine learning. *The Journal of Chinese Sociology*, *8*(1), 15. https://doi.org/10.1186/s40711-021-00152-z

Chertoff, M. (2017). A public policy perspective of the Dark Web. *Journal of Cyber Policy*, *2*(1), 26–38. https://doi.org/10.1080/23738871.2017.1298643

Choi, J., Kruis, N. E., & Choo, K.-S. (2021). Explaining Fear of Identity Theft Victimization Using a Routine Activity Approach. *Journal of Contemporary Criminal Justice*, *37*(3), 406–426. https://doi.org/10.1177/10439862211001627

Choi, K.-s., Lee, C., & Cadigan, R. (2018). Spreading Propaganda in Cyberspace: Comparing Cyber-Resource Usage of Al Qaeda and ISIS. *International Journal of Cybersecurity Intelligence & Cybercrime*, *1*(1), 21–39. https://vc.bridgew.edu/ijcic/vol1/iss1/4

Cohen, L. E., & Felson, M. (1979). Social Change and Crime Rate Trends: A Routine Activity Approach. *American Sociological Review*, *44*(4), 588–608.

Conway, M. (2003). Hackers as terrorists? Why it doesn't compute. *Computer Fraud & Security*, *2003*(12), 10–13.

Conway, M. (2014). Reality Check: Assessing the (Un)Likelihood of Cyberterrorism. In T. M. Chen, L. Jarvis, & S. Macdonald (Eds.), *Cyberterrorism: Understanding, Assessment, and Response* (pp. 103–121). Springer. Retrieved January 24, 2022, from https://doi.org/10.1007/978-1-4939-0962-9_6

Conway, M. (2017). Determining the Role of the Internet in Violent Extremism and Terrorism: Six Suggestions for Progressing Research. *Studies in Conflict & Terrorism*, *40*(1), 77–98. https://doi.org/10.1080/1057610X.2016.1157408

Crenshaw, M. (2011). *Explaining terrorism: Causes, processes, and consequences* [OCLC: 457164350].

Csardi, G., & Nepusz, T. (2022). Igraph: Network Analysis and Visualization. Retrieved March 4, 2022, from https://CRAN.R-project.org/package=igraph

Danilak, M. M. (n.d.). Langdetect: Language detection library ported from Google's language-detection. Retrieved February 24, 2022, from https://github.com/Mimino666/langdetect

Deflem, M. (Ed.). (2004). *Terrorism and counter-terrorism: Criminological perspectives*. Elsevier.

Denning, D. E. (2000). Cyberterrorism: The logic bomb versus the truck bomb. *Global Dialogue*, *2*(4), 29.

Denning, D. E. (2001). Activism, hacktivism, and cyberterrorism: The Internet as a tool for influencing foreign policy. *Networks and netwars: The future of terror, crime, and militancy*, *239*, 288.

Deveaud, R., SanJuan, E., & Bellot, P. (2014). Accurate and effective latent concept modeling for ad hoc information retrieval. *Document numérique*, *17*(1), 61–84. https://doi.org/10.3166/dn.17.1.61-84

Diamond, B., & Bachmann, M. (2015). Out Of The Beta Phase: Obstacles, Challenges, And Promising Paths In The Study Of Cyber Criminology. https://doi.org/10.5281/ZENODO.22196

DiMaggio, P., Nag, M., & Blei, D. (2013). Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding. *Poetics*, *41*(6), 570–606. https://doi.org/10.1016/j.poetic.2013.08.004

Dion, M. (2010). Advance Fee Fraud Letters as Machiavellian/Narcissistic Narratives. *4*(1), 13.

Duijn, P. A. C., Kashirin, V., & Sloot, P. M. A. (2014). The Relative Ineffectiveness of Criminal Network Disruption. *Scientific Reports*, *4*(1), 4238. https://doi.org/10.1038/srep04238

Edwards, C., & Gribbon, L. (2013). Pathways to Violent Extremism in the Digital Era. *The RUSI Journal*, *158*(5), 40–47. https://doi.org/10.1080/03071847.2013.847714

Erez, E., Weimann, G., & Weisburd, A. A. (2011). Jihad, Crime, and the Internet: Content Analysis of Jihadist Forum Discussions, 179.

Europol. (2021). Internet Organised Crime Threat Assessment (IOCTA) 2021. Retrieved February 7, 2022, from https://www.europol.europa.eu/publications-events/main-reports/internet-organised-crime-threat-assessment-iocta-2021

Evan, T., Leverett, E., Ruffle, S., Coburn, A. W., Bourdeau, J., Gunaratna, R., & Ralph, D. (2017). Cyber Terrorism: Assessment of the Threat to Insurance. Retrieved September 24, 2020, from https://www.researchgate.net/publication/328530893_Cyber_Terrorism_Assessment_of_the_Threat_to_Insurance

Everton, S. S. (2008). Tracking, Destabilizing and Disrupting Dark Networks with Social Networks Analysis. Retrieved March 2, 2022, from https://calhoun.nps.edu/handle/10945/34415

FBI. (2021). Strategic Intelligence Assessment and Data on Domestic Terrorism. Retrieved February 9, 2022, from https://www.fbi.gov/file-repository/fbi-dhs-domestic-terrorism-strategic-report.pdf/view

Ferguson, A. G. (2017). *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*. NYU Press.

Finklea, K. (2017). Dark Web. *Congressional Research Service*, 19.

Fortunato, S., & Hric, D. (2016). Community detection in networks: A user guide. *Physics Reports*, *659*, 1–44. https://doi.org/10.1016/j.physrep.2016.09.002

Forum, W. E. (2020). The Global Risks Report 2020. Retrieved February 11, 2022, from https://www.weforum.org/reports/the-global-risks-report-2020/

Foster, J. G., Foster, D. V., Grassberger, P., & Paczuski, M. (2010). Edge direction and the structure of networks. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(24), 10815–10820. https://doi.org/10.1073/pnas.0912671107

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, *1*(3), 215–239. https://doi.org/10.1016/0378-8733(78)90021-7

Freilich, J. D., Chermak, S. M., Belli, R., Gruenewald, J., & Parkin, W. S. (2014). Introducing the United States Extremis Crime Database (ECDB). *Terrorism and Political Violence*, *26*(2), 372–384. https://doi.org/10.1080/09546553.2012.713229

Furnell, S. M., & Warren, M. J. (1999). Computer hacking and cyber terrorism: The real threats in the new millennium? *Computers & Security*, *18*(1), 28–34. https://doi.org/10.1016/S0167-4048(99)80006-6

Furnell, S. (2003). Cybercrime: Vandalizing the Information Society. In J. M. C. Lovelle, B. M. G. Rodríguez, J. E. L. Gayo, M. del Puerto Paule Ruiz, & L. J. Aguilar (Eds.), *Web Engineering* (pp. 8–16). Springer. https://doi.org/10.1007/3-540-45068-8_2

Furnell, S., Emm, D., & Papadaki, M. (2015). The challenge of measuring cyber-dependent crimes. *Computer Fraud & Security*, *2015*(10), 5–12. https://doi.org/10.1016/S1361-3723(15)30093-2

Gartenstein-Ross, D., & Barr, N. (2016). Fixing How We Fight the Islamic State's Narrative. Retrieved March 11, 2022, from https://warontherocks.com/2016/01/fixing-how-we-fight-the-islamic-states-narrative/

Gaudette, T., Scrivens, R., & Venkatesh, V. (2020). The Role of the Internet in Facilitating Violent Extremism: Insights from Former Right-Wing Extremists. *Terrorism and Political Violence*, *0*(0), 1–18. https://doi.org/10.1080/09546553.2020.1784147

Gies, S. V., Piquero, N. L., Piquero, A. R., Green, B., & Bobnis, A. (2021). Wild, Wild Theft: Identity Crimes in the Digital Frontier. *Criminal Justice Policy Review*, *32*(6), 592–617. https://doi.org/10.1177/0887403420949650

Gill, P., Corner, E., Conway, M., Thornton, A., Bloom, M., & Horgan, J. (2017). Terrorist Use of the Internet by the Numbers. *Criminology & Public Policy*, *16*(1), 99–117. https://doi.org/10.1111/1745-9133.12249

Goffman, E. (1959). *The Presentation of Self in Everyday Life*. Anchor.

Gottfredson, M. R., & Hirschi, T. (1990). *A general theory of crime*. Stanford University Press.

Grabosky, P. (2016). The evolution of cybercrime, 2006–2016. *Cybercrime Through an Interdisciplinary Lens*. Routledge.

Granovetter, M. S. (1973). The Strength of Weak Ties. *American Journal of Sociology*, *78*(6), 1360–1380. Retrieved March 3, 2022, from https://www.jstor.org/stable/2776392

Greenberg, K. J. (2016). Counter-Radicalization via the Internet. *The ANNALS of the American Academy of Political and Social Science*, *668*(1), 165–179. https://doi.org/10.1177/0002716216672635

Greene, K. T., & Lucas, C. (2020). Once more, with feeling: Using sentiment analysis to improve models of relationships between non-state actors. *International Interactions*, *46*(1), 150–162. https://doi.org/10.1080/03050629.2019.1684913

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*, *101*(Supplement 1), 5228–5235. https://doi.org/10.1073/pnas.0307752101

Gruen, B., & Hornik, K. (2011). Topicmodels: An R Package for Fitting Topic Models. *Journal of Statistical Software*, *40*(13), 1–30. https://doi.org/10.18637/jss.v040.i13

Gupta, A., Maynard, S. B., & Ahmad, A. (2021). The Dark Web Phenomenon: A Review and Research Agenda. *arXiv:2104.07138 [cs]*. Retrieved March 2, 2022, from http://arxiv.org/abs/2104.07138

Hamm, M. S. (2007). *Terrorism As Crime: From Oklahoma City to Al-Qaeda and Beyond*. NYU Press.

Hamm, M. S., & Van de Voorde, C. (2005). Crimes committed by terrorist groups: Theory, research, and prevention. *Trends in Organized Crime*, *9*(2), 18–50. https://doi.org/10.1007/s12117-005-1023-y

Hardy, K., & Williams, G. (2014). What is 'Cyberterrorism'? Computer and Internet Technology in Legal Definitions of Terrorism. In T. M. Chen, L. Jarvis, & S. Macdonald (Eds.), *Cyberterrorism: Understanding, Assessment, and Response* (pp. 1–23). Springer. https://doi.org/10.1007/978-1-4939-0962-9_1

Hay, C., Meldrum, R., & Mann, K. (2010). Traditional Bullying, Cyber Bullying, and Deviance: A General Strain Theory Approach. *Journal of Contemporary Criminal Justice*, *26*, 130–147. https://doi.org/10.1177/1043986209359557

Hayward, K. J., & Maas, M. M. (2021). Artificial intelligence and crime: A primer for criminologists. *Crime, Media, Culture*, *17*(2), 209–233. https://doi.org/10.1177/1741659020917434

Hernandez-Suarez, A., Sanchez-Perez, G., Toscano-Medina, K., Martinez-Hernandez, V., Perez-Meana, H., Olivares-Mercado, J., & Sanchez, V. (2018). Social Sentiment Sensor in Twitter for Predicting Cyber-Attacks Using 1 Regularization. *Sensors*, *18*(5), 1380. https://doi.org/10.3390/s18051380

Higgins, G. E., Wolfe, S. E., & Marcum, C. D. (2008). Music Piracy and Neutralization: A Preliminary Trajectory Analysis from Short-Term Longitudinal Data. *2*(2), 13.

Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing [Publisher: American Association for the Advancement of Science]. *Science*, *349*(6245), 261–266. https://doi.org/10.1126/science.aaa8685

Hirschi, T. (1969). *Causes of Delinquency*. University of California Press.

Hodwitz, O. (2019). Recognizing & Resolving Issues in Terrorism Research, Data Collection, & Analysis. Retrieved March 13, 2022, from https://www.resolvenet.org/research/recognizing-resolving-issues-terrorism-research-data-collection-analysis

Hoffman, B. (2006a). *Inside Terrorism*. Columbia University Press.

Hoffman, B. (2006b). The Use of the Internet by Islamic Extremists. Retrieved September 24, 2020, from https://www.rand.org/pubs/testimonies/CT262-1.html

Hofman, J. M., Watts, D. J., Athey, S., Garip, F., Griffiths, T. L., Kleinberg, J., Margetts, H., Mullainathan, S., Salganik, M. J., Vazire, S., Vespignani, A., & Yarkoni, T. (2021). Integrating explanation and prediction in computational social science. *Nature*, *595*(7866), 181–188. https://doi.org/10.1038/s41586-021-03659-0

Holt, T. J. (2007). Subcultural evolution? examining the influence of on- and off-line experiences on deviant subcultures. *Deviant Behavior*, *28*(2), 171–198. https://doi.org/10.1080/01639620601131065

Holt, T. J. (2010). Examining the Role of Technology in the Formation of Deviant Subcultures. *Social Science Computer Review*, *28*(4), 466–481. https://doi.org/10.1177/0894439309351344

Holt, T. J. (2012). Exploring the Intersections of Technology, Crime, and Terror. *Terrorism and Political Violence*, *24*(2), 337–354. https://doi.org/10.1080/09546553.2011.648350

Holt, T. J. (2013). *Cybercrime and criminological theory: Fundamental readings on hacking, piracy, theft, and harassment.*

Holt, T. J. (2016). *Cybercrime Through an Interdisciplinary Lens*. Taylor & Francis Group.

Holt, T. J., & Bossler, A. M. (2009). Examining the applicability of lifestyle-routine activities theory for cybercrime victimization. *Deviant Behavior*, *30*(1), 1–25. https://doi.org/10.1080/01639620701876577

Holt, T. J., & Bossler, A. M. (2014). An Assessment of the Current State of Cybercrime Scholarship. *Deviant Behavior*, *35*(1), 20–40. Retrieved February 4, 2022, from https://heinonline.org/HOL/P?h=hein.journals/devbh35&i=21

Holt, T. J., Cale, J., Brewer, R., & Goldsmith, A. (2021). Assessing the Role of Opportunity and Low Self-Control in Juvenile Hacking. *Crime & Delinquency*, *67*(5), 662–688. https://doi.org/10.1177/0011128720978730

Holt, T. J., Freilich, J. D., & Chermak, S. M. (2017a). Exploring the Subculture of Ideologically Motivated Cyber-Attackers: *Journal of Contemporary Criminal Justice*. https://doi.org/10.1177/1043986217699100

Holt, T. J., Freilich, J. D., & Chermak, S. M. (2017b). Internet-Based Radicalization as Enculturation to Violent Deviant Subcultures. *Deviant Behavior*, *38*(8), 855–869. https://doi.org/10.1080/01639625.2016.1197704

Holt, T. J., Freilich, J. D., Chermak, S. M., Mills, C., & Silva, J. (2019). Loners, Colleagues, or Peers? Assessing the Social Organization of Radicalization. *American Journal of Criminal Justice*, *44*(1), 83–105. https://doi.org/10.1007/s12103-018-9439-5

Holt, T. J., & Graves, D. C. (2007). A Qualitative Analysis of Advance Fee Fraud E-mail Schemes. *1*(1), 18.

Holt, T. J., Lee, J. R., Freilich, J. D., Chermak, S. M., Bauer, J. M., Shillair, R., & Ross, A. (2020). An Exploratory Analysis of the Characteristics of Ideologically Motivated Cyberattacks. *Terrorism and Political Violence*, 1–16. https://doi.org/10.1080/09546553.2020.1777987

Holt, T. J., Navarro, J. N., & Clevenger, S. (2020). Exploring the Moderating Role of Gender in Juvenile Hacking Behaviors. *Crime & Delinquency*, *66*(11), 1533–1555. https://doi.org/10.1177/0011128719875697

Holt, T. J., & Steinmetz, K. F. (2020). Examining the Role of Power-Control Theory and Self-Control to Account for Computer Hacking. *Crime & Delinquency*, 0011128720981892. https://doi.org/10.1177/0011128720981892

Holt, T. J., Strumsky, D., Smirnova, O., & Kilger, M. (2012). Examining the Social Networks of Malware Writers and Hackers. *6*(1), 13.

Holt, T. J., Turner, N. D., Freilich, J. D., & Chermak, S. M. (2021). Examining the Characteristics That Differentiate Jihadi-Associated Cyberattacks Using Routine Activities Theory. *Social Science Computer Review*, 08944393211023324. https://doi.org/10.1177/08944393211023324

Howell, C., Burruss, G., Maimon, D., & Sahani, S. (2019). Website Defacement and Routine Activities: Considering the Importance of Hackers' Valuations of Potential Targets. *EBCS Articles*. https://scholarworks.gsu.edu/ebcs_articles/10

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews, 168–177. https://doi.org/10.1145/1014052.1014073

Hussain, G., & Saltman, E. (2016). Jihad Trending: A Comprehensive Analysis of Online Extremism and How to Counter it. Retrieved March 11, 2022, from https://preventviolentextremism.info/jihad-trending-comprehensive-analysis-online-extremism-and-how-counter-it

Hutchings, A., & Chua, Y. T. (2016). Gendering cybercrime. *Cybercrime Through an Interdisciplinary Lens*. Routledge.

Hutto, C., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*(1), 216–225. Retrieved April 14, 2022, from https://ojs.aaai.org/index.php/ICWSM/article/view/14550

Hvitfeldt, E., & Silge, J. (2021). *Supervised Machine Learning for Text Analysis in R*. Chapman; Hall/CRC. https://doi.org/10.1201/9781003093459

IC3. (2020). Internet Crime Report 2020. https://www.ic3.gov/

ICE. (2020). Global disruption of 3 terror finance cyber-enabled campaigns. Retrieved February 10, 2022, from https://www.ice.gov/news/releases/global-disruption-3-terror-finance-cyber-enabled-campaigns

Jaishankar, K. (2007). Establishing a Theory of Cyber Crimes. *International Journal of Cyber Criminology*, *1*(2), 3.

Jang, H., Song, J., & Kim, R. (2014). Does the offline bully-victimization influence cyberbullying behavior among youths? Application of General Strain Theory. *Computers in Human Behavior*, *31*, 85–93. https://doi.org/10.1016/j.chb.2013.10.007

Jarvis, L., & Macdonald, S. (2014). Locating Cyberterrorism: How Terrorism Researchers Use and View the Cyber Lexicon. *Perspectives on Terrorism*, *8*(2), 52–65. Retrieved February 4, 2022, from https://www.jstor.org/stable/26297136

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Jordan, T., & Taylor, P. (1998). A Sociology of Hackers. *The Sociological Review*, *46*(4), 757–780. https://doi.org/10.1111/1467-954X.00139

Kemp, S. (2019). Digital 2019: Global Digital Overview. Retrieved February 11, 2022, from https://datareportal.com/reports/digital-2019-global-digital-overview

Kemp, S. (2022). Digital 2022: Global Overview Report. Retrieved February 23, 2022, from https://datareportal.com/reports/digital-2022-global-overview-report

Kerstens, J., & Veenstra, S. (2016). Cyber Bullying In The Netherlands: A Criminological Perspective. https://doi.org/10.5281/ZENODO.55055

Keuschnigg, M., Lovsjö, N., & Hedström, P. (2018). Analytical sociology and computational social science. *Journal of Computational Social Science*, *1*(1), 3–14. https://doi.org/10.1007/s42001-017-0006-5

Kigerl, A. (2011). Routine Activity Theory and the Determinants of High Cybercrime Countries. *Social Science Computer Review*, 0894439311422689. https://doi.org/10.1177/0894439311422689

King, T. C., Aggarwal, N., Taddeo, M., & Floridi, L. (2020). Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions. *Science and Engineering Ethics*, *26*(1), 89–120. https://doi.org/10.1007/s11948-018-00081-0

Kleinberg, J. (2001). Small-world phenomena and the dynamics of information. *Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic*, 431–438.

Koehler, D. (2014). The Radical Online: Individual Radicalization Processes and the Role of the Internet. *Journal for Deradicalization*, (1), 116–134. Retrieved January 24, 2022, from https://journals.sfu.ca/jd/index.php/jd/article/view/8

Kopp, C., Layton, R., Sillitoe, J., & Gondal, I. (2016). The Role Of Love Stories In Romance Scams: A Qualitative Analysis Of Fraudulent Profiles. https://doi.org/10.5281/ZENODO.56227

Koschade, S. (2006). A Social Network Analysis of Jemaah Islamiyah: The Applications to Counterterrorism and Intelligence. *Studies in Conflict & Terrorism*, *29*(6), 559–575. https://doi.org/10.1080/10576100600798418

Krebs, V. E. (2002). Mapping Networks of Terrorist Cells. *Connections*, *24*(3), 10.

LaFree, G., & Ackerman, G. (2009). The Empirical Study of Terrorism: Social and Legal Research. *Annual Review of Law and Social Science*, *5*(1), 347–374. https://doi.org/10.1146/annurev.lawsocsci.093008.131517

LaFree, G., & Dugan, L. (2009). Research on Terrorism and Countering Terrorism. *Crime and Justice*, *38*(1), 413–477. https://doi.org/10.1086/599201

LaFree, G., & Dugan, L. (2015). How Has Criminology Contributed to the Study of Terrorism since 9/11? *Terrorism and Counterterrorism Today* (pp. 1–23). Emerald Group Publishing Limited. Retrieved September 24, 2020, from https://doi.org/10.1108/S1521-613620150000020002

Laqueur, W. (2004). *No End to War: Terrorism in the Twenty-First Century*. A&C Black.

Lazer, D., Pentland, A., Watts, D., Aral, S., Athey, S., Contractor, N., Freelon, D., Gonzalez-Bailon, S., King, G., Margetts, H., Nelson, A., Salganik, M., Strohmaier, M., Vespignani, A., & Wagner, C. (2020). Computational social science: Obstacles and opportunities. *Science (New York, N.Y.)*, *369*, 1060–1062. https://doi.org/10.1126/science.aaz8170

Ledesma, J. (2021). What is RaaS and why is it so dangerous? Retrieved February 7, 2022, from https://businessinsights.bitdefender.com/what-is-raas-and-why-is-it-so-dangerous

Lee, J. R., & Holt, T. J. (2020). Assessing the Factors Associated With the Detection of Juvenile Hacking Behaviors. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.00840

Leukfeldt, E. R., & Yar, M. (2016). Applying Routine Activity Theory to Cybercrime: A Theoretical and Empirical Analysis. *Deviant Behavior*, *37*(3), 263–280. https://doi.org/10.1080/01639625.2015.1012409

Li, C. K. W., Holt, T. J., Bossler, A. M., & May, D. C. (2016). Examining the Mediating Effects of Social Learning on the Low Self-Control—Cyberbullying Relationship in a Youth Sample. *Deviant Behavior*, *37*(2), 126–138. https://doi.org/10.1080/01639625.2014.1004023

Ligon, G. S., Logan, M., Hall, M., Derrick, D., Fuller, J., & Church, S. (2017). The Jihadi Industry: Assessing the Organizational, Leadership, and Cyber Profiles. *START*.

Liu, B. (2012). Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies*, *5*(1), 1–167. https://doi.org/10.2200/S00416ED1V01Y201204HLT016

Liu, L., Tang, L., Dong, W., Yao, S., & Zhou, W. (2016). An overview of topic modeling and its current applications in bioinformatics. *SpringerPlus*, *5*(1), 1608. https://doi.org/10.1186/s40064-016-3252-8

Louderback, E. R., & Antonaccio, O. (2020). New Applications of Self-Control Theory to Computer-Focused Cyber Deviance and Victimization: A Comparison of Cognitive and Behavioral Measures of Self-Control and Test of Peer Cyber Deviance and Gender as Moderators. *Crime & Delinquency*, 0011128720906116. https://doi.org/10.1177/0011128720906116

Lucas, C., Nielsen, R. A., Roberts, M. E., Stewart, B. M., Storer, A., & Tingley, D. (2015). Computer-Assisted Text Analysis for Comparative Politics. *Political Analysis*, *23*(2), 254–277. https://doi.org/10.1093/pan/mpu019

Maalem Lahcen, R. A., Caulkins, B., Mohapatra, R., & Kumar, M. (2020). Review and insight on the behavioral aspects of cybersecurity. *Cybersecurity*, *3*(1), 10. https://doi.org/10.1186/s42400-020-00050-w

Macdonald, S., Correia, S. G., & Watkin, A.-L. (2019). Regulating terrorist content on social media: Automation and the rule of law. *International Journal of Law in Context*, *15*(2), 183–197. https://doi.org/10.1017/S1744552319000119

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., Pfetsch, B., Heyer, G., Reber, U., Häussler, T., Schmid-Petri, H., & Adam, S. (2018). Applying LDA Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology. *Communication Methods and Measures*, *12*(2-3), 93–118. https://doi.org/10.1080/19312458.2018.1430754

Maimon, D., & Louderback, E. R. (2019). Cyber-Dependent Crimes: An Interdisciplinary Review. *Annual Review of Criminology*, *2*(1), 191–216. https://doi.org/10.1146/annurev-criminol-032317-092057

Maimon, D., Sobesto, B., & Cukier, M. (2013). Restrictive Deterrent Effects of a Warning Banner in an Attacked Computer System. *EBCS Articles*. https://scholarworks.gsu.edu/ebcs_articles/7

Malik, N. (2018). Terror in The Dark: How Terrorists use Encryption, the Darknet and Cryptocurrencies. Retrieved February 17, 2022, from https://henryjacksonsociety.org/publications/terror-in-the-dark-how-terrorists-use-encryption-the-darknet-and-cryptocurrencies/

Manning, C., & Schutze, H. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press.

Marcum, C. D., & Higgins, G. E. (2012). Battle of the sexes: An examination of male and female cyber bullying. *6*(1), 8.

Marcum, C. D., Higgins, G. E., Ricketts, M. L., & Wolfe, S. E. (2014). Hacking in High School: Cybercrime Perpetration by Juveniles. *Deviant Behavior*, *35*(7), 581–591. https://doi.org/10.1080/01639625.2013.867721

Marwick, A. E., & boyd danah, d. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, *13*(1), 114–133. https://doi.org/10.1177/1461444810365313

McCauley, C., & Moskalenko, S. (2008). Mechanisms of Political Radicalization: Pathways Toward Terrorism. *Terrorism and Political Violence*, *20*(3), 415–433. https://doi.org/10.1080/09546550802073367

Milla, M. N., Hudiyana, J., Cahyono, W., & Muluk, H. (2020). Is the Role of Ideologists Central in Terrorist Networks? A Social Network Analysis of Indonesian Terrorist Groups. *Frontiers in Psychology*, *11*. Retrieved February 28, 2022, from https://www.frontiersin.org/article/10.3389/fpsyg.2020.00333

Miller, B., & Morris, R. G. (2016). Virtual Peer Effects in Social Learning Theory. *Crime & Delinquency*, *62*(12), 1543–1569. https://doi.org/10.1177/0011128714526499

Mirea, M., Wang, V., & Jung, J. (2019). The not so dark side of the darknet: A qualitative study. *Security Journal*, *32*(2), 102–118. https://doi.org/10.1057/s41284-018-0150-5

Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a Word-Emotion Association Lexicon. *arXiv:1308.6297*. Retrieved December 2, 2018, from http://arxiv.org/abs/1308.6297

Molina, M., & Garip, F. (2019). Machine Learning for Sociology. *Annual Review of Sociology*, *45*(1), 27–45. https://doi.org/10.1146/annurev-soc-073117-041106

Monk, B., Mitchell, J., Frank, R., & Davies, G. (2018). Uncovering Tor: An Examination of the Network Structure. *Security and Communication Networks*, *2018*, e4231326. https://doi.org/10.1155/2018/4231326

Morgan, S. (2021). Cybercrime To Cost The World $10.5 Trillion Annually By 2025. Retrieved February 11, 2022, from https://cybersecurityventures.com/cyberwarfare-report-intrusion/

Morris, R. G. (2011). Computer Hacking and the Techniques of Neutralization: An Empirical Assessment. Retrieved February 5, 2022, from https://www.igi-global.com/

chapter/computer-hacking-techniques-neutralization/www.igi-global.com/chapter/computer-hacking-techniques-neutralization/46417

Morris, R. G., & Higgins, G. E. (2010). Criminological theory in the digital age: The case of social learning theory and digital piracy. *Journal of Criminal Justice*, *38*(4), 470–480. https://doi.org/10.1016/j.jcrimjus.2010.04.016

Mullins, S. (2013). Social network analysis and terrorism: An introduction to the special issue. *Behavioral Sciences of Terrorism and Political Aggression*, *5*(2), 67–69. https://doi.org/10.1080/19434472.2012.731697

Németh, R., & Koltai, J. (2021). The Potential of Automated Text Analytics in Social Knowledge Building. In T. Rudas & G. Péli (Eds.), *Pathways Between Social Science and Computational Social Science: Theories, Methods, and Interpretations* (pp. 49–70). Springer International Publishing. https://doi.org/10.1007/978-3-030-54936-7_3

Newman, M. (2006). Modularity and community structure in networks. Retrieved March 3, 2022, from https://www.pnas.org/doi/abs/10.1073/pnas.0601602103

Nielsen, F. Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *arXiv:1103.2903 [cs]*. Retrieved April 14, 2022, from http://arxiv.org/abs/1103.2903

Nikita, M., & Chaney, N. (2020). Ldatuning: Tuning of the Latent Dirichlet Allocation Models Parameters. Retrieved February 24, 2022, from https://CRAN.R-project.org/package=ldatuning

Nodeland, B., Belshaw, S., & Saber, M. (2018). Teaching Cybersecurity to Criminal Justice Majors. *Journal of Criminal Justice Education*, *30*, 1–20. https://doi.org/10.1080/10511253.2018.1439513

Papadopoulos, S., Kompatsiaris, Y., Vakali, A., & Spyridonos, P. (2012). Community detection in Social Media. *Data Mining and Knowledge Discovery*, *24*(3), 515–554. https://doi.org/10.1007/s10618-011-0224-z

Payne, B. K., & Hadzhidimova, L. (2020). Disciplinary and Interdisciplinary Trends in Cybercrime Research: An Examination. https://doi.org/10.5281/ZENODO.3741131

Pelletier, I. R., Lundmark, L., Gardner, R., Ligon, G. S., & Kilinc, R. (2016). Why ISIS's Message Resonates: Leveraging Islam, Sociopolitical Catalysts, and Adaptive Messaging. *Studies in Conflict & Terrorism*, *39*(10), 871–899. https://doi.org/10.1080/1057610X.2016.1139373

Perliger, A., & Pedahzur, A. (2011). Social Network Analysis in the Study of Terrorism and Political Violence. *PS: Political Science and Politics*, *44*(1), 45–50. Retrieved February 25, 2021, from https://www.jstor.org/stable/40984482

Pete, I., Hughes, J., Chua, Y. T., & Bada, M. (2020). A Social Network Analysis and Comparison of Six Dark Web Forums. *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS PW)*, 484–493. https://doi.org/10.1109/EuroSPW51379.2020.00071

Phillips, E., Nurse, J., Goldsmith, M., & Creese, S. (2015). Extracting Social Structure from DarkWeb Forums.

Piazza, J. A., & Guler, A. (2019). The Online Caliphate: Internet Usage and ISIS Support in the Arab World. *Terrorism and Political Violence*, *0*(0), 0–20. https://doi.org/10.1080/09546553.2019.1606801

Pollitt, M. M. (1998). Cyberterrorism — fact or fancy? *Computer Fraud & Security*, *1998*(2), 8–10. https://doi.org/10.1016/S1361-3723(00)87009-8

Pons, P., & Latapy, M. (2005). Computing communities in large networks using random walks (long version). *arXiv:physics/0512106*. Retrieved February 25, 2022, from http://arxiv.org/abs/physics/0512106

Pons, P., & Latapy, M. (2006). Computing Communities in Large Networks Using Random Walks. *J. Graph Algorithms Appl.*, *10*, 191–218. https://doi.org/10.7155/jgaa.00124

Powell, A., Stratton, G., & Cameron, R. (2018). *Digital Criminology: Crime and Justice in Digital Society*. Routledge. https://doi.org/10.4324/9781315205786

Pozzi, F., Fersini, E., Messina, E., & Liu, B. (2016). *Sentiment analysis in social networks*. Morgan Kaufmann.

Puschmann, C., & Powell, A. (2018). Turning Words Into Consumer Preferences: How Sentiment Analysis Is Framed in Research and the News Media. *Social Media + Society*, *4*(3), 2056305118797724. https://doi.org/10.1177/2056305118797724

Queiroz, G. D., Fay, C., Hvitfeldt, E., Keyes, O., Misra, K., Mastny, T., Erickson, J., Robinson, D., & Silge, J. (2021). Tidytext: Text Mining using 'dplyr', 'ggplot2', and Other Tidy Tools. Retrieved February 24, 2022, from https://CRAN.R-project.org/package=tidytext

Raab, J., & Milward, H. B. (2003). Dark Networks as Problems. *Journal of Public Administration Research and Theory: J-PART*, *13*(4), 413–439. Retrieved March 2, 2022, from https://www.jstor.org/stable/3525656

Reyns, B. W., Henson, B., & Fisher, B. S. (2011). Being Pursued Online: Applying Cyberlifestyle–Routine Activities Theory to Cyberstalking Victimization. *Criminal Justice and Behavior*, *38*(11), 1149–1169. https://doi.org/10.1177/0093854811421448

Roberts, M. E., Tingley, D., Stewart, B. M., & Airoldi, E. M. (2013). The Structural Topic Model and Applied Social Science. *Advances in Neural Information Processing Systems Workshop on Topic Models: Computation, Application and Evaluation.*, 4.

Romaniuk, P. (2015). Does CVE Work? Lessons Learned From the Global Effort to Counter Violent Extremism. Retrieved March 11, 2022, from https://www.globalcenter.org/publications/does-cve-work-lessons-learned-from-the-global-effort-to-counter-violent-extremism/

Rosenfeld, R. (2002). Why Criminologists Should Study Terrorism. *The Criminologist*, *27*(6), 1–4.

Rosen-Zvi, M., Griffiths, T., Steyvers, M., & Smyth, P. (2004). The author-topic model for authors and documents. *Proceedings of the 20th conference on Uncertainty in artificial intelligence*, 487–494.

Rowe, M., & Saif, H. (2016). Mining Pro-ISIS Radicalisation Signals from Social Media Users. *Proceedings of the Tenth International AAAI Conference on Web and Social Media (ICWSM 2016)*, 329–338. Retrieved November 5, 2020, from http://www.aaai.org/ocs/index.php/ICWSM/ICWSM16/paper/view/13023/12752

Roy, A., & Rambo-Hernandez, K. E. (2021). There's So Much to Do and Not Enough Time to Do It! A Case for Sentiment Analysis to Derive Meaning From Open Text Using Student Reflections of Engineering Activities. *American Journal of Evaluation*, *42*(4), 559–576. https://doi.org/10.1177/1098214020962576

Rudner, M. (2008). Misuse of Passports: Identity Fraud, the Propensity to Travel, and International Terrorism. *Studies in Conflict & Terrorism*, *31*(2), 95–110. https://doi.org/10.1080/10576100701812803

Rudner, M. (2017). "Electronic Jihad": The Internet as Al Qaeda's Catalyst for Global Terror. *Studies in Conflict & Terrorism*, *40*(1), 10–23. https://doi.org/10.1080/1057610X.2016.1157403

Sageman, M. (2004). *Understanding Terror Networks* (First Printing edition). University of Pennsylvania Press.

Sageman, M. (2011). *Leaderless Jihad: Terror Networks in the Twenty-First Century*. University of Pennsylvania Press.

Salganik, M. J. (2019). *Bit by Bit: Social Research in the Digital Age*. Princeton University Press.

Saltman, E., Kooti, F., & Vockery, K. (2021). New Models for Deploying Counterspeech: Measuring Behavioral Change and Sentiment Analysis. *Studies in Conflict & Terrorism*, *0*(0), 1–24. https://doi.org/10.1080/1057610X.2021.1888404

Sap, M., Card, D., Gabriel, S., Choi, Y., & Smith, N. A. (2019). The Risk of Racial Bias in Hate Speech Detection. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 1668–1678. https://doi.org/10.18653/v1/P19-1163

Scanlon, J., & Gerber, M. (2014). Automatic detection of cyber-recruitment by violent extremists. *Security Informatics*, *3*. https://doi.org/10.1186/s13388-014-0005-5

Scanlon, J. R., & Gerber, M. S. (2015). Forecasting Violent Extremist Cyber Recruitment. *IEEE Transactions on Information Forensics and Security*, *10*(11), 2461–2470. https://doi.org/10.1109/TIFS.2015.2464775

Schell, B. H., & Dodge, J. L. (2002). *The hacking of America: Who's doing it, why, and how*. Greenwood Publishing Group Inc.

Schmid, A. P. (2004). Frameworks for Conceptualising Terrorism. *Terrorism and Political Violence*, *16*(2), 197–221. https://doi.org/10.1080/09546550490483134

Schmid, A. P. (2014). Al-Qaeda's "Single Narrative" and Attempts to Develop Counter-Narratives: The State of Knowledge. *International Centre for Counter-Terrorism*, 40. http://www.jstor.org/stable/resrep29395

Schofield, A., & Mimno, D. (2016). Comparing Apples to Apple: The Effects of Stemmers on Topic Models. *Transactions of the Association for Computational Linguistics*, *4*, 287–300. https://doi.org/10.1162/tacl_a_00099

Schuurman, B. (2019). Topics in terrorism research: Reviewing trends and gaps, 2007-2016. *Critical Studies on Terrorism*, *12*(3), 463–480. https://doi.org/10.1080/17539153.2019.1579777

Schweinberger, M. (2022). Sentiment Analysis in R. Retrieved April 19, 2022, from https://slcladal.github.io/sentiment.html

Scott, J., & Spaniel, D. (2016). *The Anatomy of Cyber-Jihad: Cyberspace is the New Great Equalizer*. CreateSpace Independent Publishing Platform.

Scott, J. (2000). *Social Network Analysis: A Handbook*. SAGE.

Scott, J., & Carrington, P. (2014). *The SAGE Handbook of Social Network Analysis*. SAGE Publications Ltd. https://doi.org/10.4135/9781446294413

Scrivens, R., Chermak, S. M., Freilich, J. D., Wojciechowski, T. W., & Frank, a. R. (2021). Detecting Extremists Online: Examining Online Posting Behaviors of Violent and

Non-Violent Right-Wing Extremists. Retrieved September 17, 2021, from https://www.resolvenet.org/research/detecting-extremists-online-examining-online-posting-behaviors-violent-and-non-violent

Scrivens, R., Davies, G., & Frank, R. (2018). Searching for signs of extremism on the web: An introduction to Sentiment-based Identification of Radical Authors. *Behavioral Sciences of Terrorism and Political Aggression*, *10*(1), 39–59. https://doi.org/10.1080/19434472.2016.1276612

Shehabat, A., & Mitew, T. (2018). Black-boxing the Black Flag: Anonymous Sharing Platforms and ISIS Content Distribution Tactics. *Perspectives on Terrorism*, *12*(1), 81–99. Retrieved September 21, 2020, from http://www.jstor.org/stable/26343748

Silge, J., & Robinson, D. (2017). *Text Mining with R: A Tidy Approach* (1st edition). O'Reilly Media.

Silke, A. (Ed.). (2004). *Research on Terrorism*. Routledge.

Singer, P. W., & Friedman, A. (2013). *Cybersecurity and Cyberwar: What Everyone Needs to Know*. Oxford University Press.

Skillicorn, D. (2010). Applying Interestingness Measures to Ansar Forum Texts. *Association for Computing Machinery*, 7. https://doi.org/10.1145/1938606.1938613

Skinner, W. F., & Fream, A. M. (1997). A Social Learning Theory Analysis of Computer Crime among College Students. *Journal of Research in Crime and Delinquency*, *34*(4), 495–518. https://doi.org/10.1177/0022427897034004005

Smith, G., Moses, L., & Chan, J. (2017). The Challenges of Doing Criminology in the Big Data Era: Towards a Digital and Data-driven Approach. *British Journal of Criminology*, *57*, 259–274. https://doi.org/10.1093/bjc/azw096

Steinmetz, K. F., Holt, T. J., & Holt, K. M. (2020). Decoding the Binary: Reconsidering the Hacker Subculture through a Gendered Lens. *Deviant Behavior*, *41*(8), 936–948. https://doi.org/10.1080/01639625.2019.1596460

Stockton, P. N., & Golabek-Goldman, M. (2014). Prosecuting Cyberterrorists: Applying Traditional Jurisdictional Frameworks to a Modern Threat. *Policy Review*, *25*, 58.

Su, C., & Holt, T. J. (2010). Cyber bullying in Chinese Web Forums: An examination of nature and extent. *4*(1), 13.

Testa, A., Maimon, D., Sobesto, B., & Cukier, M. (2017). Illegal Roaming and File Manipulation on Target Computers: Assessing the Effect of Sanction Threats on System Trespassers' Online Behaviors. *Criminology & Public Policy*, *16*. https://doi.org/10.1111/1745-9133.12312

Tuck, H., & Silverman, T. (2016). The Counter-narrative Handbook. Retrieved March 11, 2022, from https://www.isdglobal.org/isd-publications/the-counter-narrative-handbook/

Turkle, S. (2005). *The Second Self: Computers and the Human Spirit*. MIT Press.

Vajjala, S., Majumder, B., Gupta, A., & Surana, H. (2020). *Practical Natural Language Processing: A Comprehensive Guide to Building Real-World NLP Systems*. "O'Reilly Media, Inc."

Verton, D., & Brownlow, J. (2003). *Black Ice: The Invisible Threat of Cyber-Terrorism* (1st ed.). Osborne.

Wall, D. S. (2001a). *Crime and the Internet*. Routledge.

Wall, D. S. (2001b). Cybercrimes and the Internet. *Crime and the Internet*, 1–17.

Wall, D. S. (2015). Dis-Organised Crime: Towards a Distributed Model of the Organization of Cybercrime. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2677113

Walter, D., & Ophir, Y. (2019). News Frame Analysis: An Inductive Mixed-method Computational Approach. *Communication Methods and Measures*, *13*(4), 248–266. https://doi.org/10.1080/19312458.2019.1639145

Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications* (1st edition). Cambridge University Press.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*(6684), 440–442. https://doi.org/10.1038/30918

Weimann, G. (n.d.). Going Darker? The Challenge of Dark Net Terrorism. *Wilson Center*, 13.

Weimann, G. (2004). *Cyberterrorism: How real is the threat?* (Vol. 31). United States Institute of Peace.

Weimann, G. (2005). Cyberterrorism: The sum of all fears? *Studies in Conflict & Terrorism*, *28*(2), 129–149.

Weimann, G. (2006). *Terror on the Internet: The New Arena, the New Challenges*. US Institute of Peace Press.

Weimann, G. (2011). Cyber-Fatwas and Terrorism. *Studies in Conflict & Terrorism*, *34*(10), 765–781. https://doi.org/10.1080/1057610X.2011.604831

Weimann, G. (2016a). Going Dark: Terrorism on the Dark Web. *Studies in Conflict & Terrorism*, *39*(3), 195–206. https://doi.org/10.1080/1057610X.2015.1119546

Weimann, G. (2016b). Terrorist Migration to the Dark Web. *Perspectives on Terrorism*, *10*(3), 40–44. Retrieved September 24, 2020, from http://www.jstor.org/stable/26297596

Welch, T. (2018). Theology, heroism, justice, and fear: An analysis of ISIS propaganda magazines Dabiq and Rumiyah. *Dynamics of Asymmetric Conflict*, *11*(3), 186–198. https://doi.org/10.1080/17467586.2018.1517943

Weulen Kranenbarg, M., Holt, T. J., & van der Ham, J. (2018). Don't shoot the messenger! A criminological and computer science perspective on coordinated vulnerability disclosure. *Crime Science*, *7*(1), 16. https://doi.org/10.1186/s40163-018-0090-8

Weulen Kranenbarg, M., Holt, T. J., & van Gelder, J.-L. (2019). Offending and Victimization in the Digital Age: Comparing Correlates of Cybercrime and Traditional Offending-Only, Victimization-Only and the Victimization-Offending Overlap. *Deviant Behavior*, *40*(1), 40–55. https://doi.org/10.1080/01639625.2017.1411030

Whittaker, J. M., & Button, M. (2020). Understanding pet scams: A case study of advance fee and non-delivery fraud using victims' accounts. *Australian & New Zealand Journal of Criminology*, *53*(4), 497–514. https://doi.org/10.1177/0004865820957077

Wickham, H. (2021). Tidyverse: Easily Install and Load the 'Tidyverse'. Retrieved February 24, 2022, from https://CRAN.R-project.org/package=tidyverse

Wilson, T., Maimon, D., Sobesto, B., & Cukier, M. (2015). The Effect of a Surveillance Banner in an Attacked Computer System: Additional Evidence for the Relevance of Restrictive Deterrence in Cyberspace. *Journal of Research in Crime and Delinquency*, *52*(6), 829–855. https://doi.org/10.1177/0022427815587761

Winter, C., Neumann, P., Meleagrou-Hitchens, A., Ranstorp, M., Vidino, L., & Fürst, J. (2020). Online Extremism: Research Trends in Internet Activism, Radicalization, and Counter-Strategies. *International Journal of Conflict and Violence (IJCV)*, *14*, 1–20. https://doi.org/10.4119/ijcv-3809

Wozniak, J. S. G., Woods, J., & Lee, Y. S. (2020). The evolving self-presentation of the Islamic State, from Dabiq to Rumiyah. *The Social Science Journal*, *0*(0), 1–13. https://doi.org/10.1080/03623319.2020.1727242

Yannakogeorgos, P. A. (2014). Rethinking the Threat of Cyberterrorism. In T. M. Chen, L. Jarvis, & S. Macdonald (Eds.), *Cyberterrorism: Understanding, Assessment, and Response* (pp. 43–62). Springer. Retrieved May 12, 2021, from https://doi.org/10.1007/978-1-4939-0962-9_3

Zelin, A. Y., & Fellow, R. B. (2013). The State of Global Jihad Online: A Qualitative, Quantitative, and Cross-Lingual Analysis, 24.

Zhang, S., Leidner, D., Cao, X., & Liu, N. (2021). Workplace cyberbullying: A criminological and routine activity perspective. *Journal of Information Technology*, 02683962211027888. https://doi.org/10.1177/02683962211027888

Appendix

Nodes	**Centrality**	**Eigen**	**Closeness**	**Betweenness**
Tayeb	338	1	0.679266	61034.9
Lulua	255	0.896858	0.621542	29190.0
OmMohammed	207	0.775857	0.590056	20844.7
Asif	199	0.814905	0.585661	13367.7
lubna	188	0.646704	0.577594	20910.7
Rasha	182	0.73589	0.573382	13920.5
tbahrain	160	0.643549	0.56362	10398.4
BinZiad	145	0.667047	0.554674	10093.8
Nzingha	135	0.608455	0.549345	9062.2
Netcurtains3	123	0.569296	0.543178	5031.5

Table 5.1: Top 10 Nodes Centrality Scores – Myiwc Interaction Network

Nodes	**Centrality**	**Eigen**	**Closeness**	**Betweenness**
Asadullah Alshishani	294	1	0.854369	7388.6
Insurgent	247	0.90149	0.768559	5173.2
Asad'Allah	235	0.930799	0.747346	2647.5
ANSAR 007	221	0.889646	0.727273	2925.5
wa1slama	209	0.863618	0.708249	3188.9
Abu Fatima	199	0.855034	0.69428	1332.2
Guest	192	0.851984	0.684825	1106.1
abdulrahman al muhajir	191	0.819765	0.683495	1813.0
tarbiya	184	0.82232	0.67433	1057.4
Abu Umar	175	0.784745	0.6629	995.4

Table 5.2: Top 10 Nodes Centrality Scores – Ansar1 Interaction Network