DISSERTATION

A GRAPH-BASED, SYSTEMS APPROACH FOR DETECTING VIOLENT EXTREMIST RADICALIZATION TRAJECTORIES AND OTHER LATENT BEHAVIORS

Submitted by

Benjamin W. K. Hung

College of Engineering

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2017

Doctoral Committee:

Advisor: Anura P. Jayasumana

Edwin K.P. Chong
Indrajit Ray
Ronald M. Sega

ABSTRACT

A GRAPH-BASED, SYSTEMS APPROACH FOR DETECTING VIOLENT EXTREMIST RADICALIZATION TRAJECTORIES AND OTHER LATENT BEHAVIORS

The number and lethality of violent extremist plots motivated by the Salafi-jihadist ideology have been growing for nearly the last decade in both the U.S and Western Europe. While detecting the radicalization of violent extremists is a key component in preventing future terrorist attacks, it remains a significant challenge to law enforcement due to the issues of both scale and dynamics. Recent terrorist attack successes highlight the real possibility of missed signals from, or continued radicalization by, individuals whom the authorities had formerly investigated and even interviewed. Additionally, beyond considering just the behavioral dynamics of a person of interest is the need for investigators to consider the behaviors and activities of social ties vis-á-vis the person of interest. We undertake a fundamentally systems approach in addressing these challenges by investigating the need and feasibility of a radicalization detection system, a risk assessment assistance technology for law enforcement and intelligence agencies. The proposed system first mines public data and government databases for individuals who exhibit risk indicators for extremist violence, and then enables law enforcement to monitor those individuals at the scope and scale that is lawful, and account for the dynamic indicative behaviors of the individuals and their associates rigorously and automatically. In this thesis, we first identify the operational deficiencies in efforts by current law enforcement and intelligence agencies, investigate the environmental conditions and stakeholders most salient to the development and operation of the proposed system, and

address both programmatic and technical risks with several initial mitigating strategies. We codify this large effort into a radicalization detection system framework.

The main thrust of this effort is the investigation of the technological opportunities for the identification of individuals matching a radicalization pattern of behaviors in the proposed radicalization detection system. We frame our technical approach as a unique dynamic graph pattern matching problem, and develop a technology called INSiGHT (Investigative Search for Graph-Trajectories) to help identify individuals or small groups with conforming subgraphs to a radicalization query pattern, and follow the match trajectories over time. INSiGHT is aimed at assisting law enforcement and intelligence agencies in monitoring and screening for those individuals whose behaviors indicate a significant risk for violence, and allow for the better prioritization of limited investigative resources. We demonstrated the performance of INSiGHT on a variety of datasets, to include small synthetic radicalization-specific data sets, a real behavioral dataset of time-stamped radicalization indicators of recent U.S. violent extremists, and a large, real-world BlogCatalog dataset serving as a proxy for the type of intelligence or law enforcement data networks that could be utilized to track the radicalization of violent extremists.

We also extended INSiGHT by developing a non-combinatorial neighbor matching technique to enable analysts to maintain visibility of potential collective threats and conspiracies and account for the role close social ties have in an individual's radicalization. This enhancement was validated on small, synthetic radicalization-specific datasets as well as the large BlogCatalog dataset with real social network connections and tagging behaviors for over 80K accounts. The results showed that our algorithm returned whole and partial subgraph

matches that enabled analysts to gain and maintain visibility on neighbors' activities. Overall, INSiGHT led to consistent, informed, and reliable assessments about those who pose a significant risk for some latent behavior in a variety of settings. Based upon these results, we maintain that INSiGHT is a feasible and useful supporting technology with the potential to optimize law enforcement investigative efforts and ultimately enable the prevention of individuals from carrying out extremist violence.

Although the prime motivation of this research is the detection of violent extremist radicalization, we found that INSiGHT is applicable in detecting latent behaviors in other domains such as on-line student assessment and consumer analytics. This utility was demonstrated through experiments with real data. For on-line student assessment, we tested INSiGHT on a MOOC dataset of students and time-stamped on-line course activities to predict those students who persisted in the course. For consumer analytics, we tested the performance on a real, large proprietary consumer activities dataset from a home improvement retailer. Lastly, motivated by the desire to validate INSiGHT as a screening technology when ground truth is known, we developed a synthetic data generator of large population, time-stamped, individual-level consumer activities data consistent with an a priori project set designation (latent behavior). This contribution also sets the stage for future work in developing an analogous synthetic data generator for radicalization indicators to serve as a testbed for INSiGHT and other data mining algorithms.

ACKNOWLEDGEMENTS

I am profoundly grateful to many people for the role they played in my life and academic journey. I have been blessed with tremendous teachers and mentors throughout middle school and high school, at the United States Military Academy, at the Massachusetts Institute of Technology, and now at Colorado State University.

Special thanks go to Dr. Stephan Kolitz and Professor Asuman Ozdaglar for their instruction and mentorship during my first graduate research endeavor. I will always be grateful for their kindness, patience, and generosity with their time.

I am deeply indebted to Professor Jayasumana for taking me on as a student, especially knowing that it was going to be through part-time distance initially, and for the confidence he had in me in supporting my transition to complete the program on a full-time basis over a very short period authorized by the U.S. Army. I am, of course, indebted to him for his mentorship and advisorship throughout this research effort. I am also tremendously grateful to my committee members, Professors Chong, Ray, and Sega for their invaluable support, belief in our work, and their guidance along the way.

I would also like to give thanks to Dr. Vidarshana Bandara, my advisor's former student, for the countless hours he assisted me in learning Big Data processing techniques and in reviewing my work. Throughout this research period, I have also benefited tremendously from the scholarship of and collaboration with Professor Klausen at Brandeis University and the lifetime of study she has in terrorism and intergroup conflict.

None of this would have been possible without the love and support from my family including my parents Frank and Vera, my brothers Andrew and Eric, and my in-laws John

and Ann, particularly for their generous hospitality in Colorado for the last 18 months. To my incredible children Nathanael, Stephen, and Anna, thank you for making me smile and laugh and for putting up with my absences during the many hours I was undertaking this important work. Lastly, to my darling wife Elizabeth, your love and support has meant everything and enabled everything. Thank you from the bottom of my heart.

I am grateful to my Lord and Savior who has blessed us with every good gift and given me the strength and perseverance to complete this work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

DISCLAIMER

The views expressed in this article are those of the author and do not reflect the official policy or position of the U.S. Army, Department of Defense, or the U.S. Government.

CHAPTER 1

Introduction

Radicalized violent extremists seeking to support or commit terrorist acts continue to pose a serious threat in the United States and abroad. Those extremists motivated by the Salafi-jihdaist ideology have perpetrated 26 such attacks in the U.S. alone since 9/11, including the 2009 Fort Hood attack, the 2013 Boston Marathon Bombings, the 2015 San Bernardino attack, and the 2016 Orlando night club attack. In Western Europe, radical violent extremists who adhere to the same ideology[1] carried out numerous other attacks including the 2015 Paris attacks, the 2016 Brussels airport and metro attacks, the 2016 Nice attack, the 2016 Berlin market attack, the 2017 London Parliament attack, and most recently the 2017 Manchester concert bombing. In order to thwart other violent extremists from carrying out future attacks, law enforcement agencies in their investigative capacities are effectively called upon to monitor and make continuous risk assessments on a large number of individuals for the likelihood of violence. To prevent people from becoming attracted to violent extremist groups in the first place, non-governmental organizations (NGOs), working either in private or partnered with government entities in prevention and de-radicalization programs, make these risk assessments as well [221]. The task is fraught with challenges, which was aptly summarized recently by the former FBI Director following the terrorist attack at an Orlando night club in June 2016: "We are looking for needles in a nation-wide haystack, but we're also called upon to figure out which pieces of hay might someday become needles" [56]. His remark highlights that both *dynamics and scalability* are key interrelated issues involved in the detection of radicalization and the prevention of future attacks.

[1]Throughout this work, we focus our investigation on the violent extremists motivated by the Salafi-jihadist ideology and will often refer to them as just 'violent extremists' for brevity.

Current research suggests that radicalization, while complex, may be understood as a *dynamic* and phased-based process where individuals exhibit indicative behaviors or psychological states along pathways to violence.[2] While the importance of kinship or other social ties to ones involvement in terrorism is well established [133, 262], there is neither consensus on all the components (phases or indicators) of these models [165, 214], nor on how long the process itself takes. The latter has been posited to range from years to weeks, and to even days [123, 165]. Yet law enforcement agencies recognize the dynamics of the problem and utilize all lawful investigative techniques and methods, including both physical and electronic surveillance, in order to detect indicators of violent radicalization at the earliest opportunity and to be postured to rapidly foil plots [57, 123]. However, given the commitment of personnel and technical assets necessary, employment of these full-on techniques do not *scale* well to the caseload. These resource constraints have subsequently forced the agencies to make tough, subjective decisions on the level of surveillance and monitoring that suspected individuals receive [88, 123, 255]. Moreover, recent terrorist attack successes highlight the real possibility of missed signals from, or continued radicalization by, individuals whom the authorities had formerly investigated and even interviewed.[3] Additionally, beyond considering just the behavioral dynamics of a person of interest is the need for investigators to consider the behaviors and activities of social ties vis-á-vis the person of interest. As in the conspiracy behind the San Bernardino attack in 2015, it is only when behaviors are viewed

[2]See Chapter 2 or [31, 164] for some thorough surveys of the different models and [165] for a good discussion on the operationalization of these conceptual models.

[3]Recent U.S. cases include Tamerlan Tsarnaev (Boston Marathon bombings, 2013) [106], Omar Mateen (Orlando night club shooting, 2016) [56], and Ahmad Khan Rahami (New York and New Jersey bombings, 2016) [217]. Recent Western European cases include several involved in the Paris terrorist attacks in 2015 [96], Anis Amri (Berlin market truck attack, 2016) [77], and Salman Abedi (Manchester concert bombing, 2017)[205]. Likewise, NGOs and practitioners working in de-radicalization programming also fear mistakenly taking someone on who is in fact highly radicalized and poses a risk to society [169].

collectively that one may be able to detect the many indicators of a violent extremist plot are present.[4]

1.1. Challenges and Motivation

Due to the devastating effects of 9/11, the U.S. Department of Justice and the FBI has since strategically shifted efforts from prosecution of terrorism to preventive counterterrorism efforts through the investigation of those on suspected pathways of radicalization to violent extremism.[5] Although law enforcement face a number of challenges in identifying radicalization and preventing violent extremist attacks, they are all related to either issues of dynamics or scalability and can be grouped into two principle and interrelated categories: 1) the insufficiency of current radicalization risk assessment protocols to anticipate the imminent risk of violence, and 2) the shortfalls related to the monitoring and surveillance of those considered at risk for committing extremist violence.[6]

In order to determine whom to initially investigate and the risk level of individuals currently under investigation, law enforcement agencies generally use some form of a risk assessment protocol or structured professional judgment instrument, which can be laborious and require the manual assessment of dozens of indicator items for each individual of interest. Furthermore, the process does not sufficiently distinguish those truly on pathways to extremist violence and those who are not [264, p. 11], is not designed to consider behavioral dynamics of individuals [168] and their relevant social ties, and has no known associated automated methods to keep up with the evolving individual-level indicative behaviors.

[4]See Section 3.2.3 (Complexity of Assessing Risk through Social Ties) and [299] for more details.

[5]See the 2002 DOJ fact sheet [79] and [10, 284] for analysis of this change.

[6]See Section 3.2 (Operational Deficiencies) for more details.

The other related group of challenges all involve shortfalls with monitoring and surveillance of those considered at risk for extremist violence. First, law enforcement faces shortfalls in terms of resources for full-on surveillance. For example, while it was estimated that the FBI had over 900 active investigations in 2015 [27] and around 1000 investigations in 2016 [123] of ISIS-related homegrown violent extremists, it had the ability to thoroughly surveil the activities of only "dozens" of individuals [123]. Furthermore, the current state of the law enforcement and intelligence information sharing enterprise is not conducive to tracking individual radicalization indicators of individuals and their associates due to the stove-piped agency databases [211] and the reliance on the query/response pattern [229] for each individual of interest. Lastly, federal and state-level law enforcement agencies face limitations with the utilization of available indicators on social media due to proprietary restrictions and privacy protections [197]. Overall, these shortfalls have the led the FBI, in particular, to seek help from local law enforcement [244] as well as make tough, resource-constrained decisions on whom it would select for surveillance based on insufficiently reliable risk assessment protocols. Unfortunately, this has allowed some of those whom the Bureau had previously investigated to eventually go on and carry out an attack. See, for example, Tamerlan Tsarnaev [106], Omar Mateen [56], and Ahmad Khan Rahami [217].

Addressing the aforementioned challenges is of great interest to both law enforcement and intelligence agencies, as well as the researchers who desire to support them. The latter, however, are also challenged by a deficit in individual-specific data on behavioral cues for study. A more detailed discussion on the "lack of comprehensive and reliable data" impeding scholarship in terrorism studies is in [264, p. 6-7]. The problem stems from the fact that such data is often considered sensitive or even classified, and is closely held by law enforcement and

intelligence agencies. With the exception of the Klausen dataset recently available through the National Institute of Justice [167],[7] there is a paucity of large datasets for training and testing that contain individual-level, time-stamped activities containing behavioral cues for data scientists to test algorithms for detecting radicalization.

1.2. Research Purpose and Questions

The overarching purpose of this research is to assist US law enforcement and intelligence agencies in identifying domestic radicalization to violent extremism and preventing future violent extremist attacks. Based upon the aforementioned challenges in the previous section, in this thesis, we investigate the need and feasibility of risk assessment assistance technologies that enables law enforcement monitoring at the scope and scale that is lawful, and that rigorously and automatically considers the dynamic indicative behaviors of individual persons of interest as well as their associates. During this investigation, we are guided by the following questions related to the issues of both *dynamics and scalability* in assessing the risk of violent extremist radicalization:

(1) Which risk assessment indicators for violent extremism in the extant literature are detectable via automated or semi-automated technologies, and what databases and datasets must be integrated to facilitate this detection? (*scalability*)

(2) Can computationally efficient tools be used to mine these databases (existing and streaming data) for the specific purposes of monitoring and screening for in near real-time those individuals who pose a significant risk for violence? Do these tools allow for the better prioritization of limited investigative resources? (*scalability*)

[7] See Section 2.4.4.1 for a description and use of this dataset in this work.

(3) Can tools that rigorously examine and account for the activities of close associates better assess the risk that an individual engages in violent extremism? (*dynamics*)

(4) Beyond the established indicators in extant literature, are there any more discerning, dynamic patterns of behavioral indicators that could help law enforcement better assess the risk of violent extremist radicalization? (*dynamics*)

Understanding which risk assessment indicators are detectable through automated means (Question 1) concerns scalability and has the potential to address the resource gaps that law enforcement agencies face in tracking the radicalization trajectories of a large number of individuals. It also involves the prescription of including currently under-utilized data sources such as social media, and better integrating extant law enforcement and governmental databases where specific indicators can be found.

The development of computationally efficient tools to mine the databases (Question 2) is the main thrust of this present research effort and addresses several law enforcement challenges with dynamically assessing risk at scale. The technology must be efficient enough to sift through voluminous open and restricted (government) databases for patterns of radicalization indicators for a large number of persons of interest, and to assist in dynamically updating law enforcement and intelligence agencies when these individuals exhibit additional indicators through their evolving behaviors. Additionally, given the paucity of available large-population datasets for testing and validating such tools, developing a process for anonymized, synthetic data generation would be an important supporting effort.

Accounting for the activities and risk indicators in associates (Question 3) directly addresses a deficiency in current individualized risk assessment protocols by not rigorously accounting the dynamic influences of close associates. A technology which can account for

the activity and behaviors of linked individuals has the potential to better assessing the risk of an individual's risk for radicalization and the risk of conspiratorial plots.

Lastly, the search for more discerning, dynamic patterns of behavioral indicators (Question 4), is focused on contributing to developing better risk assessment protocols for law enforcement use. By utilizing a technology that can consider sets of indicators and their occurrence rates, investigators and analysts may be better able to distinguish those at risk for extremist violence.

Although the prime motivation of this research is the detection of the violent extremist radicalization, similar problems exist in the other domains such as consumer analytics, on-line student assessment, behavioral health, and cybersecurity. This is primarily because, in each of these domains, there is a compelling interest to detect the presence of an individual's latent behavior utilizing time-stamped indicator data.[8] Investigating the applicability and utility of our work to each of these domains is an additional research purpose. In consumer analytics, for example, businesses are interested in using an individual's on-line activities and previous purchases over time to track his or her place on the customer journey and determine the potential for future purchases [89, 90, 309]. Likewise, in the field on-line student assessment, Massive Open Online Course (MOOC) teachers and course designers are interested in predicting student performance and persistence through the time-stamped course-related behavioral data [8, 32, 98, 173]. In behavioral health too, we find that family members and caregivers are interested in identifying those who may be showing indicators of suicide risk or veterans who may exhibit signs of post-traumatic stress [152, 232, 248]. Each of these risk behaviors has their own set of unique, identifiable signs which may be exhibited over time. Lastly, in cybersecurity, organizations continually seek to prevent insider threats

[8]We define a latent behavior as a hidden or emergent activity exhibited by an entity [109].

by detecting risk potential using performance-related and technical indicators recorded over time [45, 73].

1.3. Solution Approaches

Given the size and complexity of the problem set, we undertake a systems-based approach for the need and feasibility analysis of a risk assessment assistance technology. We henceforth refer to this objective technology as a 'radicalization detection system' oriented towards assisting law enforcement and intelligence analysts in screening for, in near real-time, those individuals who are on the pathway towards extremist violence. Additionally, as part of a nested effort at demonstrating technological opportunities for a radicalization detection system, we pursue the development of a computationally efficient tool that can mine, monitor and screen for the occurrence of radicalization indicators in large heterogeneous databases in order to provide early warnings of individuals or groups on behavioral trajectories toward extremist violence. For this technical portion, we utilize a dynamic graph pattern matching approach.

1.3.1. A systems-based approach. Systems engineering, defined as "an interdisciplinary approach and means to enable the realization of successful systems" [147], is a well suited approach to investigate the need and feasibility of a radicalization detection system. It is an ordered process that 1) focuses on framing the right (albeit possibly complex) problem to solve through holistic, associative thinking, 2) integrates the expertise from a multitude of relevant disciplines to solve the problem, 3) analyzes and seeks to meet stakeholder needs to ensure optimal value for the resources [239].

A systems approach is critical because the problem area straddles a multitude of complex issues, including

- Adequately enabling and empowering law enforcement and intelligence agencies to address a real and growing threat from violent extremists and to prevent future attacks.
- The ability of law enforcement to access and utilize the early warning behavioral indicators available in on-line activities of persons of interest, while ensuring the protection of civil liberties and the privacy to the broader population.
- The ability of law enforcement to access and utilize the early warning behaviors and indicators in disparate government databases, while still preserving classification levels and authorities for access.
- The proper roles of social media companies and internet service providers to balance the fostering of free communities, the policing of content, and the legal requirements of supporting law enforcement investigative efforts.
- Addressing public concerns over the possibility of perceived discriminatory policies and actions against minorities.

Involved in each of these issues are stakeholders with competing interests as well as some adversarial groups who are opposed to any significant increase in law enforcement capabilities or to any erosion of individual privacy.[9] As will be described later in Chapter 2, the threat posed by violent extremists is real and growing, as well as the resource and technical problems posed to law enforcement and intelligence agencies. Real too is the ability of social media companies to cause the shut down of all or portions of a social media monitoring company's operations by denying application program interfaces (APIs) over concerns of the assistance given to law enforcement agencies.

[9]These groups have been shown to have high salience, which is defined as possessing one or more either the power to influence the system, legitimacy in relation to the system, and urgency of the claim on the system [239]. See Section 2.5.3 for more details.

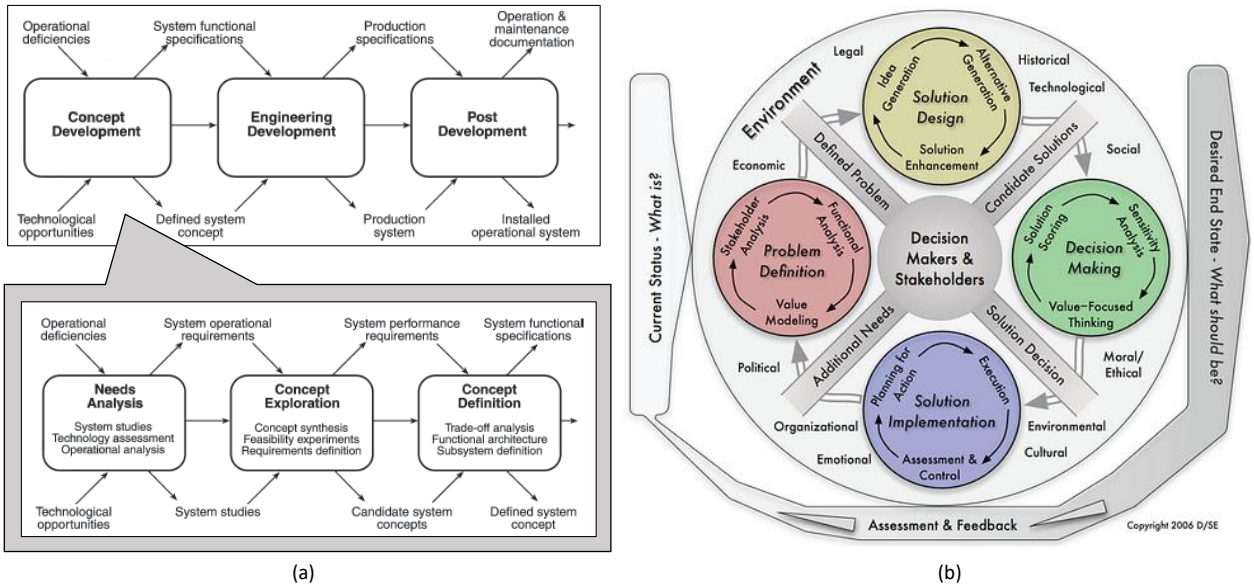


FIGURE 1.1. (a) Principle stages of the systems engineering life cycle with the expansion of concept development phases [174]. (b) Systems Decision Process [239].

This thesis takes an integrated analysis approach utilizing both foundational systems engineering theory in [174] and the systems decision process (SDP) in [239]. The former is depicted in Fig. 1.1a, which details the three stages of the systems engineering life cycle model with a more in-depth look at the sub-stages and activities within concept development. The latter is depicted in Fig.1.1b, which details the disciplined, cyclical process involving stakeholders and decision makers throughout the life cycle. While the systems engineering life cycle model focuses more on the inputs, outputs, and functions of each stage, the SDP better emphasizes the requisite analyses of environmental factors and stakeholders in all the stages [239].

Therefore, in this stage of research effort for a radicalization detection system, we focus on the components of the Needs Analysis sub-stage of the systems engineering life cycle, where one primarily asks, "Is there a valid need for a new system?" and "Is there a practical approach to satisfying such a need?" [174]. Our analysis in this thesis includes identification

of the operational deficiencies of current law enforcement efforts, and a robust set of system studies that covers the measures currently being taken at the operational level to address the threat from violent extremism.

The preponderance of this thesis is devoted to the study of technological opportunities by developing and testing the utility of a novel dynamic graph pattern matching technique for the identification of individuals matching a radicalization pattern of behaviors. However, consistent with the analytical framework of the SDP, a significant portion of this research effort was also devoted to the description of the environmental context and a framework for a radicalization detection system, the full development of which requires engagement of multiple stakeholders as well as sizable associated efforts by the government and other researchers to realize. Lastly, we also assess both high-level programmatic and technical risks of this system.

For reasons of specificity and thoroughness of analysis, we define the boundary of our systems-level examination of a radicalization detection system as proposed for implementation within the U.S., but acknowledge that the principles can seemingly be applied to comparable efforts in Western Europe and elsewhere. We also focus on supporting law enforcement (over NGOs involved in de-radicalization), because of the imminence of the threat that these organizations face, their access to restricted governmental and intelligence-related data sources, and the scale of technological effort that we are advocating.

Lastly, as is characteristic of the systems approach, the multi-disciplinary aspects of this research effort are also prominent. We surveyed and regularly revisited the extensive scholarship of social scientists and terrorism experts to better grasp an understanding of radicalization processes and early warning indicators of violence. Later, we were fortunate

to also consult with Professor Jytte Klausen, founder of the Western Jihadist Project at Brandeis University and a highly-respected, cited, and published researcher in the areas of jihadist radicalization and violent extremism. These efforts provided the foundation for the investigation and development of a technological opportunity to assist law enforcement in screening for individuals at risk for extremist violence.

1.3.2. A DYNAMIC GRAPH PATTERN MATCHING APPROACH. For the specific task of detecting whether an individual or group of individuals exhibit a pattern of radicalization indicators, we utilize a dynamic graph pattern matching approach which we call *investigative graph search*. While traditional graph pattern matching is well-studied and has been used extensively in a variety of applications to include complex object identification, software plagiarism detection, traffic route planning, and recommender systems [23, 100, 113, 193], it relies on the certainty of specific types of connections or attributes in the query pattern. Since the detection of latent behaviors such as radicalization may involve less certainty about the query structure, or that the entities of interest may not exhibit all of the possible behaviors or attributes, we devised investigative graph search to search for and prioritize persons of interest who may exhibit part or all of a pattern of suspicious behaviors or connections.

We framed our technical approach as a unique dynamic graph pattern matching problem and introduced a technology called INSiGHT (Investigative Search for Graph-Trajectories) to help identify individuals or small groups with conforming subgraphs to a radicalization query pattern and follow the match trajectories over time. We tested our software implementation of INSiGHT on small synthetic and real datasets related to radicalization. The small, synthetic dataset was a stylized set of five persons of interest and their on- and off-line behaviors arranged in a heterogeneous graph of 61 nodes and 59 time-stamped edges. Each

of the five individuals fit some profile determined a priori that included violent extremists of varying numbers and types of exhibited behaviors, as well as a non-extremist and a former violent extremist. We also validated our work on a real dataset from [167], which consisted of 135 U.S. homegrown violent extremists who were arrested for terrorism-related offenses between 2001 and 2015 or had died in the commission of their offenses, and 1,326 combined behavioral indicators exhibited over time. Due to the lack of large-scale radicalization-related datasets, we also tested our implementation of INSiGHT on the BlogCatalog dataset [320], a heterogeneous graph consisting of over 382K nodes and 4 million edges which served as a data proxy because it contained structural and behavioral parallels to intelligence-related networks.

Importantly, beyond the stated purpose of detecting radicalization, we demonstrated in this thesis that the basic technical approach of investigative graph search is also very applicable to several other domains that involve the detection of latent behaviors utilizing time-stamped indicator data. As discussed earlier, such domains include consumer analytics, on-line student assessment, behavioral health, and cybersecurity. In this work, we applied investigative graph search and provided results for problems in both the consumer analytics and on-line student assessment domains. We tested INSiGHT on a real, large proprietary consumer activities dataset from a home improvement retailer with 60K customers and over 11 million transactions over a two-year period as well as a synthetically generated dataset of 1K customers and 25K transactions. INSiGHT was indeed useful in the detection of customers undertaking certain home improvement projects based on this time-stamped purchase data. For on-line student assessment, we tested INSiGHT on a portion of the Knowledge Discovery and Data Mining (KDD) Cup 2015 competition dataset [160] of approximately

1K students and 19K on-line activities to predict those students who persisted in the MOOC course.

1.4. Contributions and Outcomes

The following summarizes our main contributions in this thesis. We first investigate at a system-level the fundamental need and feasibility of a violent extremist radicalization detection system and correspondingly propose an overarching analyst-in-the-loop framework for a tool that would mine pubic data and government databases for individuals who pose a significant risk for extremist violence. We also describe the environmental conditions and most salient stakeholders related to the proposed system, and address programmatic and technical risks with several initial mitigating strategies.

Towards the development of a computationally efficient tool to mine these databases, we formalize the technical approach of investigative graph search and develop INSiGHT as an algorithmic implementation that performs it. INSiGHT is a vectorized, multi-hop class similarity graph pattern matching technique that tracks full or partial matches of subgraphs to a query graph over time. Tailorable to the detection of radicalization indicator patterns, enhancements were also developed to account for the re-occurrence of indicators, the time decay of indicator significance, and the incorporation of red flag and other conditional filters. The applicability of INSiGHT to detect latent behaviors in other domains such as consumer analytics and on-line student assessment was demonstrated through experiments with real data. We also further enhance INSiGHT with an algorithmic implementation for non-combinatorial graph neighbor matching that accounts for the activities of close associates to better reveal the presence of suspicious individuals or conspiratorial plots.

Additionally, as a means to address the paucity in large population, time-stamped, individual-level activities data with ground truth in both the radicalization domain and other domains, we provide a synthetic data generator software implemented in MATLAB. The generator produces large population, time-stamped, individual-level consumer activities data consistent with an a priori project set designation (latent behavior). Importantly, this formulation sets the stage for future work in developing an analogous synthetic data generator for radicalization indicators to serve as a testbed for INSiGHT and other data mining algorithms.

The last significant contribution contained in this thesis is the development of a novel discrete dynamical process to model violent extremist radicalization and use of state transition analysis to find more discerning patterns among the behavioral indicators. Our efforts centered on analyzing the unique, restricted-use Klausen dataset [167] that contained known U.S. violent extremists and their behavior indicators coded in time. Significantly, we found highly frequented indicator transitions that could subsequently be incorporated in a system to better detect radicalization trajectories. Additionally, our analysis showed that while indeed perpetrators took widely various paths in total (as is commonly characterized in the literature), an overwhelming majority of them followed at least some highly common *segments* of paths. The identification of these few frequent pair-wise sequences could prove useful to law enforcement and intelligence analysts.

1.5. Thesis Outline

This thesis is organized as follows. In Chapter 2 we provide a primer on the threat from violent extremists motivated by the Salafi-jihadist belief system, a description of governmental strategies to counter this threat, and a survey of the most recent research on radicalization

processes and early warning behaviors. This chapter motivates our efforts in this problem area and establishes a foundation into the scholarship of largely social-psychological process and phenomenon.

In Chapter 3 we first study of the current operational deficiencies that law enforcement and intelligence agencies face in the very challenging work of preventing future extremist violence and protecting the public. Then we provide a description of a radicalization detection system framework, which is an overarching approach designed to address the existing deficiencies. In line with the systems development process, we also describe the environmental conditions in which this proposed system would be developed and operate, provide an analysis of the most salient stakeholders, and address programmatic and technical risks with several initial mitigating strategies.

Chapter 4 introduces our foundational technical approach to detecting individual-level radicalization trajectories through investigative graph search, a novel dynamic graph pattern matching process. In this chapter, we also develop a categorization of indicators for violent extremist radicalization that is consistent with the threat assessment literature and important in both reducing false positives and enabling alerts for red flag behaviors.

Chapter 5 covers the development of our principal technical contribution, the INSiGHT dynamic graph pattern matching algorithm and its use in detecting latent behaviors through the mining of large graph databases for indicators that match a query pattern. We test INSiGHT on an array of datasets, both real and synthetic and of varying sizes.

In Chapter 6 we provide a synthetic data generator for large population, time-stamped, individual-level consumer activities data consistent with an apriori latent behavior. This enabled further validation of INSiGHT as a screening technology when ground truth is

known, as well as sets the stage for future work in developing an analogous synthetic data generator for radicalization indicators to serve as a testbed for INSiGHT and other data mining algorithms.

In Chapter 7 we extend our formulation of INSiGHT to enable the non-combinatorial neighborhood matching on graphs to identify potential threats from clusters of individuals in possible terrorist conspiracies. We also develop a match goodness function as a quantitative means to prioritize the investigation of matches (i.e., potential threats), and test the application of INSiGHT on small synthetic radicalization datasets and one real world proxy dataset on a benign domain with structural parallels to radicalization.

In Chapter 8 we develop a discrete dynamic model of radicalization by utilizing a real indicator dataset offered by the Western Jihadism Project at Brandeis University [167], and relate the results back to refining the query pattern for INSiGHT with more discerning behavioral transitions.

Finally, in Chapter 9 we summarize the contributions of our work and provide several key areas of future work on the develop and testing of INSiGHT as well as the extension of the methodology for detecting other types of latent behaviors such as suicide risk and insider threats.

We also include the following in the appendices: Appendix A, a reference table for the recent prominent violent extremist attacks (not limited to Salafi-jihadism ideology); Appendix B, a set of case students of U.S. homegrown violent extremists; and Appendix C, Klausen's original codebook for radicalization indicators that was utilized in [165, 168, 171].

CHAPTER 2

A Primer on the Violent Extremist Threat and Radicalization Processes

2.1. Introduction

This chapter serves as a primer on the threat from violent extremists and existing research on radicalization processes. An understanding of the nature and scope of the problem is necessary to better understand its complexity, the motivation of our efforts, and where our research fits in the context of other existing efforts to prevent future attacks. It first covers the trends primarily in the U.S. in violent extremists plots and successful attacks, followed by an overview of the strategies and methods governmental officials are proposing to counter the threat of violent extremism. Later, it provides a survey of the latest research advances in understanding the violent extremist population, radicalization process and the associated indicators or early warning behaviors. Lastly, it presents the specific system studies of measures currently being taken at the operational level to address the threat from violent extremism.

We first note that violent extremism is subject to terminological differences. The U.S. government primarily refers to it in the broadest sense of "supporting or committing violent acts to achieve political, ideological, religious, or social goals" and include under this umbrella groups such as "white supremacists, anti-government groups, and groups with extreme views on abortion, animal rights, the environment, and federal ownership of public lands; and radical Islamist entities, such as the Islamic State of Iraq and Syria (ISIS)" [122, p. 1]. However, given the vast differences in motivations and ideologies and indicators [168], in this

thesis we narrow the scope of violent extremism as others have done[10] to the supporting or commitment of violent acts motivated by the Salafi-jihadist ideology.

2.2. Growing Threat from Violent Extremism

Over the last decade and a half, we have seen a precipitous rise in the threat of violent extremists inspired by the Salafi-jihadist ideology who have been radicalized and seek to commit acts of terrorism in the United States and abroad. As opposed to the terrorist attacks on 9/11 which were planned and conducted by foreigners, the recent growing threat has been from 'homegrown violent extremists' who 'act alone or in small groups on behalf of al-Qa'ida and now ISIS without any direct or formal connection to the foreign terrorist organizations" [171].[11] The homegrown terrorist threat in the U.S. has been facilitated and fueled by extremist organizations utilizing the internet and social media for recruitment and radicalization [166]. This technique seeks to inspire others in their home countries to conduct decentralized attacks and do not necessarily require the organizational hierarchy that once facilitated transnational terrorism.[12]

[10]Researchers and practitioners who have scoped their focus particularly to the aforementioned radical Islamist entities have either specified the extremists' motivational ideology (i.e., Klausen specifies "violent extremists motivated by the Salafi-jihadist ideology" in her work [165, 167]), or, as was done in CSIS's seminal work "Turning Point: A New Comprehensive Strategy for Countering Violent Extremism," stated upfront that use the term violent extremism "[referred] to the subset of violent extremist organizations that claim that religion of Islam as their motivating source to justify their nefarious goals" [126, p. 2].

[11]The characterization of 'homegrown' is particularly apt in the U.S., where every attack successfully carried out since 9/11 was committed by citizens or permanent legal residents [16]. We note that Klausen has recently advocated that such a concept is misleading when referring to terrorist violence in Europe. She points to the empirical evidence of a large European terrorist network of 85 individuals who responsible for the 2016 Brussels and Paris attacks were also responsible for 9 other plots in Europe (4 of which were successful) from 2014 to 2016.

[12]The term 'lone wolf terrorist' is another frequently used term for 'homegrown terrorist,' but is also commonly-debated among terrorism researchers and practitioners. According to [116, 292], 'lone wolf terrorist' potentially helps glorify the offenders while also ambiguously setting unclear limitations on either the number of actors and their connections to any broader networks.

2.2.1. Data sources. Deciding on a data source is the first step in quantifying the previous acts of violent extremism. Various organizations to keep track of terrorist-related incidents, but the numbers seem to vary based upon the definition of terrorism or extremist violence, the time window of analysis, and the geographical scope (e.g., US, Western Europe, international). They also vary in the level of reported detail and the plot/incident characteristics coded. For instance, the University of Maryland's National Consortium for the Study of Terrorism and Responses to Terrorism (START) hosts the Global Terrorism Database [178] of both U.S. and international terrorist incidents from 1970 to 2015, while Esri has produced a crowd-sourced database and visualization of worldwide terrorist incidents for 2016 and 2017 [97]. START also produced two databases with a U.S. focus. First, the U.S. Extremist Crime Database [112] contains criminal incidents from 1990-2015 by far rightists, Islamist radicals, animal and environmental extremists. Second, START had maintained an informative database of the characteristics of U.S. violent and non-violent extremists called Profiles of Individual Radicalization in the United States (PIRUS) [179]. Importantly, it contains individual-level details of offender demographic backgrounds and radicalization characteristics of nearly 1500 individuals motivated by varying ideologies from 1948-2013. While the START databases are widely cited for raw incident numbers, we chose to use the New America database[13] in this primer to provide an overview of the trends in terrorist attacks. This think tank's database 1) is focused only on U.S. incidents since September 11, 2001, 2) includes important coded details on the incident as well as their perpetrators, 3) contains up-to-date information and is observed to be updated fairly quickly following recent incidents, and 4) is well-researched and documented with the sources of information.

[13] https://www.newamerica.org/in-depth/terrorism-in-america/part-i-overview-terrorism-cases-2001-today/

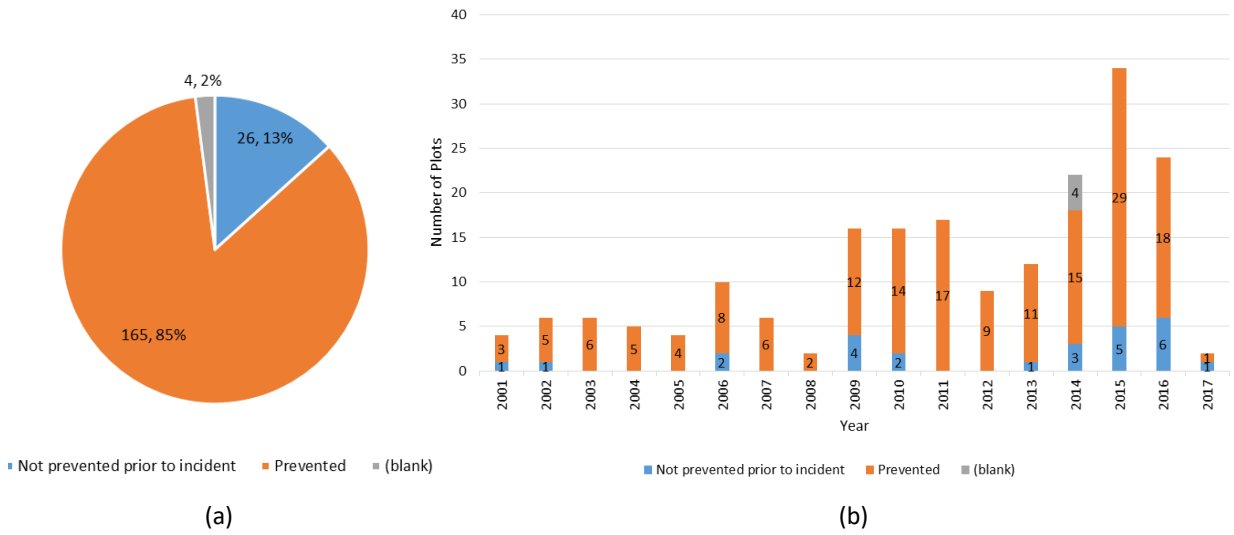


FIGURE 2.1. (a) Percentage of U.S. terrorist plots from jihadist ideology that were prevented or not prevented. The (blank) categories in the dataset were mostly attributed to U.S. individuals who successfully traveled or attempted to travel to become foreign fighters. (b) The breakdown of plots by type over time. Data provided by [17].

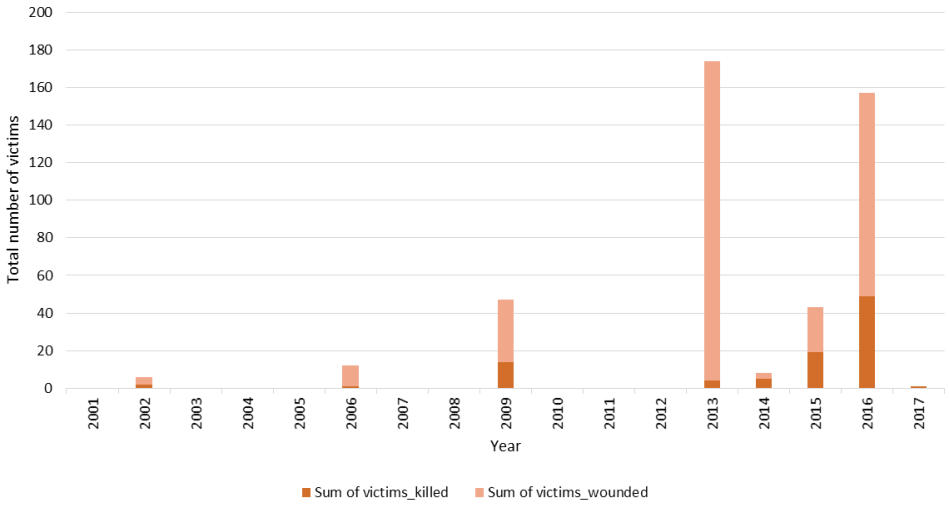


FIGURE 2.2. The total number of victims (killed and wounded) by year in U.S. terrorist plots from the jihadist ideology. Data provided by [17].

2.2.2. THE TRENDS IN BOTH SUCCESSFUL AND PREVENTED ATTACKS. Within the U.S. both the number and lethality of violent extremist plots have been growing. While Fig. 2.1a shows a vast majority of terrorist plots inspired by the Salafi-jihadist ideology have been prevented, Fig. 2.1b shows that the number of plots has grown particularly since 2009 and

peaked to 34 in 2015 alone.[14] Beyond simply the number of plots, Fig. 2.2 shows the yearly totals of those killed and wounded by these terrorist attacks that were successfully carried out. While one could argue that the number of incidents has been relatively small since 9/11, the timeline is punctuated by spikes of a large number of victims for each successful attack. Notably, the Fort Hood shooting by Nidal Hasan occurred in 2009, the Boston Marathon bombing by Tamerlan and Dhzokhar Tsarnaev occurred in 2013, the San Bernardino shooting by Sayed Farook and Tashfeen Malik occurred in 2015, and most recently the Orlando night club shooting by Omar Mateen occurred in 2016. The costliness of the operational deficiencies of existing counter-radicalization and counter-terrorism efforts is clear: the 26 attacks carried out in the U.S. since 9/11 had a collective total of 95 killed and 353 wounded[15]

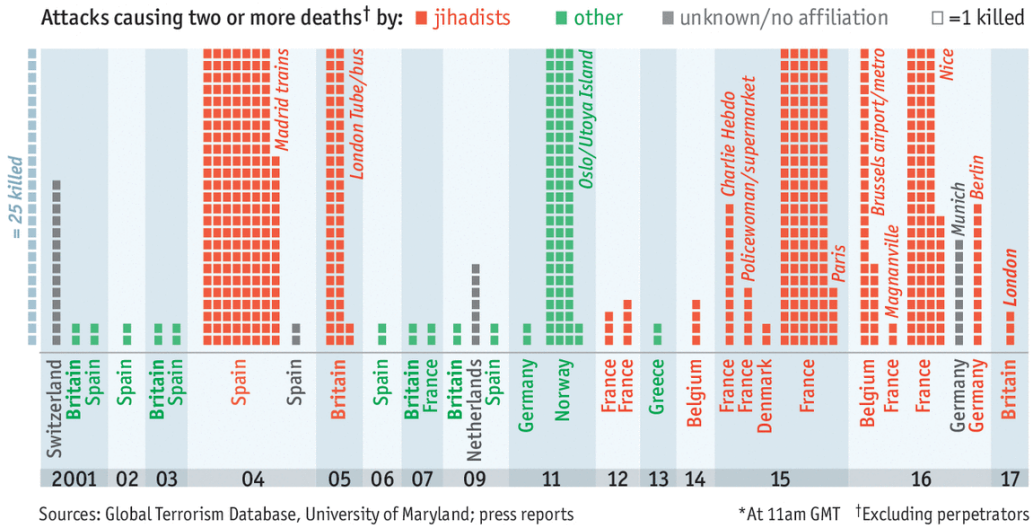


FIGURE 2.3. Major terrorist attacks in Western Europe from September 11, 2001 to March 23, 2017. Source: [87].

[14]Plots according to the database are broadly defined to include anyone charged with a terrorist-related offense, and can include those attempting to travel to Iraq or Syria to fight for the Islamic State, or those providing material support to terrorists, or those plotting and attempting to carry out their own attacks.

[15]While a statistical outlier, it is also important to recall that the terrorist attacks on 9/11 killed 2996 people and wounded over 6000.

It is also important to put the violent extremist threat in the U.S. in context with other areas around the world. Of serious concern is the noticeably larger threat by Salafi-jihadists to Western Europe. The infographic in Fig. 2.3 shows the number of killed in each of the major terrorist attacks since 2001, with a focus on the data points in red since those are due to jihadists. Several factors have been posited to explain this difference, to include Europe's geographic proximity to the conflicts in Iraq and Syria and the relative ease by which trained and radicalized foreign fighters can return, as well as the lack of intelligence sharing among European countries [36, 96]. Law enforcement agencies in Western Europe too are facing resource constraints and the inability to track all those radicalizing individuals they deem are potential threats [88, 255].

Another benefit of the New America dataset is that it provides, where possible, the identification of the primary method by which each terrorist attack or threat was prevented based upon publicly available information. While the method of prevention was unclear for a significant portion of cases (28%), we do gain insights on the prevalence of the known methods of prevention in aggregate in Fig. 2.4 as well as over time in Fig. 2.5.

The data source did not provide exact definitions for each of these prevention methods, but we provide general descriptions below.

- Informant: A confidential human source such as a former extremist now cooperating with law enforcement or an undercover law enforcement officer who provided intelligence about a particular individual and plot [10].
- Militant self-disclosed: The data sources [17] seems to utilize this code when the individual openly discussed his/her views in an open forum (through social media or some other periodical).

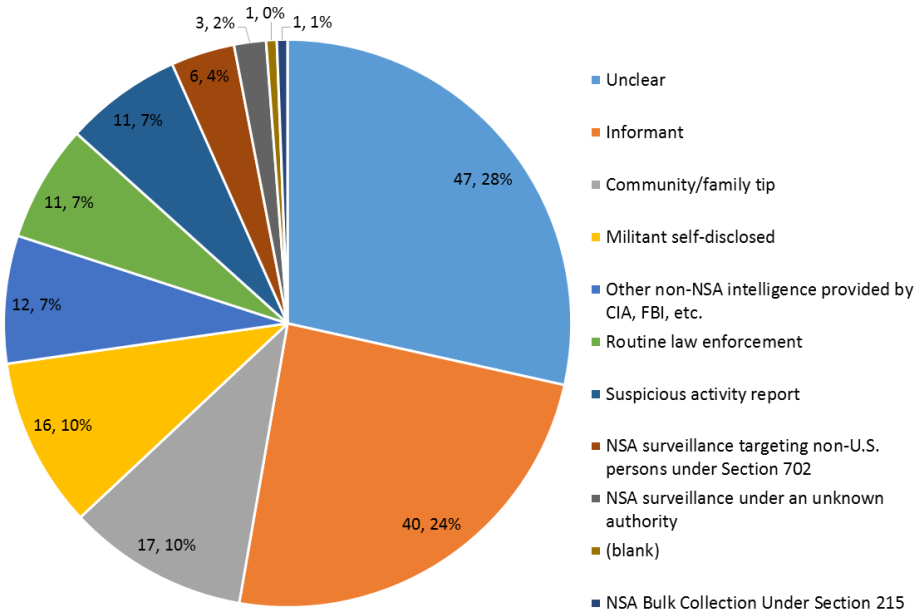


FIGURE 2.4. Method of prevention of U.S. terrorist plots from the jihadist ideology since 9/11. Data provided by [17].

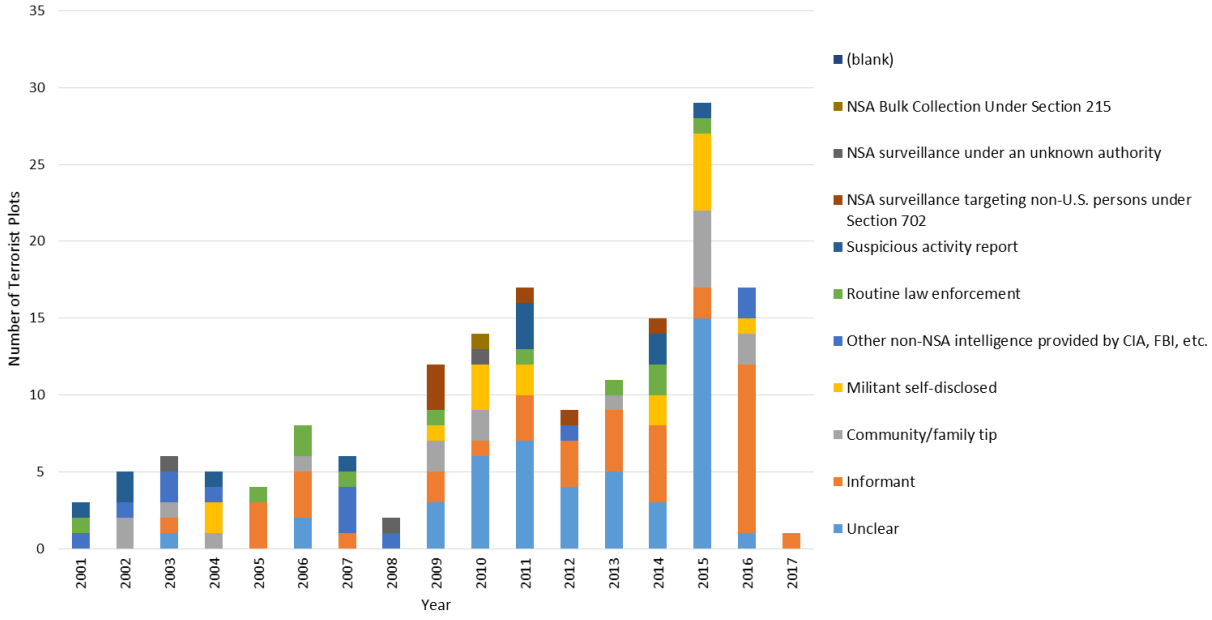


FIGURE 2.5. Method of prevention of U.S. terrorist plots from the jihadist ideology since 9/11 over time. Data provided by [17].

- Routine law enforcement: The data sources [17] seems to utilize this code when individuals are discovered primarily through the investigative work of local law enforcement and observing a suspicious social media posting.

- Suspicious activity report (SAR): A report of "observed behavior reasonably indicative of preoperational planning related to terrorism and other criminal activity" [222]. The report is usually made by local law enforcement and sent for further analysis to either the FBI or an intelligence fusion center.
- NSA surveillance targeting non-U.S. persons under Section 702 (of the Foreign Intelligence Surveillance Act- FISA): This is the targeted collection of communications of non-U.S. persons located abroad for foreign intelligence purposes. The revelation and use of the identities of U.S. persons who are part of those communications are only allowed "under narrowly defined circumstances" [257]. Based upon a 2013 NSA fact sheet, the circumstances are when "it is necessary to understand the intelligence or assess its importance, is evidence of a crime, or indicates a threat of death or serious bodily harm" [218].
- NSA bulk collection under Section 215 (of the PATRIOT Act): This is the telephone metadata program that allows the NSA to selectively query telephone carriers to reveal the date and time of the call, the calling and called numbers, and the duration of the call [136].

As mentioned earlier, the demand on law enforcement to prevent attacks has compelled them to increase the use of confidential human sources (informants). Since 9/11, this method accounted for about 24% of the total number of preventions (the most of any known prevention method) and is seen in the Fig. 2.5 as generally increasing since 2009. But as pointed out earlier, these methods are also resource-intensive [123] and therefore limited. It is also worth highlighting that the proportion of preventions as a result of community or family tips is only 10% since 9/11 and has occurred intermittently without a noticeable increase over

time. While more analysis most assuredly needs to be done, when this statistic is paired with another that others were aware of lone-actors terrorist planning in over 69% of Al-Qaeda related cases in U.S. and Europe from 1990-2012 [116, p. 431],[16] it calls into question the efficacy of the current emphasis on law enforcement engagement of local communities for increased information sharing.[17]

2.2.3. Role of social media and the widespread dissemination of propaganda. While there are a variety of socio-economic issues that may be factors in the radicalization process, the prevalence of jihadist themes and messages on social media most recently promulgated by the Islamic State (commonly known as ISIS) serves as a significant driver by inspiring recruits, garnering support, and provoking homegrown attacks or foreign fighter activities. ISIS has shown its preference to disseminate propaganda and messages on Twitter because of its low barrier to entry and the high message reach [166].

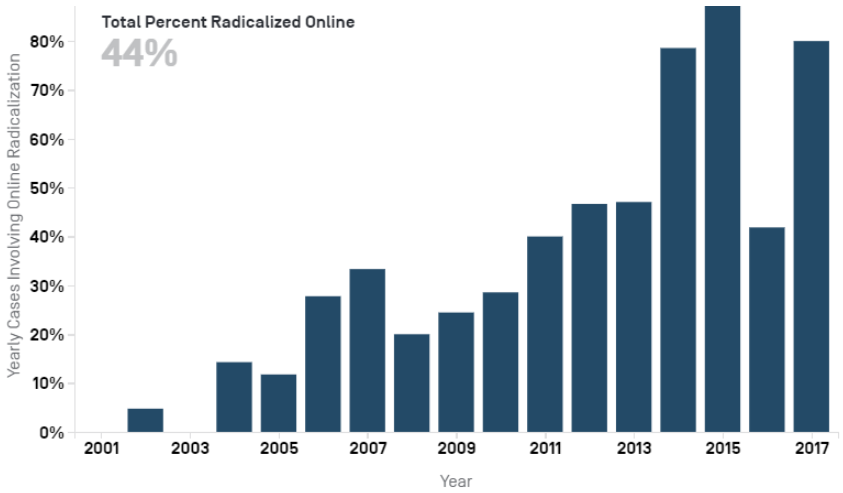


Figure 2.6. Percentage by year since 9/11 of violent extremists inspired by the jihadist ideology who "maintained a social media profile with jihadist material or utilized encryption for plotting" [17]. Source: [17].

[16]See Section 2.4.3 for more details.

[17]See Section 2.3.1 for more details.

Fig. 2.6 depicts the percentage of violent extremists who had some form of online or digital presence, which [17] defined as those who "maintained a social media profile with jihadist material or utilized encryption for plotting." It is clear that among the perpetrators of violent radicalization, this behavior has become more prevalent.[18]

During testimony before the Senate in 2016, the former FBI Director observed,

> [W]e are confronting an explosion of terrorist propaganda and training available via the Internet and social networking media. Due to online recruitment and indoctrination, foreign terrorist organizations are no longer dependent on finding ways to get terrorist operatives into the U.S. to recruit and carry out acts. Terrorists in ungoverned spacesboth physical and cyberreadily disseminate poisoned propaganda and training materials to attract easily influenced individuals around the world to their cause. They encourage these individuals to travel, but if they cannot travel, they motivate them to act at home. This is a significant change and transformation from the terrorist threat our nation faced a decade ago [57].

An in-depth treatment of social media radicalization and recruitment is found in [187]. The specific usage of Twitter and other social media by ISIS is described in [19], while an analysis of the social media networks of foreign fighters in Iraq and Syria is available in [166].

Other social media services such as Telegram and WhatsApp are being used for even more nefarious purposes because they proffer end-to-end encryption of communications and claim that it is impossible to give access to the data even with a court warrant. This is what the FBI Director calls, "going dark," and is a tactic that is increasingly being used by violent extremists especially in the near-term execution of their attacks [58]. See the 2015 Paris terrorist attack [245] and 2017 UK Parliament attack [70] as case studies.

[18]We note that [134] seemed to contradict this by determining that only 12.6% of 183 individuals convicted for terrorist offenses in the U.S. from 1995-2012 had viewed videos or websites dedicated to online extremist material. However, the study admitted that their estimate was likely more conservative due to data limitations (i.e., they principally used as sources indictments and sufficient quantities of open source reporting, neither of which may have contained details of this non-illegal behavior.)

2.2.4. Trends in known associates and co-conspirators. As discussed previously, recent cases of violent extremism have demonstrated that perpetrators can operate in a conspiracy to commit terrorist acts. As shown in Fig. 2.7, nearly a third of all plots in the U.S. since 9/11 involved more than 1 individual, meaning that attacks were carried out by more than 1 person, or that authorities were able to bring charges to associates for providing material support or a conspiring to commit terrorist acts. Fig. 2.8 is a histogram of the total number of plots involving a varying number of individuals. For example, the Boston Marathon bombing by the Tsarnaev brothers is counted in the '2' bar, whereas the San Bernardino shooting involving Farook, Malik, and Marquez is counted in the '3' bar. The point is that a significant portion of terrorist plots involve associates, and therefore it might be important to explore how clusters of individuals could be exhibiting indicators for planned violence.

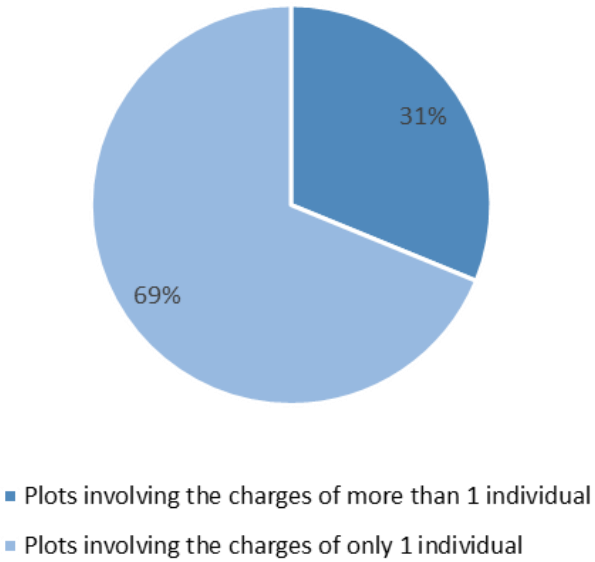


Figure 2.7. Percentage of U.S. terrorist plots since 9/11 involving the charges of more than one individual. Data source: [17].

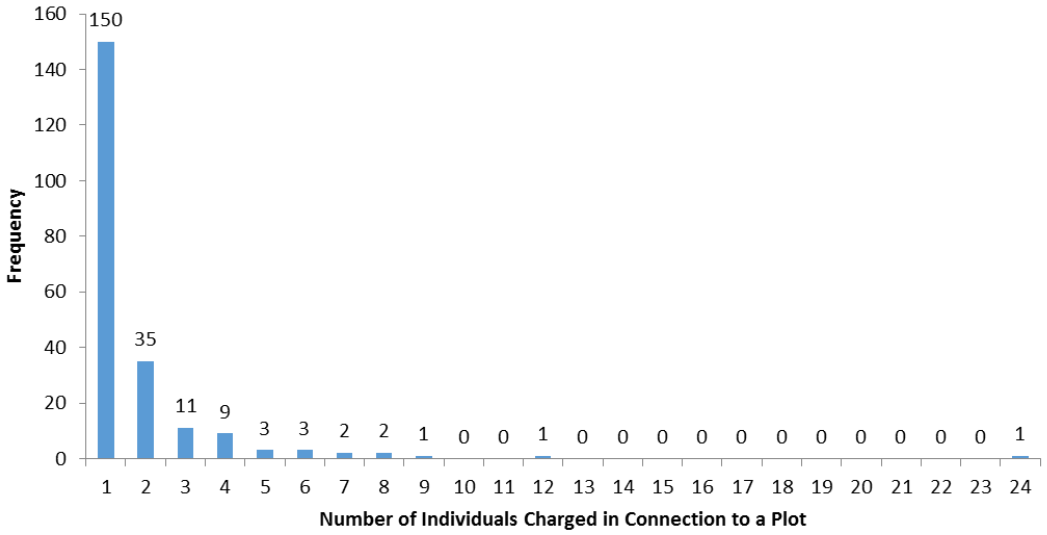


FIGURE 2.8. Histogram of the number of individuals charged in U.S. terrorist plots from the jihadist ideology since 9/11. Data source: [17].

There is further support for the prevalence and relevance of small groups of perpetrators among violent extremist attacks [116, 292]. Researchers in [116] sought the inclusion of "isolated dyads," which they defined as "pairs of individuals who operate independently of a [terrorist or extremist]group" and "may become radicalized to violence on their own (or one may have radicalized the other), and they conceive, develop, and carry out activities without direct input from a wider network" [116, p. 426]. According to the same researchers, isolated dyads constitute around 27% of the 119 'lone-actor terrorists' from the United States and Europe from 1990-2012 (which include those motivated by right-wing and Ismalist ideologies, as well animal, environmental, and anti-abortion issues) [116].

In summary, this section quantified the threat from violent extremists motivated by Salafi-jihadism in the U.S. as well as investigated the empirical support for various characteristics of their radicalization (namely, the online/social media component and the role of associates and co-conspirators).

2.3. Strategies to Counter the Threat from Violent Extremists

Given the scale and multi-faceted aspects of the threat posed by violent extremists and violent extremism, many recognize that comprehensive strategies must be developed and implemented to adequately counter the threat. Particularly useful in each of the strategies is also an inherent position on the gaps or shortfalls present in current effort. In this section, we cover the current U.S. strategy to counter threats from violent extremists, as well as elements of the Center for Strategic and International Studies (CSIS) multi-national strategy to countering violent extremism and the U.S. House of Representatives Homeland Security Committee's strategy against Islamic Terror.

Figure 2.9. Graphic from the GAO Report to Congressional Inquiries depicting how countering violent extremism is different from counterterrorism. Source: [122, p. 7].

2.3.1. Current U.S. strategy. While a legacy from the previous Administration, the current U.S. strategy to counter the threat from violent extremists can be viewed as bifurcated into two efforts: countering violent extremism (CVE) and counterterrorism (CT).

The former is defined as "proactive actions to counter efforts by extremists to recruit, radicalize, and mobilize followers to violence" [227, p. 2], while the latter, lacking an available explicit definition, can by aptly summarized as actions to prevent terrorist attacks and "to disrupt, dismantle, and eventually defeat al-Qa'ida and its affiliates and adherents" [226]. The Government Accounting Office (GAO) most recently released a report assessing the implementation of the Federal Government's strategy to counter violent extremism [122] and included a graphic to depict the distinction between CVE and CT. See Fig. 2.9.

The U.S. Government's strategy to counter violent extremism, as expressed in the 2016 Strategic Implementation Plan for Empowering of Local Partners to Prevent Violent Extremism in the United States, fundamentally seeks to "address the conditions and reduce the factors that most likely contribute to recruitment and radicalization by violent extremists" [227, p. 2]. Operating on the assumption that "strong and resilient local communities are the most effective means of safeguarding individuals in the United States against violent extremist recruitment and radicalization," the strategy is focused entirely on engaging and supporting local communities and stakeholders. Where law enforcement is addressed, it is limited to expanding the use of community policing strategies to "build trust, mutual respect, and collaboration between police and the communities they serve" [227, p. 9], while also limiting CVE efforts from including the "gathering intelligence or performing investigations for the purpose of criminal prosecution" [227, p. 2]. The Department of Homeland Security's Strategy to Counter Violent Extremism echoed this prohibition as one of its guiding principles: "intelligence and law enforcement investigations are not part of CVE activities" [74,

p. 2]. Part of this may be attributed to a lesson learned from the UK Prevent counter-radicalization program, which was accused of targeting Muslim communities and ultimately lost the trust of many British Muslims [24].

However, while the Department of Homeland Security and the Department of Justice jointly lead the federal efforts on countering violent extremism [122], each also has their own investigative arms (Homeland Security Investigations and the Federal Bureau of Investigations, respectively). The possibility of a conflict of interests has already been addressed in [291].[19] The problem is exacerbated at the State, Local, and Tribal law enforcement levels where many organizations do not necessarily have the personnel or resources to keep separate responsibilities between CVE and normal investigations [267].

Whether such agencies can truly separate their CVE efforts and not involve any aspect of their investigative authorities has yet to be established. However, recent research states that this prescription is only part of the set of broader recommendations to help improve effective community policing to counter violent extremism [267]. Overall, the recommendations are focused on a genuine commitment by law enforcement leaders for deeper community engagement, the quality and expanse of law enforcement engagement efforts with the community, training for officers in "outreach techniques and cultural competency," and "finding ways to

[19]Dr. Southers, the director of Homegrown Violent Extremism Studies at the University of Souther California Sol Price School of Public Policy, wrote in an Op-Ed in the *Los Angeles Times*, "There was a fundamental error in charging U.S. attorneys with managing the CVE pilots. A successful program cannot be run by the same arm of government that prosecutes terrorism cases. That demands an unrealistic level of trust on the part of the community. Why would someone participate in such an initiative if they fear their questions, comments and concerns could lead to an FBI agent knocking on their door?" [291].

divert individuals away from the criminal justice system when possible by providing them the resources and assistance they need" [267, p. 6].[20]

Beyond the U.S. White House strategy to deal with the threat from violent extremism, we deem it important to discuss two more because of their integrated approach to address all aspects of the problem and their prescriptions for the integration of both hard and soft power.

2.3.2. CSIS STRATEGY FOR COUNTERING VIOLENT EXTREMISM. In 2015 the Center for Strategic and International Studies (CSIS) established the Commission on Countering Violent Extremism to assess the worldwide problem and provide specific recommendations to the next U.S. administration and its governmental and non-governmental partners on "dimish[ing] the appeal of extremist ideologies and narratives" [126]. The comprehensive strategy represented the consensus of numerous public- and private-sector leaders from academia, civil society, the faith community, and technology companies. This report outlined the following eight components:

(1) Strengthening resistance to extremist ideologies.

(2) Investing in community-led prevention.

(3) Saturating the global marketplace of ideas.

(4) Aligning policies and values.

(5) Deploying military and law enforcement tools.

(6) Exerting White House leadership.

[20]Schanzer acknowledges that many police departments "have been interacting with their Muslim American constituents for years or are forging ahead with substantial efforts to build relationships of trust with Muslim American communities [267, p. 13]. For example, following the release of this CVE strategy, the FBI described their own CVE strategy implementation as involving the positive engagement of local communities in Minnesota leading to law enforcement's ability to "charge - locally or at the federal level - some really bad actors and recruiters of young people for nefarious purposes" [107].

(7) Expanding countering violent extremism (CVE) models.

(8) Surging funding [126].

This strategy to counter violent extremism was comprehensive and included efforts to support and bolster communities to resist the extremist ideologies and the need to employ additional law enforcement tools as well. The CSIS Commission agreed that CVE and CT must be separated in terms of "tactics, agencies, and actors involved" but also stated that "effective strategy will require soft and hard power operating at scale an in tandem" [126, p. 49]. For instance, the Commission stressed the need for "codified protocols for referrals" where law enforcement can recommend cases for NGO/community off-ramps, and where communities can refer cases to law enforcement [126, p. 41].

We stake that our proposal of a radicalization detection system is part of the 'developing military and law enforcement tools' component in this overarching strategy. We also envision that our system could be helpful in the assessment efforts of non-governmental organizations working in CVE [165]. These organizations are actively trying to identify those radicalizing individuals short of any criminal activity and steer them to suitable "off ramps."

2.3.3. U.S. House of Representatives strategy. Another strategy document worth mentioning is the U.S. House of Representatives Homeland Security Committee's "A National Strategy to Win the War Against Islamist Terror" [197]. Citing the obsolescence of existing U.S. strategies, the Homeland Security Committee developed this strategy in consultation with other national security experts. Their prescriptions included:

(1) Thwart attacks and protect our communities.

(2) Stop recruitment and radicalization at home.

(3) Keep terrorists out of America.

(4) Take the fight to the enemy.

(5) Combat terrorist travel and cut off financial resources.

(6) Deny jihadists access to weapons of mass destruction.

(7) Block terrorists from returning to the battlefield.

(8) Prevent the emergence of new networks and safe havens.

(9) Win the battle of ideas [197].

Our work also addresses several aspects of this House Committee's strategy. For instance, to thwart attacks the committee recommended 1) "robust, real-time information sharing" without the loss of even a "data-point", and 2) that "social media should be better incorporated into investigations as well as routine criminal screening and other background checks in order to identify suspects who have openly broadcasted their support of foreign terrorist organizations" [197, p. 9]. In the category of stopping recruitment and radicalization, the Committee recommended expanding confidential tip line for citizens within communities to report not just suspicious activity but other concerns related to "possible terrorist radicalization" in their neighborhoods. Our vision for a common, fused graph database of the on- and off-line indicators and tips to law enforcement about those on the radicalization pathways would be a major step towards each of these prescriptions.

2.4. Radicalization Research

In this section, we provide a brief survey of the large body of radicalization research within the last 15 years. We first discuss two opposing views of the population of individuals undergoing radicalization. Then we provide an overview of the main radicalization models,

frameworks, and early warning behavior research currently available. Ultimately, our proposal for a radicalization detection system relies on the advancements on this still expanding literature.

2.4.1. Radicalization of beliefs versus actions. In order to understand a particular radicalization model, we identified two main conceptualizations of the radicalization process. Much of the research until recently was based on the notion that radicalization is a social-psychological process that individuals underwent toward increasingly threatening stages that did not seem to distinguish beliefs, views, or opinions from behaviors. See Fig. 2.10, which visually depicts the population along the spectrum from the vulnerable to the imprisoned. Some critical of this conceptualization have characterized it as a "conveyer belt" [200, p. 211], or worse as a "religious conveyer belt" [241, p. 3] implying the pathway wrongfully rests on Islamic stereotypes.

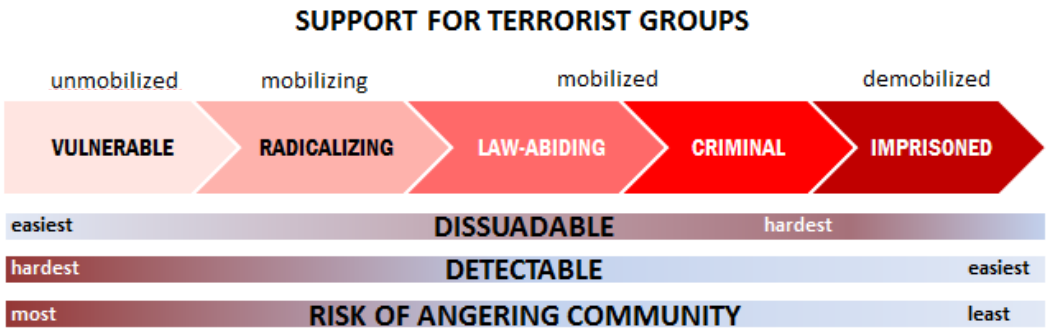


Figure 2.10. The various populations of support for terrorist groups as conceptualized by Berger [2012]. We take the support for terrorist groups, especially with the distinctions of "law-abiding" and "criminal" to be synonymous with violent extremist radicalization. Source: [18].

More recently, researchers beginning with Bartlett and Miller [12] began distinguishing those who held radical views versus those who turned to violence while holding those views. McCauley and Moskalenko conceptualized this in [199] by proposing a two pyramid model

that separates radicalization of opinion and radicalization of action. See Fig. 2.11. The key insight is that violent radicalization (towards actions) is a distinct process from non-violent radicalization (of opinions and beliefs) [200, 264, 266] and can explain how many people may hold radical beliefs or opinions, but only a few relatively will undertake radical actions.

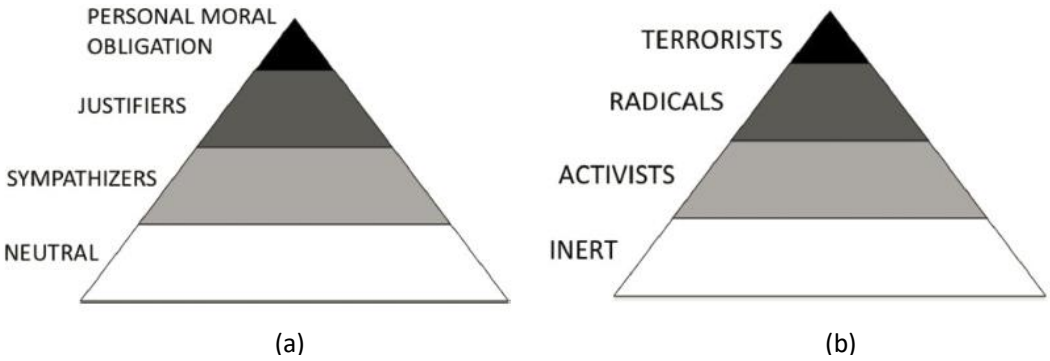


FIGURE 2.11. The two pyramid model of radicalization. (a) is the opinion radicalization pyramid and (b) is the action radicalization pyramid. Source: [199]

Researchers in [199, 200] clearly advocate for the two pyramid model to support the statistical observation that radicalization of opinion rarely leads to radicalized action. On the other side of the debate are those who see beliefs and actions as inextricably interwoven in radicalization conceptualizations. There are even others like Dean in [67] who focus on the assessment of beliefs and attitudes as a critical step in conducting a threat risk assessment.

Regardless of who is right, we think that our approach, more than anything, calls for the fusion of the ideology identification with the behavioral detection in order to come up with a more informed risk assessment of extremist violence.

McCauley and Moskalenko also state that "means and opportunity" are important in evaluating the risk of engaging in terrorist action, again as a way to explain why so many people with radical views never commit radical actions [199, p. 4]. But our approach of utilizing a heterogeneous graph database, in fact, seeks to find the connections between

individuals, actions, and circumstances that would reveal such means and opportunities to an analyst. For example, our data model would capture the activity of support networks who give people the last push to participate in some action, or the new social influence or a suspicious travel behavior that may have opened doors and social networks for individuals. It may even capture the existence of alienation or desperateness (loss of family, divorce, threat of deportation, etc) through text analysis of social media posts or court filings.

Lastly, of particular interest is that McCauley and Moskalenko recently posited "at least five trajectories of radicalization to terrorist action:"

(1) Lone wolf (an individual undertakes "political violence alone without group or organizational support")

(2) Foreign fighters (an individual undertakes violence "by joining an already violent group")

(3) Suicide bomber (an individual undertakes violence "by volunteering as a suicide bomber for an already violent group")

(4) Small, isolated group terrorist plot (akin to Sageman's "bunch of guys" [263])

(5) "Small group within a larger activist movement" turning to "violence as part of intergroups competition" [200]

The importance of distinguishing and specifying what behavior one is tracking the radicalization is also discussed in [134, 214, 266].[21] Our methodology, although presently limited in empirical testing of a generic extremist violence trajectory, could also potentially handle separate graph patterns for each of the other different types of trajectories.

[21]For example, Monahan stated that risk assessments need to be clear on what they are assessing: "the risk of terrorism in the aggregate, or of specific types of terrorism, or of specific phases in the process of becoming a terrorist, or of specific roles in terrorist activity" [214, p. 167].

2.4.2. Radicalization mechanisms. There are several works which focus on understanding the various mechanism of radicalization, meaning those factors which might cause someone to increasingly participate in violent extremism. For instance, [198] provides an overview of the individual and group mechanisms (psychological and social) that would cause an individual to radicalize. This included a list seven individual-level mechanisms: personal grievance, group grievance, slippery slope of small increments in action, love for someone in a militant group, escape from a situation more risky than terrorism, thrill and status seeking, and seeking new friends after losing social connections (unfreezing) [199, p. 3]. Another example is [302], which provides a root cause model of Islamist Radicalization. We refer the reader to these works but do not provide additional details in this thesis because we are less focused on explanatory models for radicalization than the detection of the observable behavioral indicators when one is undergoing violent radicalization.

2.4.3. Antecedent and early warning behavior research. Beyond just understanding what brings someone further along the radicalization process, there are quite a number works which focus on early warning behaviors. Gill in [116] conducted an in depth study the statistically significant characteristics and early warning behaviors that are present among 119 lone-actors terrorists who had engaged in or planned to engage in violence within the U.S. and Europe and were either convicted of their offenses or had died as a result of their offenses. The important research better informs both CVE organizations focused on identifying and off-ramping individuals on radicalization paths, as well as law enforcement organizations seeking to prevent future attacks. See Fig. 2.12 for these features in Al-Qaeda related lone-actors (radical extremists inspired by Salafi-jihadist ideology) as compared to other lone-actors.

	Right Wing (n = 40)	Single Issue (n = 21)	Al-Qaeda Related (n = 52)
Town size <20,000	37.5%***	28.6%	9.6%***
University experience	15%***	52.4%	50%**
Worked in construction	12.5%***	0%	0%**
Worked as a professional	2.5%*	14.3%	11.5%
Student at time of event	2.5%*	4.8%	17.3%**
Unemployed	50%*	38.1%	30.8%
Verbal statements to friends/family about intent or beliefs	52.5%**	71.4%	71.2%
Religious convert	2.5%***	19%	36.5%***
Sought legitimization	7.5%**	9.5%	28.8%***
Lived away from home when ideology adopted	15%**	19%	38.5%***
Others helped procure weaponry	10%***	33.3%	32.7%*
Engaged in dry runs	17.5%**	47.6%**	30.8%
Recently joined a wider group/movement	47.5%**	38.1%	23.1%**
Evidence of command and control links	5%**	4.8%	30.8%***
Based in the United States	52.5%	71.4%***	28.8%***
In a relationship	20%	52.4%***	21.2%
Previous criminal conviction	50%	61.9%**	26.9%***
Previously imprisoned	27.5%	47.6%**	19.2%*
Provided a pre-event warning	17.5%	38.1%*	21.2%
Spouse/partner part of a wider movement	5%	19%**	3.8%
Learned through virtual sources	37.5%	19%***	65.4%***
History of mental illness	30%	52.4%**	25%
Others aware of individual's planning	52.5%	38.1%**	69.2%**
Children	15%**	42.9%*	28.8%
University degree	5%	4.8%	17.3%**
Average age	36.3 years	36.8 years	26.7 years***
Successful execution of terrorist attack	57.5%	66.7%	40.4%**

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

FIGURE 2.12. Table comparing the characteristics and antecedent behaviors of right-wing, single-issue, and Al Qaeda-related lone-actors. Statistical analysis of 119 individuals who were convicted or dies in the commission of their crimes in the United States and Europe from 1990-2012. Source: [116, p. 431].

We include these tables for reference specifically because of the antecedent behaviors and not for the personal characteristics that would profile individuals. In particular, the statistically significant behaviors which may help indicate radicalization include:

- More likely to seek legitimization from "religious, political, social, or civic leaders prior to their terrorist event or plot" [116, p. 431].
- More likely to have others help procure weaponry.
- More likely to have command and control links with a terrorist group, which means that the individual was "trained and equipped by a group- which may also choose their targets" but "attempt[ed] to carry out their attacks autonomously" [116, p. 431].
- Less likely to have a previous criminal conviction or been previously imprisoned.

- More likely to have learned through virtual (online) sources.
- More likely that others are aware of individual's planning.

Furthermore, these researchers also provided analysis of the network-related antecedent behaviors that provide the supporting motivation for our graph-based approach. See Fig. 2.13. Even for those individuals without command and control links, there are many statistically significant pre-attack behaviors such as the consumption of propaganda and learning to conducting various aspects of the terrorist plot through virtual sources.

	Individuals Without Command and Control Links (n = 87) (%)	Individuals With Command and Control Links (n = 21) (%)	Isolated Dyads (n = 11) (%)
Based in the United States	55.2***	4.8***	36.4
Previous military experience	31***	4.8**	9.1
Previous criminal conviction	47.1**	19**	36.4
Held a PhD	2.3	0	18.2**
Lived alone	40.2	38.1	9.1**
Lived away from home when ideology adopted	23	42.9*	27.3
Received training	20.7	33.3	0*
Learnt through virtual sources	40.2***	66.7*	72.7*
History of mental illness	35.6**	19	9.1
Socially isolated	57.5*	33.3*	45.5
Recently joined a wider group/movement	27.6**	47.6	45.5
Noticeable increase in religiosity	23***	61.9***	27.3
Family/close associates involved in political violence/crime	27.6***	57.1**	63.6**
Interacted face-to-face with wider network	39.1***	61.9	90.3***
Interacted virtually with wider network	28.7***	57.1**	63.6*
Others helped procure weaponry	17.2***	38.1*	45.5*
Others helped build IED	6.9***	33.3***	27.3
Others aware of individual's planning	42.5***	100***	100***
Attempted to recruit others	27.6**	33.3	81.8***
Consumed propaganda from a wider movement	65.5*	85.7*	72.7
Al-Qaeda related	33.3***	76.2***	63.6
Single issue	23**	4.8*	0
Right wing	39.1**	9.5**	36.4
Successfully executed an attack	57.5**	33.3*	27.3

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

FIGURE 2.13. Table comparing the characteristics and antecedent, network-related behaviors of individuals who had or did not have command and control links as well as isolated dyads. Source: [116, p. 432].

The above results counter the common notion that homegrown violent extremist attacks are "virtually undetectable" [116] due to their social isolation [10, p. 1652], and rather suggest that "many lone-actor terrorists regularly interact with wider pressure groups and movements either face-to-face or virtually" and lends credence to the use of "traditional

counterterrorism measures (such as counterintelligence, HUMINT, interception of communications, surveillance of persons, etc)" in the detection of those on pathways towards violence [116, p. 434]. In our work, we also intend to exploit these network connections in order to conduct risk assessments on those who may be on pathways towards violent extremism.

Another related study is a proposal for the typology of early warning behaviors found in various forms of targeted violence by Meloy [203]. These eight early warning behaviors (not just antecedent behaviors) shown in Fig. 2.14 were derived by extensive empirical study by a forensic psychologist who consults for the FBI. These indicators were intended to be

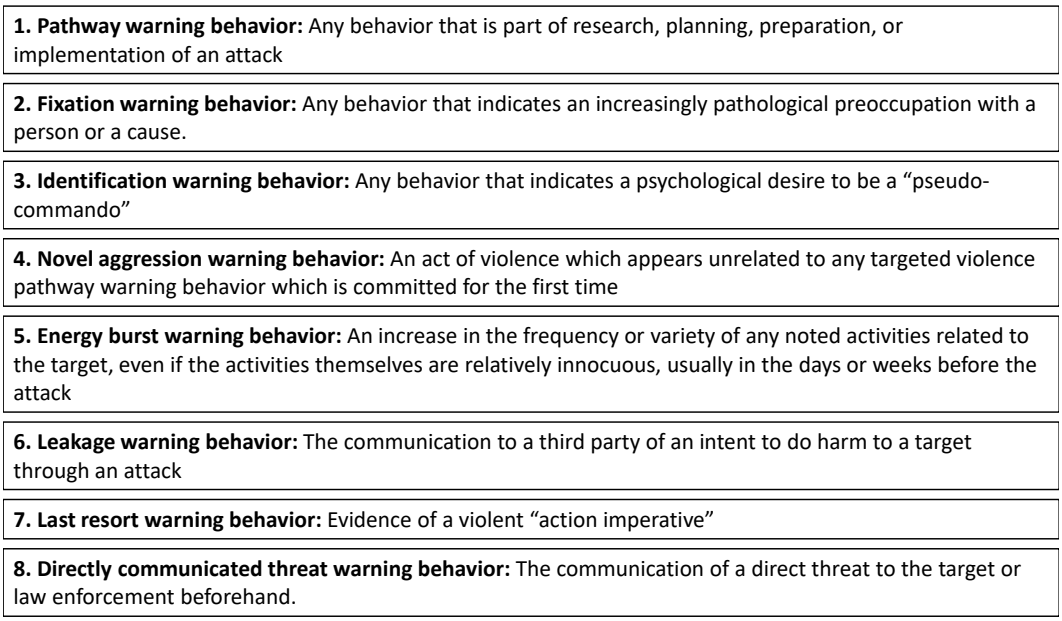


FIGURE 2.14. The eight warning behaviors of targeted violence proposed Meloy [2012]. Graphic created from content in [203].

determined through a threat risk assessment by a trained psychologist taking into account the totality of the available verbal communications and behaviors observed to subjectively determine the threat. However, we note that at least two of them— leakage and directly communicated threat warning behaviors— might be directly available to law enforcement

prior to an attack through confidential source reports or public social media posts. Additionally, under the Electronic Communications Protection Act, a law enforcement entities may be able to get access to very specific electronic activities (email content, internet searches, etc) with a subpoena, warrant and other court order [34, 301] that might corroborate these behaviors as well as reveal more about pathway warning behaviors.[22] Such requests, however, face tremendous scrutiny and often require demonstration of probable cause that a crime has already been committed or justification of emergency circumstances to save lives. See for example Google's explanation of the law enforcement requirements in [124].

Another work worth mentioning due to its unique methodology is the work by Bartlett and Miller in [12], who set out to find out what characteristics of violent radicals were distinct from non-violent radicals (as a means to avoid selecting only on the dependent variable). Through extensive interviews with non-violent radicals and convicted violent radicals, the researchers provided empirical evidence for the distinguishing indicators found only in violent ones and dispelled generalized indicators common to both populations. For instance, the researchers pointed out that both groups are familiar and agreed that the term and concept of "kafir" (non-believer) can describe non-Muslims, but only the violent radicals use the term as a means to "dehumanize non-Muslims and Muslims who disagree with their views" [12, p. 10]. They also found that while many violent and non-violent radicals viewed violent films, it was mostly violent radicals who would watch them in groups. Also, only violent radicals distributed videos about jihad, engaged in debates between "do-ers" and "talkers." [12, p. 17]. This work has tremendously important implications in the development of more discerning indicators of violent radicals. However, as part of a radicalization detection

[22]Researchers in [118] describe empirical digital evidence for past lone actors who signaled attacks, selected targets, and conducted other pre-attack planning and preparation through internet-related activity.

system, it would very valuable to detect the presence of these concepts automatically through machine learning and natural language processing.

Lastly, in this subsection, we highlight the work by Schuurman and Eijkman [271], a relevant paper which attempts to identify the seven distinct (possibly concurrent) phases of terrorist preparation for an attack. The study is based upon seven cases of homegrown jihadism in Western Europe from 2004 to 2007. The researchers essentially examined the empirical evidence to formulate a pre-attack process framework. Particularly useful was the detailed behavior types provided for each of the phases and threat stages, and the potential basis the collection served as a future risk indicator typology.

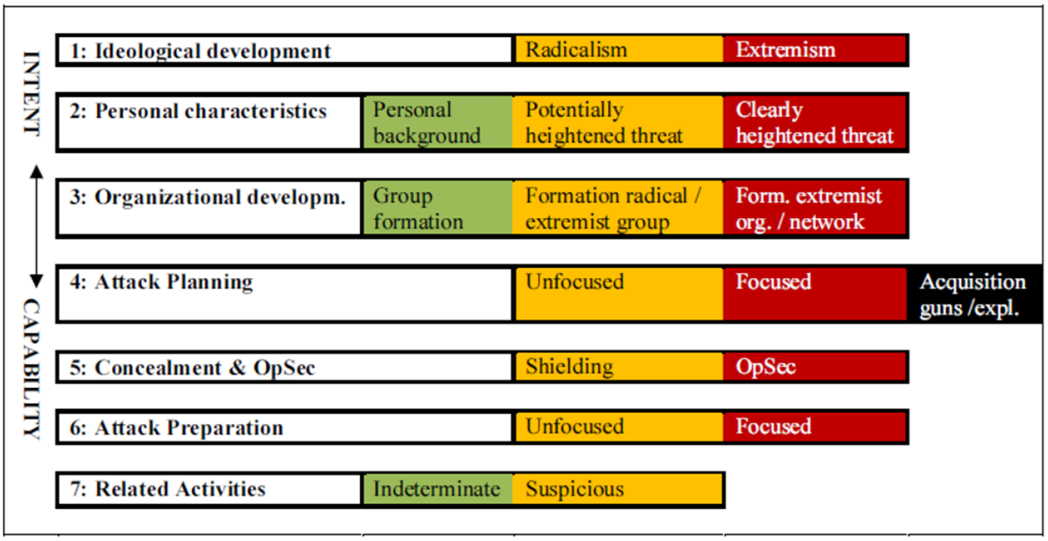


FIGURE 2.15. The conceptual framework of pre-attack activities of terrorist proposed by Schuurman and Eijkman [2015]. Source: [271].

2.4.4. STAGE-BASED RADICALIZATION MODELS. There is quite a number of conceptual, phase-based models for violent radicalization proposed in the literature that provide indicative behaviors or psychological states of individuals along the process. See [31, 86, 164] for some thorough surveys and Fig. 2.16, which shows some detail on the five commonly cited models to give the reader a sense for the types of the progression forwards terrorist

violence. According to [164], linear means that the progression must occur in ordered stages, while Sageman's non-linear model attempts to describe the four factors which combine (not necessarily in order) to motivate extremist violence. Later in this subsection, we intend to discuss the radicalization model proposed by Klausen [165, 167], which is based in part on the NYPD's Silber and Bhatt model.

Author	Type of model	Stages or factors
Borum 2003	Linear, progressive	1. Social and economic deprivation 2. Inequality and resentment 3. Blame and attribution 4. Stereotyping and demonizing the enemy
Wiktorowicz 2004	Linear and emergent	1. Cognitive opening 2. Religious seeking 3. Frame alignment 4. Socialization
Moghaddam 2005–2006[9]	Linear, progressive	1. Psychological interpretation of material conditions 2. Perceived options to fight unfair treatment 3. Displacement of aggression 4. Moral engagement 5. Solidification of categorical thinking 6. The terrorist act
NYPD (Silber & Bhatt) 2007	Linear	1. Pre-radicalization 2. Self-identification 3. Indoctrination 4. Jihadization
Sageman 2008	Non-linear, emergent	1. Sense of moral outrage 2. Frame used to interpret the world 3. Resonance with personal experience 4. Mobilization through networks

FIGURE 2.16. The five stage or factor-based radicalization models analyzed in King [2011] and a summary of the various stages or factors. Source: [164].

The phase-based models that emerged between 2003 and 2008 faced two main criticisms. The first was that they are subject to selection bias because they are derived from "successful" cases of radicalization to violent extremism [264, 302]. The second critique was that they

utilize "vague" or "general" traits which could easily apply to others who are not radicalizing at all and has the potential for causing discrimination [241, 302].

We have a two-part response to the first critique. First, as mentioned previously, researchers beginning with Bartlett and Miller [11, 12] have begun to empirically distinguish those who held radical views versus those who turned to violence while holding those views. Admittedly, this research approach is still maturing, but the original Bartlett findings were significant in and of themselves. Additionally, without excusing the need for discerning indicators, our approach and the approach of many other researchers and practitioners is to screen for those suspicious behaviors that may suggest someone is on the pathway to radicalization. We are not trying to statistically predict those who will commit violent acts.

In response to the second critique, more recent work including that of Schuurman and Eijkman [271] and Klausen [165, 168] propose radicalization models and frameworks that are much more descriptive and include specific behaviors to describe the radicalization process. Notably, the Klausen radicalization model, which we will cover next, contains many of Bartlett's discerning characteristics between radicals and terrorists (violent radicals). Additionally, Klausen's behavior-based approach, which we adopt in our research, is in contrast to ones that are reliant on either physical appearance or socio-demographic profiles, and have been dismissed as lacking statistical basis [116] or as potentially misleading [223].

2.4.4.1. *Klausen dynamic radicalization model.* In this work, we specifically seek the advancement and eventual operationalization of the Klausen's novel dynamic radicalization model and its associated behavioral indicators found in [165, 168]. In a project funded by the National Institute of Justice, Klausen and her research team undertook a multi-year study to empirically test their dynamic model for radicalization based upon the NYPD Silber and

Stage:	Pre-Radicalization	Stage 1: Detachment	Stage 2: Peer-Immersion and Training	Stage 3: Planning and Execution of Violent Action
Description:	Searching behavior indicative of cognitive opening.	Detachment from previous life; e.g. by spending inordinate amounts of time with online extremist peers.	Leaves home to become closer to a peer group of like-minded individuals.	Attempts or enacts violent action—or joins a terrorist group abroad or attempts to join a group.
This could include:	Expressions of disillusionment with world affairs or with religious or political authorities.	Actively seeking to get closer to new authority figures, or engaging in *Da'wah* online or to proselytize in public.	Attempting to go abroad to join an organization or a network to "live" as prescribed by the ideology.	Actively supporting another person carrying out violent action on behalf of the ideology.
	Behavior indicative of a personal crisis in response to personal events, e.g. a family crisis, drug addiction, or being arrested.	Experiencing a revelation or making changes to lifestyle such as dropping out of school or work.	Behavior indicative of a desire to permanently join the militant community, e.g. by finding a spouse through the extremist community.	Issuing threats online or real-life, or in other ways supporting immediate violent action, e.g. by engaging in online fraud.
	Seeking out information in venues outside the individuals' established social milieu, either online or real-life, from new authority figures.	Picking fights with local mosque or teachers, colleagues, and family—or otherwise trying to convince others to change, e.g. by starting a blog or a website.	Seeking out ways to demonstrate commitment to the new ideological community and its mission, e.g. by acquiring practical training in the use of firearms or other skills considered important to the mission of the extremist community.	Joining a foreign terrorist organization or taking practical steps to carry out an attack, e.g. by acquiring materials needed to fabricate a bomb or purchasing firearms.

FIGURE 2.17. Klausen's Dynamic Risk Assessment Model showing the behavioral indicators of state progression in radicalization trajectories [167].

Bhatt model discussed previously [285]. The model contains four stages: Pre-Radicalization, Detachment, Peer-Immersion and Training, and Planning and Execution of Violent Action. The description and possible indicators associated with each stage are shown in Fig 2.17.

The researchers then compiled a dataset that contained actual or inferred dates of the behavioral indicators of 135 US Al Qaeda-inspired violent extremists who committed offenses between 2001 and 2015. This dataset [167] was constructed entirely from publicly available sources (court documents and investigations into their activities conducted by the United States government and news media, which may have included online communications posted by the terrorist offenders). Appendix C contains the codebook they used in this process to determine if the indicators were present in each individual's detailed history.

The researchers then empirically assessed a dynamic model for radicalization, found a large number of behavioral cues that occurred in the right sequence in a high percentage of cases, and arrived at a predictable structure/template for the radicalization process. They also made some important findings on the duration of radicalization trajectories. This study was a significant advancement in the field, especially for the accessible dataset based on publicly available material and the important conclusion that radicalization based on Salafi-jihadist ideology did often follow an empirically-assessed structure or template. Moreover, its findings on radicalization durations should be very informative to both practitioners and policy-makers alike, and may even signal necessary changes to policies and procedures on how long investigations should stay open and how long any form of surveillance is authorized to continue [168].

It is important to emphasize that these conclusions were limited to violent extremists motivated by the Salafi-jihadist belief system. Klausen wrote:

> The question is often raised whether radicalization to violent extremism is perhaps not "the same" across ideologies, and the essential factor here is simply some pathology of extremism. This may be true in so far as networks and detachment from ordinary life are the essential elements of the recruitment to cults and sects - or gangs. But if the process may appear to be similar, the drivers are different. At this level of abstraction little can be learned that is of practical use for crafting intervention programs addressing homegrown terrorism. There are common features, such self-alienation from family and productive engagement with institutions of education or employment. However, other crucial features of behavior vary because of the different action scripts for followers advocated by the ideologies [169].

The U.S. government's broad definition of violent extremism and its umbrella strategy of countering violent extremism based upon all ideologies and motivations[23] had fostered generalizable radicalization research detached from specifics to any one ideology. Klausen's

[23]See Section 2.3.1 for discussion of the U.S. government strategy.

conclusion at least suggests the most effective detection and subsequent intervention programs would depend on knowledge of ideology-specific indicators.

Later in Chapter 8, we utilize the Klausen dataset to model radicalization as a discrete dynamical process in order to find more discerning patterns of behavioral indicators for those on path trajectories towards extremist violence.

2.4.5. Case Studies of Violent Extremists. In Appendix B, we detail three specific case studies of violent extremist plots planned or carried out in the U.S. While no case is considered "typical," we chose these specific ones because they highlight the realistic complexity of law enforcement prevention efforts, a retrospective look at the possible indicators exhibited, and insights into the level of data collection and analysis needed to improve detection efforts.

2.5. System Studies of Operational Level Research and Tools

In this section, we provide specific system studies of research or measures taken at the operational level to counter the violent extremist threat. These efforts include 1) proposal for early warning detection systems of targeted violence primarily through social media, 2) practitioner instruments and tools to enable a more consistent but still manual assessment of individual risk for violence, and 3) commercialized social media monitoring systems.

2.5.1. Proposed early warning detection systems of targeted violence. There have been several preliminary efforts for threat-based monitoring systems in the literature in recent years, including [35, 157, 274, 276]. In [274] and [276], researchers devised the architecture for a three-component monitoring system (consisting of a crawler, repository, and analyzer), as well as formalized the necessary social media site data ontology. Their main

application was the detection of potential school shooters. Researchers in [35] developed an approach to detect the weak signals of lone wolf terrorists by analyzing for intent, capability, and opportunity. This effort 1) emphasized individual potential terrorist identification over terrorist group identification, and 2) focused on developing semi-automated (human analyst in-the-loop) tools rather than fully-automated tools. Their proposed methodology employed a web-crawler to find extremist forums/websites and algorithms to identify the potential actors/aliases who are active them. From there, they identified potential methods to estimate the components of a lone wolf hypothesis through tailored natural language processing techniques. Lastly, in [157] researchers proposed SEMCON to calculate the similarity of time-stamped social media posts to an empirically derived criminal ontology to identify potential criminals.

Our work differs in two ways. First, while previous works focus only on utilizing social media posts for indicators of threat-behavior, our overarching framework is premised on the idea that analyses of both social media and linkages to off-line behaviors and activities apparent in governmental databases are necessary for increased accuracy and the reduction of false positives. Secondly, our work differs because we chose a graph-based representation of the data to capture the richer relationships between individuals as well as the important context of various behaviors.

2.5.2. Practitioner instruments and tools. In terms of manual tools, there is a prevalence of structured professional judgment instruments for law enforcement agencies. See, for example, VERA 2 [249], Radar-iTE used by the German BKA [15], and SAVE 30 [67, 68]). We highlight only the first and last for purposes of discussion.

VERA 2 is a structured professional judgment tool of 31 indicators for the identification of risk specific to terrorists and violent political extremists. See Fig. 2.18. Developed by psychologists in consultation with other professionals in law enforcement, corrections, and forensic psychology, it was designed for those who were responsible for "assessing individual risk for terrorist-related violence" [249, p. 244]. The researchers advocate for its use with all "violent offenders for whom ideological motivation was involved in criminal action" (i.e., those already convicted)[249, p. 247] as a means to determine risk for future violence as well as help in bail, placement and security classifications [249, p. 238]. The researchers also allow for its use in pre-crime settings and for individuals under surveillance but recommend caution based on "ethical and empirical [complexity] concerns" [249, p. 244]. The VERA 2 researchers made it a point to state the importance of gathering and integration of "all available facts and knowledge accessible from intelligence, legal, law enforcement, correctional and other reports" for the assessor prior to the use of the tool for risk specification [249, p. 247]. This fusion of information is clearly no small task in many cases and serves as one of the main motivations of our research. As will be discussed in further detail later in Section 3.3.2, even after considerable investment and effort, the fusion of intelligence and law enforcement data is still challenged. It is not clear how quickly and easily knowledge about the individual being assessed could be obtained, let alone considering if the knowledge about her or her associates could be integrated into the analysis. For instance, indicator HC.2 "Network (family, friends) involved in violent action" presumably requires knowledge of the individuals (possibly many) associations and their behaviors. To implement the recommended process to scale, it is clear that a novel database structure and query tools would be required.

Table I The VERA 2 indicators

	VERA 2 (Pressman and Flockton) Indicator items	*Low*	*Moderate*	*High*
BA.	Beliefs and attitudes			
BA.1	Commitment to ideology justifying violence			
BA.2	Victim of injustice and grievances			
BA.3	Dehumanization/demonization of identified targets of injustice			
BA.4	Rejection of democratic society and values			
BA.5	Feelings of hate, frustration, persecution, alienation			
BA.6	Hostility to national collective identity			
BA.7	Lack of empathy, understanding outside own group			
CI.	Context and intent			
CI.1	Seeker, consumer, developer of violent extremist materials			
CI.2	Identification of target (person, place, group) in response to perceived injustice			
CI.3	Personal contact with violent extremists			
CI.4	Anger and expressed intent to act violently			
CI.5	Expressed desire to die for cause or martyrdom			
CI.6	Expressed intent to plan, prepare violent action			
CI.7	Susceptible to influence, authority, indoctrination			
HC.	History and capability			
HC.1	Early exposure to pro-violence militant ideology			
HC.2	Network (family, friends) involved in violent action			
HC.3	Prior criminal history of violence			
HC.4	Tactical, paramilitary, explosives training			
HC.5	Extremist ideological training			
HC.6	Access to funds, resources, organizational skills			
CM.	Commitment and motivation			
CM.1	Glorification of violent action			
CM.2	Driven by criminal opportunism			
CM.3	Commitment to group, group ideology			
CM.4	Driven by moral imperative, moral superiority,			
CM.5	Driven by excitement, adventure			
P.	Protective items			
P.1	Re-interpretation of ideology less rigid, absolute			
P.2	Rejection of violence to obtain goals			
P.3	Change of vision of enemy			
P.4	Involvement with non-violent, de-radicalization, offence-related programs			
P.5	Community support for non-violence			
P.6	Family support for non-violence			
SPJ	VERA final judgment	Low	Moderate	High

Note: Rating differences for protective items: high rating = more mitigation and less risk

FIGURE 2.18. VERA 2 Indicators Source: [249, p. 245].

SAVE 30 is an empirically-validated, neurocognitive structured professional judgment tool to quantify the predictive risk assessment of potential violent extremists [67, 68]. It requires a trained analyst to complete an inventory on 30 perceptions and beliefs of each suspected individual and is supported by a software tool that helps the assessor visualize the risk [69]. For proprietary reasons, the specific 30 indicators are not available. This work, while promising, differs from ours in three areas. First, our risk assessment protocol

specifically utilizes overt and observable behavioral cues that bystanders (as well as law enforcement) can identify [168], rather than cognitive states or perceptions which may be much more difficult to discern. Second, unlike Dean, the initial dynamic risk assessment protocol in [168] that we intend to build upon does not cover all forms of violent extremism. As previously stated, Klausen's conclusion is that "the ideology and the behavioral changes and adaptations required by the Salafi-jihadist belief system [lends] a predictable structure to the radicalization process" [168, p. i]. Lastly, beyond structured professional judgment tools, we are rather seeking advancements towards a semi-automated risk assessment system that can scale with both the number of potentially radical and radicalizing violent extremists and the voluminous amount of behavioral and activities data they generate.

Some including [214, 249] proffer the use of structured professional judgment instruments and make the conclusion that they are "clearly preferable [over other methods]" because they can "jog the assessor's memory [of relevant indicators]" while still allowing for "informed clinical judgment" [214]. However, structured professional judgment tools clearly do not address the aforementioned challenges related to the dynamics and scale of the radicalized violent extremist problem. First, their use does not *scale* well because they require the manual assessment of dozens of indicator items for each individual (in the case of VERA 2, 31 indicators [249]). This process can be laborious and requires careful study of a suspect. Furthermore, such instruments are not designed to consider behavioral dynamics of individuals or their social ties, nor any measure of their trending.[24]

2.5.3. Social media monitoring systems. It is important to mention the commercial systems that have been utilized by a variety of local and state law enforcement agencies to

[24]For example, while VERA 2 has the indicator "HC.2: network (family, friends) involved in violent action," [249] for which an assessor could rate "low,", "moderate," or "high," this would likely oversimplify any specific network influences or the unique roles that associates could perform in a conspiratorial plot.

glean real time insights and alerts from social media. Prominent ones include Geofeedia, Digital Fly, Social Sentinel, Digital Stakeout, Snaptrends, and Media Sonar. These services have also found root in school and university safety [314]. According to [53] published by Brennan Center for Justice, law enforcement agencies have used these services to:

- Gather evidence in criminal investigations.
- Decrease response times to incidents by alerting agencies of incidents.
- Alert police to potential threats.
- Detect trends in activities.
- Analyze the negative sentiment levels in social media posts [53].

None of these services have published their proprietary methods, but analyses of their product websites indicates that the techniques include: 1) keyword filters for threats based upon client input and security word libraries, 2) sentiment analysis, 3) establishment of geo-fencing that provides search results localized to an area or establishment (i.e., school or patrol area) of interest, 4) allowing clients to add context to searches to help officials know what is really being discussed. Snaptrends, one service reportedly in use by both the Department of Homeland Security and Department of Justice [39], also reportedly has the capability for social media account entity resolution to allow the development of individual profiles (including location information) across multiple social media accounts [40].

All services seem to provide alerts to administrators and clients of suspicious activity. But two key shortcomings are that 1) insights are drawn primarily from social media and no other data sources, and 2) alerts are only trigger-based, and may not explicitly include analysis on individual behavioral trajectories through indicators over time.

It is also important to note of that a number of these services, including Geofeedia, Snaptrends, Media Sonar had some of their application program interfaces (APIs) with popular social media sites suspended due to revelations about their uses by U.S police departments [39–41, 60]. For example, Facebook, Twitter, and Instagram all cut off both Geofeedia and Media Sonar between October 2016 and Janurary 2017 following revelations that the services helped police track social media posts of BlackLiveMatter activists [40, 243]. Additionally, Snaptrends had its Twitter access suspended and reportedly shut down its law enforcement support [53]. Twitter updated its platform policy in November 2016 to state, "We prohibit developers using the Public APIs and Gnip data products from allowing law enforcement – or any other entity – to use Twitter data for surveillance purposes. Period" [213]. Facebook also updated its platform policy in March 2017 to state, "Don't use data obtained from us to provide tools that are used for surveillance" [281].

One other commercial product worth mentioning is Beware®, a public safety personnel information service created by West Safety Solutions (formerly Intrado). Designed to better inform law enforcement of the potential threats when they respond to a house-call or other emergency, the system is unique because it fuses not only social media [155] but publicly available commercial records, data which includes "vehicle registrations, criminal records, warrants, property records, and known associates before arrival at the scene" [308] to produce a "threat" score for residents of the premises. This service also received public backlash. The most current product fact sheet has removed any mention of social media usage, and it is no longer clear if it still has access to data from the popular sites.

Lastly, it is important to mention signs that other countries may be having successes with their use of social media monitoring services to prevent terrorist attacks. A recent article in

the *The Economist* revealed that the Israeli Defense Forces reportedly employed specially developed algorithms to monitor the social media activity of Palestinians on sites such as Twitter and Facebook [87] to look for the presence of indicators found in past attackers.[25] It is unclear the how such monitoring occurs in the international setting or whether such surveillance continues given the later updates to platform policies. Interestingly, the Australian police forces also utilize some form of social media analytics and touted some cases of successful disruption of cases in 2016 [237]. According to a recent poll, the public there seems to be favorable of surveillance for counter terrorism purposes, while being quite unfavorable of the use of social media data for commercial and advertiser purposes [315] (which ironically seems the opposite of the sentiment found in the U.S. [66, p. xi]).

2.6. Summary

This primer was provided to first detail the trends in both the number and lethality of violent extremists plots particularly in the U.S. but also Western Europe. Included also was a review of the large body of literature on the latest understanding of violent radicalization. Existing research affirms the notion that personal (albeit widely various) behavioral trajectories towards violent radicalization exist, and that many of these behavioral indicators were indeed are detectable and observable. This conclusion forms a firm foundation for the rest of the efforts in this thesis.

As shown in each of the stage-based models, there is also support for the importance of social influences on one's involvement in violent extremism. Personal and online relationships

[25] Indicators include: "allegations that Israel is 'desecrating' the al-Aqsa mosque on Temple Mount in Jerusalem, complaints about the Palestinian leadership, and declarations of how they belong to a 'lost generation' or are personally enraged by a relative, friend or neighbour having been killed by Israel" [149]. The algorithms also look for the presence of "personal problems", such as "forced marriages, debt and social exclusion" [149].

play a complex role in the radicalization of terrorists, and there is a tendency for extremist individuals to leverage preexisting friendship or kinship ties or immerse themselves among like-minded peers attachment to other jihadists. These observations also propel us later in this thesis to consider the behaviors and influences of neighbors on the assessment of risk in potential violent extremists.

This chapter also provided an overview of the various governmental strategies proposed to counter the violent extremist threat. While these strategies emphasized a holistic approach to countering the underlying Salafi-jihadist ideology and empowering local religious and civic groups to provide off-ramps for susceptible individuals, it is also important that they included statements about investing in law enforcement investigative tools, improved interagency information sharing, and proper utilization leveraging social media for the identification of radicalization indicators.

CHAPTER 3

A Radicalization Detection System Framework

3.1. Introduction

In this chapter, describe our solution approach as embodied in a proposed radicalization detection system framework to enable law enforcement and intelligence to mine, monitor, and screen for the occurrence of radicalization indicators in large heterogeneous databases in order to provide early warnings of individuals or groups on behavioral trajectories toward extremist violence. We first describe the principle operational deficiencies in efforts at detecting the radicalization of violent extremists. Then we propose the radicalization detection system framework. Following this, we describe the most salient environmental factors that would impact the development of such a system, as well as an identification of the most relevant stakeholders. Lastly, we discuss initial risks associated with the proposal of this system as strategies to mitigate the risk.

3.2. Operational Deficiencies

In Chapter 2, we detailed the concerning trends in both the number and lethality of plots by violent extremists motivated by Salafi-jihadism in the U.S. and Western Europe. This leads us to investigate the present operational deficiencies in efforts to prevent extremist violence as a first step in needs analysis. As discussed in Section 2.4, current research suggests that radicalization, while complex, may be understood as a dynamic and phased-based process where individuals exhibit indicative behaviors or psychological states along pathways to violence. To prevent future extremist violence, the tasks to law enforcement can be aptly summarized in two steps. First, it needs to determine whom to investigate based upon an

individualized risk assessment using knowledge of observed or inferred indicators. Second, it needs to decide on both the level of investigative resources to be devoted and the type of investigative techniques and methods (including both physical and electronic surveillance) to be used on the individual. In this section, we discuss where current approaches by law enforcement and intelligence agencies are deficient in both these principal tasks.

3.2.1. Deficiencies in current risk assessment protocols. As was previously discussed in Section 2.5.2, in order to determine whom to investigate, law enforcement agencies generally use some form of a risk assessment protocol or structured professional judgment instrument such as the VERA 2 [249] or SAVE 30 [67, 68]. However, their use can be laborious and requires careful study of a person of interest and the manual assessment of dozens of indicator items for each individual (in the case of VERA 2, 31 indicators [249]). These current risk assessment protocols are deemed insufficient to distinguish those truly on a pathway to extremist violence and those who are not (i.e., distinguishing true positives from false positives). According to Sageman in [264, p. 11], "law enforcement agencies complain that they are drowned by an ocean of false alarms, which overwhelm their resources.[26] Furthermore, such instruments are not designed to consider behavioral dynamics of individuals or their social ties, nor any measure of their trending, and ultimately do not scale well to the number of threats that law enforcement face.

3.2.2. Physical and electronic surveillance is resource intensive. For those with the highest risk of extremist violence, law enforcement would desire both physical and

[26]Morever, Sageman writes, "The major request from the field is help to distinguish the very few true positives that will turn to violence from the vast majority of false positives- young people who brag and pretend that they are tough and dangerous, but, in fact, just talk, talk, talk, and do nothing"[264, p. 11].

electronic surveillance in order to detect indicators of radicalization at the earliest opportunity and be postured to rapidly foil plots [57, 123]. However, the employment of these full-on techniques does not *scale* well to the caseload. In November 2015, the FBI revealed that it had at the time over 900 active investigations related to homegrown violent extremists in the U.S. [27]. The Bureau also maintains records of, but cannot possibly continually monitor or investigate, the hundreds of thousands of individuals in the Terrorist Screening Database (TSDB) [55, 123]. Furthermore, recent law enforcement successes in intercepting homegrown terrorists have relied upon a limited and resource-intensive approach with confidential human sources. It has been estimated in the open media that between 30-40 personnel (FBI agents, technicians, and analysts) are required to thoroughly surveil a single individual and that the Bureau has enough resources for only "dozens" of people [123].

Reports suggest that law enforcement agencies in Western Europe like Germany, France, and Belgium are likewise overwhelmed by the number of people they need to track [88]. For instance, following a terrorist attack in Berlin in December 2016, a German Interior Ministry official said on the condition of anonymity, that their law enforcement agencies were having difficulty keeping tabs on nearly 600 gefärder ("someone deemed likely to endanger the state") [88]. Following the November 2015 coordinated Paris terrorist attacks that killed 129 people, French intelligence officials revealed they they simply did not have the resources to monitor all whom they consider a threat. A CNN reporter wrote:

> [I]t takes 15 to 20 people to monitor one suspect 24 hours a day. [The French] have 11,000 people on their 'fiche S' list, used to flag individuals considered a threat to national security and who they believe are radicalized. Of those, 5,000 have been elevated to an additional level of concern. Adding to that hundreds, perhaps more than 1,000, who have gone to Syria and Iraq, of whom about half have returned [255].

It is also interesting to note that following the January 2015 Charlie Hebdo attack, France announced a $493 million plan to add 2680 new counterterrorism security posts, as well as "[provide] more tools and technology for monitoring, such as phone-tapping and Internet surveillance" [25].

In the end, resource constraints have forced the agencies to make tough, subjective decisions on the level of surveillance and monitoring that suspected individuals receive [123]. Moreover, recent terrorist attack successes highlight the real possibility of missed signals from, or continued radicalization by, individuals whom the law enforcement agencies had formerly investigated and even interviewed. Recent U.S. cases include Tamerlan Tsarnaev (Boston Marathon bombings, 2013) [106], Omar Mateen (Orlando night club shooting, 2016) [56], and Ahmad Khan Rahami (New York and New Jersey bombings, 2016) [217]. Recent Western European cases include several involved in the Paris terrorist attacks in 2015 [96], Anis Amri (Berlin market truck attack, 2016) [77], and Salman Abedi (Manchester concert bombing, 2017) [205].

3.2.3. Complexity of assessing the risk for violent extremism through social ties. Complicating matters more, the *dynamics* for an individual's social ties, are also important for assessing the risk for violent extremism. These dynamics may include activities by jihadist recruiters as well as the activities of co-conspirators in terrorist plots. As shown by the statistics in Section 2.2.4, about 30% violent extremist plots in the last 20 years have involved two or more co-conspirators, each of whom we can reasonably suspect might have exhibited indicators. For example, the FBI investigation after the San Bernardino terrorist attack in California revealed that the indicators of the homegrown violent extremist threat were dispersed among Syed Farook, wife Tashfeen Malik, and Farook's associate,

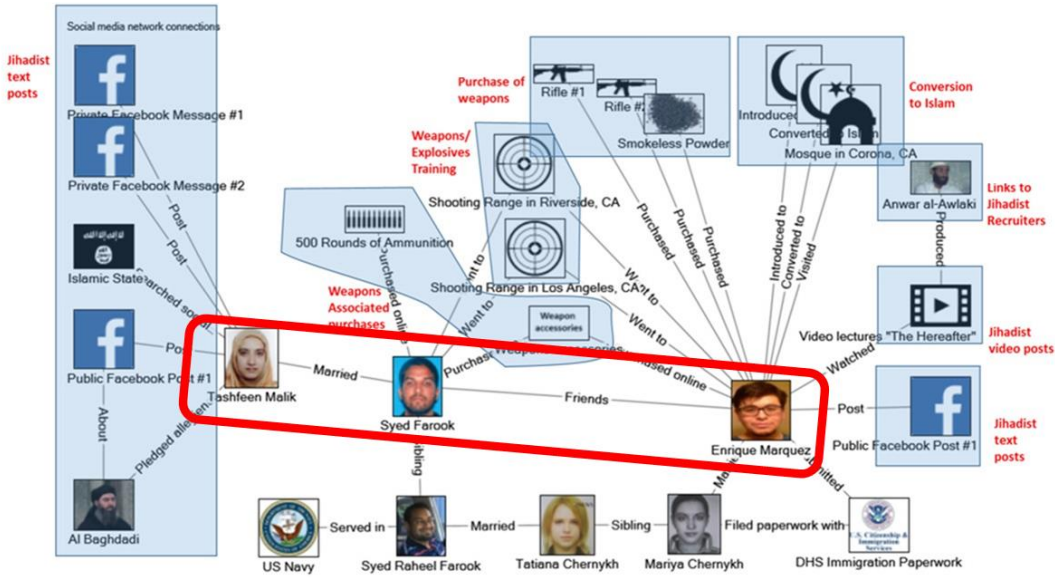


FIGURE 3.1. San Bernardino Terrorist Attack, 2015. Behavioral indicator and association graph of Farook, Malik, and Marquez showing the indicators and signals of their collective radicalization and preparations for the attack, consolidated from investigative findings [299].

Enrique Marquez [299]. Fig. 3.1 depicts that Marquez served as Farook's straw purchaser of 2 weapons, which were ultimately used by Farook and Malik in the attack. Both Malik and Marquez had suspicious social media activity, the details of which are available in Appendix B.3 and [299]. It is only when their behaviors are viewed collectively, as a conspiracy, that one can see the many indicators present.

3.2.4. CONTINUING CHALLENGES IN THE SHARING AND FUSION OF AVAILABLE INTELLIGENCE AND LAW ENFORCEMENT DATA. The U.S. government reforms following the 9/11 terrorist attacks included initiatives for better sharing of information across all levels of government to prevent future attacks. These reforms included the establishment of 1) the Office of the Program Manager for the Information Sharing Environment (ISE) (which had just published its framework in 2014 [229]), 2) the National Network of Fusion Centers to receive, analyze and share such threat-related information across levels of government (federal,

state, local, and tribal), and 3) the National Suspicious Activity Reporting (SAR) Initiative (NSI) to facilitate information sharing of suspicious activity reports (SARs).[27] A multitude of additional information sharing portals was established, to include the Homeland Security Information Network (HSIN) [228], eGuardian [33], and several others [44].

However, interagency cooperation and sharing continue to be challenging. The failure to share information about Tamerlan Tsarnaev, one of the perpetrators of the 2013 Boston Marathon Bombings, was investigated by an interagency team of inspectors general [148]. The final unclassified report found that Russian Federal Security Service (FSB) had provided intelligence to both the FBI and the CIA that Tsarnaev was a follower of radical Islam and was "preparing to travel to Russia to join unspecified underground groups in Dagestan and Chechnya" [148, p. 1], but neither agency knew of the report to the other agency. Additionally, despite Tsarnaev being put on two separate watchlists, the inspectors general could not confirm that the Department of Homeland Security notified the FBI when Tsarnaev departed and returned from Russia less than a year before the bombings. This precluded the FBI from conducting any follow-up investigation on Tsarnaev for his travel abroad [148]. A 2017 Inspectors General report still identified gaps and shortfalls in information sharing [228].

The ISE Framework outlines three principle patterns of information sharing: query/response, broadcast (of alerts, warnings, or notifications), and workflow. However, our examination of the description and capabilities of each of the existing information sharing portals found that most only utilize query/response in the form of a document repository or database that has to be searched. While economical from an information consumption

[27] See NSI website `https://nsi.ncirc.gov/about_nsi.aspx` and the DHS website `https://www.dhs.gov/topic/information-sharing` for more details.

standpoint, it presupposes that law enforcement analysts know precisely whom they need to query about. Given the dynamic nature of the threat of radicalizing violent extremists, this is a likely a very tall order. The broadcast model seems the most beneficial for tracking individual-level radicalization indicators which may only initially be visible to a particular government agency (such as local law enforcement or the TSA). However, to the extent that broadcast is available in the existing portals, it was for notification of imminent, actionable threats or share-point style notifications of new documents in a particular document repository [75].[28]

We also note that down at the State, Local, and Tribal (SLT) law enforcement level, there is a multitude of available systems by which personnel can obtain sensitive intelligence or information on counterterrorism issues. A recent survey of SLT law enforcement listed six different information sharing services with varying levels of use [44]. Moreover, the report noted many investigators and analysts will only access one of the systems due to convenience and simplicity even though different information may appear on each system [44, p. 16].

In summary, the present threat information sharing environment is not conducive to tracking violent extremist radicalization indicators that may be apparent only initially to disparate government entities. There are a multitude of existing portals, which are too

[28]NSI: It is unclear what information besides SARs are shared and whether fusion is done manually or automatically. HSIN: A secure, web-based portal used by fusion centers across the country to share Sensitive But Unclassified terrorist-related information and intelligence that includes a document repository, messaging and instant messaging services, and GIS mapping capabilities [75]. However, the alerts and notifications in this system are either only bulletins pushed to all users, or SharePoint-like notifications when documents or document folders have been updated. The portal collects disparate pieces of information and intelligence but does not necessarily represent a true 'fusion' like a graph database. eGuardian: Part of the FBI's Law Enforcement Enterprise Portal (LEEP), it is a database system which pools new SARs related to terrorist or other potential criminal offenses with a legacy SARs reporting system and feeds a separate database for Joint Terrorism Task Forces to utilize [33]. Besides standardizing reporting formats and providing a triage-like system for SARs, it is not entirely clear what type of fusion (linkages, entity resolution, etc) is available in the eGuardian system [33].

reliant on the query/response pattern and selective, human-driven fusion of information about individuals and their related data points.

3.2.5. Lack of rigorous utilization of available indicators on social media. It is acknowledged by both the U.S. House of Representatives Homeland Security Committee [197] and the National Fusion Center Association that law enforcement agencies are presently under-utilizing the indicators available on social media.[29] As Sections 2.2.3 (Role of Social Media) and 2.4.3 (Antecedent Behaviors) described, there is an increasing involvement of the internet and social media in the radicalization of individuals and the possibility of finding early warning behaviors online. Consumption of propaganda and learning to conducting various aspects of the terrorist plot through virtual sources were among the statistically significant pre-attack behaviors. Additionally, warning behaviors such as leakage and direct communication of threats may also be available through electronic sources as well. In Appendix A, we provide empirical evidence in recent cases of extremist violence in which the perpetrators left early warning digital signals, many of which were undetected by law enforcement.

There are several factors that make the utilization of social media challenging to law enforcement. First, the open monitoring of social media for pre-crime indicators has significant push-back from civil liberty and privacy groups as well as the social media companies themselves. This state of affairs was previously described in Section 2.5.3. Second, as previously described in Section 2.4.3, even when conducting investigations of a specific individual, U.S.

[29]There is a new initiative called the Real-time Open Source Analysis of Social Media (ROSM) led by the National Fusion Center Association with the goal of determining "how law enforcement agencies can and should analyze and share social media information and related criminal intelligence to help identify common indicators that can support intervention with potentially violent extremists and thereby prevent and/or disrupt attacks" [228, p. 8]. However, nothing is publicly available on this initiative and is deemed a seemingly novel effort to which our research efforts could eventually help contribute to.

law enforcement is subject to the Electronic Communications Protection Act and is required to present a subpoena, warrant, or another court document to the social media company to request preservation and release of electronic records. Lastly, even if law enforcement could readily monitor social media, sorting through the immense amount of data is a problem. While machine learning and natural language processing algorithms are improving in this area (see Section 3.5.3), there are still significant challenges to automatically identifying genuine threats and discerning indicators of violent radicalization.

The confluence of issues involving the authorities for use of social media data and the technological support to utilize the signals was most recently addressed by the former Acting Director of the Defense Intelligence Agency David Shedd, who said:

> I have been a strong advocate that the U.S. Department of Justice must give new authorities to the FBI to conduct something between a full and open investigation on a potential attacker...These in-between collection authorities would require court oversight but would use technology to track and then flag anomalies related to the kind of correspondence, postings, or activities of a suspicious individual. This approach would make it far less human resource-intensive [280].

The present gaps and shortcomings are indeed limiting the early warning signals available for law enforcement analysis and straining agency resources, but Shedd is proposing that changes in legal authorities and technological innovation can improve the situation. We go further by including in the scope of our proposed system the activities that are apparent in government databases (like the TSA and weapons background checks).

3.2.6. Lack of accessible training data. Lastly, many have noted the paucity in available training data for researchers who want to assist law enforcement and intelligence agencies. In 2013 LaFree, the Director of the National Center for the Study of Terrorism and Responses to Terrorism (START), summarized that, "[C]ompared with collecting data

on other types of crime, collecting data on terrorism has been especially challenging" [177]. He went on to cite the 3-stage progress that START and associated researchers had done to improve the empirical study of terrorist attacks including 1) "development of international databases of terrorism– assaults where a group or an individual from one country attacks targets in another country," 2) "collection of domestic as well as international data on terrorist attacks," and 3) "development of specialized data sets on specific subsets of terrorism cases," [177] such as ones segmented by motivation/issue [116, 117]. However, beyond the three advances previously mentioned is a forth initiative: the development of large population, synthetic but empirically-based datasets on the early-warning behavioral indicators of violent extremists. Such an advancement would specifically address Sageman's concern of a "lack of comprehensive and reliable data" and the impediment it is to scholarship in terrorism studies due to the difficulty gaining access to sensitive and even classified details of individuals which are held closely by law enforcement and intelligence agencies [264, p. 6-7], and greatly benefit the growing body of researchers supporting law enforcement with analyzing and weighing the pre-incident behavioral indicators of terrorism and violent extremism. Specifically, a large-population synthetic dataset of people and potential radicalization indicators could serve as a validation testbed for risk assessment tools. In order to be effective, the dataset would likely have to contain synthetically generated samples of 1) terrorist offenders and their behavioral characteristics based on anonymized, empirical case studies, and 2) non-violent radicals and their behavioral characteristics. The latter could potentially be derived from the datasets found in [12], which were acquired through extensive field interviews with non-violent radicals.

Similar efforts appear in other domains, including [7] for generating synthetic data mimicking real time-varying multi-attribute characteristics of computing nodes and [120] for the behavioral indicators of cybersecurity insider threats. Most recently, efforts such as [242] attempt to produce synthetic datasets in a variety of domains that protect privacy and closely resemble the original source data.

3.3. A Radicalization Detection System Framework

Ultimately, there is a compelling need to address the deficiencies in law enforcement processes and capabilities so that they can handle both the *scale* and *dynamics* of the violent extremist threat. In [142, 143] we codified this need in the definition of the *radicalization detection problem*, which is "the automated task of dynamically detecting and tracking behavioral changes in individuals who undergo the process of increasingly espousing jihadist beliefs, and who transition to the use of violent action in support of those beliefs." In this section, we present a radicalization detection system framework as a proposed overarching approach to addressing this problem and enabling law enforcement to monitor persons of interest at the scope and scale that is lawful, and to rigorously and more automatically account for the dynamic indicative behaviors of individual persons of interest as well as their associates. It first involves the development of a robustly capable data management system that fuses and processes all the person-centric data from governmental databases that law enforcement would have an interest in tracking. Additionally, the data management system would fuse social media data by individuals whom law enforcement would currently be approved to surveil. Second, the framework calls upon a multi-disciplinary team of law enforcement practitioners, terrorism experts, and data scientists to derive discerning patterns of violent radicalization and develop computationally efficient tools to mine the databases

(of both existing and streaming data) for these patterns. These tools would have the specific purpose of identifying radicalization indicators and then monitoring and screening for in near real-time those individuals who pose a significant risk for extremist violence.

The framework is depicted in Fig. 3.2. We discuss below the main components of 1) data requirements and processing, 2) query pattern development, 3) the foundational investigative graph search approach, and 4) the analyst's role in the interpretation and use of the results.

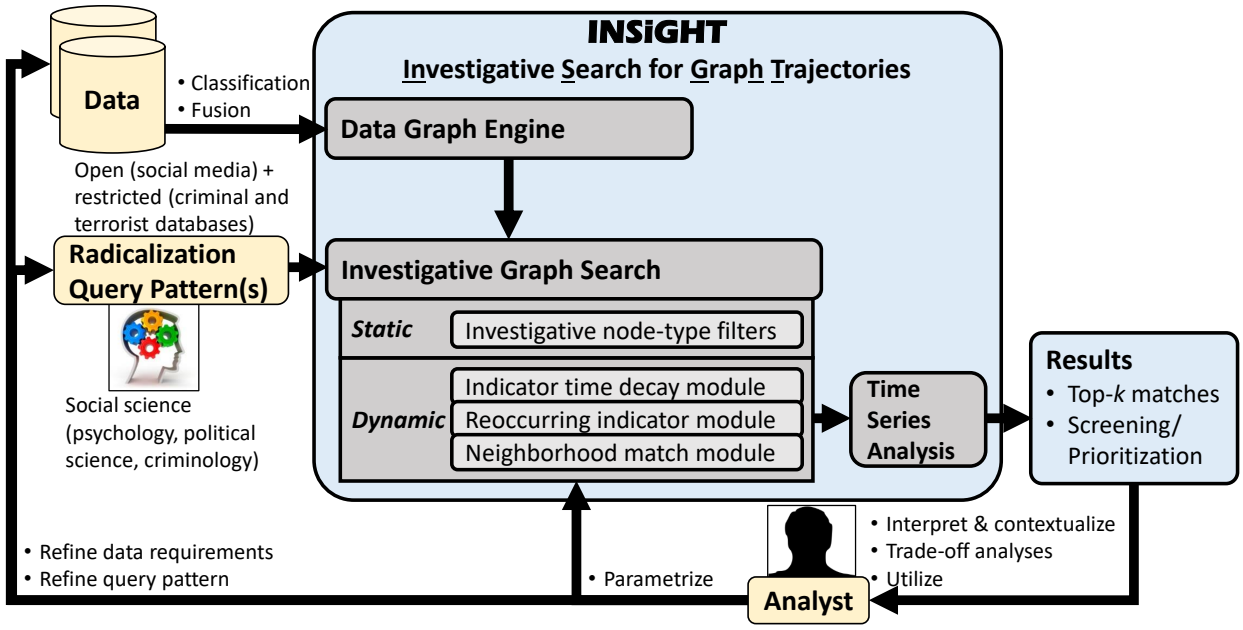


FIGURE 3.2. Radicalization detection system framework. This figure depicts a system for law enforcement and intelligence analysts to detect the radicalization trajectories of individuals using INSiGHT.

3.3.1. QUERY PATTERN DEVELOPMENT. As previously mentioned, our approach is predicated upon the development of effective query graph patterns of indicators for violent radicalization or some other latent behavior of interest. The creation of indicator-based screening patterns for violent radicalization that are empirically-derived is a necessary contribution from social scientists (i.e., psychologists, criminologists, political scientists). Based

upon several existing radicalization models, it is likely that the pattern will have an 'intent' or 'ideology' indicators as well as a 'capability' or 'operational preparation' indicators [35, 223]. Obviously, the pattern subsequently drives what types of data are necessary and what types of automated classifiers are needed for semantic analysis.

3.3.2. Data Fusion and Processing. Critically, our approach also calls for continued efforts at improving the law enforcement and intelligence data management systems to enable automated methods for fusing time-stamped data from a variety of domains and activities. RAND recently analyzed the vital role of this step and its potential value in any behavioral indicator threat detection system in [66, p. xxxiv and xi]. The numerous systems available today for analysts to *reference* data on records of persons of interest were summarized in Section 3.2.4. Given the dynamic nature of the threat of radicalizing violent extremists, we advocate for a dynamic, continuously updated graph database of individuals and their linked person-centric data points. It is this type of database on which graph pattern matching algorithms could operate and return meaningful results with of individuals' radicalization indicators available across disparate government databases.

It is important to mention a potential system by which this improved data management system could be based. The FBI also hosts as part of LEEP the National Data Exchange (N-DEx), an "unclassified national information sharing system that enables criminal justice agencies to search, link, analyze, and share local, state, tribal, and federal records" [105]. An ambitious effort that is designed to enable detectives and investigators from any law enforcement agency in the U.S. access available data on individuals and on-going investigations. Unlike other databases, N-DEx specifically touts a link-analysis tool that allows investigators to examine "associations between people, places, things, and events" in the data available

[105]. This type of database is the beginning of the type architecture on which INSiGHT can run on.

Short of true fusion of a multitude of data and intelligence sources, one possible proposed technology is the Person-Centric Identify Management (PCIM) capability which recently appeared in a MITRE white paper in 2017 [211]. Confirming the state of affairs, the paper stated, "The vast majority of the data collected [by U.S. government agencies] is stored in stand-alone and stovepiped systems with limited or no data sharing capabilities, no standard correlation capability to resolve or link identifies, and no substantive capability to link additional peripheral encounter data (e.g., information gleaned from social media, checkpoint screening results, or other relevant agency records)" [211, p. 1]. It affirms the need for fused information across disparate databases for counterterrorism and other security purposes and proposes that the Department of Homeland Security lead a cross-government initiative to develop this capability. What it envisions is the ability to obtain, upon query, a federated view of an individual's interactions with various government entities. We suspect that such a technology could be a utilized with INSiGHT if queries of a large number of individuals could be returned by efficiently accessing separate databases and combining the information into a person-centric graph database.

Based upon case studies of radical extremist cases, we surmise that at a minimum we need the fusion of local, state, and federal criminal and terrorist databases (e.g., SARs, FBI's Tripwire and 'FBI Tips' programs and TSA's Automated Targeting System and Secure Flight programs), firearm background check databases (e.g., National Instant Criminal Background Check- NICS), and publicly available social media data (e.g., Twitter, Facebook). The fusion of such data has technical challenges, such as the construction and universal adoption of an

ontology, as well as entity resolution between social media accounts and individuals. There are also privacy and civil liberties concerns as well, but as stated in RAND's study "the irony is that commercial organizations (and even political parties) are already far ahead in exploiting the relevant technologies and forever changing notions of privacy," [66, p. xi]. In our present work, we intend to demonstrate that such a fusion of such data would greatly help law enforcement and intelligence officials connect the dots among different types of behavioral indicators.

Moreover, as previously mentioned in the query description, the data must undergo binary or multi-class classification for each of the features or labels prescribed in the query pattern. There are already a few machine learning approaches put forth for the detection of ideological indicators of radicalization [35, 51]. But clearly, much advancement is required in this important area for the larger radicalization detection system to the feasible.

3.3.3. Investigative Graph Search. Once the requisite inputs are in place, we run our parser to produce a time-stamped, heterogeneous data graph. The nodes in this graph represent entities or their activities, while the directed, time-stamped edges represent the connections between nodes at the specified time-step. Investigative graph search is the novel process of searching for and prioritizing persons of interest who may exhibit part or all of a pattern of suspicious behaviors or connections. It is particularly relevant to search for those on radicalizing towards violent extremism and has both static and dynamic variants which we will cover in detail in Chapter 4 (Section 4.2) and Chapter 5, respectively. The dynamic variant is the technique is called INSiGHT, which is our algorithmic graph pattern matching approach that calculates the multi-hop class similarities between nodes in the query and

data graphs over time. Presently, our methodology collects data and performs the analysis on the entire network at periodic times.

The investigative graph search has four enhancements to make the technique particularly applicable to helping law enforcement identify individuals or small groups with conforming subgraphs to a radicalization query pattern. The first improvement made to static investigative graph search is the incorporation of filters for certain types of investigative indicators (see forthcoming Section 4.3) to minimize false positives for stand-alone individually innocuous activities. The next three enhancements were made to dynamic investigative graph search (INSiGHT) and will be covered in detail in Section 5.7. One is the parameterized method of decaying an indicator's significance over time. Another is the parameterized method to weigh the re-occurrence of each class of indicator. The last is a non-combinatorial method for neighbor matching to reveal the presence of suspicious individuals with nearby connections and the possibility of collective threats if individuals happen to be working together. The resulting time-series of similarity scores from investigative graph search readily enables the monitoring and screening of those individuals whose behaviors indicate a significant risk for violence.

3.3.4. Role of the Analyst. Several works have previously highlighted the importance of analysts in threat detection systems or risk-assessments [66, 91]. As outlined in Fig. 3.2, we envision the analyst needing to interpret and contextualize the results from INSiGHT, perform trade-off analysis of the sensitivity and selectivity of the technology, adjust the parameters allowed in the weighting and decay modules, and intelligently utilize the results for possible law enforcement intervention.

Moreover, we also envision that the analyst is someone with formal psychological training to make risk assessments of individuals. Recently, Sarma in [266] outlined the needed training and the skills required of individuals performing such assessments. This included "knowledge of the hazard and associated risks, analytic skills, and confidence in completing risk assessments," as well as knowledge of applicable theories of radicalization and others generally related such as social movement theory [266, p. 285]. The researcher also mentions the importance of a supervisory authority over the analyst to ensure that support and feedback are present [266].

3.4. Environmental Analysis

The proposal for a novel analyst-in-the-loop system to mine, monitor, and screen for the occurrence of radicalization indicators sits at the nexus of some key issues and debates in the U.S. and abroad. Namely, the issues affected include the pronounced and high-profile threat from violent extremists; the legal, ethical, and civil liberties issues involved in countering the threat from violent extremism; the on-going debate between privacy and security; and on-going struggle of inter-agency and federal, state, local, and tribal law enforcement intelligence sharing. In this section, we discuss some of the most pertinent aspects these issues by examining each of the environmental dimensions in a complex ecosystem of government agencies, non-governmental organizations, private companies, and civic and religious groups.

3.4.1. Social. Many studies support the idea that local communities and organizations are critical to preventing the radicalization of susceptible individuals and that law enforcement engagement efforts with these communities need to be a centerpiece of any CVE strategy [24, 126, 267, 313]. However, Muslim communities both nationally and internationally have been suspecting of such engagement efforts as stigmatizing or discriminating against

Muslims, disproportionately focusing on Muslim communities, or serving as a ploy to surveil or otherwise gain intelligence on Muslims for terrorist investigations [267]. Any effort at developing tools for law enforcement must consider these social factors.

3.4.2. Economic. The development of any tools to assist law enforcement in the detection of radicalization trajectories must consider the economic factors associated with social media companies, the explosive growth of users, and the pervasiveness and variety of technologies, services, and communication mediums.

3.4.3. Emotional. There are several aspects of the emotional factors to consider in the environment related to violent extremism. A terrorist's intention is to induce an emotional response of fear and susceptibility from its victims and the intended targeted populations. These responses have ripple effects that include the public's overestimation of the true threat, as well as the irrational reactions to stigmatize or discriminate the ethnic or religious groups of some of the offenders en masse. Others, however, may make light of the situation and underestimate the true threat [126, 176].[30]

Another crucial consideration in the area of the development of government use of a detection-type system for potential terrorists is privacy and civil liberties. The arguments are wide and various but principally centered on an emotional aversion to the government knowledge of aspects of a person's life, regardless of whether it was self-disclosed or not to a public forum. This demand for privacy is not without exception, however. People are readily willing to give up many aspects of their privacy for other social or economic utility. For example, Facebook continually collects a large number of personal features data on its

[30]According to the 2016 CSIS countering violent extremism assessment, "The U.S. policymakers have severely underestimated the allure of violent extremism, which has constrained the allocation of funding and manpower to deal with it" [126].

users [82] and willingly share it with data brokers and advertisers [254], but unwilling to have its data for use by law enforcement agencies to prevent terror attacks [281] or by insurance companies to glean auto accident risk [59].

3.4.4. Legal. There are two major aspects of this legal factors surrounding the development of violent extremist detection technologies. The first is the court-ordered warrant for social media and cloud data from the accounts of suspected or known perpetrators of extremist violence. Many of the popular social media companies such as Facebook and Twitter have been known to rightfully comply [233]. There are others, however, such as Telegram and WhatsApp, which proffer end-to-end encryption of communications and claim that it is impossible to give access to the data even with a court warrant[85]. This is what the FBI Director calls, "going dark," and is a tactic that is increasingly being used by violent extremists especially in the near-term execution of their attacks [58, 70, 245]. There are currently no known legal means to compel these companies to create a 'backdoor' [57]

The second issue is the bulk collection and analysis of social media platforms to track the statements of suspected radical or radicalizing individuals. The Fourth Amendment to the U.S. Constitution, which guards against the government's "unreasonable searches and seizures," should protect an individual's right to privacy while giving the "government the necessary means to solve crimes, keep order, and safeguard national security" [63]. Up until recently, U.S. courts have relied on the third-party doctrine, which suggests information in the hands of a third party receives no Fourth Amendment protection [63]. However, such distinctions have blurred in light of the prevalence of on-line and cloud-based technologies that have made many aspects of people's lives electronically available. While many of an individual's posts are publicly available, it is currently a gray area legally of the expectation

of privacy associated with an individual's posts from the government [63]. In this case, it seems that the social media companies can dictate the terms and conditions of the data usage in their platform and privacy policies. Most recently, several prominent social media sites suspended data access to application program interfaces (API) to social media monitoring companies when it was discovered that law enforcement agencies were using these companies to conduct some tacit form of surveillance on suspected individuals [39–41].

3.5. Technology Assessment

In this section, we provide a brief overview of the latest technical innovations which give promise to the realization of a scalable radicalization detection system on large heterogeneous databases.

3.5.1. Graph databases and distributed systems. Graph databases are now widely used in the commercial sector to leverage the complex and dynamic relationships between entities. There is a growing list of graph databases on the market such as Neo4j [256]. There are also recent advancements in translating text documents into graphs [153], using distributed systems to search big graphs [140, 193], the storage and analysis of temporal graph data [48, 207]. We by no means provide a sufficient survey of the literature in these areas, but only attempt to demonstrate how viable is the use of and searches that can be conducted on large graph databases through distributed systems.

3.5.2. Entity resolution. Entity resolution is a significant problem in database fusion, especially considering the array of disparate databases from which individual records will be extracted. The state of the art techniques in entity resolution in SQL-based and graph database settings are ably covered in [304]. Additional efforts at entity resolution

with a particular focus on online account contexts for security applications is offered in [35, 188, 170].

3.5.3. Machine learning and natural language processing advancements. A key challenge in our radicalization detection framework is the automated or semi-automated classification of data as specific indicators of a radicalization query pattern. We envision that the data points could be in the form of social media posts, investigator case notes in a document repository, or simply an entry in a specific database (such as a flight manifest). Without denying the difficulty, it is also important to point out that researchers have been making advances in the use of artificial intelligence and machine learning algorithms on a variety of data sources to identify depression [61] as well as predict and prevent suicide [212]. This includes uses of medical records, social media post content (see also [38, 206]), and auditory signals through cell phone usage.

3.6. Identify Stakeholders and Their Initial Interests

Here we consider the ecosystem of stakeholders in the development of risk assessment enabling technologies for law enforcement. This includes the perspectives of law enforcement, the public whom they serve and protect, the governmental agencies at the federal, state, local, and tribal levels who may have relevant data on individuals to share, and the commercial companies that develop the communication technologies or manage the online data. Throughout this research effort, we take the principal viewpoint that law enforcement and intelligence agencies are prospective clients in the exploration and development of technologies that can aide them in more effectively utilizing investigative resources to prevent violent extremist threats. All the while, we integrate the multitude of interests by other stakeholders and consider their impact on the feasibility and acceptance of such technologies.

The 2016 U.S. Strategic Implementation Plan for Empowering Local Partners to Prevent Violent Extremism in the United States identified a broad list of stakeholders who have "expressed or identified role in countering violent extremism and include, but are not limited to: Federal, state, tribal, territorial, and local governments and law enforcement; communities; non-governmental organizations; academia; educators; social services organizations; mental health providers; and the private sector" [227, p. 1]. We adopt this initial list. The full process of stakeholder analysis, to include identification, engagement, interviewing, surveying, and value modeling is an involved, multi-step process [239] which we intend to conduct in future work. In this thesis, we begin by describing the initial interests or role each may have broadly in the development of the proposed system.

3.6.1. Law enforcement and Intelligence Enterprise. The entire enterprise includes law enforcement agents and officers, intelligence analysts, and technicians at every level (federal, state, local, and tribal); fusion centers and every agency involved in the Information Sharing Environment (ISE). Within the U.S., the Departments of Homeland Security and Justice are leading the interagency effort to counter violent extremism [228], and we expect within these organizations' leadership are the decision makers for any future system. While we believe that this enterprise is united in their efforts to protect the homeland from future terrorist attacks, we also acknowledge that each will have their own parochial interests.

3.6.2. Communities. "Communities" is a broad term that encompasses the localities across the country that have been, are currently, or may be in the future affected by extremist violence, and the subsequent results or consequences of the implementation of any part of a radicalization detection system. While communities all value broad assurances of safety, specific ones will have different valuations and priorities among public safety, privacy, and

law enforcement and intelligence actions. Some of these concerns were addressed in Section 3.4.

We highlight one aspect of these communities which plays a role in the impetus for the continued efforts to counter violet extremist motivated by Salafi-jihadists. According to the polls conducted on U.S. residents by Pew Research for the last 10 years, there has been sustained concern over Islamic extremists both at home and abroad. See Fig. 3.3.[31]

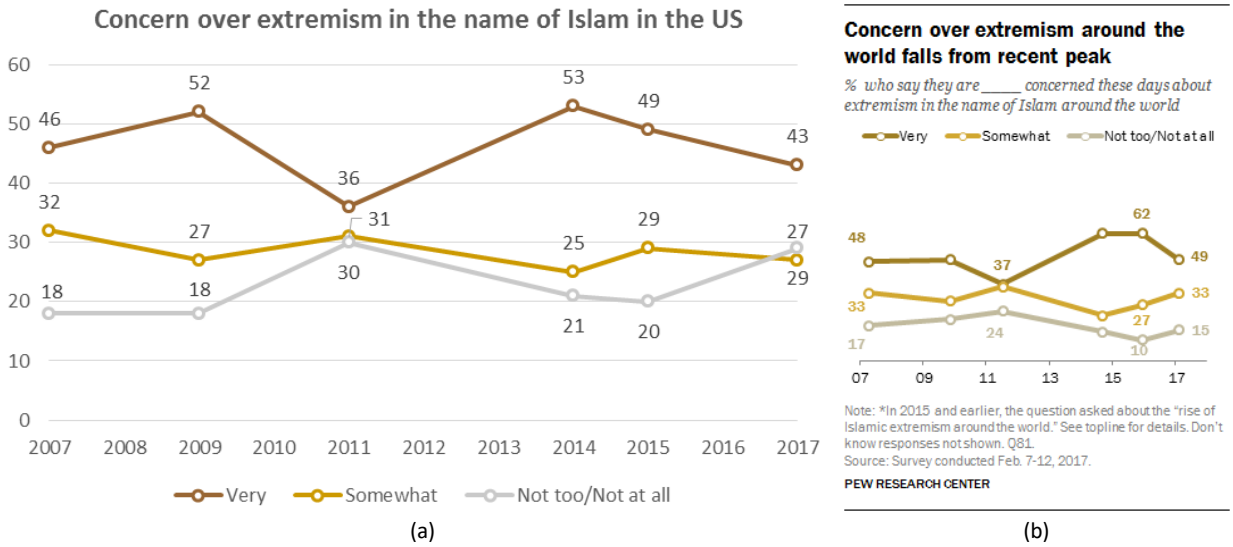


FIGURE 3.3. (a) Pew Research polls in the U.S. about the concern for extremism in the name of Islam in the U.S. Data source: [246].(b) Pew Research polls in the U.S. about the concern for extremism in the name of Islam in the world. Source: [246].

3.6.3. NON-GOVERNMENTAL ORGANIZATIONS (NGOs). There are quite a number of NGOs involved specifically in countering violent extremism by working with local communities and providing suitable off-ramps through various forms of social and economic assistance. Anecdotally, these NGOs and practitioners working in de-radicalization programming are admittedly fearful of being unable to discern highly radicalized individuals from those

[31]We note the drop in 2011 in both Fig. 3.3a and 3.3b. Although it was never discussed in the poll analysis, it is possible some of the decrease in concern during the survey in July 2011 is due to the U.S. killing of Osama bin Laden on May 2, 2011.

individuals open to their efforts and would likely value a tool that can help law enforcement and intelligence analysts screen for the most dangerous individuals [169]. There are also NGOs who are focused on protecting civil liberties and privacy. See [241, p. 25] for a description of the roles these organizations have played in resisting any added capabilities to law enforcement in this area.

3.6.4. Academia. Academia obviously plays a role in the continued rigorous research into the radicalization process, any more discerning early warning behaviors associated with those on pathways to violent extremism, and the development of technological systems and tools to assist law enforcement. See [264] for a description of broad survey of their efforts and limitations.

3.6.5. Social services organizations. This is a broad term for agencies and organizations who would provide social and economic assistance to various segments of society, such as helping refugees or immigrants. Some at least may come in contact with potential perpetrators or even victims of extremist violence and are likely to have knowledge of susceptible individuals that could benefit CVE NGOs or law enforcement.

3.6.6. Mental health providers. Like social services organizations, these providers who may come into contact with those susceptible to violent radicalization are likely to have knowledge that would benefit CVE NGOs or law enforcement.

3.6.7. Private sector. This includes a broad listing of companies offering services in the online domain, to include social media companies, other technology companies such as cell phone manufacturers, application (app) developers, and search engine/email providers. Their interest is primarily in the economics, any legal or regulatory requirements placed on

them, and their accountability to shareholders. Their positions on law enforcement use of their data and platforms were covered broadly in Section 3.4.

3.7. Identify Risk Factors and Initial Risk Management Plan

Risk factors are categories of "uncertain events or conditions whose occurrence will have a negative impact on system cost, schedule, value, technical performance, or safety" [239, p. 79]. The identification of risk factors is a critical step throughout the systems development process. INCOSE recognizes four main categories of risk: technical risk, schedule risk, cost risk, and programmatic risk [130]. In this thesis, we cover broadly only technical and programmatic risks, but not schedule or cost risks explicitly. This is due to the anticipatory and forward-thinking nature of this research and the discussion of schedules and budgets would be premature. For both technical and programmatic risks, we also provide an initial discussion of the mitigation strategies.

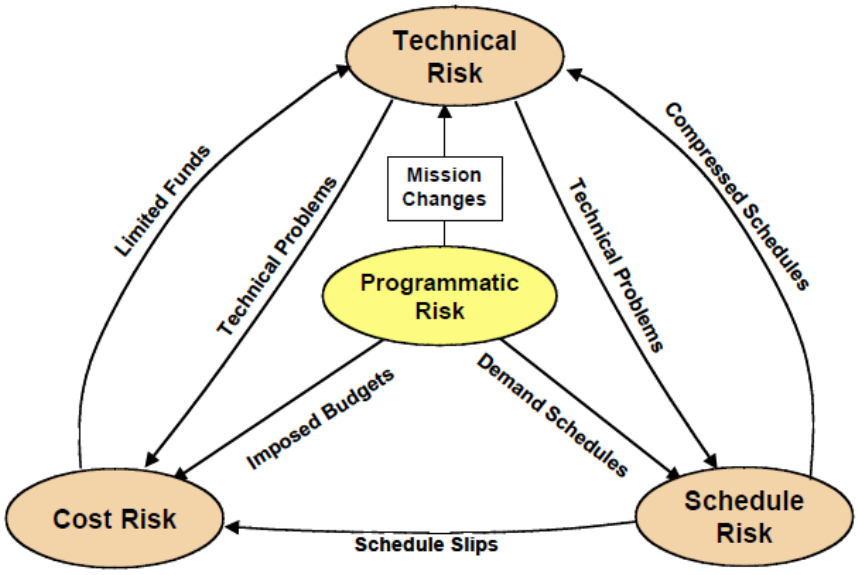


Figure 3.4. Graphic depicting the four main risk categories and the typical effects on each other. Source: [130, p. 7.14].

3.7.1. Programmatic Risks. Programmatic risks are those risks due to external factors or decisions that may adversely impact the successful development of a system [130, 239]. As covered in Section 3.4 there are numerous external realities in the environment.

3.7.1.1. *Multitude of high-salience stakeholders complicates the approval/adoption process.* The first significant programmatic risk for the development of the system is a multitude of government agencies which serve as stakeholders in the counter-terrorism and countering violent extremism domain, and the resulting difficulty that would result even if the project's sponsor was clear. As mentioned previously, the U.S. government designated both the Department of Justice and the Department of Homeland Security as the lead agencies for countering violent extremism. But there are numerous other agencies involved in the effort including the Departments of State, Education, Health and Human Services as well as the the National Counterterrorism Center (NCTC) and the U.S. Agency for International Development (USAID) [184].

3.7.1.2. *Inability to access restricted government and law enforcement data.* Another major programmatic risk is the inability to access restricted government and law enforcement data that would be needed to inform the development and truly validate the system. This research calls upon the integration and fusion of several restricted databases including the data accessible to the Joint Terrorism Task Forces and the network of Fusion Centers. Part of the larger research effort is a data management system that would interface with our graph querying procedure.

3.7.1.3. *Cut-off from access to social media data via APIs.* Another risk for the development of the system is the cut-off to access to social media data from various companies due

to platform policies. As discussed in Section 2.5.3, several commercial social media monitoring companies have been negatively affected in this way due to the assistance they provided to law enforcement. This results in large business risk and costs when a for-profit company needs to re-orient towards tailoring its services for other sectors and domains.

An important mitigation strategy that best handles these aforementioned programmatic risks is that this research and system development should primarily be carried out by a Federally Funded Research and Development Center (FFRDC). These unique, independent, not-for-profit organizations that are sponsored by a U.S. Government agency (or agencies) and principally receive funding from the federal government to conduct responsive research and development.[32] Sageman in [264] called for greater integration between academic researchers and government agencies, but neglected to mention such close cooperation already exists in these institutions. There are a total of 41 centers currently, but the few with the likely specialized domain expertise and associated relevant government sponsor are shown in Table 3.1.

TABLE 3.1. Select FFRDCs in related fields. Source: [210].

FFRDC	**Administrator**	**Government Sponsor**
Lincoln Laboratory	Massachusetts Institute of Technology	Department of Defense
National Security Engineering Center	The MITRE Corporation	Department of Defense
Software Engineering Institute	Carnegie Mellon University	Department of Defense
Homeland Security Systems Engineering and Development Institute	The MITRE Corporation	Department of Homeland Security

[32]The Defense Acquisition University defined a Federally Funded Research and Development Center (FFRDC) as "an activity sponsored under a broad charter by a Government agency (or agencies) for the purpose of performing, analyzing, integrating, supporting, and/or managing basic or applied research and/or development, and that receives 70 percent or more of its financial support from the Government; and – 1) A long-term relationship is contemplated; 2) Most or all of the facilities are owned or funded by the Government; and 3) The FFRDC has access to Government and supplier data, employees, and facilities beyond that common in a normal contractual relationship [72].

One of the merits of this FFRDC-based approach is that the work would have a clear government agency sponsor and have greater credibility with other government stakeholders than compared to a private company responding to Broad Area Announcements (BAAs). Additionally, much of the programmatic risk is absorbed by the sponsoring agency for the research and systems development.

Relatedly, FFDRC's established relationships with the federal customers and stakeholders will also assist in the development and validation of any proposed system. This includes the close partnership with law enforcement analysts, whom the developers could draw upon for expertise, guidance and processing protocols. It is also likely that a combined effort would allow the developers at the FFRDC to access generally restricted data for testing and validation purposes [72].

Lastly, an FFRDC's unique role and associations with federal law enforcement agencies will also help mitigate the last significant risk in relation to social media data access. The support from a law enforcement and legal team who are experts in obtaining warrants and court orders to compel social media companies to release the requested data would greatly help in the development and continued operation of the proposed system [34, 301].

3.7.2. Technical Risks. A technical risk concerns the failure to achieve some system requirement, including a performance objective or a functional, operability, producibility, testability, or integration requirement [130, 239]. At this early conceptual development phase, we have not yet developed all the system or subsystem requirements. But based upon the performance objectives, we identify the following risks.

3.7.2.1. *Inability to achieve data integration/fusion due to incompatible systems.* As described in Section 3.3.2, there is presently a multitude of systems by which law enforcement

agencies share information. Many of these systems were developed independently, and it is unclear at this stage if and how data from each could be integrated and fused.

3.7.2.2. *State-of-the-art natural language processing and machine learning algorithms not sufficient to accurately classify specific textual indicators.* While natural language processing (NLP) and machine learning algorithms continue to improve in capability, no research that we are aware of has specifically addressed the detection of certain social-psychological indicators of radicalization. As discussed in Section 3.5.3, progress has been made in related areas such as the detection of suicide indicators and warning signs.

3.7.2.3. *Radicalization query patterns not sufficient at screening for individuals.* Given the variance among social scientists on the indicators of radicalization and the various forms of resulting violent actions (i.e., foreign fighter, homegrown attack, material support, etc.), it is likely that this technology needs to be able to accept sets of patterns for testing and validation.

The key mitigating strategy for the first two technical risks is the pursuit of Commercial-Off-The-Shelf (COTS) technologies. We anticipate advancements in both the database technologies as well as NLP and machine learning algorithms to address these risks. The COTS strategy has been pursued by the Department of Defense since 1999 in the development of systems [114] and is currently prominently integrated into the Department of Homeland Security Science and Technology Innovation Strategy [76]. This out-sources some of the technical and cost risks to researchers in the commercial industry and academia, but generally requires at least some effort adopting and transitioning the technology to meet the government's needs. Of course, this transition has its own risks that can be identified with

the Software Engineering Institute's COTS Usage Risk Evaluation (CURE) [42]. Specific security related issues are also explored in [83].

As mentioned in Section 3.3.2, there are also two mitigation strategies specifically for the first technical risk of the inability to achieve true data fusion and integration. The FBI's extant information sharing portal N-DEx touts a link-analysis tool for analysts to investigate the "associations between people, places, things, and events" in the data available [105]. Leveraging this accomplishment would greatly assist in the future development of a similar exchange for the radicalization detection system. Another mitigation strategy is reliance on MITRE's proposed Person-Centric Identify Management (PCIM) capability which obtains through query a federated view of an individual's interactions with various government entities [211]. If such a technology were efficient, law enforcement agencies could periodically query for a large number of persons of interest to access separate databases and then subsequently combine the information into a person-centric graph database for INSiGHT's use.

3.8. Conclusion

In this chapter, we described a radicalization detection system framework as an overarching approach to mine, monitor, and screen for the occurrence of radicalization indicators in heterogeneous databases. It was derived from analysis of the existing operational deficiencies in law enforcement and intelligence agency efforts. We also analyzed the potential environment in which such a system would be developed and in use as well as the stakeholders that have an interest in supporting or resisting such a system. Importantly, we also identified initial programmatic and technical risks and offered several mitigating strategies that need to be strongly considered before any future steps in the systems design process.

CHAPTER 4

Investigative Graph Search- A Technical Approach

4.1. Introduction

This chapter introduces the foundational technical approach of the proposed radicalization detection system framework: *investigative graph search* and the employment of graph pattern matching for use in law enforcement investigations and intelligence analysis to find a pattern of indicators for a latent behavior in a large heterogeneous graph. Recalling the radicalization detection system framework described in Section 3.3 and depicted in Fig. 3.2, we note that this approach is predicated on the implementation of the entire framework, including the access to and proper classification/labeling data from open and restricted sources to produced a large heterogeneous data graph as well as the radicalization query pattern from criminologist and terrorism study experts.

In the era of big data and user-generated content, one of the most pressing needs in many applications continues to be filtering unnecessary/irrelevant data and finding the desired information so that one can make timely and accurate decisions [193]. Since much of the data in a variety of domains can be conveniently represented as heterogeneous data graphs, graph pattern matching is of ever growing importance to find such information. While a recent overview [193] lists complex object identification, software plagiarism detection, and traffic route planning as some additional applications, a bulk of the research in this field is oriented towards social search and recommender systems [13, 49, 100, 103, 162, 192, 193, 247, 297]. In social search, for instance, one may utilize graph pattern matching to find an entity with specific types of connections or attributes, while recommender systems help individuals form collaboration networks with people with specific skills and expertise.

In the application domains of law enforcement and intelligence analysis, we make the analogous extension of a graph pattern matching framework to finding radicalizing individuals on large heterogeneous graphs who exhibit indicators through their on- and off-line behaviors and associations with other individuals and may be on pathways to carryout extremist violence.

This chapter includes a survey of the related graph pattern matching literature, the introduction of a necessary node categorization for investigative indicators, and the development of Investigative Simulation (InvSim), an extension to an existing graph pattern matching scheme to make it appropriate for intelligence analysts and law enforcement officials.

4.2. Investigative Graph Search

While graph pattern matching approaches are efficient and robust in application, most rely on the certainty of specific types of connections or attributes in the query pattern. In reality, one may be much less certain about the query structure, or the entities of interest may not exhibit all of the possible behaviors or attributes. This is especially true in the search for those undertaking latent behaviors, which we define as hidden or emergent activities exhibited by an entity [109]. We, therefore, introduce the concept of *investigative graph search*, which is the process of searching for and prioritizing persons of interest who may exhibit part or all of a pattern of suspicious behaviors or connections. It is particularly relevant to search for those on radicalizing towards violent extremism. Some distinguishing characteristics of investigative graph search from other searches include:

(1) Nodes in a query pattern are hypothesized indicators of a latent behavior of interest; all indicators may not or need not appear in the matched result to make a partial match worthy of further investigation.

(2) Some indicators nodes are only significant in the context/presence of other indicators.

(3) The ranked full or partial match results should help analysts prioritize among potentially many matches based upon the presence of red-flag indicators as well as the similarity of the matches to the query.

(4) While the hypothesized indicators may or may not have a known sequence of occurrence, the rate and trajectory by which someone exhibits the indicators in a query pattern are often of interest to investigators.

4.2.1. A MOTIVATING EXAMPLE IN THE STATIC SETTING. A small example problem related to the investigative search for homegrown violent extremists is shown in Fig. 4.1a. The top left graph is a simplified query graph Q of some possible indicators of a homegrown violent extremist. The pattern is a person who 1) posted radical- and extremist- labeled n-grams from a social media account, 2) underwent suspicious travel to a foreign country and received terrorist-related training, and 3) purchased a firearm. This is only an example for illustration, and the behaviors characterizing such queries need to be generated by experts on radicalization behaviors. The top right is a simplified data graph G of 4 people each with various on- and off-line activities.[33] The problem is to find all whole or partial matches of the query Q in the data graph G and present results according to some intuitive ranking scheme.

[33]As noted previously, our approach is predicated on access to and proper classification/labeling data from open and restricted sources to produce a large heterogeneous data graph. Details of these aspects of the proposed radicalization detection system framework were discussed in Section 3.3. For this example, it is clear that social media data (e.g, Twitter, Facebook) was envisioned to be fused with firearm background check databases and local/state/federal criminal and terrorist databases (including data from the FBI's Tripwire and 'FBI Tips' programs as well as the TSA's Automated Targeting System and SecureFlight programs).

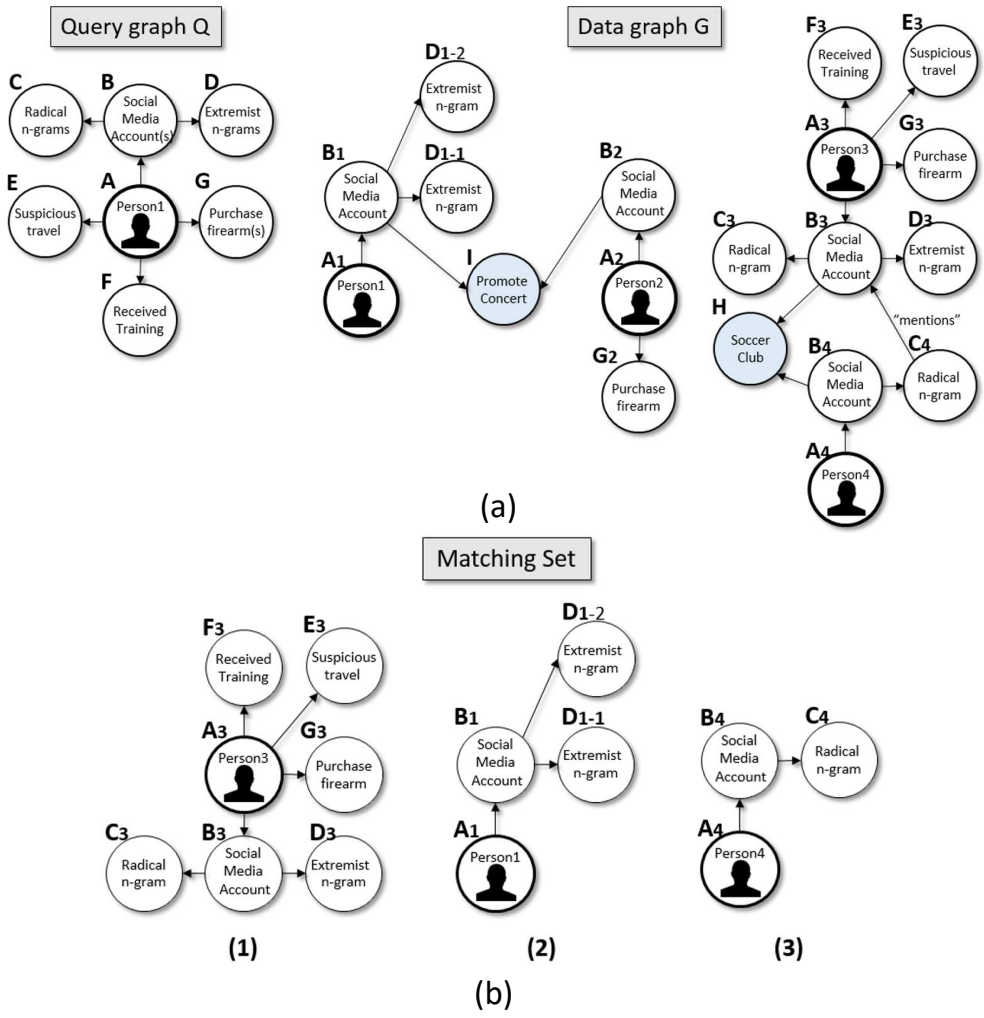


FIGURE 4.1. (a) An example graph query of a potential homegrown violent extremist and a fictitious data graph of 4 people with on and off-line activities. Nodes in the data graph represent distinct entities (person or social media account) or behaviors (posting extremist n-gram, purchasing a firearm, etc.) with the class label shown inside the node. The letter of the label outside the node is a code for the class and the number (if applicable) denotes the person responsible for that entity or behavior. The query graph represents the pattern of nodes (by class) that may help identify potential homegrown violent extremists. (b) Desired matching set that includes full and partial matches in rank order.

The ideal matching results from investigative graph search are shown in Fig. 4.1b. Person 3 (and his related activities) is the only complete match for all indicators and is returned as the top suspect. Person 1 and 4 (and their related activities) should also be returned as partial matches due to the posting of radical and extremist n-grams. Note that Person 2,

despite having a social media account and purchasing a firearm, should not be returned because neither indicator is important unless there are other suspicious indicators of motivation or intent to commit targeted violence.

This simplistic, static example, captures the essence of the technical problem we propose to assist law enforcement and intelligence analysts effectively screen for and prioritize individuals who may be on pathways to extremist violence. Of course, there are remains challenges with both scale and dynamics, which we will address in subsequent sections in this thesis.

4.3. Categorical Node Labeling for Investigations

In order to return such a match results above, it was necessary to recognize the categories of indicators consistent with the threat assessment literature and to impose this categorical structure on the nodes. We discuss this important node weighting in this section.

Specifically, investigative queries may contain nodes that are representative of perfectly legal and innocuous activities that are only potential indicators of a latent behavior of interest when they occur with other indicators. For example, purchasing a firearm may only serve as a targeted violence threat indicator if it is accompanied by an overt communication of threats to others. At the same time, it is possible to identify indicators, should they occur, which may *be individually sufficient* to warrant further investigation. While we could use a numerical node weighting scheme to ensure such indicator/node differences (as in [13] and [318]), we suspect that node weights may change when in the context of other indicators and as such complicate the matching process. For example, while we may initially want the indicator for purchasing a firearm to have a relatively low weight when it is the only indicator

TABLE 4.1. Labels for investigative node-types

Node type	**Definition**	**Source**	**Examples**	**Sufficient for further investigation**
Query focus (QF)	Subjects of an investigative query (i.e., we want individuals who match a particular pattern).	[103]	1) Person	No
Individually innocuous but related activity (IIRA)	Behavior that is individually not harmful or threatening (and often common), but in this context may be indicative of a threat when combined with something else.	N/A	1) Buying a gun 2) Having a social media account	No
Indicator (IND)	Term used to broadly classify those unusual behaviors which may suggest that an individual is a threat and work more as potential building blocks towards a threat assessment. Can encompass behaviors, traits, characteristics, risk factors, and warning signs.	[81, 271]	1) Fascination with weapons 2) Purchasing large quantities of fertilizer 3) Downloading the Anarchist Cookbook 4) Re-tweeting a violent jihadist video 5) Watching a video from a violent extremist preacher	No
Red flag indicator (RF)	Risk factors which, if present, will singly determine that a case ranks as a high risk or concern until proven otherwise [203].	[203, 310]	1) Motives for violence [310] 2) Homicidal ideas 3) Fantasies or preoccupations 4) Violent intentions or expressed threats 5) Pre-attack planning and preparation 6) Received terrorist training overseas	Yes

of a person as a threat, such an indicator would likely be weighted much more heavily when it also occurred with the same person making overt threats [81].

We thus propose a categorical weighting of nodes to the set of node/edge labels Σ based upon research from the threat assessment and homegrown violent extremist radicalization literature, which has generally advocated for condition-based weighting. See Table 4.1 for the definitions and examples provided. For clarity, we also show in Fig. 4.2 the labeling of each of the query nodes in the pattern original shown in Fig. 4.1.

The first category *query focus (QF)* is used to label nodes which are the subject of the investigative query– namely the people in the data graph. The second category *individually innocuous but related activity (IIRA)* is for activities which need to occur in conjunction with other (more suspicious) indicators to be worth further examination. The third category *indicator (IND)* is a broad term to classify unusual behaviors which may suggest a person is a threat. Lastly, the fourth category *red flag indicator (RF)* are for activities which are *individually sufficient* to warrant further investigation. If a node does not fall in any of these categories, it can be labeled as *no category (NC)*.

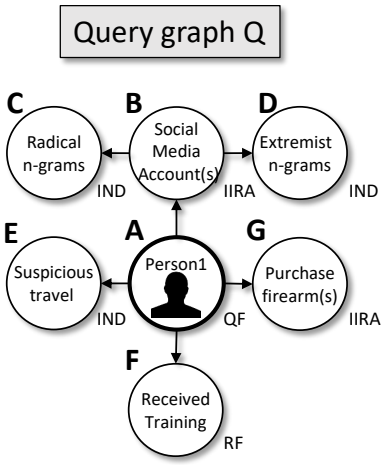


FIGURE 4.2. Consistent with the node category definitions and examples, we label the nodes in the example graph query pattern Q as follows: query focus (A), individually innocuous but related activity (B,G), indicator (C, D, and E), and red flag indicator (F).

Overall, the purpose of these investigative node categories is either to avoid over-matches to query patterns based on individually innocuous behaviors or to facilitate alerts when a match includes a red-flag that is sufficient for further investigation. This modeling extension, as will be shown in subsequent sections, is a critical component in making existing graph pattern matching schemes useful to law enforcement and intelligence analysts as well as developing our own novel graph pattern matching technique.

4.4. Investigative Simulation- Static Graph Pattern Matching

4.4.1. Introduction. Although investigative graph search is largely based social search and graph-based recommender systems that depend are quite ably handled with existing graph pattern matching settings, we find that these techniques fall short of achieving desirable results in our setting. We return to the motivating example in Fig. 4.1 in Section 4.2 and recall that we identified the ideal match set for the query Q. For example, dual simulation [192, 193], which represents the near state of the art in simulation-based graph pattern matching approaches, technically returns only Person 3 (and his related activities) as the only matching connected subgraph. Because the algorithmic implementation [192] [193] of dual simulation also returns remnant individual node matches, the n-grams from Person 1 and 4 (D_{1-1}, D_{1-2}, C_4) as well as the 'purchase firearm' node (G_2) from Person 2 are partial matches. From this, we identify three shortcomings of dual simulation for investigative search: 1) the requirement for every node in the query to have some match (and no allowance for partial matches) is too restrictive when an investigator may include indicator nodes in the query which need not nor may not all be associated with every person of interest), 2) any remnant node matches with valid matching indicators do not contain the

subject of the search, and 3) remnant node matches may contain nodes that are innocuous activities except when observed with other suspicious indicators.

In this section, we develop our modification to dual simulation, which we call *investigative simulation*. It is designed to address these aforementioned shortcomings and returns the ideal match result in Fig. 4.1b. Person 3 (and his related activities) is still the only complete match for all indicators. However, Person 1 and 4 (and their related activities) are now also returned as partial matches due to the posting of radical and extremist n-grams. Note that Person 2 is *no longer* returned as a match despite having a social media account and purchasing a firearm because neither indicator is important unless there are other suspicious indicators of motivation or intent to commit targeted violence.

4.4.2. Related Work. Our work with investigative simulation builds upon advances in graph pattern matching in the static setting. Several surveys exist, including [113] and [23]. Of the two principal types of matching, exact and inexact, we focus our efforts on the state of the art in inexact matching due to its flexibility for returning results in the presence of noise or errors in the data [113]. The notable works in static inexact matching include *best-effort matching* [297], TALE [295], SIGMA [215], NeMa [162], and MAGE [247]. The 'inexact' component of these works primarily involves the allowance for finding nearby matches for nodes in which an exact match does not exist.

Of these, the work most closely related to ours in intention is [247], which first introduced a graph pattern matching method that supports exact and inexact queries on both node and edge attributes as well as wildcard matches. This matching notion specifically cites intelligence analysis as a use-case and offers great flexibility in the query construction that would allow analysts to explore the unknown or uncertain connections. However, this matching

scheme still does not truly support uncertain indicator-type matches nor innocuous nodes which become significant only in the context of other indicators.

Equally important are simulation-based matching schemes, starting with bounded graph simulation [100, 102] to find meaningful matches given a pattern graph with arbitrary or specified path lengths in the connections, and *dual and strong simulations* [102, 192, 193] by preserving query graph topology through enforcement of both parent and child relationships in the match and imposing locality constraints.

For the purposes of investigative search where a person may exhibit an indicator behavior one or more times (and each instance is counted as an appropriate match), we find that simulation-based approaches may be most appropriate due to the allowance of each query node to be matched to multiple nodes in the data graph as long as match labels are preserved at the match-level, as well as with the parent- and child-levels. However, as previously mentioned in Section 4.4, dual simulation has several shortcomings with investigative search including the lack of allowance for partial matches, incomplete remnant node matches, and lack of ability to handle matches of innocuous activities that become important only when observed with other suspicious indicators.

Lastly, because most graph queries end up returning many matches given a large graph, researchers have also devised ways to rank the most relevant matches using various goodness functions. Such criteria include social impact [103], social diversity [103], structural similarity [13, 162, 297], weighted attribute similarity [13], or label similarity [162]. As sophisticated as these ranking methods are, we find that none account for intuitive red-flag indicators (i.e., those matches which demand the immediate attention of an analyst) that are relevant in investigative searches.

4.4.3. Technical Preliminaries and Notation. Before we define investigative simulation in the next section, we first review graph terminology and notation, as well as the dual simulation graph pattern matching notion.

Definition 1. Graph [192]. In our work, both the query graph Q and data graph G are identically defined as a directed graph of the form $G(V, E, L)$, where V is a set of nodes, $E \subseteq V \times V$ is a set of edges, in which (u, u') denotes an edge from node u to u'; and $L : V \cup E \rightarrow \Sigma$ is a labeling function which assigns nodes and edges to a set of labels Σ.

Definition 2. Dual Simulation [192, 193] Graph G matches a pattern Q via dual simulation if there exists a binary match relation $S_D \subseteq V_Q \times V_G$ such that:

- for all nodes $u \in V_Q$ there exists a node $v \in V_G$ such that $(u, v) \in S$; and
- for each pair $(u, v) \in S, u \sim v$ (i.e., $L_Q(u) = L_G(v)$), and for each edge $(u, u') \in E_Q$ there exists an edge $(v, v') \in E_G$ such that $(u', v') \in S_D$, and for each edge $(u', u) \in E_Q$ there exists an edge $(v', v) \in E_G$ such that $(u', v') \in S_D$.

We then refer to S_D as a match (via dual simulation) to Q.

Dual Simulation was a significant advancement over graph simulation and previous notions because it preserved not only parent-child relationships but also child-parent relationships in the match (and thus produced more sensible matches). However, as described in Sections 4.4 and 4.4.2, there are shortcomings when performing investigative search. We summarize the notations we utilize in this paper in Table 5.1.

4.4.4. Investigative Simulation. The proposed investigative simulation approach is described next, first by providing a formal definition, and then devising an algorithm for this new matching notion.

TABLE 4.2. Summary of notations

Notation	Description/Meaning
G	Data graph $G(V_G, E_G, L_G)$
Q	Query graph $Q(V_Q, E_Q, L_Q)$
S_D	Binary Match relation (via dual simulation)
S_{InvSim}	Match relation (via investigative simulation)
(u, u')	Directed edge from node u to u'
$u \in V_Q$, $v \in V_G$	Nodes with index u (v) are in graph V_Q (V_G), respectively.
$R_{(u,v)}$	Relevant set of matching node $v \in V_G$ w.r.t. query node $u \in V_Q$.

DEFINITION 3. Investigative simulation: An extension of dual simulation for investigative search. Graph G contains (partial or complete) matches of pattern Q if there exists a binary match relation $S_{InvSim} \subseteq V_Q \times V_G$ such that:

- for all nodes $u \in V_Q : L(u) =$ *'QF'* and at least 1 node $u \in V_Q : L(u) =$ *'IND'* or *'RF'* there exists a node $v \in V_G$ such that $(u, v) \in S_{InvSim}$;
- for each pair $(u, v) \in S_{InvSim}$, where $u \in V_Q$ and $v \in V_G$, $u \sim v$ (i.e., $L_Q(u) = L_G(v)$), and
- for each edge $(u, u') \in E_Q$ there exists an edge $(v, v') \in E_G$ such that $(u', v') \in S_{InvSim}$, and for each edge $(u', u) \in E_Q$ there exists an edge $(v', v) \in E_G$ such that $(u', v') \in S_{InvSim}$.

We then refer to S_{InvSim} as a match (via investigative simulation) to Q.

Instead of all nodes in Q needing a match in G (as in dual simulation), investigative simulation allows for partial matches by only requiring all 'QF' nodes and at least 1 indicator node (regular 'IND' or red-flag 'RF') to have a match in G. This keeps matching results specific to 'QF' nodes with at least a single indicator-type that would may make it worthy of further analysis or investigation.

4.4.5. INVSIM- EXTENSION OF DUAL SIMULATION ALGORITHM. One of the merits of the 'DualSim' algorithm for dual simulation found in [192, 193] was that it returns the entire binary match relation S_D, which contains not only complete matches (connected component

subgraphs for all nodes in Q) but also remnant node matches (those nodes whose parent and/or child were pruned away as a result of their connections). While [192, 193] never use these remnant node matches in the construction of the maximum subgraph, they in fact form the basis of the partial matches that are informative for investigative searches. However, they are by nature incomplete matches (e.g., in the case of the network schema in the motivating example problem in Fig. 4.1, the query focus nodes associated with remnant indicator nodes were not in the match relation). We develop a post-processing extension to the dual simulation algorithm (Algorithm 1: InvSim, short for Investigative Simulation) that corrects both issues specific to indicator-type patterns. In it we utilize a modified 2-hop concept of a relevant set from [103]. Given a match v of a query node u in V_Q, the relevant set of v w.r.t. u (denoted as $R_{(u,v)}$) includes all matches v' of u' for up to 2-hop descendants u' of u in V_Q.

Lines 1-9 are those lines found in [192] and [193], which is the implementation of the dual simulation algorithm and results in the intermediate match relation S_D. Our post-processing extension to this algorithm begins in Line 10, when we iterate through all nodes in the matching set in data graph G which are labeled as 'QF' or query focus (i.e. persons). If the intersection of the relevant set of v and the match relation S_D for nodes $\tilde{v}$ of type 'IIRA' or individually innocuous but related activity, then we remove this node from the match relation and remove its query focus parent if it was in the relation (Lines 11-12). This effectively removes matching nodes that are considered benign without the presence of other indicators, as well as the associated person from further consideration.

Next, we ensure that the parent query focus nodes of other matching nodes of type 'indicator' are included in the match relation S_D. We first search for all nodes $\tilde{v}$ that are

Algorithm 1: InvSim (for Investigative Simulation)

Input: Query graph Q with investigation category node labels, and data graph G
Output: The match relation S_{InvSim} of Q and G

1 **foreach** $u \in V_Q$ **do**
2 $\quad sim(u) := \{v \mid v \in V_G$ **and** $L_Q(u) = L_G(v)\}$
3 **while** there are changes **do**
4 $\quad$ **foreach** edge $(u, u') \in E_Q$ and **each** node $v \in sim(u)$ **do**
5 $\quad\quad$ **if** there is no edge (v, v') in G with $v' \in sim(u')$ **then** $sim(u) := sim(u) \backslash \{v\}$;
6 $\quad$ **foreach** edge $(u', u) \in E_Q$ and **each** node $v \in sim(u)$ **do**
7 $\quad\quad$ **if** there is no edge (v', v) in G with $v' \in sim(u')$ **then** $sim(u) := sim(u) \backslash \{v\}$;
8 $\quad$ **if** $sim(u) = \emptyset$ **then return** $\emptyset$;
9 $S_D := \{(u, v) | u \in V_Q, v \in sim(u)\}$
10 **foreach** node $v \in S_D$ where $L(v) =$'QF' **do**
11 $\quad$ **if** $L(\tilde{v}) =$'IIRA' for all $\tilde{v} \in R_{(u,v)} \cap S_D$ **then**
12 $\quad\quad sim(\tilde{u}) := sim(\tilde{u}) \backslash \{\tilde{v}\}$; **if** node $v \in S_D$ **then** $sim(u) := sim(u) \backslash \{v\}$;
13 $\quad$ **if** there exists a node $\tilde{v} \in R_{(u,v)} \cap S_D$ **and** $v \notin S_D$ **then**
14 $\quad\quad sim(u) := sim(u) \cap \{v\}$; **if** every node $\check{v}$ along the shortest path from $(v, \tilde{v})$ is not in S_D **then** $sim(\check{u}) := sim(\check{u}) \backslash \{\check{v}\}$, $\forall(\check{u}, \check{v})$;
15 $S_{InvSim} := \{(u, v) | u \in V_Q, v \in sim(u)\}$
16 **return** S_{InvSim}.

both in the relevant set of v and the match relation S_D but whose parent query focus node v is not yet in S_D (Line 13). We join this node v to the match relation (Line 14) as well as add all nodes in the shortest path from node v to $\tilde{v}$ in the match relation (Line 14). Finally, we consolidate and return the modified match relation S_{InvSim} (Line 15-16).

4.4.6. Results.

4.4.6.1. *Real dataset for a proxy investigative search.* In order to test investigative simulation on real data, we utilized the BlogCatalog dataset,[34] which is a scrape taken in July 2009 of a social media site that allows users to register/promote their own blog and connect with other bloggers. The graph had over 470,000 nodes and over 4 million edges; it is further detailed in Table 4.3. The network schema shown in Fig. 4.3 describes the node types and

[34]Available at http://dmml.asu.edu/users/xufei/datasets.html

connections present the network. In essence, an ID owns a User_Id, which in turn both authors blogs with a Weblog_Id as well as forms directed friendship connections with other User_Ids. Lastly, each weblog will provide one or more user-specified tags.

TABLE 4.3. BlogCatalog Graph Characteristics

Characteristics	Value
Total Nodes	**471,267**
Number of ids	88,781
Number of userids	80,949
Number of weblogs	127,227
Number of unique tags	174,310
Total Edges	**4,098,290**
Number of links from id to userid	88,784
Number of links from userid to userid	3,223,640
Number of links from userid to weblog	127,227
Number of links form weblog to tags	658,639

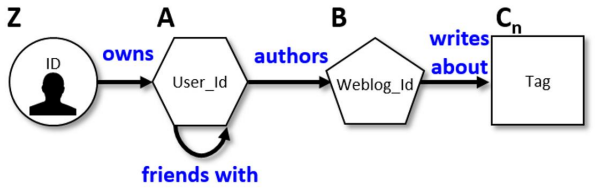


FIGURE 4.3. Network schema of the BlogCatalog graph. IDs own account User_Ids. User_Ids author one or more Weblog_Ids, and are friends with other User_Ids. Weblog_Ids write about one or more tags (which are user specified).

4.4.6.2. *Query Description.* To test the performance of the matching scheme and algorithm, we devised a proxy query on a benign subject matter with structural parallels to investigations. The query focus is for user IDs who had been writing blogs about Microsoft Windows operating systems (XP and/or Vista) and subsequently also began to write about Windows 7 when it was released in July 2009 (the month in which the data was collected). Node Z is a true person ID, node A is the user ID query focus, and node B is the weblog with certain tags. All C nodes are meant to be seen as labels of a post or blog entry (i.e.,

determined through machine-classified semantic analysis). The labels 'computer' (C535) and 'windows' (C2033) are IIRA (i.e., relatively frequent labels which help provide context or additional clarity on the true topic set), and labels 'xp' (C23136) and 'vista' (C20693) are indicators that the blog is about Windows operating systems (i.e., necessary but not sufficient for trajectory behavior). Finally, label 'windows 7' (C20684) is considered a red flag indicator.

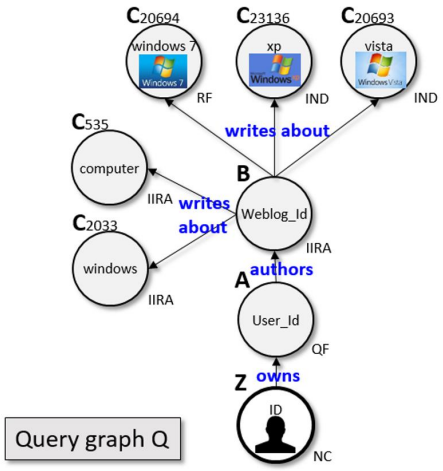


FIGURE 4.4. Experiment query for BlogCatalog. Query focus is for User_Ids who had been writing blogs broadly related to 'computers' and 'windows', and specifically to Windows operating systems. In this example, we treat the tag 'windows 7' as a red flag indicator.

4.4.6.3. *Ranking Method and Analysis.* As expected, investigative simulation returned meaningful partial matches to the query. Our intuitive ranking scheme for the top-k results was to 1) first order by the presence of any 'QF' nodes with red-flag ('RF') indicators, and 2) followed by the size of the relevant matching set for each 'QF' node (i.e., in decreasing order of $|R_{(u,v)} \cap S_{InvSim}|$, where $L_Q(u) = L_G(v) =$ *'QF'*). This method effectively highlights to analysts those first who have red-flag indicators, followed by the those who have the most indicators towards the latent behavior of interest.

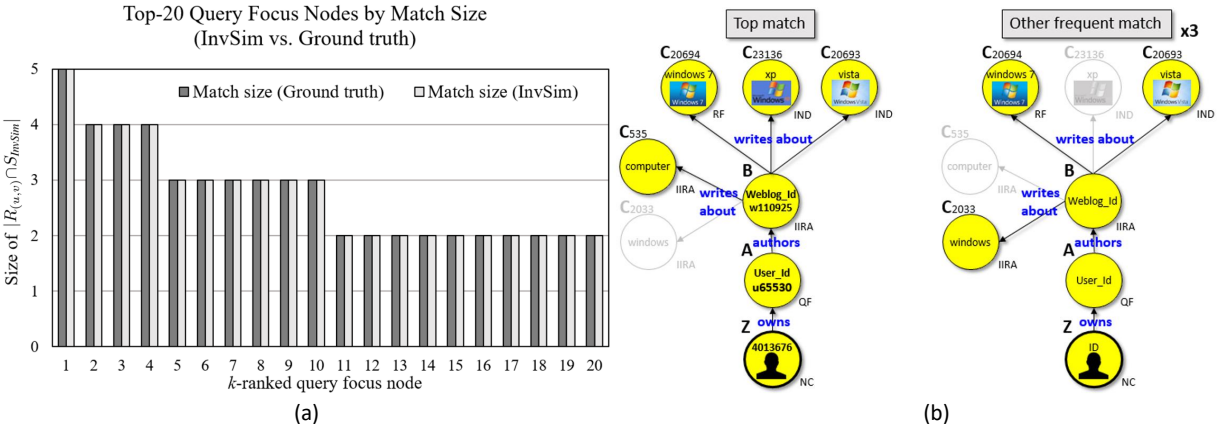


FIGURE 4.5. (a) A paired bar graph showing the exact correspondence of the top-20 query focus nodes by match size between both InvSim and exhaustive search, where $|R_{(u,v)} \cap S_{InvSim}|$is the number of matching nodes in the relevant set of the query focus node. (b): Top-4 results of investigative simulation on the BlogCatalog dataset with the query in Fig. 4.4. The top-match is User_Id 'u65530' with 5 indicator nodes matching in the relevant set (2 directed hops from Node A). Note the presence of the red flag indicator 'windows 7' in each of these matches. The grayed-out nodes were the original query nodes not matched.

We find that investigative simulation performed well in the matching, as measured with both quantitative and qualitative methods. First, we quantitatively measured the similarity between the top-20 results with the original query pattern by using Jaccard similarity and compared it with the top-20 ground truth results acquired through an exhaustive search. Qualitatively, we performed subjective validation of the sensibility of each of the top-10 match results. The top-4 partial matches to the query are shown in Fig. 4.5.

4.5. Discussion of the Technical Approach Over Other Possible Approaches

In this section, we briefly discuss the cause for investigative graph search as a proposed technical approach over other possible approaches.

4.5.1. Pattern detection over anomaly detection detection approach. To identify and screen for individuals at risk for violent extremism, we chose to utilize the

investigative graph search technique based on graph pattern matching rather than anomaly detection. Anomaly detection techniques are methods to identify outliers or "patterns in data that do not conform to expected behavior" [46]. Used traditionally in credit card fraud and computer network intrusion detections among other applications [46], it has been most recently utilized as well in insider threat detection. [240, 265, 277]. Anomaly detection seems particularly suited for identifying insider threats for a number of reasons including:

- Companies, organizations, or network monitoring services have access to vast amounts of data related to all aspects of computer usage (login and logout times, normal work flow, USB usage, etc.) [201]
- Companies and organizations have prescribed computer-use policies from which policy violations are detectable [216]
- Companies and organizations generally have hierarchical roles and associated norms for users in the same role. In these cases, it is generally known who should have access to certain files or programs, and peer activities that can serve as a baseline of behaviors (See [216] as well as the survey [265] of related work in this area).

Such norms for work procedures or computer/network use as well as the availability of employee roles for base-lining behaviors do not transfer to the population of individuals at large outside of established organizations. Rather, in the domain of risk assessments for violent extremists and the radicalization process, there is a large body of work proffering the likely early warning indicators and patterns of suspicious behavior that lend to more direct pattern matching techniques. Moreover, outside of established organizations and their closed, monitored systems, it is much harder to characterize on and off-line behaviors that would be

considered "normal." Lastly, [66] also suggests that anomaly detection approaches to identify the unusual behaviors from normal behaviors are predicated on the broad, "behavioral monitoring" that would have serious civil liberty issues [66, p. 150]. Because of these reasons, we initially pursued a pattern matching approach over an anomaly detection approach.

4.5.2. Analyst decision aid using risk indexes over Bayesian inference techniques. As stated from the outset, the proposed technology is designed to mine, monitor, and screen for those individuals who exhibit behavioral indicators of violent extremist radicalization. Inherent in this is technology is the role of enabling non-actuarial structured professional judgment instruments at scale and through the integration of specific databases. On the other hand, Bayesian inference is a key concept that is central to machine learning and prediction because in many cases, it is easier to get conditional probabilities in one 'direction' of inference and Bayes rule can recover the conditional probabilities in the other direction. For example, in the study of radicalization [116, 168] identified the prevalence of certain pre-incident indicators or warning behaviors in known violent extremists (i.e., conditioned upon his/her being a violent extremist, this is the probability that this indicator or behavior is present). Ultimately, however, one is trying to determine whether an individual wants to commit extremist violence, given that the presence of certain indicators or behaviors. Bayesian inference theoretically allows us to determine this through the following formulation, which is a modified version of the one found in a RAND's "Using Behavioral Indicators to Help Detect Potential Violent Acts" [66, p. 191-195].

Let V signify that a person is a violent extremist, and $\boldsymbol{I} = \{I_1, I_2, ...I_m\}$ be the set of m indicators of violent extremism. Then $P_t(V|\boldsymbol{I})$ is the conditional probability that a person is a violent extremist at time t given the set of indicators $\boldsymbol{I}$ can be determined by Eq. 1.

$$P_t(V|\boldsymbol{I}) = \frac{P_{t-1}(\boldsymbol{I}|V)P_{t-1}(V)}{P_{t-1}(\boldsymbol{I})} \tag{1}$$

While researchers have provided insights into $P_{t-1}(\boldsymbol{I}|V)$ (the set of indicators present when one is a violent extremist) and an estimate for $P_{t-1}(V)$ (the base rate for violent extremists), no research exists that we have found to address $P_{t-1}(\boldsymbol{I})$ (the base joint probability of indicators present in the population whether one is a violent extremist or not). The latter is complicated not only because it would involve the estimate for the prevalence of a particular indicator in the entire population (for example, whether one posts a 'radical' statement on social media), but also because we cannot assume independence of the indicators of extremist violence. While several researchers [35, 231] have proposed the use of Bayesian inference in violent extremist detection, the proposals were never implemented on real data. In the end, researchers in [66] were skeptical of the true utility of Bayesian inference in this application and stated that the methods, "are unlikely to go very far...except for the simplest of instances" [66, p. 195].[35]

The key point is that this research does not aim for prediction. Our proposed radicalization detection system has both a data management component as well as a dynamic pattern recognition technology. The former emphasizes data fusion and push alerts to help law enforcement agencies maintain awareness of the activities of persons of interest, while the latter tracks indicator behaviors over time to aid law enforcement agencies in screening for those at higher risk for extremist violence. Ultimately, the system we propose can identify potential

[35]For instance, the formulation in Eq. 1 would be more realistic and complex if we had to distinguish different types of violent extremists (such as foreign fighters, those who provide material support, or those seeking to commit domestic plots) each of whom could have their own set of partially overlapping indicators.

threats and lower the number of false positives with the benefits of artificial intelligence and human-factors [201, p. 1].

4.5.3. Assessing risks over providing prediction. Prediction of any human behavior is an ambitious goal, especially when we are considering individual intentions and behaviors rather than group or locality outcomes as is now prevalent in "predictive policing." Consider the area of predicting an individual's physical path trajectory when enabled by the global positioning systems in smart phones. One paper observed, "Why is there any hope that good predictions can be computed? Typically, people try not to waste time and move at least partially on shortest or quickest paths. It is only this observation that allows us to come up with any prediction at all" [93, p. 7]. However, there is a paucity of terrorist and radicalization research that suggests any cognitive or behavioral equivalent, and we conclude that estimating likelihoods probabilistically is very difficult. Others have outright dismissed actuarial-based tools that establish a procedure for combining risk factor scores or producing a likelihood estimate for violence as "patently infeasible" due to the statistical power needed given the small sample sizes of terrorists [214].

Meloy stated, "The problem of describing risk of intended or targeted violence in any given individual (as opposed to a group) is the very low base rate in any population under consideration, and the guarantee of an unacceptably high false-positive rate" [203, p. 257].[36] Moreover, the problem is hard because of the terrible consequences of false negatives as well (i.e., those individuals whom law enforcement deem as low risk, and then who end up successfully carrying out an attack). False negatives first and foremost put the safety and

[36]The use of the term "unacceptably" is obviously a subject of debate. Clearly, false positives are less favorable in the terrorism context than in comparison to, for instance, credit card fraud detection. While the latter could be resolved with a (sometimes automated) call from the credit card company to the customer to verify a transaction, the former might involve the use of additional law enforcement time and resources to resolve and possibly cause significant ill will when one is falsely investigated)[268].

security of innocent lives at risk, but also erode the trust and confidence in law enforcement and causes further scrutiny of their actions [266].

Rather, we take a much more established threat risk assessment approach and seek to assist law enforcement in the prevention task without necessarily needing to *predict* violent extremism. Meloy, a consultant for the FBI Behavioral Analysis Units, wrote extensively about this subject. First, he stated that "Risk factors allow the separation of individuals into risk groups, typically high, medium, or low. Typing someone as high risk is not a probability estimate that he or she will behave in a violent way; rather, it is a statement that the subject shares important statistical associations with that group of people from which the few individuals who will go on to commit the behavior are most likely to emanate"[203, p. 257]. Furthermore, he stated that "prevention [of terrorist violence] does not require prediction... detect[ing] the proximal indicators of concern for law enforcement [can] narrow the focus of an investigation, prioritize cases, and help plan a timely risk-management intervention" [204]. In summary, an empirically-tested dynamic radicalization risk assessment protocol that more reliably anticipates violent action would greatly assist law enforcement and NGOs, but a supporting technology that can automatically detect the likely presence of those radicalization indicators would help them all even further by screening and providing alerts for those most at risk, and allowing for better decision making under resource constraints.

4.5.4. Dynamic risk assessment over one-time determinations. We take a dynamic approach to risk assessment primarily because risk assessments, particularly for the fluid processes of radicalization, must be repeated to consider the latest intelligence or behavioral information. Meloy wrote that "Risk is a dynamic process and it is necessary to repeat

the consideration of risk factors (the "threat assessment") in the light of new information, as each occasion that risk is considered constitutes simply a "snapshot" of a moving scene- a still frame in movie [203, p. 259]. This observation, supported by [271], has empirical support where unfortunately individuals who were deemed not a risk (or at least not an imminent risk) by law enforcement agencies previously went on to carry out acts of extremist violence. See for instance Tamerlan Tsarnaev [106], Omar Mateen [56], and Ahmad Khan Rahami [217].

4.5.5. Discussion of law base-rates. The base rate fallacy is the "the fallacy of allowing indicators to dominate base rates in your probability assessments" [9, p. 212]. The caution against this fallacy, often called base-rate neglect, has been applicable to many different areas, to include most recently in the detection of potential terrorists or violent extremists. For example, [264] states that analysts in the intelligence community "suffer from low base rate neglect for very rare [terrorist] events" and that "[m]uch of their time is spent investigating obvious false alarms, sometimes losing track of important developments" [262, p. 10]. Others, such as [266, 268], argue vehemently against the use of quantitative approaches to conducting these risk assessments, stating "actuarial risk assessment systems cannot work" [266, p. 283] and "data-mining systems won't uncover any terrorist plots until they are very accurate, and that even very accurate systems will be so flooded with false alarms that they will be useless" [268].

Sarma's presentation and analysis of the low base rate problem in the risk assessments of terrorists in a leading psychology journal [266, p. 283] were misleading. He proposed a hypothetical example of a population of 100 individuals where 80 are non-terrorists and 20 are terrorists (the percentage of true terrorists is inflated to 20%) and a risk assessment tool

with 90% sensitivity and 70% specificity. He correctly calculated that the tool would detect 18 true positives and 24 false positives with a resulting positive predictive value (PPV) of $18/(18+24) = 0.43$. But then he wrote that this percentage was "worse than chance," implying that a random guessing heuristic would be better at finding terrorists.[37] This can clearly mislead a reader. Indeed, 0.43 is less than 0.50, which is the random chance for success in an experiment with binary outcomes (i.e. flipping a coin). However, Sarma fails to mention that 0.43 should, in fact, be compared to 0.20, the original percentage of terrorists in the population. In fact, the use of this hypothetical risk assessment system effectively increased the odds of finding the terrorist from 2 in 10 to over 4 in 10, a boost of 215%. Taken another way, Sarma's mention of chance is not applicable to the subset of 42 (since those were acquired by the merits of the tool), but to the entire original population of 100. Law enforcement following a random guessing heuristic would in expectation identify as positives 10 terrorists and 40 non-terrorists. The PPV here is $10/(10+40) = 0.20$. One must also keep in mind that this random guessing heuristic fails to detect 10 terrorists who presumably go on to conduct some (devastating) attack, whereas the tool mentioned misses 2.

This critique emphasizes the points by Koehler in [172], who wrote that a decision's maker's goals, values, and task assumptions must be considered in the analysis of low base rate problems. Law enforcement agencies would most likely value a system that makes it two times more likely for them to catch a perpetrator of terrorism. On the other hand, if

[37] A similar misleading argument about data mining technologies being worse than"flipping a coin" using a Bayesian probabilistic framework can be found in a blog by a research fellow at Harvard University [269].

the false positives are an unbearable burden for the agencies to sift through,[38] then perhaps those issues can be addressed with other means such as greater funding, changes in policy, and investment in technologies such as the continuous monitoring system proposed in this thesis.

Others who, rather than becoming naysayers, proposed other techniques in addressing low base rate problems. Researchers in [66] point to the low base rates of people of violent intentions to highlight the importance of future screening technologies that result in population samples in which this propensity is higher. Studies such as [182] have rigorously examined the effect of the base-rate fallacy in screening tests for child mental health problems, analyzed the sustainability impacts of even good psychometric tests, and discussed how sequential screening can be helpful. Once again, as in the case for counterterrorism policy, the efficacy of procedures is inherently tied to some measure of the costs associated with both missed identification and incorrect identification.

In this research effort, we do not neglect the base rates of the latent behaviors we are trying to detect. For example, in Chapter 6 we specifically test the effectiveness of our methods against a ground truth of low-rate behaviors. In all cases, we analyze the contribution of the proposed detection system in how effective it is in screening for individuals at higher risk by comparing the proportions of true positives among the entire population and among the subset of individuals identified for further investigation as a result of the technology.

[38]According to Sageman in [264, p. 11], "law enforcement agencies complain that they are drowned by an ocean of false alarms, which overwhelm their resources. Moreover, he writes, "The major request from the field is help to distinguish the very few true positives that will turn to violence from the vast majority of false positives- young people who brag and pretend that they are tough and dangerous, but, in fact, just talk, talk, talk, and do nothing"[264, p. 11].

4.6. Conclusion

In this chapter, we have introduced investigative graph search as the process of searching for and prioritizing persons of interest who may exhibit part or all of a pattern of suspicious behaviors or connections. We also described our radicalization detection system framework as a holistic analyst-in-the-loop framework to assist law enforcement and intelligence agencies in detecting those on trajectories of violent extremist radicalization.

Lastly, we developed investigative simulation and corresponding matching algorithm as an extension of dual simulation for investigative searches. We show that this form of graph pattern matching produces more sensible matches, more complete partial matches, and less false positives through the imposition of categorical node labels related to indicators

CHAPTER 5

INSiGHT: A Novel Dynamic Inexact Graph Pattern Matching Technique

5.1. Introduction

In this chapter, we operationalize the previously defined concept of investigative graph search in a dynamic variant and formulate the problem of searching for individuals undertaking latent behaviors such as violent extremist radicalization as a unique dynamic graph pattern matching problem on a large heterogeneous graph of individuals and their on- and off-line behaviors. To solve this problem, we develop a dynamic inexact graph pattern matching technique, called INSiGHT (Investigative Search for Graph-Trajectories) that identifies individuals or small groups with conforming subgraphs to a radicalization query pattern and follow the match trajectories over time. INSiGHT is aimed at developing tools for assisting law enforcement and intelligence agencies in monitoring and screening for those individuals whose behaviors indicate a significant risk for violence, and allow for the better prioritization of limited investigative resources.

The outline of this chapter is as follows. In Section 5.2 provides an overview of related work. Section 5.3 provides definitions and notation and reviews technical preliminaries. Section 5.4 outlines our multi-hop class similarity approach for graph pattern matching over time. Sections 5.5 and 5.6 present the results of basic INSiGHT technique applications on two datasets. Section 5.7 details the improvements we made to filter by indicator node-type classes, as well as account for the parameterized time-based decay of indicators and the value

of repeated indicators. In Sections 5.8-5.11 we present first results on some small synthetic graphs, and then extensive experimental results on real datasets.

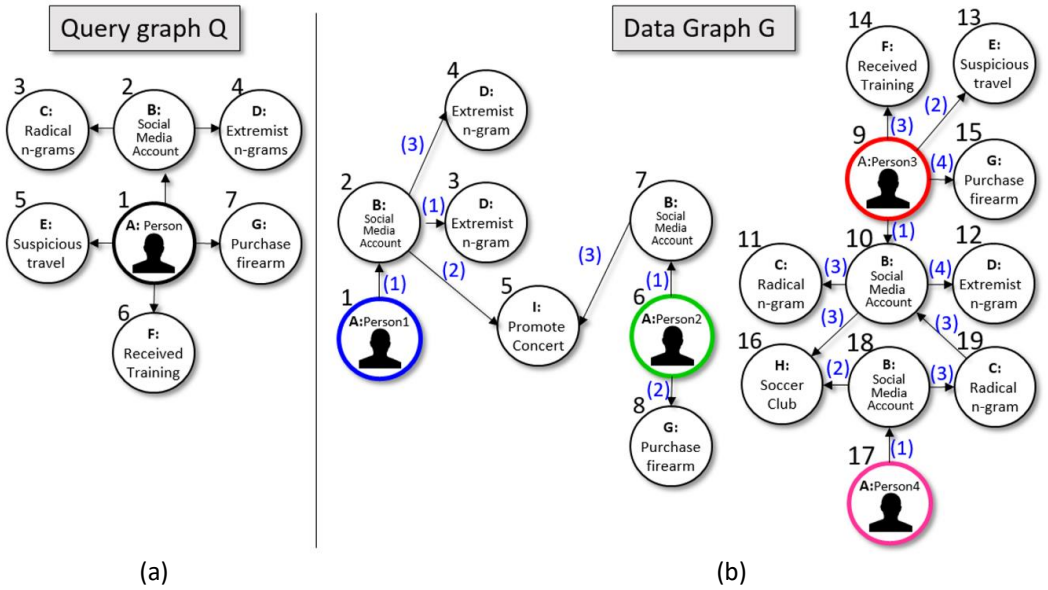


FIGURE 5.1. Motivating example for detecting trajectories of homegrown violent extremists. A small example problem related to the investigative search for homegrown violent extremists. (a) Query graph Q- an example graph query of some possible indicators of a homegrown violent extremist. (b) Data Graph G- a fictitious data graph of 4 people with various associated indicators as on- and off-line activities. The node class labels are inside the node, and the node IDs are outside the node. Each edge has a timestamp (in blue) that denotes the time in which the edge was formed.

5.1.1. A MOTIVATING EXAMPLE RADICALIZATION DETECTION PROBLEM. We begin with a small dynamic example problem to demonstrate the basic approach of finding potential homegrown violent extremists among a group of people and following their radicalization trajectory over time (Fig. 5.1). It is a dynamic variant of the example introduced in Section 4.2. It was also previously used in [144] and is based on the radicalization trajectory framework and methodology from [165] and a sample of behavioral indicators proposed in

[35, 220]. Nodes in the data graph represent distinct entities such as a person or social media account, or behaviors such as posting extremist n-gram or purchasing a firearm.

The query graph Q (Fig. 5.1a) represents the pattern of nodes by class that may help identify potential homegrown violent extremists, but there is no specification on the sequence of occurrence of those indicators due to its variability. The pattern is a person who 1) posted radical- and extremist- labeled n-grams from a social media account, 2) underwent suspicious travel to a foreign country and received terrorist-related training, and 3) purchased a firearm. As before, this is only an example for illustration, and the behaviors characterizing such queries need to be generated by experts on radicalization behaviors. The data graph G (Fig. 5.1b) depicts 4 people each with various on- and off-line activities performed over a period of 4 timesteps. The problem is to find all whole or partial matches of the query Q in G, and to present results of the top matches and the changes to their match of the query over time.

5.1.2. Assumptions. It is important to note that our approach is predicated on a graph model where entities perform behaviors and may be connected to other entities through a limited number of path types. This graph model is exemplified in the previous motivating example. As a result, we make the following modeling assumptions:

- Data graph is a set of distinct, directed, acyclic, entity-based conforming subgraphs of the query pattern.
- Distinct behaviors for each entity are modeled as distinct nodes.
- Relationships between entities are periodically assessed and modeled as bi-directional edges of known path types (such as known direct relations between entities, or known relations via on-line social networks).
- The query pattern contains at most one occurrence of each node class.

5.2. Related Work

The related work in the static setting was already covered in Section 4.4.2 when we developed investigative simulation. Here, we highlight several state-of-the-art *dynamic* graph pattern matching methods worth mentioning for further comparison. Researchers in [49] developed an *exact* subgraph incremental search algorithm for continuous queries. Their system relies on partitioning the query graph, tracking and combining matches with small subgraphs, and specifying a join order in which the small subgraphs are combined. However, this approach is limited to finding exact matching subgraphs and also does not track the full or partial matches over time.

Additionally, [290] proposed a time-based extension of the dual bounded simulation form of graph pattern matching. The data graph was enriched with timestamped edges, which is a modeling practice we adopted. The query graph was also expanded to include a strict allowable edge sequence for the matches to occur. However, we note that the strict edge sequence constraints are likely too restrictive for noisy social graphs that might have connections that appear out of an anticipated sequence. Additionally, this framework still does not return partial matches nor track the trajectory of the matches to the query over time.

Additionally, continuous subgraph pattern search proposed in [47] is also closely related to our work. They developed an approach to check for approximate subgraph isomorphism to a query pattern in graph streams by using a Node Neighbor Tree filtering technique. While continuous pattern search aspect is very similar, our approach calculates similarity scores to a query pattern and is not explicitly searching for subgraph isomorphisms, which is a condition too strict in many real-world applications.

Building upon the many recent advances in dynamic graph pattern matching, we take a unique vectorized approach of investigating the dynamics in multi-hop class similarities between nodes in query graph and data graph over time. By tracking partial match trajectories, we provide another dimension of analysis in investigative graph searches to highlight entities on a pathway towards a pattern of a latent behavior. To our knowledge, no other comparable scheme exists. In future work, we seek to modify some of the aforementioned graph pattern matching approaches to return and track dynamic match similarity scores.

5.3. Technical Preliminaries and Notation

Overall, given a time-independent query graph Q and a time-dependent data graph G, INSiGHT conducts inexact graph pattern matching defined in Definition 4 to find all (including possibly partially conforming) subgraphs G_S and calculates and tracks the multi-hop class similarity between nodes.

Definition 4. Inexact Graph Pattern Matching. The graph matching process to find the binary match relation $S \subseteq V_Q \times V_G$ such that:

- for each of as many nodes $u \in V_Q$ as possible (but at least one), there exists a node $v \in V_G$ such that $(u, v) \in S$, and
- for each pair $(u, v) \in S$, $u \sim v$, and
- for each of as many edges $(u, u') \in E_Q$ as possible (but at least one) there exists an edge $(v, v') \in E_G$ such that both $(u, v) \in S$ and $(u', v') \in S$.

We first review the definitions of the heterogeneous information network (or data graph) as well as the class membership matrix.

Definition 5. Heterogeneous Information Network (or Data Graph). A heterogeneous information network (or data graph) is defined as a directed graph $G = (V, E, f_A, f_T)$, where

- V is a finite set of nodes, and n is number of vertices ($|V| = n$);
- $E \subseteq V \times V$, in which (v, v') denotes an edge from node v to v';
- $f_A(\cdot)$ is a function which associates a node $v \in V$ or an edge $e \in E$ with a tuple $f_A(v) = (A_1 = a_1, ..., A_m = a_m)$, where a_i is a constant, and A_i is referred to as an object/edge class of v, and m is the number of classes.
- $f_T(\cdot)$ is a function which assigns a timestamp from the set Γ to an edge $e \in E$. A timestamp indicates the beginning of the existence of the edge.

It is important to point out that node v or edge e can be members of one or more classes. The presence of more than one class or edge type makes the network heterogeneous. We also acknowledge that graph G is also a function of time t, but refer to it as G for readability. Our approach relies on t to be discrete, but graph updates can be of fixed or varying intervals based upon the application and desire for analysis at regular times or by event.

Definition 6. Class Membership Matrix. A class membership matrix $\mathbf{A}$ is an $n \times m$ integer matrix made up of the tuple class membership vectors associated where every node $v \in V$ has an associated class membership vector $\vec{a_v} = [a_1, ..., a_m]^T$ [297] [102]. In the case of binary class membership, $a_{vk} = 1$ if node v is labeled with the k^{th} class; 0 otherwise. Of course, the class membership value could be weighted as well (to signify the strength of association or membership in a particular class).

In the field of graph pattern matching, class membership of neighboring nodes is of great importance. This is seen in the simulation-based pattern matching approaches [100] [192], where the class memberships of both parent and child nodes are examined before a node is considered 'matching' to a node in the query. Additionally, the inclusion of timestamps in edges is a relatively new but important development in the field of graph pattern matching

TABLE 5.1. Summary of notations

Notation	Description/Meaning
G	Data graph $G(V_G, E_G, f_{A,G}, f_{T,G})$.
Q	Query graph $Q(V_Q, E_Q, f_{A,Q}, f_{T,Q})$.
(u, u')	Directed edge from node u to u'.
$u \in V_Q$, $v \in V_G$	Nodes with index u (v) are in graph V_Q (V_G), respectively.
$\mathbf{W}_h(t)_G$, $\mathbf{W}_h(t)_Q$	h-hop Adjacency Matrices at time t for G and Q, respectively.
$\mathbf{A}_G$, $\mathbf{A}_Q$	Class Membership Matrices for G and Q, respectively.
$\mathbf{M}_{G,Q}$	Sparse node class match matrix between G and Q.
$\mathbf{C}_h(t)$	Parent-to-Child h-hop Class Adjacency Matrix at time t, where $\mathbf{C}_h(t) = \mathbf{W}_h(t)\mathbf{A}$.
$\mathbf{P}_h(t)$	Child-from-Parent h-hop Class Adjacency Matrix at time t, where $\mathbf{P}_h(t) = \mathbf{W}_h(t)^T\mathbf{A}$.
$\mathbf{S}_h(t)$	Multi-hop class membership similarity matrix at hop h and time t between the query nodes and data graph nodes which are class-matching.
$\tilde{\mathbf{S}}(t)$	Aggregate score vector for all nodes $V_n \in G$ for all hops h for time t and α decay parameter between $[0, 1]$.

[290], and reflects real-life graph dynamics. Given all this, we introduce a modification to the h-hop connectivity matrix from [185] [208], and define a new structure that we will use throughout the paper that succinctly captures the class memberships of h-hop neighbors over time. In this work, an h-hop neighborhood of node v is the set of nodes that are reachable from v in h hops (following a path of exactly length h, without backtracking).

We note that the value of h leads to a trade-off between the depth of similarity one desires to consider to matches and computational complexity. Its determination is largely dependent upon the size of the query graph as well as the knowledge of the structure and size of the data graph's network schema.

DEFINITION 7. h-hop Adjacency Matrix. A h-hop adjacency matrix $\mathbf{W}_h(t)$ is an $n \times n$ integer matrix with element $w_h(t)_{i,j}$ representing the existence at time t of an h-hop relation

(possibly weighted and/or directed) between vertex i to vertex j in the graph G, where n is the number of vertices ($|V| = n$).

Note that this structure is in part a simplification of the h-hop connectivity matrix in [208] [185], where each entry $(i,j) > 0$ denotes not only the existence of the relation between at h-hops but also the h hops necessary to make the connection. However, our h-hop Adjacency Matrix also is a function of time t.

The h-hop Adjacency Matrix is calculated algorithmically as shown in **Algorithm 2**. With this structure established, we can now calculate the h-hop Class Adjacency Matrix.

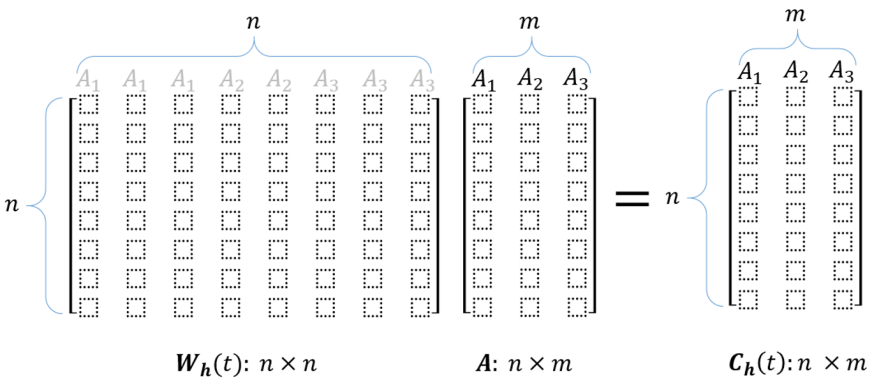


FIGURE 5.2. Depiction of the h-hop Adjacency Matrix $\mathbf{W}_h(t)$ and the Parent-to-Child h-hop Class Adjacency Matrix $\mathbf{C}_h(t)$

DEFINITION 8. Parent-to-Child h-hop Class Adjacency Matrix. A parent-to-child h-hop class adjacency matrix $\mathbf{C}_h(t)$, an $n \times m$ integer matrix with element $c_h(t)_{v,k}$ representing the existence at time t of a h-hop relation (possibly weighted and/or directed) between vertex v and the class k in the graph G, where n is number of vertices ($|V| = n$), m is the number of classes, and $m \ll n$. This matrix is the result of the product of the h-hop adjacency matrix $\mathbf{W}_h(t)$ and the Class Assignment Matrix $\mathbf{A}$, i.e. $\mathbf{C}_h(t) = \mathbf{W}_h(t)\mathbf{A}$. Each entry $c_h(t)_{v,k}$ is essentially the (possibly weighted) sum of the number of nodes of a particular class k adjacent by 1 to h hops to node v at time t. See Fig 5.2.

In the state of the art for graph pattern matching through simulation, researchers consider both the match of both parent-to-child relationships as well as child-from-parent relationships. To build on this advancement, we also develop the children-from-parent $\boldsymbol{h}$-hop class adjacency matrix denoted as $\mathbf{P}_h(t)$, where $\mathbf{P}_h(t) = \mathbf{W}_h(t)^T\mathbf{A}$.

Algorithm 2: h-hop Adjacency Matrix algorithm

Input: $\mathbf{W}_1(t)_G$ (the 1-hop adjacency matrix of graph G of size $n \times n$), $t : t_{\text{start}} \leq t \leq t_{\text{end}}$, and $h_{\max}$ (the desired number of hops)

Output: $\mathbf{W}_h(t)_G$ (h-hop Adjacency Matrices of G), $t : t_{\text{start}} \leq t \leq t_{\text{end}}$ and $h : 2 \leq h \leq h_{\max}$

```
1 foreach t = t_start to t_end do
2     foreach h = 2 to h_max do
3         W_h(t)_G =binary matrix of (W_1(t)_G)^h, where each entry w_h(t)_{G,i,j} = 1 when
           w_1(t)^h_{G,i,j} = 1, i ≠ j, and w_{h-prev}(t)_{G,i,j} ≠ 1 for h-prev < h; and 0 otherwise.
4 return W_h(t)_G
```

5.4. Approach

Given a time-independent query graph Q and a time-dependent data graph G, we investigate the multi-hop class similarity between nodes by following the approach below. Figure 5.3 is a graphical depiction of steps 4-6 outlined below.

(1) Construct the class membership matrices $\mathbf{A}_Q$ and $\mathbf{A}_G$ for query graph Q and data graph G, respectively.

(2) Construct the (sparse) node class match matrix $\mathbf{M}_{G,Q}$ between query graph Q and data graph G, where each entry $m_{ij} = 1$ if the class of data graph node $i \in V_G$ exactly matches the class of the query graph node $j \in V_Q$, and 0 otherwise.

(3) Construct the h-hop adjacency matrices $\mathbf{W}_h(t)_Q$ and $\mathbf{W}_h(t)_G$ (h of each of them for each time t) for graph Q and G, respectively (**Algorithm 2**).

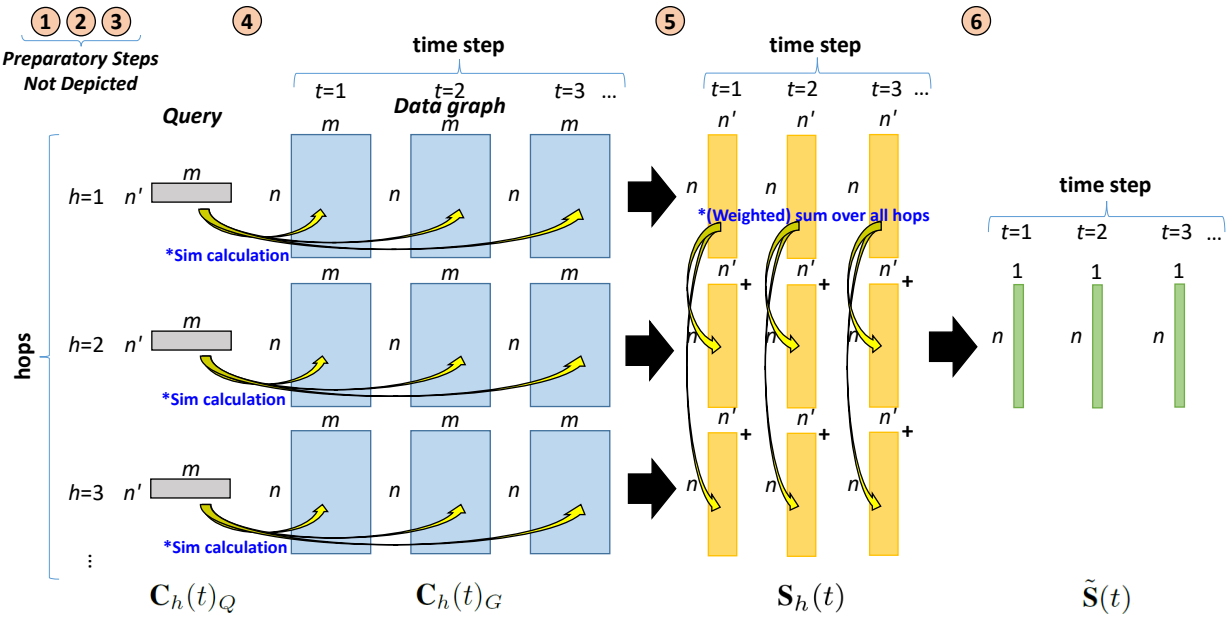


FIGURE 5.3. Graphical depiction of baseline INSiGHT algorithm. This graphic depicts the basic steps in the INSiGHT algorithms described in [144]. Here n' is the number of nodes in the query graph Q, n is the number of nodes in the data graph G, and m is the number of classes. The steps shown are only for the construction of the parts of $\mathbf{S}_h(t)$ that correspond to the Parent-to-Child class membership similarity; the steps for the Parent-from-Child portions are not depicted.

(4) Calculate the parent and child h-hop class adjacency matrices $\mathbf{C}_h(t)_Q$ and $\mathbf{P}_h(t)_Q$, and $\mathbf{C}_h(t)_G$ and $\mathbf{P}_h(t)_G$ (h of each of them for each time t) for graph Q and G, respectively (**Algorithm 3**).

(5) Calculate the multi-hop class membership similarity matrices $\mathbf{S}_h(t)$ at each hop h and each time t between the query nodes and data graph nodes which are class-matching. Such conditioning nodes reduces the number of similarity calculations from $\mathcal{O}(|V_G||V_Q|)$ to $\mathcal{O}(|V_G|)$ for each hop and each timestep. Using the entries $m_{i,j} = 1$ in $\mathbf{M}_{G,Q}$, we calculate the h-th hop class membership similarity between each i-th row of the h-hop query graph class adjacency matrices $\mathbf{C}_h(t)_Q$ and $\mathbf{P}_h(t)_Q$ and the corresponding j-th row of the h-hop data graph class adjacency matrices $\mathbf{C}_h(t)_G$ and $\mathbf{P}_h(t)_G$ (**Algorithm 4**).

(6) Summarize the class membership similarity over multiple hops at each time t by calculating the weighted sum of similarity scores $\tilde{\mathbf{S}}(t)$.

The above approach produces structures which would enable analysts to explore numerically and through visualizations the dynamics of changing in similarity scores by hop over time.

Algorithm 3: Parent and Child h-hop Class Adjacency Matrices algorithm

Input: $\mathbf{W}_h(t)$ (h-hop Adjacency Matrices of a graph for a given t), and $h_{\max}$ (the desired number of hops).

Output: $\mathbf{C}_h(t)$ (Parent-to-Child h-hop Class Adjacency Matrices of a graph), and $\mathbf{P}_h(t)$ (Child-from-Parent h-hop Class Adjacency Matrices of a graph), $t : t_{\text{start}} \leq t \leq t_{\text{end}}$ and $h : 1 \leq h \leq h_{\max}$

```
foreach t = t_start to t_end do
    foreach h = 1 to h_max do
        C_h(t) = W_h(t)A
        P_h(t) = W_h(t)^T A
return C_h(t), P_h(t)
```

Algorithm 4: h-hop Class Membership Similarity algorithm

Input: $\mathbf{M}_{G,Q}$ (sparse node class match matrix between query graph Q and data graph G), $\mathbf{C}_h(t)_Q$ and $\mathbf{P}_h(t)_Q$, and $\mathbf{C}_h(t)_G$ and $\mathbf{P}_h(t)_G$ (parent and child h-hop class adjacency matrices for all time $t_{\text{start}} \leq t \leq t_{\text{end}}$ for graph Q and G, respectively), and $h_{\max}$ (the desired number of hops).

Output: $\mathbf{S}_h(t)$ (h-hop Class Similiarity Matrices between class-matching query and graph nodes at time t using some similarity metric) where $t : t_{\text{start}} \leq t \leq t_{\text{end}}$ and $h : 1 \leq h \leq h_{\max}$.

```
define zero matrices S_h(t) for each hop h..h_max and each time t_start ≤ t ≤ t_end of size
  n × l, where n and l are the number of nodes in G and Q, respectively.
foreach t = t_start to t_end do
    foreach h = 1 to h_max do
        foreach index pair (i, j) : m_{i,j} = 1 in M_{G,Q} do
            set {s_h(t)_{i,j} in S_h(t)} =
              similarity(c_h(t)_{G,i·} and c_h(t)_{Q,j·}) + similarity(p_h(t)_{G,i·} and p_h(t)_{Q,j·})i
return S_h(t)
```

5.4.1. Similarity Metric. In our analysis, we initially utilized two common similarity measures— Jaccard and Sorensen-Dice indices— to measure the similarity between two sets. However, upon further analysis, we realized that the calculated similarities with such metrics had improperly penalized good matches to the original query because of the presence of classes in the data graph that were not in the query. In fact, we assume in our problem that the query graph is generalized with the complete list of all possibly interesting connections or indicators. Thus, it only makes sense that our similarity metric should be a recall-based index Q_R (2), where A and B are the nodes in G and Q, respectively.

$$Q_R(A, B) = \frac{|A \cap B|}{|B|} \tag{2}$$

In other words, we are interested in finding the subset of nodes in G which are also in Q, divided by the total nodes in Q. Nodes with classes that are not in Q do not affect the similarity calculation and reflect noisy, un-related activities. Despite this, our work is compatible with other existing metrics or others to be developed in the future. See Section **??** for intended improvements.

5.4.2. Aggregating Similarity Scores. As a way of compactly distinguishing between both the parent and child similarity between the query and data graphs, we use a complex number structure where the real component and imaginary components are the parent and child similarity, respectively (see **Algorithm 4**, Line 5).

Additionally, $\mathbf{S}_h(t)$ for the $h_{\max}$ hops and time from t_{start} to t_{end} give us a hypercube of class membership similarity scores. How are we to aggregate the scores by hop and over time such that trajectories may be easily determined without losing information?

We propose a method to aggregate the similarity scores for each node $V_n \in G$ over multiple hops h at each time t that is analogous to a technique for exponentially weighted moving averages (EWMA). The aggregate scores are stored in a $n \times 1$ vector $\tilde{\mathbf{S}}(t)$ where each entry $\tilde{s}(t)_n$ is shown in (3).

$$\tilde{s}(t)_n = \sum_{h=1}^{h_{\max}} \alpha^h \cdot \max(s_h(t)_{n\cdot}) \tag{3}$$

Here α is a decay parameter between $[0, 1]$, and $\max(s_h(t)_{n\cdot})$ is the maximum value of the n-th row (corresponding to node n) of the matrix $\mathbf{S}_h(t)$. Note that when each node is a member of only one class, at most one of the entries in the n-th row will be non-zero. When $\alpha = 1$, the summation equally weights each term, while $\alpha < 1$ discounts the weight of each successive hop in the total value of $\tilde{s}(t)_n$.

5.4.3. Complexity Analysis. Our collective algorithm to return the h-hop class similarity scores between the query Q and data G graphs over t timesteps has a complexity bound of $\mathcal{O}(th^2(|V_G|^3 + |V_Q|^3))$. Thus, the run times are cubic with the number of nodes in the data and query graphs, quadratic with the number of hops, and linear with the number of timesteps. Due to the sparsity of both the adjacency $(n \times n)$ and membership $(n \times m)$ matrices, the utilization of sparse linear algebra packages to implement the matrix multiplications would take much less than $\mathcal{O}(n^3)$ and $\mathcal{O}(n^2 m)$ operations, respectively [319].

Algorithm 2 has an overall complexity bounded of $\mathcal{O}(th^2|V_G|^3)$. This is because the matrix power (Line 3) takes $\mathcal{O}(h|V_G|^3)$. We then do this for each timestep and hop (Lines 1 and 2, respectively) for both the query and data graphs.

Algorithm 3 has a complexity bound of $\mathcal{O}(th|V_G|^3)$. This is because each matrix multiplication (Lines 3 and 4) is upper bounded by $\mathcal{O}(n^2 m)$, where n is number of nodes and m

is the number of classes, but we can approximate this to $\mathcal{O}(|V_G|^3)$ when we do not account for the distinction of classes. For a query Q, complexity is $\mathcal{O}(th|V_Q|^3)$.

Algorithm 4 has a complexity bound of $\mathcal{O}(2m|V_G|)$. There is a total of $2|V_G|$ similarity calculations between the Q and G class adjacency matrices when we filter operations with the matrix of class matches $\mathbf{M}_{G,Q}$. Each similarity calculation would take m operations because it is equivalent to finding the cosine similarity of each class adjacency vector and calculating the dot product over the m number of classes.

When we combine all three algorithms and take the highest order terms for both $|V_Q|$ and $|V_G|$ we get an overall complexity bound of $\mathcal{O}(th^2(|V_G|^3 + |V_Q|^3))$.

In the following section, we describe the application of our technique on a real, large dataset in the domain of social media activity detection.

5.5. Results for Motivating Example Problem

The approach detailed in this paper produces Fig. 5.4, a graph of the multi-hop class membership similarity scores of the 4 persons of interest over time in relation to the query graph shown in Fig. 5.1. Here we utilized 3 hops and 4 timesteps ($h_{\max} = 3$ and $t = [1, 4]$). As expected through visual inspection of the original data graph G, Person 3 after 4 periods of time exhibited all the radicalization indicators in query graph Q. However, our time-based approach also returns the multi-hop class membership dynamics and shows the trajectory of each of the 4 persons towards exhibiting the indicators of homegrown violent extremism over time. This analysis could be useful for law enforcement or intelligence analysts who may be interested not only in the extent to which someone exhibits indicators in a pattern of potential violence, but also whether an accelerating occurrence of indicators or behaviors constitutes a trajectory towards violence [203].

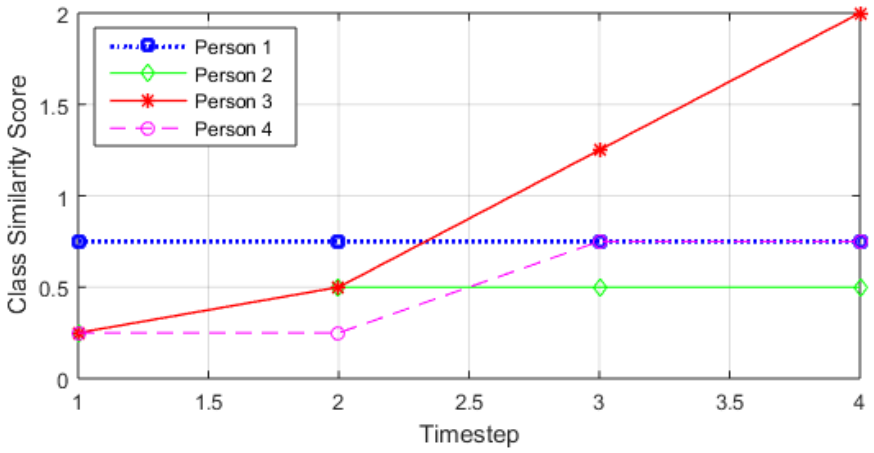


FIGURE 5.4. Plot of the multi-hop class similarity scores over time $\tilde{\mathbf{S}}(t)$ for time-based data graph G for example problem in Fig 5.1. We used $\alpha = 1.0$ (non-weighted sum over each hop). We focus on the indicators/activities associated with each of the persons of interest. For each of time t between 1 and 4, we show the changes in class similarity over 3 hops from each person of interest node.

However, these results also highlight a few shortcoming of the initial version of INSiGHT for this radicalization detection application. Specifically, one can see that Person 1's similarity score did not change despite posting a second extremist n-grams at timestep 3. Given their importance as an indicator, an analyst may desire specific reoccurring indicators to be accounted for in the score. Additionally, we notice that Person 2 had non-zero scores for radicalization due to partial matches with perfectly legal activities (e.g., purchasing a firearm or establishing a social media account). In fact, these activities are only potential indicators of radicalization when they occur with other indicators. Furthermore, in this example, the value of each indicator's contribution to the similarity score does not decay with time. Accounting for these discounts are important especially in light of the on-going non-governmental and governmental counter-radicalization efforts. Lastly, we notice that a person's score is not linear with the number of indicators, but rather weighted by the total

number of indicators at that hop distance away from the Person node in the query. For instance, Person 3 had 2 behaviors (Social Media Account and Suspicious Travel) by timestep 2 for a score of 0.50 and another 2 behaviors (Radical n-gram and Received Training) by timestep 3 for a score increase of 0.75. This is because each of the indicators 1-hop away from the Person node are valued at 0.25 (4 indicators total), and each of the 2-hop indicators is valued at 0.50 (2 indicators total). In the subsequent sections, we systematically address all these shortcomings.

5.6. Real Data Application: Online Blog Behavior Detection

Beyond detecting trajectories for radicalization, we assert that INSiGHT is applicable in a variety of domains involving other latent behaviors. As a stylized behavioral detection example involving a real, large dataset, we return to utilizing the real BlogCatalog dataset [320]. The full description was previously provided in Section 4.4.6.1, and the network schema was shown in Fig. 4.3.

The original graph had over 470,000 nodes and over 4 million edges (see Table 5.2). However in preprocessing, we filtered out 98.56% of nodes and 99.65% of edges which had no connection to indicators in our query. Specifically, we constructed a subgraph from the user IDs which had tags that contained the words 'computer,' 'windows,' or 'windows 7.' The routine we used produced the subgraph that included all such user IDs and any interconnections between them, as well as their respective weblog IDs and all tags. The resulting graph had only about 6,800 nodes and 14,4000 edges and is further detailed in Table 5.2.

5.6.1. Query Description. To test the performance of the matching scheme and algorithm, we devised a proxy query on a benign subject matter with structural parallels to

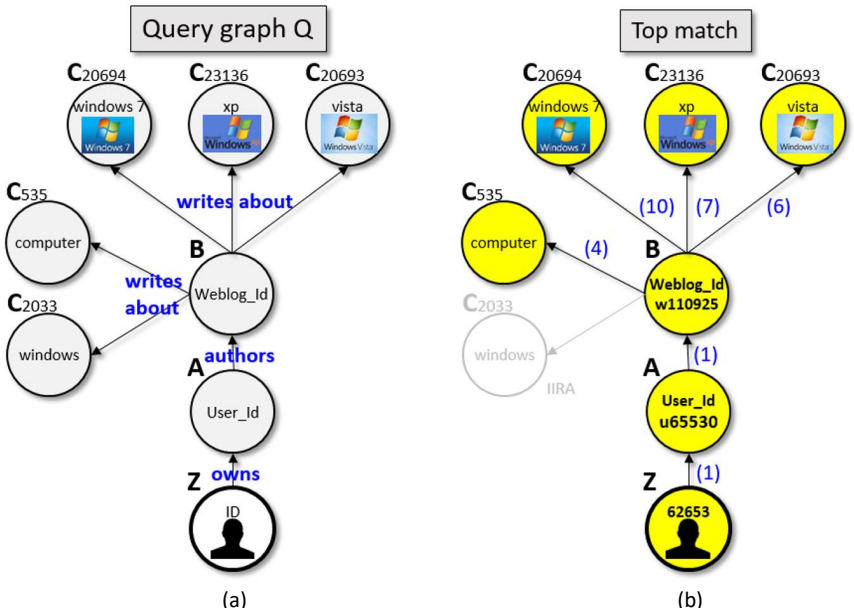


FIGURE 5.5. Experiment query for BlogCatalog (a). Query focus is for User_Ids who had been writing blogs broadly related to 'computers' and 'windows', and specifically to Windows operating systems. In this example, we treat the tag 'windows 7' as a red flag indicator. The top-match in graph G shown in (b) is User_Id 'u65530' with 5 indicator nodes matching in the relevant set (2 directed hops from node class A). Each edge is labeled with a timestamp. The grayed-out node is one of the original query nodes not matched.

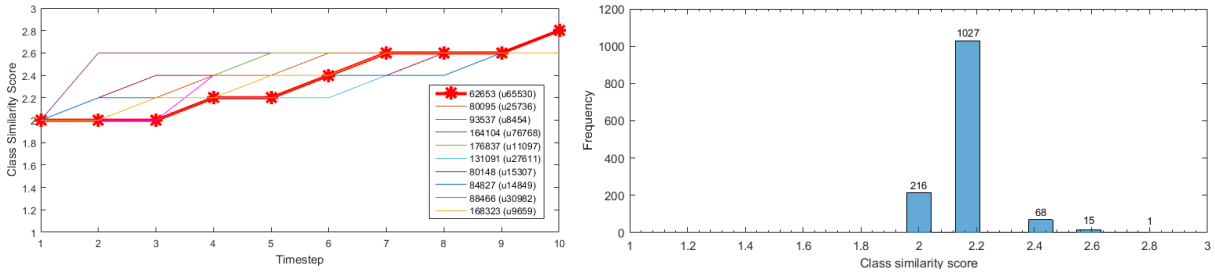


FIGURE 5.6. (a) Plot of the class similarity scores over time $\tilde{s}(t)_n$ for the top 10 nodes in the BlogCatalog data graph using $\alpha = 1.0$ (non-weighted sum over each hop). For each of the timesteps t between 1 and 10, we show the changes in class similarity over 3 hops for the top user IDs of interest. The top scoring User_Id was 'u65530' and the multi-hop parent-child class similarity score over time is shown in bold red. (b) Histogram of aggregated class similarity scores for the 1327 ID nodes in the subgraph. The number above each bar is the number of ID nodes with that respective class similarity score.

TABLE 5.2. BlogCatalog full and subgraph characteristics

Characteristics	$G_{\mathbf{full}}$	$G_{\mathbf{subgraph}}$
Total Nodes	**471,267**	**6,805**
Number of ids	88,781	1,327
Number of userids	80,949	1,327
Number of weblogs	127,227	1,455
Number of unique tags	174,310	2,696
Total Edges	**4,098,290**	**14,483**
Number of id-userid links	88,784	1,327
Number of userid-userid links	3,223,640	3,452
Number of userid-weblog links	127,227	1,451
Number of weblog-tag links	658,639	8,253

an investigation for a latent behavior (Fig. 5.5, left). The query's focus is for user IDs who had been writing blogs about Microsoft Windows operating systems (XP and/or Vista) and subsequently also began to write about Windows 7 when it was released in July 2009, which is the month in which the data was collected. Node class Z is a true person ID, node class A is the query focus user ID, and node class B is the weblog with certain tags. All C class nodes are meant to be seen as labels of a post or blog entry which were determined through machine-classification and semantic analysis. The labels 'computer' (C535) and 'windows' (C2033) are relatively frequent labels which help provide context or additional clarity on the true topic set, and labels 'xp' (C23136) and 'vista' (C20693) are indicators that the blog is about Windows operating systems. These latter nodes are necessary but not sufficient for the latent behavior of interest. Finally, label 'windows 7' (C20684) is considered a latest-occurring red flag indicator.

TABLE 5.3. Synthetic Timestamps for BlogCatalog data

Edge Type	**Timestamp**
ID to User_Id	1
User_Id to User_Id	Unif(1,3)
User_Id to Weblog_Id	1
Weblog_Id to Other Tags	Unif(1,8)
Weblog_Id to 'windows 7' Tag	Unif(9,10)

5.6.2. Adding Edge Timestamps. While the BlogCatalog dataset has the desirable attributes of size and heterogeneity as well as the social network connectivity between users, it lacked the fidelity in dynamics of tags for each blog. Specifically, in addition to the blog tags themselves, knowledge of *when* user IDs added or deleted tags is itself important particularly when one is trying to identify a trajectory towards a latent behavior. In the case of our query, this behavior is described as augmenting an existing interest in blogging about Microsoft operating systems with the latest version of Windows. In order to test our approach, we devised a timestamp labeling function $f_T(\cdot)$ to add randomized time steps to each edge in the graph. Table 5.3 shows the functional rules we utilized to generate the timestamps for particular types of edges. All IDs, User_Ids, and Weblog_Ids are existent at time $t = 1$. The social network connections between User_Ids are formed at times uniformly distributed (discrete) between timestamps [1,3], and tags for each weblog are developed at times uniformly distributed (discrete) between timestamps[1,8], except for the tag 'windows 7' which develops latest uniformly between timestamps [9,10]. The timestamps were generated in MATLAB R2105b using seed `12345`.

5.6.3. Analysis of BlogCatalog Results. Our approach not only produces the top-k full or partial matches in the large BlogCatalog graph to the query, but also generates a perspective on the multi-hop class membership similarity trajectory for those top-k matches. In Fig. 7.7a, we depict the pattern trajectories of the top 10 accounts over 3 hops and 10 timesteps ($h_{\max} = 3$ and $t = [1, 10]$). The top scoring account was User_Id 'u65530', whose matching partial subgraph is shown in Fig. 5.5b. This account utilized the 4 of the 5 indicator tags, 'computer,' 'vista,' 'xp,' and 'windows 7' over the course of time.

Additionally, in this example, we see the merits of our approach in investigative graph search by segmenting out only a fraction of the entities who are on pathways of partially or completely matching a query pattern. As the histogram in Fig. 7.7b shows, most of the accounts in the subgraph only had none or only 1 of the indicators present for the entire time frame.

5.6.4. Run times. We also report the tests on the run times of the collective set of our multi-hop algorithms in a couple of different settings. Initially, we ran an experiment on the effect of the number of nodes on the overall run time. Utilizing the same experimental setting as the performance test on the BlogCatalog subgraph dataset, we fixed the query pattern, query graph size, and hops desired (at $h_{\max} = 3$). However, in addition to the principal run on the subgraph of 6805 nodes, we also generated 4 smaller subgraphs of size 100, 500, 1000, and 3000 nodes and ran the suite of algorithms of those graphs. The results are shown in Fig. 5.7.

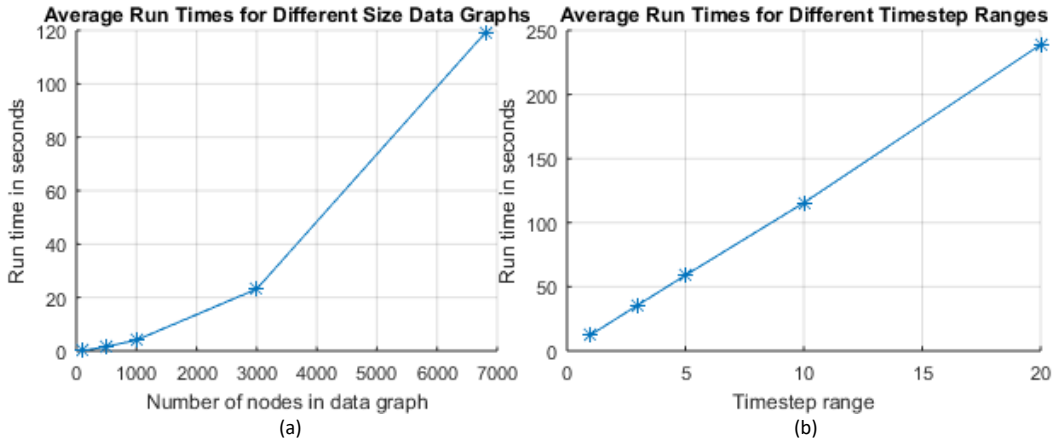


Figure 5.7. (a) Run times for varying data graph size. (b) Run times for varying timestep range.

Second, we ran an experiment on the effect of the timestep range on the overall run time. Utilizing the BlogCatalog subgraph dataset, we fixed the query pattern, query and

data graph size, and hops desired (at $h_{\max} = 3$). We then generated timestamps for the data graph edges for the ranges of 1, 3, 5, and 10 units and ran the suite of algorithms of those graphs. The results are shown in Fig. 5.7.

All experiments were carried out on a laptop computer with an Intel Core i7-4710MQ CPU @ 2.50 GHz, 8.00 GB RAM, and a 64-bit OS. Our results show that as expected, overall run times scale linearly with the timestep range and polynomially with the size of the data graph.

5.7. Enhancements to INSiGHT

In this section, we describe several enhancements we devised for INSiGHT to account for a variety of real-world dynamics associated with the analysis of radicalization pathways or trajectories. First, we modify our baseline approach to account for:

(1) Reoccurrences of each indicator.

(2) Time recency of each indicator.

(3) Individually innocuous but related activity (IIRA) indicators occurring by themselves.

These adjustments ultimately involve a transformation the Parent-to-Child h-hop Class Adjacency matrix $\mathbf{C}_h(t)$ and the Child-to-Parent h-hop Class Adjacency matrix $\mathbf{P}_h(t)$ into $\hat{\mathbf{C}}_h(t)$ and $\hat{\mathbf{P}}_h(t)$, respectively. Furthermore, we devise a novel approach to matching indicators that occur in the neighborhood of each QF node, as well as define a match goodness function $g(Q, G_S, t)$ between the query graph Q and a conforming data subgraph G_S at time t that integrates all the improvements.

TABLE 5.4. Summary of notations for INSiGHT enhancements

Notation	Description/Meaning
$\mathbf{C}_h(t)$	Parent-to-Child h-hop Class Adjacency Matrix at time t, where $\mathbf{C}_h(t) = \mathbf{W}_h(t)\mathbf{A}$.
$\mathbf{P}_h(t)$	Child-from-Parent h-hop Class Adjacency Matrix at time t, where $\mathbf{P}_h(t) = \mathbf{W}_h(t)^T\mathbf{A}$.
$\hat{\mathbf{C}}_h(t)$	Weighted Parent-to-Child h-hop Class Adjacency matrix at time t.
$\hat{\mathbf{P}}_h(t)$	Weighted Child-to-Parent h-hop Class Adjacency matrix at time t.
$F_r(x, \lambda)$	Exponential growth function for the score of an indicator based on frequency x and parameter λ.
$F_d(\tau, \beta, \xi)$	Hyperbolic tangent decay function for the score of an indicator based the time from last class occurrence τ and parameters β and ξ.
$\phi_v(t, h)$	Binary indicator variable for each QF node v at time t over h hops.
$F_{r,\text{mod}}(\phi_v, x, \lambda)$	Exponential growth function F_r modified with the indicator variable $\phi_v(t, h)$.
$\mathbf{1}$	Matrix of 1's.
$\mathbf{\Lambda}$	Diagonal matrix of parameters λ for multiple occurrences of indicators.
$\boldsymbol{\beta}$	Matrix of parameters β for decay of indicator score based on time of last activity.
$\boldsymbol{\xi}$	Matrix of parameters ξ for decay of indicator score based on time of last activity.
$\boldsymbol{\chi}_h(t)$	Last Activity Time h-hop matrix at time t for Parent-to-Child Class Adjacency matrix.
$\mathbf{\Pi}_h(t)$	Last Activity h-hop matrix at time t for Child-to-Parent Class Adjacency matrix.
$\mathbf{\Phi}(t)$	Diagonal matrix of indicator variables $\phi_v(t, h)$.
$\hat{\mathbf{S}}_h(t)$	Weight-adjusted multi-hop class membership similarity tensor at hop h at time step t between the query nodes and data graph nodes which are class-matching.

5.7.1. MULTIPLE INSTANCES OF EACH INDICATOR. Our previous recall-based metric equally weighs all nodes and limits the count of multiple instances of a connection to the same node class to only a single edge. Allowing a diverse set of indicator weights and counting repeated indicators makes intuitive sense in investigative graph search.

We provide a parameterized method to score multiple occurrences of an indicator with the exponential function F_r shown in 4. Here $x(t, h)$ is the number of occurrences of an

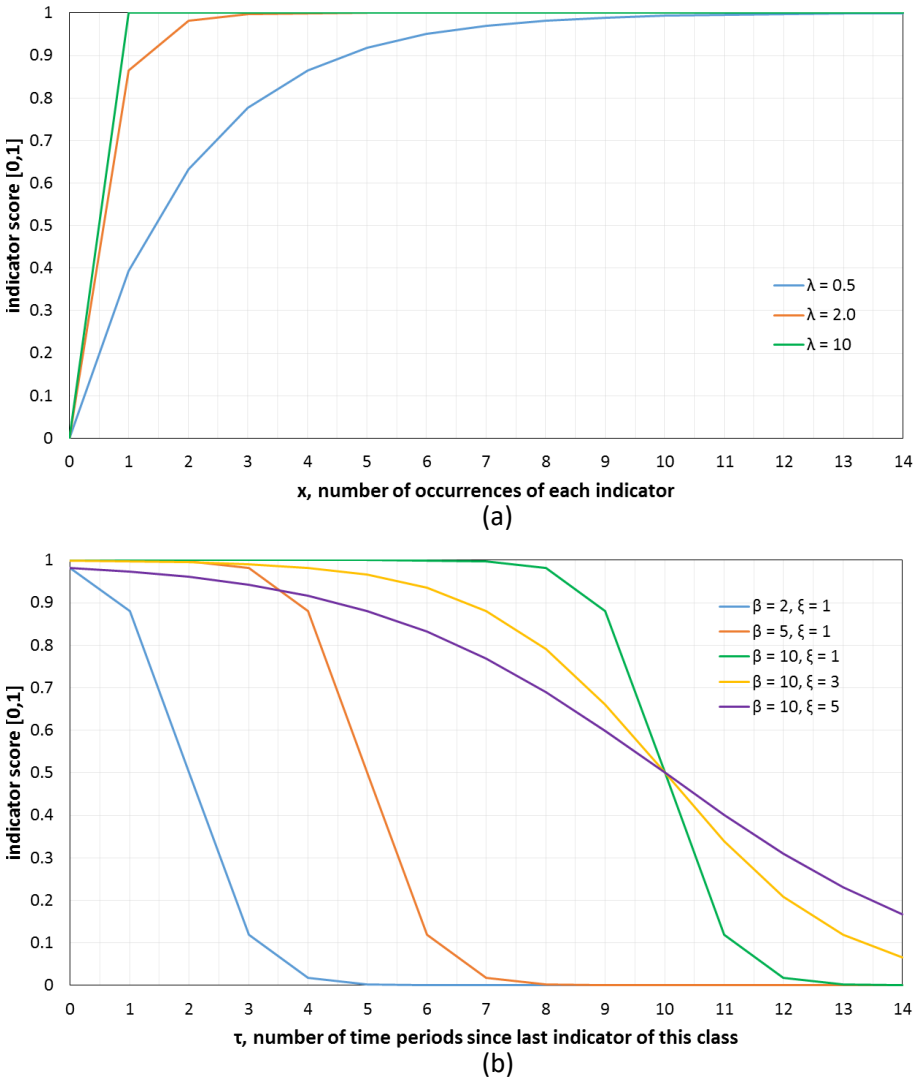


FIGURE 5.8. Sample parameterized growth and decay curves. (a) Plot of exponential growth function F_r for repeated indicators using various parameters for λ. (b) Plot of hyperbolic tangent decay function F_d for diminished significance of indicators from the last occurrence using various parameters for β and ξ.

indicator found for each person at time t and hop h, and the parameter $\lambda \geq 0$ determines how the number of occurrences accumulates to a maximum score of 1. Fig. 5.8a depicts the exponential function F_r for various parameters λ.

$$F_r(x,t,h;\lambda) = \begin{cases} 1 - e^{-\lambda x(t,h)}, & x(t,h) \geq 0 \\ \text{Undefined}, & x(t,h) < 0 \end{cases} \tag{4}$$

To assist the analyst in determining the λ parameters, we devise an equation shown in 5. The variable $x^* \geq 0$ is the number of indicator recurrences needed to achieve a score equal to $1 - \epsilon$, where $0 < \epsilon \leq 1$.

$$f_\lambda(x^*; \epsilon) = -\ln(\epsilon)/x^* \tag{5}$$

This additional modeling feature is an improvement because it provides a means for investigators to weigh the various frequencies of an indicator in the radicalization trajectory measure. It makes intuitive sense that eventually after some number of occurrences of a particular indicator, an investigator will deem the indicator sufficiently present.

5.7.2. Decay of indicator significance over time. Our previous recall-based metric considered an indicator that occurred early in time with the same weight as those that occurred most recently. Intuitively, we know that we must consider the decay in the significance of an indicator over time in the similarity calculation. In an extreme case, one would likely decide that an indicator of a certain class of activities that occurred 10 years ago should count less than if the same indicator occurred yesterday. Each indicator score's rate of decay should clearly be parameterized and controlled by an investigator or analyst, who should consider those who may have legitimately turned away from the radicalization path as well as those who may have only 'gone dark' to avoid detection through encrypted communications in the days, weeks, or months leading up to a planned attack [125, 135].

We provide a parameterized method to decay the significance of an indicator in the similarity calculation over time. Our intuition is that a sigmoid decay function could be appropriate because there is slow decay initially (thus the indicator retains a high score shortly after its occurrence) and slow decay after much time (thus the indicator retains some trace score to prevent its earlier occurrence from being completely nullified). The same idea

is posited with the value of different types of documents [65]. While we could have chosen any one in the family of sigmoid functions, we initially utilize the classical, scaled hyperbolic tangent decay function F_d shown in 6. Here $\tau(t,h)$ is the number of time periods since the last indicator of this class as of time t and hop h, the parameter $\beta \geq 0$ determines the timestep when the indicator score decays to $\frac{1}{2}$, and the parameter $\xi \geq 0$ determines over how long the indicator score decays over time. Fig. 5.8b depicts the exponential function F_d for various parameters β and ξ.

$$F_d(\tau,t,h;\beta,\xi) = \begin{cases} \frac{1}{2}\left(1-\tanh\left(\frac{\tau(t,h)-\beta}{\xi}\right)\right), & \tau(t,h) \geq 0, \\ \text{Undefined}, & \tau(t,h) < 0 \end{cases} \tag{6}$$

As opposed to the exponential growth function (4), an analyst can determine the β parameters more directly because $\beta = \tau^* \geq 0$ is the number of time periods needed to decay the score to $\frac{1}{2}$, and 3ξ is approximately the number of time periods over which the decay occurs to $\frac{1}{2}$.

5.7.3. Incorporating categorical node labels for investigations. In this section, we seek to utilize the investigative node-type classes defined in Section 4.3 and Table 4.1 to aid a law enforcement or intelligence analyst more effectively identifying those needing further investigation [142].

First, we seek to incorporate in a similarity metric the filtering of matches based on the investigative node-type classes, where individually innocuous indicators would not be counted unless they occurred with other definitive indicators of a latent behavior. This avoids overmatching by masking the matches of individually innocuous indicators occurring by themselves. We define in 7 the indicator variable $\phi_v(t,h)$ to toggle the masking or unmasking

of the v-th row in either the parent-to-child class adjacency matrix $\mathbf{C}_h(t)$ or the child-to-parent class adjacency matrix $\mathbf{P}_h(t)$.

$$\phi_v(t,h) = \begin{cases} 1, & \text{QF node } v \text{ is } h\text{-hop adjacent to 1 or more IND/RF nodes at time } t \\ 0, & \text{o/w} \end{cases} \tag{7}$$

We integrate this indicator variable in the original reoccurence function 4 and define a modified exponential function $F_{r,\text{mod}}$ in 8.

$$F_{r,\text{mod}}(\phi_v, x, t, h; \lambda) = \begin{cases} 1 - e^{-\phi_v(t,h)\lambda x(t,h)}, & x(t,h) \geq 0 \\ \text{Undefined}, & x(t,h) < 0 \end{cases} \tag{8}$$

Second, we also develop a module to annotate when each individual exhibits any one of the red flag behaviors, as designated by the RF node-type category. We accomplish this by augmenting the previous baseline Algorithm 4 with a procedure that constructs a matrix recording the timestep of occurrence for each of the Red Flag (RF) indicators. We refer to this updated version as **Algorithm 5**. For clarity, the entirety of the new algorithm is shown below.

Specifically, we added the input of A_{RF} which is the set of node classes which are designated as red flags. After performing the similarity calculations between the parent and child h-hop class adjacency matrices for the data and query graphs at each time t and hop h (Line 6), we then iterate through each of these RF node-type classes (Line 7) (still at each time t and hop h), and find all nodes who connect to the RF node-type class portion based upon its entries in the child h-hop class adjacency matrix. As long as no non-zero time was previously recorded, the algorithm assigns the time t to $\mathbf{F_h}$, the h-hop RF Timestep Tracker

Algorithm 5: h-hop Class Membership Similarity algorithm + Red Flag Detection

Input: $\mathbf{M}_{G,Q}$ (sparse node class match matrix between query graph Q and data graph G), $\mathbf{C}_h(t)_Q$ and $\mathbf{P}_h(t)_Q$, and $\mathbf{C}_h(t)_G$ and $\mathbf{P}_h(t)_G$ (parent and child h-hop class adjacency matrices for all time $t_{\text{start}} \leq t \leq t_{\text{end}}$ for graph Q and G, respectively), $h_{\max}$ (the desired number of hops), and A_{RF} (set of RF node-type classes).

Output: $\mathbf{S}_h(t)$ (h-hop Class Similiarity Matrices between class-matching query and graph nodes at time t using some similarity metric) and $\mathbf{F}_h$ (h-hop RF Timestep Tracker Matrix of size n $\times$ $|A_{RF}|$), where $t : t_{\text{start}} \leq t \leq t_{\text{end}}$ and $h : 1 \leq h \leq h_{\max}$.

define zero matrices $\mathbf{S}_h(t)$ for each hop $h..h_{\max}$ and each time $t_{\text{start}} \leq t \leq t_{\text{end}}$ of size $n \times l$, where n and l are the number of nodes in G and Q, respectively.
foreach $t = t_{\text{start}}$ to t_{end} **do**
  foreach $h = 1$ to $h_{\max}$ **do**
    foreach index pair $(i, j) : m_{i,j} = 1$ in $\mathbf{M}_{G,Q}$ **do**
      set $\{s_h(t)_{i,j}$ in $\mathbf{S}_h(t)\} =$ **similarity**$(c_h(t)_{G,i.}$ and $c_h(t)_{Q,j.}) +$ **similarity**$(p_h(t)_{G,i.}$ and $p_h(t)_{Q,j.})$i
    foreach $r \in A_{RF}$ **do**
      set $f_h(i, k) = t \; \forall$ datagraph nodes i where $c_h(t)_{G,ir} = 1$ for the first time and k is the index of r in the set A_{RF}
return $\mathbf{S}_h(t)$ and $\mathbf{F}_h$

Matrix. The size of this matrix is the number of nodes in the network by the number of RF node-type classes ($n \times |A_{RF}|$). Just like the $\mathbf{S}_h(t)$ similarity matrices, the $\mathbf{F}_h$ RF tracker matrices are also summed over all hops in a subsequent step of the overall INSiGHT approach (see Section 5.4). This record of the occurrence of red flag indicators enables subsequent visualization and alerts to analysts, as we will demonstrate in later example applications.

5.7.4. Fast similarity calculations with matrices. The matrix versions of this function that operates on the Parent-to-Child h-hop Class Adjacency matrix $\mathbf{C}_h(t)$ and the Child-to-Parent h-hop Class Adjacency matrix $\mathbf{P}_h(t)$ are shown in (9) and (10), respectively.

$$\hat{\mathbf{C}}_h(t) = \tfrac{1}{2}\left(\mathbf{1} - e^{-\mathbf{\Phi}_h(t)\mathbf{C}_h(t)\mathbf{\Lambda}}\right)\left(\mathbf{1} - \tanh\left(\frac{\boldsymbol{\chi}_h(t) - \boldsymbol{\beta}}{\boldsymbol{\xi}}\right)\right) \tag{9}$$

$$\hat{\mathbf{P}}_h(t) = \tfrac{1}{2}\left(\mathbf{1} - e^{-\mathbf{\Phi}_h(t)\mathbf{P}_h(t)\mathbf{\Lambda}}\right)\left(\mathbf{1} - \tanh\left(\frac{\mathbf{\Pi}_h(t) - \boldsymbol{\beta}}{\boldsymbol{\xi}}\right)\right) \tag{10}$$

Given n and m are the number of nodes and classes in graph G, respectively, $\hat{\mathbf{C}}_h(t)$ is the $n \times m$ weighted Parent-to-Child h-hop Class Adjacency matrix at time t whose equation is shown in 9. Here $\mathbf{1}$ is the $n \times m$ matrix of 1's, $\mathbf{\Phi}_h(t)$ is the diagonal matrix of size $n \times n$ such that each entry (i, i) is equal to 1 if i is the index for a QF node that is adjacent to 1 or more IND/RF nodes over h hops, $\mathbf{C}_h(t)$ is the $n \times m$ Parent-to-Child h-hop Class Adjacency matrix at time t, $\mathbf{\Lambda}$ is the $m \times m$ diagonal matrix of parameters λ, $\boldsymbol{\chi}_h(t)$ is the $n \times m$ Last Activity h-hop matrix at time t, $\boldsymbol{\beta}$ is the $n \times m$ matrix of parameters β, $\boldsymbol{\xi}$ is the $n \times m$ matrix of parameters ξ. When the j-th column of $\boldsymbol{\beta}$ is a vector of size $n \times 1$ of the β for class j.

Essentially, the weight-adjusted Class Adjacency matrix at time t and hop h is equal to the product of a 'multiple indicator occurrence' factor and an 'indicator decay' factor operating on filtered class adjacency matrix that masks stand-alone IIRA indicators.

We propose that we can calculate the Last Activity h-hop matrix $\boldsymbol{\chi}_h(t)$ at time t by taking the difference between the previous class adjacency matrix at t_{previous} and the current one at t, and setting all new connections per class (entries $\chi_h(t)$ to be equal to 0 and advancing all other $\chi_h(t)$ entries to be equal to time t. Thus the each entry in $\chi_h(t)$ signifies the time lapse since the last connection to that class occurred.

Additionally, rather than performing pair-wise set comparisons using the recall-based index Q_R (Equation 1 in [144]), we adopt a faster matrix-level dot product approach that is related to cosine similarity [175]. Specifically, we row-normalize and square each entry of the query Parent-to-Child h-hop Class Adjacency matrix $\hat{\mathbf{C}}_h(t)_Q$ (size $n' \times m$, where n' is the number of nodes in the query graph), and then multiply it by the transpose of the

weighted data graph h-hop Class Adjacency matrix $\hat{\mathbf{C}}_h(t)_G^T$ (size $m \times n$) (11). $\hat{\mathbf{S}}_h(t)$ is then the weight-adjusted h-hop similarity matrix of the query nodes to the data graph nodes (size $n' \times n$).

$$\hat{\mathbf{S}}_h^{(1)}(t) = \hat{\mathbf{C}}_h(t)_Q \cdot \hat{\mathbf{C}}_h(t)_G^T \tag{11}$$

$$\hat{\mathbf{S}}_h^{(2)}(t) = \hat{\mathbf{P}}_h(t)_Q \cdot \hat{\mathbf{P}}_h(t)_G^T \tag{12}$$

$$\hat{\mathbf{S}}_h(t) = \left[\hat{\mathbf{S}}_h^{(1)}(t)\ \hat{\mathbf{S}}_h^{(2)}(t)\right] \tag{13}$$

This procedure allows for the similarity score of nodes in the query graph and data graphs to be calculated quickly and expressed as a fraction over the number of nodes in the query graph.

5.8. Illustrative Applications on Synthetic Graphs

5.8.1. Small, Synthetic Radicalization Toy Graph.

We return to the motivating problem in Fig. 5.1 and show how the enhancement to INSiGHT has the potential to reduce false positives by accounting for the categorical node-types of the indicators. The results of the baseline INSiGHT approach for this problem are reproduced in Fig. 5.9a. We notice that Person 2 had non-zero scores for radicalization due to partial matches with perfectly legal activities (e.g., purchasing a firearm or establishing a social media account). Additionally, Person 4 had a non-zero score at time step 2 due to establishing a social media account only. These activities are only potential indicators of radicalization when they occur with other indicators. To correct this effect and minimize false positives for legitimate queries, we label nodes according to investigative node types established in [142].

TABLE 5.5. Class Node-Types and Parameter Sets 1 and 2 for Decay and Re-occurrence Modules for Radicalization Query

			Parameter Set 1			Parameter Set 2		
Class	**Label**	**NodeType**	$\boldsymbol{\lambda}$	$\boldsymbol{\beta}$	$\boldsymbol{\xi}$	$\boldsymbol{\lambda}$	$\boldsymbol{\beta}$	$\boldsymbol{\xi}$
A	Person	QF	10.0	1000.0	1.0	10.0	1000.0	1.0
B	Social Media Account	IIRA	10.0	1000.0	1.0	10.0	1000.0	1.0
C	Radical n-gram	IND	4.6	1000.0	1.0	2.3	6.0	3.0
D	Extremist n-gram	IND	4.6	1000.0	1.0	4.6	12.0	3.0
E	Suspicious Travel	IND	10.0	1000.0	1.0	4.6	8.0	3.0
F	Received Training	IND	10.0	1000.0	1.0	10.0	16.0	3.0
G	Purchase Firearm	IIRA	10.0	2.0	1.0	10.0	16.0	3.0
H	Promote Concert	NC	0.0	1000.0	1.0	0.0	1000.0	1.0
I	Soccer Club	NC	0.0	1000.0	1.0	0.0	1000.0	1.0

Specifically, in Table 5.5 we assign to the node classes in the radicalization graph query pattern Q a specific investigative 'node type.' The logic behind our designations was intuitive. We are trying to distinguish the establishment of a social media account and the purchase of a firearm as common, innocuous behaviors ('IIRA') while recognizing that their performance can ultimately contribute towards a pathway of radicalization when combined with indicators. Additionally, we want to designate the receipt of terrorist training as an indicator which is serious enough to individually generate an alert to analysts. By default, classes of nodes appearing in the data graph that are not part of the query pattern are labeled as 'no category' or 'NC.'

By adjustment of the class adjacency matrices with through 10, we obtain the resulting class similarity score time series plot shown in Fig. 5.9b. Notice that Person 2 now has a zero class similarity score throughout the window of analysis because its only matching indicators were of type 'IIRA' (Social Media Account and Purchase Firearm). Additionally, Person 4 has a zero class similarity score until it exhibited first indicator of type 'indicator' at time step 3 (posting a radical n-gram).

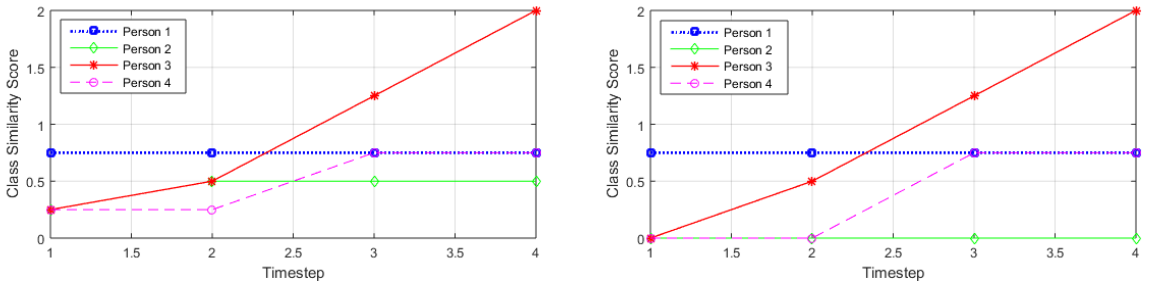


FIGURE 5.9. Multi-hop class similarity for the radicalization example with (a) and without (b) investigative indicator type filtering. Shown are the results of the efforts to minimize false positives for legitimate queries, we label nodes according to investigative node types. Notice that Person 2 now has a zero class similarity score throughout the window of analysis because its only matching indicator was of type 'IIRA.' Additionally, Person 4 has a zero class similarity score until it also posted a radical n-gram at time step 3.

5.8.2. EXTENDED TIME SYNTHETIC RADICALIZATION TOY GRAPH. The new data graph depicted in Fig. 5.10 is an expansion of the one shown in Fig. 5.1b and now has one more individual and depicts additional reoccurring indicators of the on-line behavior of some homegrown violent extremists.

Person 1 is an extremist with a smaller number of SM posts, but later indicators of suspicious travel and firearm purchase. Person 2 is a non-extremist who purchased 2 firearms. Person 3 is an extremist with a large number of posts and other indicators early-to-mid in the timeline. Person 4 is a former extremist who made a small number of radical posts early in the timeline. Finally, Person 5 is an extremist with large number of radical posts throughout, but only extremist posts late in the timeline.

As in the motivating example problem, we utilize the same query graph Q shown in 5.1a. Based upon the nature of the indicators, we chose the parameters for the time-decay and re-occurrence modules as shown in Table 5.5. A λ of 10.0 signifies that just one occurrence of that indicator class is necessary to achieve the maximum similarity score. The classes 'Person'

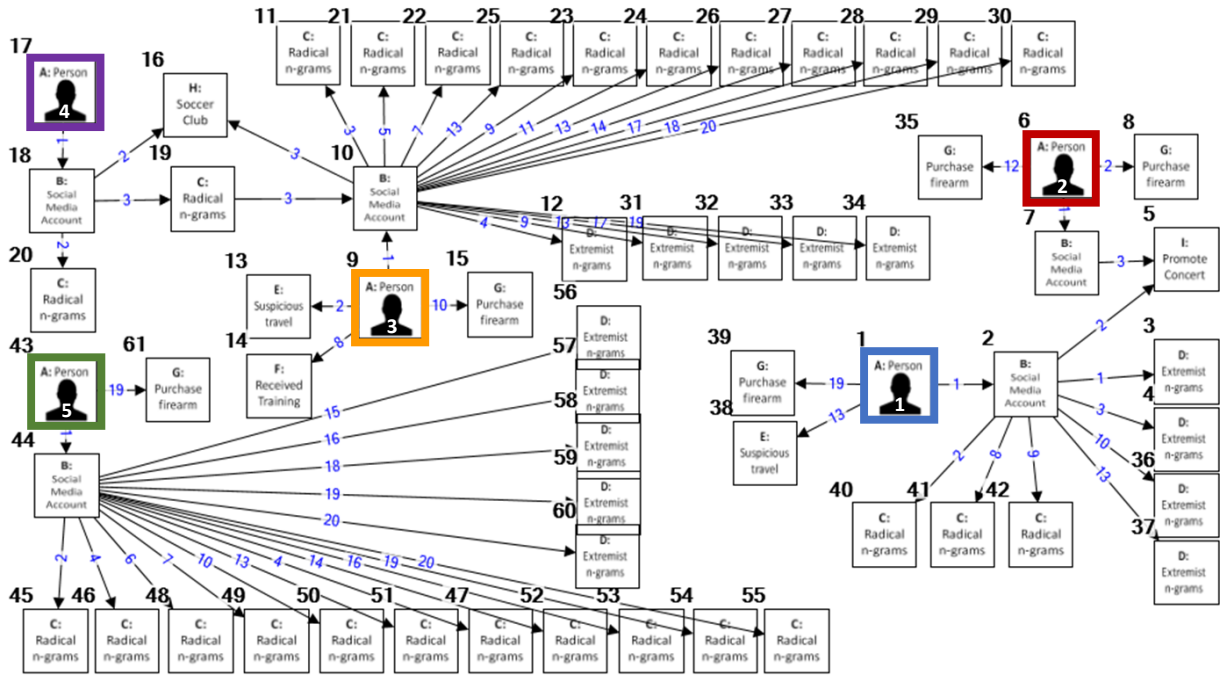


FIGURE 5.10. Expanded motivating example for detecting trajectories of homegrown violent extremists. Beyond the base example from Fig.5.1, the new data graph Fig. 5.10a now has one more individual and depicts additional reoccurring indicators indicative of online behavior of some homegrown violent extremists.

and 'SM Account' are basic nodes which need to occur once, while 'Received Training' and 'Purchase Firearm' may be more threatening indicators whose singular occurrence become important. A λ of 4.6 equates to 2 or more occurrences of an indicator to achieve a near maximum similarity score, and a λ of 2.3 equates to 5 or more occurrences of an indicator to achieve a near maximum similarity score.

For the time-decay parameter β, we selected to equate a timestep in our synthetic dataset to represent 3 months (thus the 20 timesteps in the dataset account for indicators which occurred over 5 years). A β of 1000 nearly eliminates the decay for the value of an indicator class. We chose this for the 'Person' and 'SM Account' basic classes, but also for the node classes 'Promote Concert' and 'Soccer Club' which are not indicators in the query pattern. A β of 6 results in the decay to half of the maximum score of an indicator class after 18

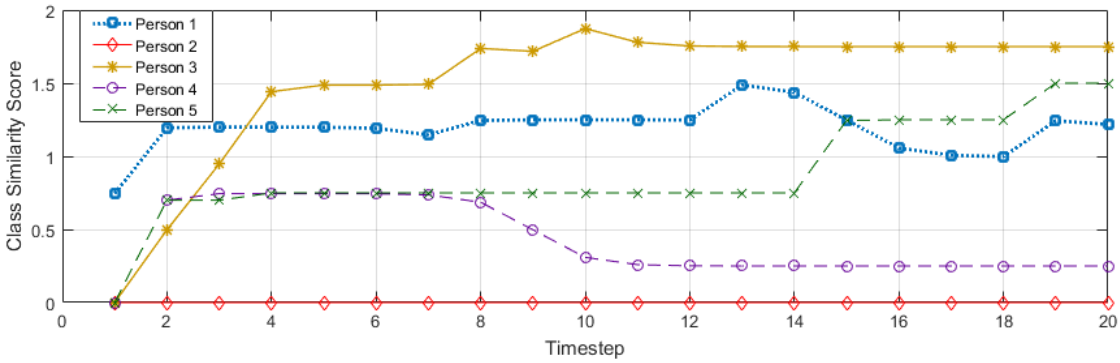


FIGURE 5.11. Class similarity score time series for the expanded radicalization example depicting the effect on the similarity score due to the reoccurring indicators and time decay from inactivity.

months ($= 6 \cdot 3$ months). The β parameters 8 and 16 obviously signifies that knowledge of their occurrence remains important for a longer period of time.

Utilizing the improvements to INSiGHT, we produce the time series of behavioral similarities to a profile for each of the individuals in the graph shown in Fig. 5.11. It is clear that Person 3 has the highest initial gradient towards radicalization and maintains his score with more instances of indicators over time. However, the trajectories of the other individuals are also worth noting. Person 1 has a relatively high similarity score throughout due to radical and extremists posts but also has visible spikes when he goes on suspicious travel (timestep 13) and purchases a firearm (timestep 19). Person 5 also clearly has a sustained spike in his score later in the time frame due to the posting of radical and extremists and the purchase of a firearm.

The decay of Person 4's earlier radical statements is evident with the shape of similarity score curve. Lastly, Person 6 now has a score of 0 throughout the entire time frame of analysis because his activities were limited to the IIRA category.

Ultimately, this experiment showed that with a small example, we can indeed detect those who may be on a radicalization trajectory towards violent extremism based upon a

simplistic query pattern. Next, we test INSiGHT on real data that contain time-based, labeled indicators of bona fide cases of radicalization that ultimately led to violent activity.

5.9. Application #1 Real Data: Radicalization Detection

In this section, we first apply INSiGHT on the Klausen dataset as proof of principle that our technology can be applied to real radicalization data. We first establish the schema for our heterogeneous data graph. Since the dataset does not contain any known network connections between individuals, the schema shown in Fig. 5.12a is straightforward: each person is connected to their attributed activities (behavioral indicators). The Klausen dataset contains 135 individuals and coded 1326 time-stamped features. Table 5.6 shows a summary of the bipartite graph's characteristics. Furthermore, we construct a bipartite graph query directly from the behavioral indicators that Klausen utilized in her research. See Fig. 5.12b. The person is the query focus node and is connected directly to each of the indicator nodes. We utilize only 23 of the original 27 indicators because Birth, Arrest.Date, and Sentencing.Date were not relevant in the early detection of violent extremists.

Table 5.6. Radicalization graph characteristics (Klausen)

Characteristics	$G_{\mathbf{full}}$
Total Nodes	**1,461**
Number of individuals	135
Number of behavioral indicators	1,326
Total Edges	**1,326**

Another critical step is the establishment of the investigative node-type categories of indicators as well as the initial parameterization. In Table 5.7, we provide these details. The Person node is the query focus (QF). We also designated Convert Date, Disillusionment, Trauma, Personal Crisis, Educational/Occupation Disengagement, Drop Out Date,

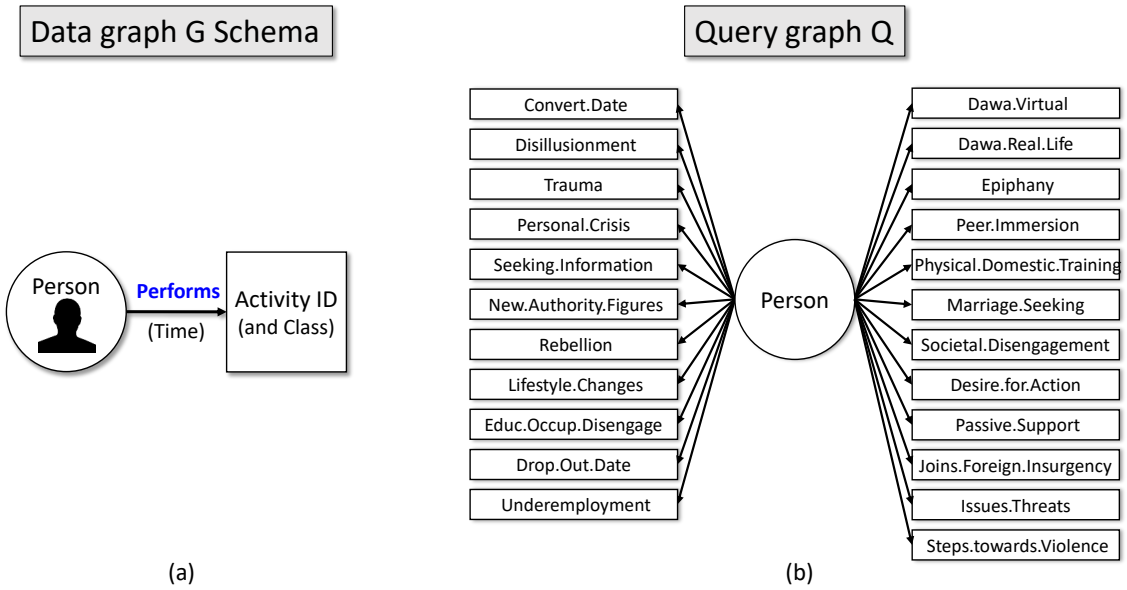


FIGURE 5.12. (a) The schema of the heterogeneous data graph G of individuals and any exhibited radicalization indicators. (b) The query graph Q of the 23 indicators of radicalization modeled as a bipartite graph.

and Underemployment as IIRA due to their innocuous nature. These all either fall under Klausen's Stage 0 (Pre-Radicalization) or Stage 1 (Detachment). We also designed the indicators from Klausen's Stage 3 (Planning and Execution of Violent Action) as red flags (RF). These were Passive Support, Joins Foreign Insurgency, Issues Threats, and Steps Towards Violence. The features Date of Criminal Action and Arrest Date, while nearly applicable to all individuals in the dataset, were designated as No Category (NC) because they are not indicators of radicalization, but just account for the overall timeline of each perpetrator. All other behaviors, we designated as IND (indicators). See Appendix C for the definitions of these indicators provided from [167].

Table 5.7 also shows the parameters we selected for use with INSiGHT. Recalling from Section 5.7, the parameter λ determines how the number of occurrences of a particular class of indicator accumulates to a maximum score of 1, while the β and ξ govern the shape of

TABLE 5.7. Klausen Radicalization Query- Class Node-Types and Parameter Set

			Parameter Set		
Class	**Label**	**NodeType**	$\boldsymbol{\lambda}$	$\boldsymbol{\beta}$	$\boldsymbol{\xi}$
A	Person	QF	10.0	1.0×10^8	1.0
B	Convert.Date	IIRA	10.0	1.0×10^8	1.0
C	Disillusionment	IIRA	10.0	1.0×10^8	1.0
D	Trauma	IIRA	10.0	1.0×10^8	1.0
E	Personal.Crisis	IIRA	10.0	1.0×10^8	1.0
F	Seeking.Information	IND	10.0	1.0×10^8	1.0
G	New.Authority.Figures	IND	10.0	1.0×10^8	1.0
H	Rebellion	IND	10.0	1.0×10^8	1.0
I	Lifestyle.Changes	IND	10.0	1.0×10^8	1.0
J	Educ.Occup.Disengage	IIRA	10.0	1.0×10^8	1.0
K	Drop.Out.Date	IIRA	10.0	1.0×10^8	1.0
L	Underemployment	IIRA	10.0	1.0×10^8	1.0
M	Dawa.Virtual	IND	10.0	1.0×10^8	1.0
N	Dawa.Real.Life	IND	10.0	1.0×10^8	1.0
O	Epiphany	IND	10.0	1.0×10^8	1.0
P	Peer.Immersion	IND	10.0	1.0×10^8	1.0
Q	Physical.Domestic.Training	IND	10.0	1.0×10^8	1.0
R	Marriage.Seeking	IND	10.0	1.0×10^8	1.0
S	Social.Disengagement	IND	10.0	1.0×10^8	1.0
T	Desire.for.Action	IND	10.0	1.0×10^8	1.0
U	Passive.Support	RF	10.0	1.0×10^8	1.0
V	Joins.Foreign.Insurgency	RF	10.0	1.0×10^8	1.0
W	Issues.Threats	RF	10.0	1.0×10^8	1.0
X	Steps.towards.Violence	RF	10.0	1.0×10^8	1.0
Y	Date.of.Criminal.Action	NC	0.0	1.0×10^8	1.0
Z	Arrest.Date	NC	0.0	1.0×10^8	1.0

the decay function to devalue an indicator's component score based upon the last known occurrence of that indicator. In this proof of concept, we designated all classes of nodes except for Date of Criminal Action and Arrest Date with $\lambda = 10.0$, $\beta = 1.0 \times 10^8$, and $\xi = 1.0$. These parameter selections effectively provide a full component score for the single occurrence of each indicator class and negate any decay. For the node classes Date of Criminal Action and Arrest Date, we designated $\lambda = 0.0$ to mask any matches (because it provides a component score of 0).

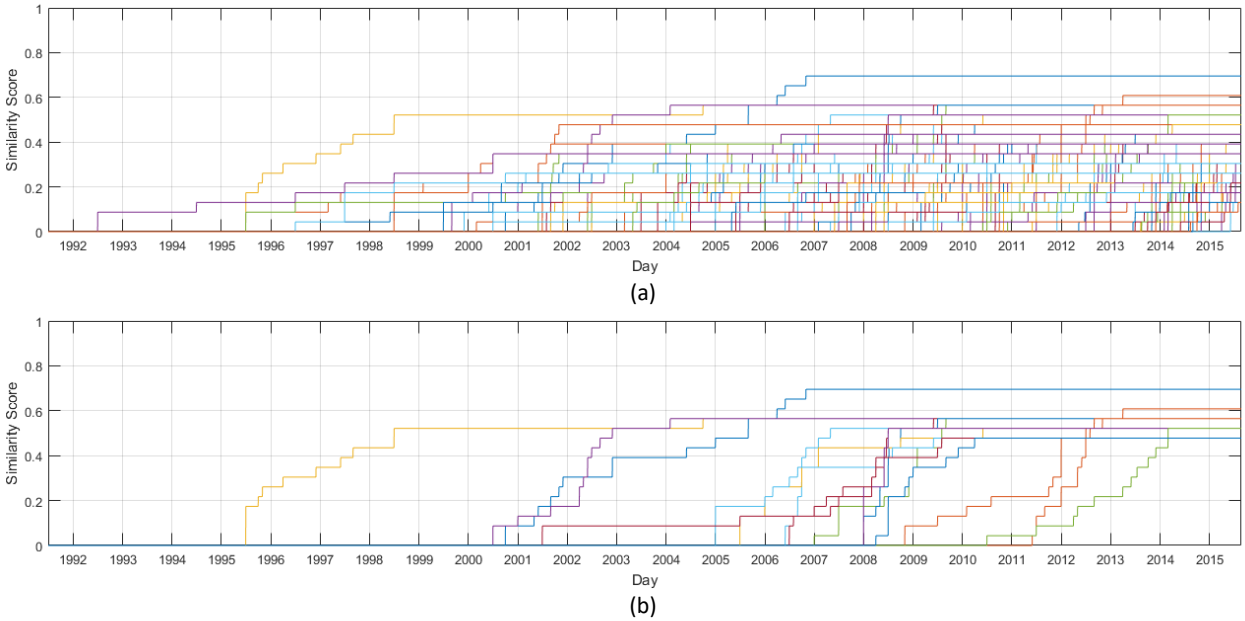


FIGURE 5.13. The radicalization time series plots for (a) all 135 U.S. violent extremists and (b) only the top 15 scoring individuals in the Klausen dataset.

INSiGHT searches for the whole or partial matches to the radicalization query pattern among each individual's behavioral indicators in the heterogeneous network and specifically enables an analyst to visualize how each is radicalizing over time. See Fig. 5.13a for the time series plots for all 135 offenders in the dataset, and Fig. 5.13b for only the top 15 offenders (determined as the 15 individuals with the highest similarity scores achieved by the last timestep). We include the latter plot simply to better depict how long the radicalization processes took for a subset of individuals.

The distribution of similarity scores with the radicalization query pattern shown in Fig. 5.14. Similarity scores ranged from 0.130 to 0.696, which reflect that as few as 3 but as many as 16 behavioral indicators were matched for individuals. It is also important to note that this dataset specifically deals with known offenders. The number of indicators associated with each individual was based on research from publicly available sources and based on a

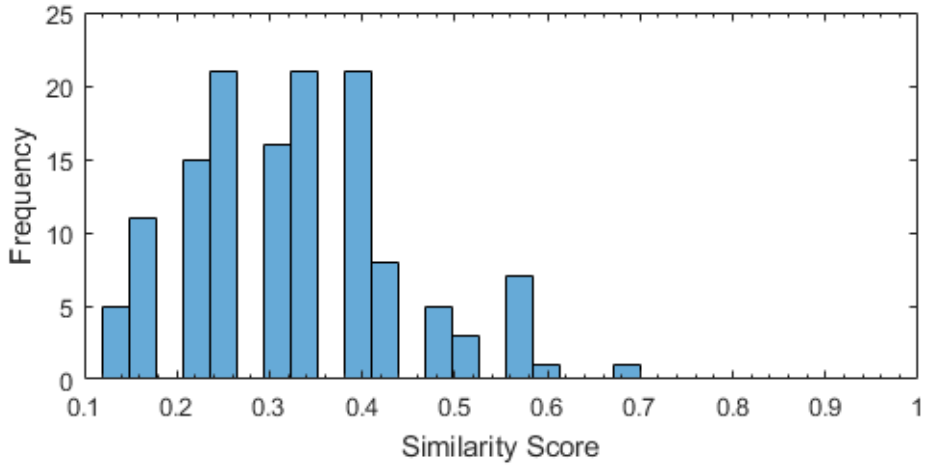


FIGURE 5.14. Histogram showing the distribution of final similarity scores for all 135 U.S. violent extremists.

subjective cut-off from a larger listing of over 300 offenders that there was a suitable amount of information to code at least several indicators (but often more).

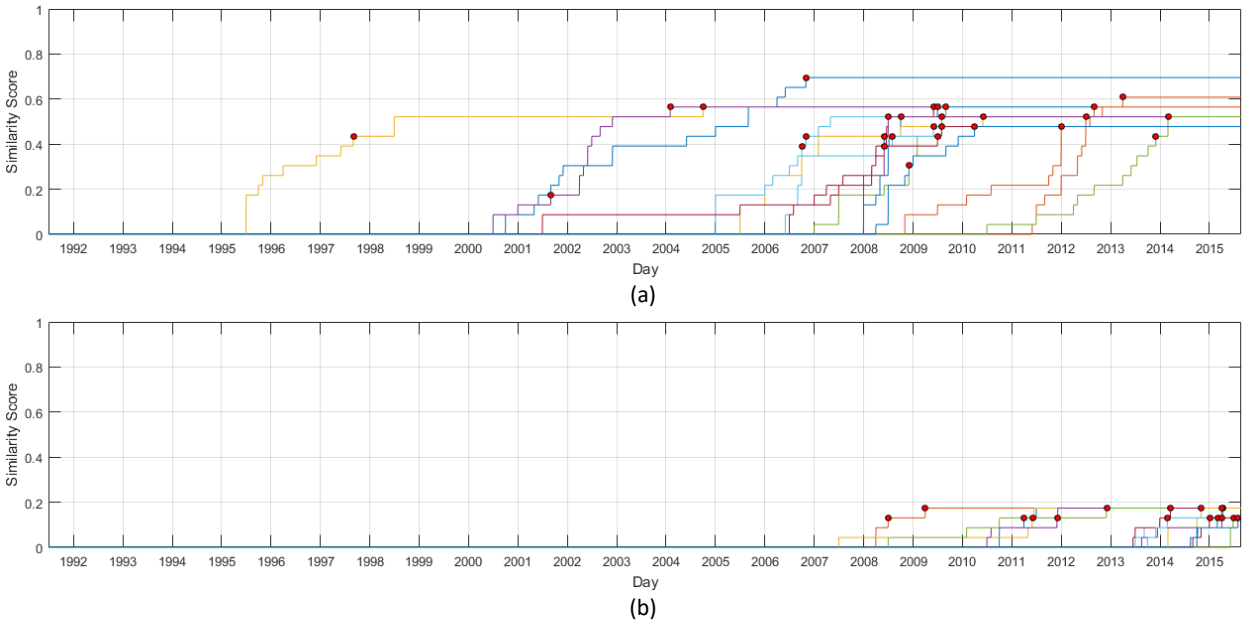


FIGURE 5.15. The radicalization time series plots for (a) only the top 15 U.S. violent extremists and (b) only the bottom 15 scoring individuals in the Klausen dataset. The red circles represent the exhibition of flags.

At this point of research and analysis, it is not yet possible to definitively determine a similarity score threshold by which analysts would be alerted for threats and screen for

high risk individuals. Absent any suitable data of non-violent radicals who exhibit some of the indicators for evaluation, we are unable to determine if say a 0.2 threshold effectively achieves a desired true positive rate and acceptable false positive and false negative rates.

TABLE 5.8. Prevalence of Red Flag Indicators and the Fit in the Klausen Radicalization Model. Source: [168].

Red Flag Indicator	**Description**	**Freq**	**Fit**
Steps Towards Violence	Procurement of materials for plot, surveillance, operation planning	64.4% (87)	100% (87)
Joins Foreign Insurgency	Travel (successful or attempted) abroad with the intention of taking part in a foreign insurgency	45.9% (62)	100% (62)
Non-violent Support	Material, logistical, or financial support to extremist individual or organization	15.6% (21)	90.5% (19)
Issues Threats	Communicates violent threats online or in real-life to specific individuals or groups	11.9% (16)	100% (16)

However, it is noteworthy to mention that red flag visualizations and alerts can assist the analyst in identifying high risk individuals even when they have relatively lower similarity scores. INSiGHT utilized the red flag module from Algorithm 5 and tracked the timestep of occurrence of the pre-designated red flag indicator. In Fig. 5.15 we show the top 15 and bottom 15 scoring individuals and their radicalization trajectories. Despite the difference in similarity scores, the timing of red flags for each individual can be used as a significant risk factor. In fact, Klausen's research team shows that nearly all these red flags occurred in the last stage of radicalization [168]. This means that very few indicators or cues theorized to occur earlier in a radicalization process actually occurred after these red flag indicators, and that the date of criminal action was imminent. See Table 5.8, which is directly extracted from [168].

5.10. Application #2 Real Data: MOOC Persistence Detection

5.10.1. Introduction. Massive Open Online Courses (MOOCs) are now a widely popular form of learning. However, institutions that have offered MOOCs are continually struggling with low completion rates. Researchers most recently have begun to investigate the possible factors associated with the completions or non-completions, to include a students' original goals as well as the quality of engagement or content in the beginning of the course [98, 253]. In this application, we intend to demonstrate the utility of INSiGHT and aim to identify the latent behaviors that may indicate whether a person is more or less likely to continue (persist in) a MOOC. Specifically, we wish to examine how predictive is staying on pace with the course by accessing the breadth of online material, and the prevalence (if any) of those who may try to "cram" and complete the course in a short amount of time at the end. We propose to test our previously developed dynamic graph pattern matching tool for another application for use on the MOOC data set.

5.10.2. Data. We acquired our dataset from the Knowledge Discovery and Data Mining (KDD) Cup 2015 competition for detecting MOOC drop-outs [160]. The original dataset had 120,542 enrollment ids (students), 5,890 different courses, 8,157,277 registered online activities as part of XuetangX, a MOOC learning platform sponsored by Tsinghua University in China. To scale our work as proof of concept, we selected 1 course with the anonymized alphanumeric ID `fbPkOYLVPtPgIt0MxizjfFJov3JbHyAi` which shall be referred to throughout the rest of this chapter as Course X. The course began on January 17, 2014, and likely ran through March 2014 based upon the release dates of course material. However, the competition only provided student activities in Course X from January 17, 2014 to February 15, 2014. The objective of the competition was to predict which MOOC students continued

in Course X (as opposed to drop-out) after February 15, 2014, which is defined as some recorded course activity within 10 days after February 15, 2014. Based upon the ground truth provided, 84 students out of 983 (8.55%) continued. As with many MOOCs, course completion is considered a low-base rate behavior [154, 253].

For Course X, the competition provided a hierarchy of course content consisting of 221 unique objects as shown in Table 5.9. There were originally 983 students enrolled in the Course X, who collectively performed 41,033 activities. However, data cleaning was required. Out of all activities, 22,121 of them were attributed to 7 objects of an unknown type which were not listed in the course content hierarchy. By removing those unknown activities from consideration, we were left with 18,912 activities that corresponded to objects of type chapter pages (8), problems (33), videos (36), or sequential chapter pages (43)(for a total of 120 objects). Additionally, Fig. 5.16a-d show the frequency of each object that was accessed by the students. As expected in MOOC content, later course materials are much less frequently accessed than the earlier materials.

TABLE 5.9. Course X content

Content	**Quantity**
course homepage	1
discussion page	1
static tab	1
course info pages	2
about pages	3
outlink	5
chapter pages	8
html	10
problems	33
videos	36
sequential chapter pages	43
vertical chapter pages	78
Total	221

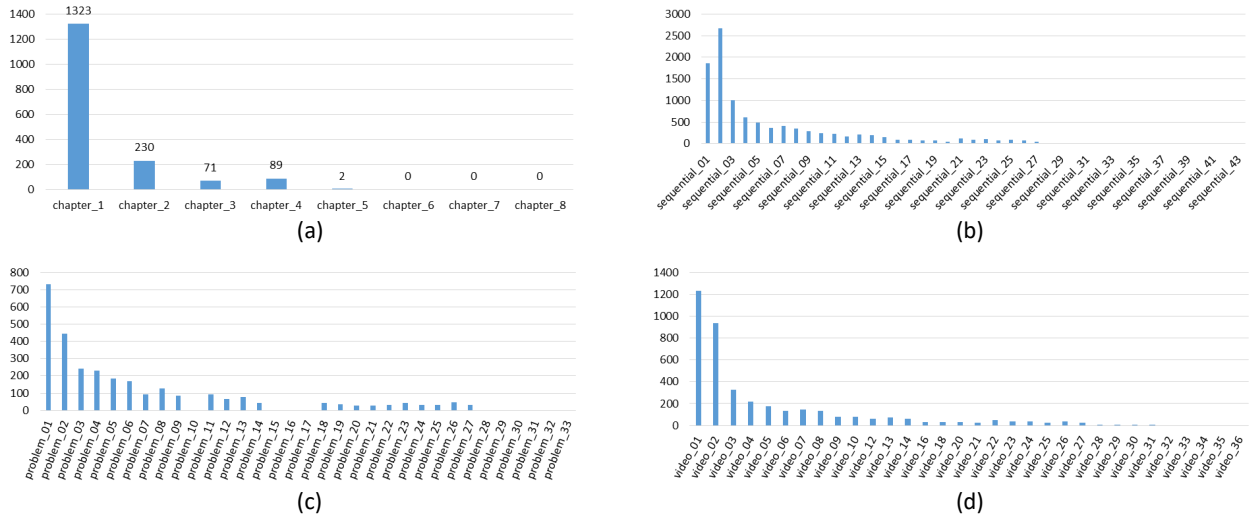


FIGURE 5.16. Activity histograms for each of the four types of activities: (a) chapters, (b) chapter-sequentials, (c) problems, and (d) videos.

5.10.3. SET-UP. Recent research suggests that participation in course activities ("specifically videos watched per week and posts and comments per week") are positively associated with completion rates [251, p. 211]. Based upon this, our intuition is that the more a student accesses the breadth of content, the higher likelihood of he or she is of continuing.

We model each student's on-line course activities a dynamic graph where nodes are the course content accessed and edges represent the hierarchical or sequential organization of the course content. An example of this hierarchy is that a student needs to access a sequential chapter landing page before accessing the problem or video included in that portion of the lesson. This is reflected in the data graph G schema shown in Fig. 5.17a. Another way to view the data model is to view the query in terms of levels

- Level 0 (root node): Student
- Level 1: Activity accessing a Sequential (43) or Chapter page (8)
- Level 2: Activity accessing a Video (33) or Problem (36)

Edges between levels were constructed from sequential data. Each time a student accessed a level 1 activity (sequential or chapter page), a new edge was created between the student a unique activity node. If the same student accessed the same sequential page twice each at different times, this would cause the creation of two nodes of the class for that sequential page and an edge between the student and each of them. When a student accesses Level 2 activity (video or problem), a new node is created with that class and an edge is created between the last sequential page associated with that Level 2 activity (determined by the course material hierarchy). The INSiGHT algorithm keeps track of the number of times a node of a certain class is accessed, but the scoring functions allow us to treat the connections as binary (occurred or did not occur) or as counts (number of times it occurred).

The resulting data graph G has the characteristics shown in Table 5.10. There is a node for all 983 students as well as a distinct activity node for each activity the students performed. The 18,912 edges connect student nodes to their activities of class sequential or chapter page, as well as the sequential pages to the associated problems and videos.

TABLE 5.10. MOOC data graph characteristics

Characteristics	G
Total Nodes	**19,895**
Number of students	983
Number of chapters	1,715
Number of problems	2,935
Number of sequentials	10,4434
Number of videos	3,988
Total Edges	**18,912**
Student to sequential/chapter activity	11,989
Sequential to problem/video activity	6,923

The query graph depicted in Fig. 5.17b is a graph of all 120 course activities. A complete match of this query would reflect that the student accessed *all* course materials (every content reading page, problem, video). The intention for INSiGHT is to find whole of partial matches

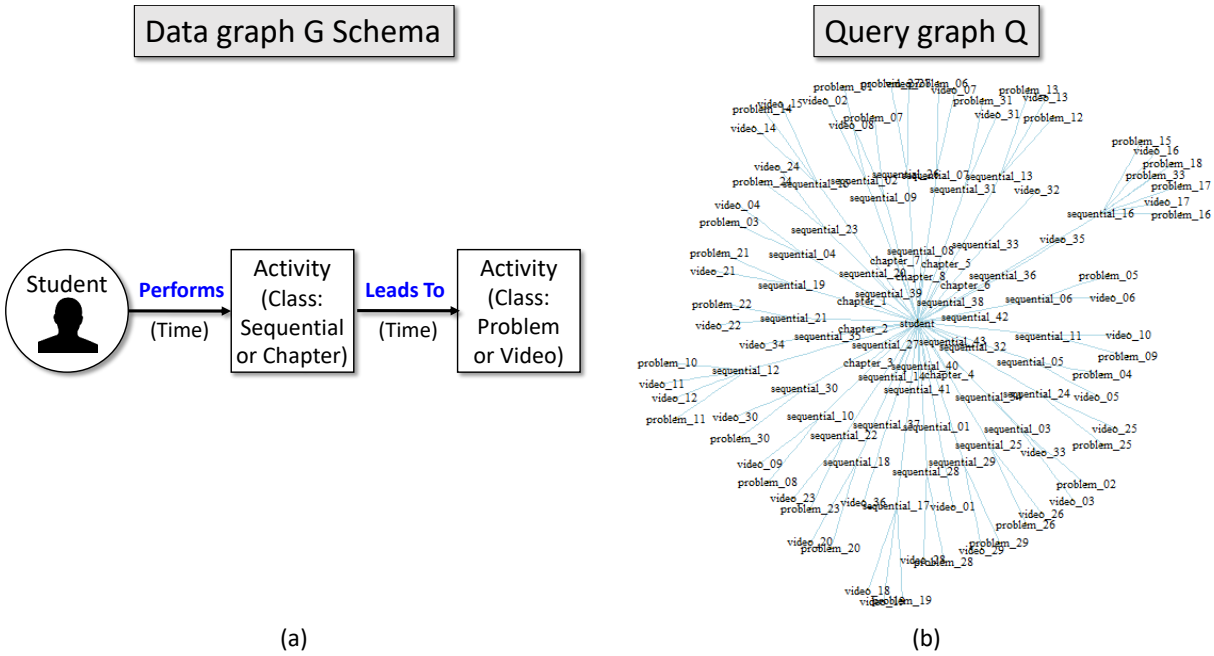


FIGURE 5.17. (a) Data graph G schema relating MOOC students and activities. (b) Query pattern for Course X, which reflects the hierarchy of course materials.

of this query pattern in the data graph of students and their activities over time. By following content access dynamically for each student, we seek to effectively determine whether each student is on a trajectory to continue with the course.

Lastly, we utilized a similar parameterization in other applications. Students were designated as query focus (QF) nodes, while all other activities were equally treated as indicator (IND) nodes. All query nodes were parameterized with $\lambda = 10.0$, $\beta = 1.0 \times 10^3$, and $\xi = 1.0$.

5.10.4. RESULTS AND ANALYSIS. INSiGHT readily found the whole or partial matches of each student to the query Q over time. The similarity score time series plot for all $n = 983$ students is shown in Fig. 5.18. Benefiting from the availability of ground truth in this dataset, the similarity score trajectories are colored blue if the student continued with the course and light red if the student dropped-out. It is clear that many of the students with the highest similarity scores throughout the window of analysis ended up continuing with

the course. This confirms our intuition that the breadth of activity is a factor to a student continuing the course. In fact, only 25 of the 983 students had final similarity scores greater than or equal to 1.00, and of these 20 were students who continued. However, 5 of these high scoring individuals did not continue.

Moreover, there are many individuals who scored low (accessed very little course content) and still persisted past February 15, 2014. The histogram in Fig. 5.19 makes this more explicit. It is important to determine what other characteristics would help us distinguish students who accessed very little course content, yet still persisted.

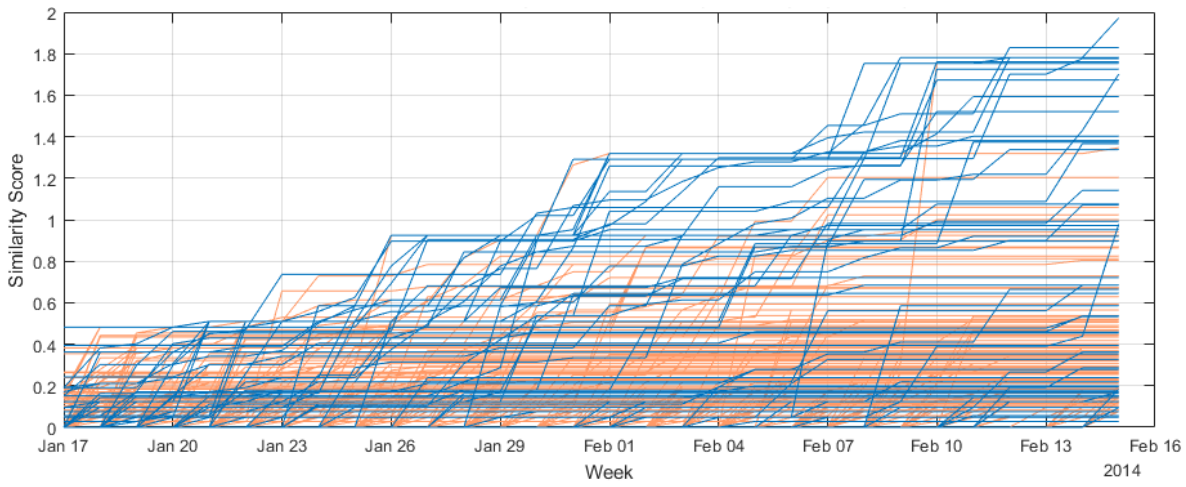


FIGURE 5.18. Similarity score time series plot for those who continued in the MOOC (blue, $n = 84$) and those who did not continue (light red, $n = 899$).

There are several limitations of the data that are worth addressing. First, as previously mentioned, nearly half of the online activities provided were attributed to unknown objects and were therefore removed from consideration. It is possible that these activities would have been more discriminating between those who continued and those who did not. For instance, 15 of the 84 'continuing' students (17%) had accessed very few course objects (final similarity score ≤ 0.20) and had been absent from the course at least 20 days. These students are depicted in the blue markers in the lower right of Fig. 5.20 who had similarity score of

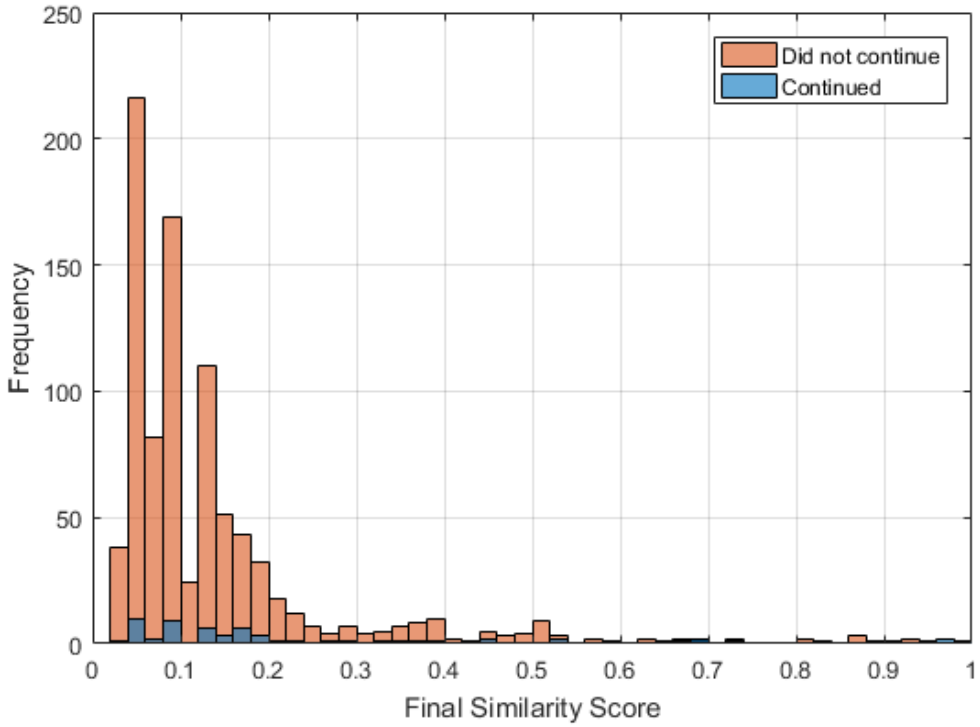


FIGURE 5.19. Histogram of the final similarity scores for the MOOC continuation query pattern.

0.2 or less and day of last activity of 10 or less. Among the disregarded activities could have been ones that rightfully belonged to these students and therefore aided in the prediction. Additionally, the competition's definition for continuing (i.e., at least one activity 10 days after February 15, 2014) is rather arbitrary and does not necessarily serve as a suitable proxy for the continuing/drop dichotomy. Considering that the course likely continued until the end of March 2014, it is possible that 10-day absence (February 16-25) from accessing course material does not necessarily mean that the student dropped. This may explain why the few who accessed a wide breadth of activities (i.e., the 5 students with final similarity scores greater than 1) still may have gone on to access the course after February 25.

We were also interested in whether the inclusion of some additional features besides the breadth of access to course content apparent in the output of INSiGHT could be used to achieve better prediction of the students who 'continued.' Specifically, we observed from the

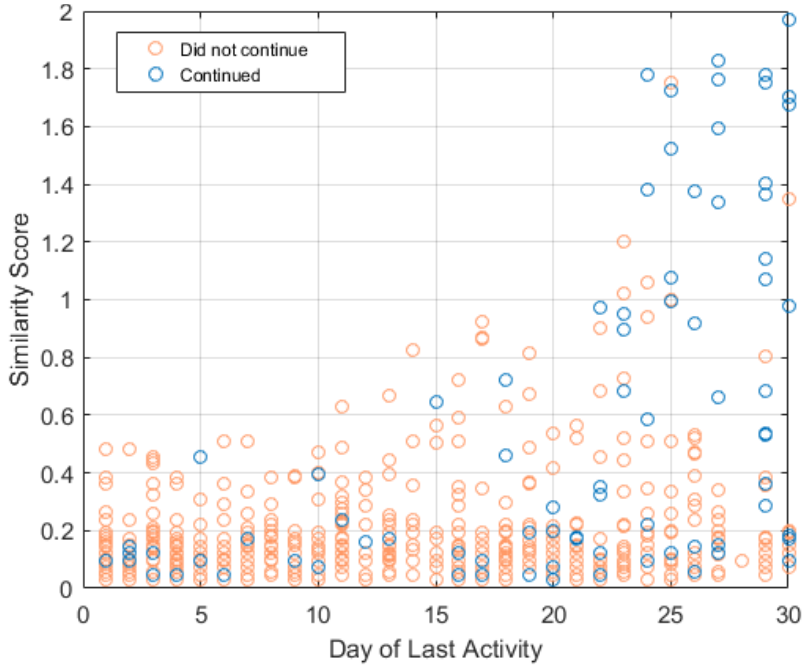


FIGURE 5.20. Scatter plot of the final similarity score and the day of last activity for both MOOC students who continued and dropped.

time series plot that those with high similarity scores who did not 'continue' seemed to last activity dates which were earlier than those who 'continued.' This led us to construct and test a logistic regression model with three features: final similarity score, the day of first activity, and the day of last activity. With 5-fold cross validation, the resulting modeling with the coefficients shown in Table 5.11 had an AUC of 77%.

TABLE 5.11. Table of terms and coefficients for a logistic regression model for MOOC continuation

Term	**Estimate**	**SE**	**tStat**	**p-value**
(Intercept)	4.1098	0.3084	13.0347	7.7699E-39
Start day	0.0481	0.0192	2.5096	0.0121
Last Day	-0.0997	0.0209	-4.7608	1.9287E-06
Final Similarity Score	-2.0646	0.4433	-4.6571	3.2077E-06

5.10.5. CONCLUSION. Overall, the use of INSiGHT in this application enabled the screening for those most likely to continue a MOOC based upon the only *one* hypothesized factor of the breadth of course material accessed. This detected over 23% of the true positives.

Notwithstanding the data limitations previously mentioned, and considering the results of the logistic regression model, it is likely that the inclusion of some additional features as outputs of INSiGHT or the development of more robust query pattern or set of patterns could be used to achieve better prediction of the students who 'continued.'

5.11. Application #3 Real Data: Consumer Project Detection

We acquired a real, large consumer activities dataset from a home improvement retailer selling products that customers use in complex and/or multi-step projects. The retailer was interested in detecting when a customer may be engaged in those projects based on shopping behavior alone. Unlike other customers who may have one-time and long-term interests and product affinities, those customers undertaking a project may be interested in purchasing more associated products but only for a short duration to get the project done. A retailer would therefore greatly benefit by quickly identifying such customers, and help them complete their projects through tailored marketing and support. In our investigative search setting, the latent behavior is a customer undertaking a specific type of project, associated purchases are indicators of that latent behavior, and the trajectory is the behavioral-temporal path of indicators towards completion of a project.

Table 5.12. Customer purchases full and subgraph characteristics

Characteristics	$G_{\mathbf{full}}$	$G_{\mathbf{subgraph}}$
Total Nodes	**11,690,738**	**271,617**
Number of customers	60,000	30,443
Number of purchase ids	11,630,738	241,174
Total Edges	**11,630,738**	**241,174**

The original dataset had 60,000 customers and over 11 million transactions from March 5, 2012 to March 4, 2014 (see Table 5.12). Each of the products purchased had a categorical hierarchy of product_id, subclass, class, and group. The retailer also labeled each customer

as either a professional (contractor) or customer (residential customer). In preprocessing, we filtered out 97.68% of nodes and 97.93% of edges in this bipartite graph which had no connection to indicators in our query. Specifically, we constructed a subgraph of those who made at least one purchase of a product (distinguished by purchase_id) whose sub-class identifier was in the query pattern. The resulting graph had only about 271K nodes and 241K time-stamped edges and is further detailed in Table 5.12 and the schema in Fig. 5.21a.

While each of the edges has a time stamp label with a date and time resolution, we binned each date into weeks. Thus all purchases made during the 2-year period were covered in 104 weekly buckets.

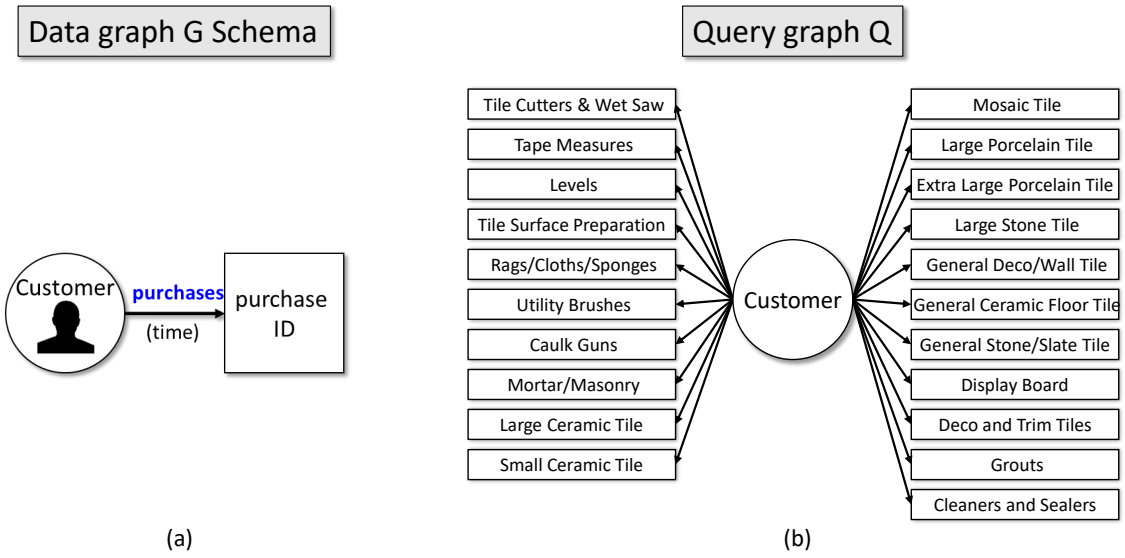


FIGURE 5.21. (a) Network schema of the Customer Activities graph. (b) Query graph Q for a tiling wall and floor project.

5.11.1. QUERY DEVELOPMENT AND DESCRIPTION. Since the home improvement retailer never specified what constituted a "project," we needed to develop a suitable set of project queries that would meet the intent. Accordingly, we decided to focus on a bathroom renovation project as a proof of principle and extracted the tools and material requirements

TABLE 5.13. Home Renovation Project Query- Class Node-Types and Parameter Set

			Parameter Set		
Class	**Label**	**NodeType**	$\boldsymbol{\lambda}$	$\boldsymbol{\beta}$	$\boldsymbol{\xi}$
A	Customer	QF	10.0	1.0×10^8	1.0
B	Tile Cutters & Wet Saw	RF	10.0	1.0×10^8	1.0
C	Tape Measures	IIRA	10.0	1.0×10^8	1.0
D	Levels	IIRA	10.0	1.0×10^8	1.0
E	Tile Surface Preparation	IND	10.0	1.0×10^8	1.0
F	Rags/Cloths/Spongers	IIRA	10.0	1.0×10^8	1.0
G	Utility Brushes	IIRA	10.0	1.0×10^8	1.0
H	Caulk Guns	IIRA	10.0	1.0×10^8	1.0
I	Mortar/Masonry	IND	10.0	1.0×10^8	1.0
J	Large Ceramic Tile	IND	10.0	1.0×10^8	1.0
K	Small Ceramic Tile	IND	10.0	1.0×10^8	1.0
L	Mosaic Tile	IND	10.0	1.0×10^8	1.0
M	Large Porcelain Tile	IND	10.0	1.0×10^8	1.0
N	Extra Large Porcelain Tile	IND	10.0	1.0×10^8	1.0
O	Large Stone Tile	IND	10.0	1.0×10^8	1.0
P	General Deco/Wall Tile	IND	10.0	1.0×10^8	1.0
Q	General Ceramic Floor Tile	IND	10.0	1.0×10^8	1.0
R	General Stone/Slate Tile	IND	10.0	1.0×10^8	1.0
S	Display Board	IND	10.0	1.0×10^8	1.0
T	Deco and Trim Tiles	IND	10.0	1.0×10^8	1.0
U	Grouts	IND	10.0	1.0×10^8	1.0
V	Cleaners and Sealers	IND	10.0	1.0×10^8	1.0

listed in [62], a how-to manual for bathroom remodels. Even within the bathroom renovation, we identified numerous sub-projects: removal and demolition, showers and tubs, sinks and vanities, toilets, lighting, and tiling walls and floors. We matched the tool and material requirements list with the product sub-class names provided by the retailer and subsequently developed query graphs for each sub-project in a bathroom renovation project. The query graph for tiling walls and floors consisting of 21 related product sub-classes is shown in 5.21b.

We found it important to further delineate those product sub-classes by type of investigative indicator according to Table 4.1 in Section 4.3, and show these labels in Fig. 5.21c. Among all the indicators, we designated the Tile Cutter and Wet Saw sub-class as a red

flag indicator of a tiling project. We also selected Tape Measures, Levels, Utility Brushes, and Caulk Guns as IIRA because they are common items that could be used in many other projects in addition to tiling. As in the Klausen radicalization parameterization, we selected $\lambda = 10.0$, $\beta = 1.0 \times 10^8$, and $\xi = 1.0$ in order to provide a full component score for the single occurrence of each indicator class and negate the effect of any score decay over time.

5.11.2. Analysis of Customer Purchase Dataset Results. We show that our approach can indeed detect customers who are likely engaged in a specific home improvement project. Fig. 5.22 shows the similarity scores over the 2-year period for the top-3 scoring residential customers. We observe that customers 625 and 646 seemed to have initially purchased a substantial number of items towards the project, but each had long periods before they purchased additional items. On the other hand, customer 685 seemed to have purchased items for the project regularly over a much longer period of time.

We first describe the purchase behavior in each of the two market segments. The average final similarity score among professionals)was statistically higher than customers (0.1 average difference had a p value < 0.0000). This difference in similarity scores is depicted visually in the histogram in Fig. 5.23a, where the professional similarity score distribution has a heavier tail than the customer distribution, as well as the box plots in Fig. 5.23b. We also found that the average duration of the project trajectories among professionals was statistically higher than customers (20 week average difference had a p value $= 0.0003$), as shown in Fig. 5.23c. Both these observations match our intuition because professionals are likely doing multiple tiling projects over the 2 year period and purchase more items over time to support them, while residential customers are likely just buying for their single home project. Also, professionals match more of the query project pattern because they use (and

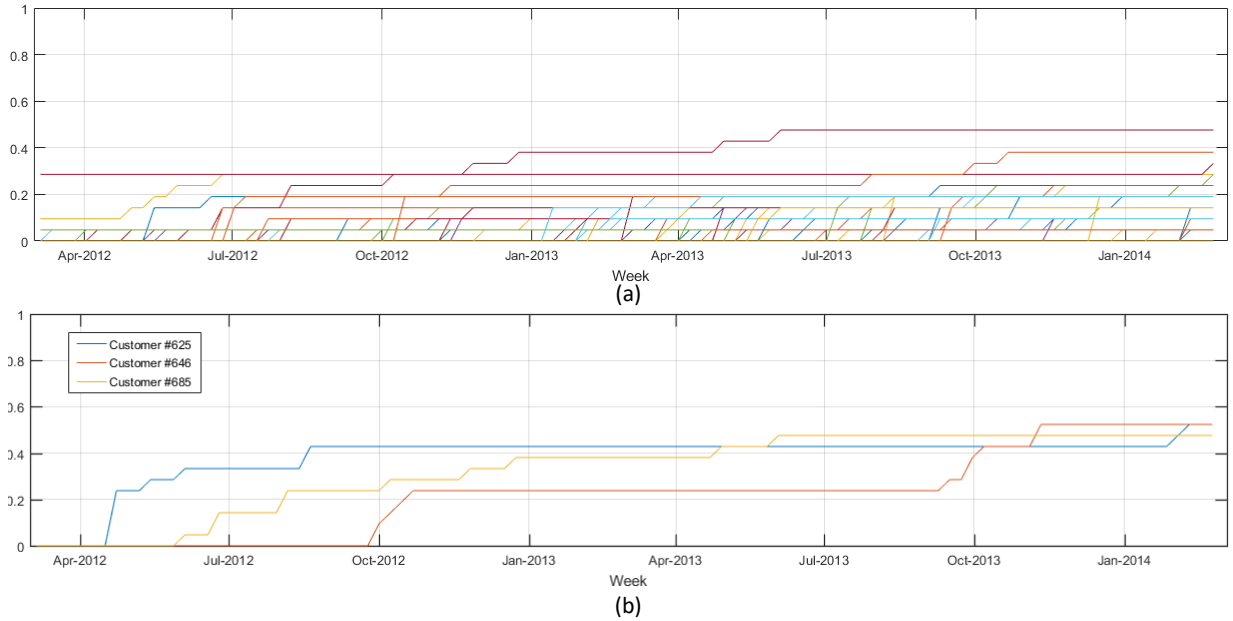


FIGURE 5.22. Plot of the class similarity scores over time $\tilde{s}(t)_n$ for (a) 150 randomly sampled non-contractor customers and (b) the top 3 non-contractor customers in the home improvement purchase data graph. For each of the weekly timesteps t between March 5, 2012 and March 4, 2014, we show the changes in class similarity.

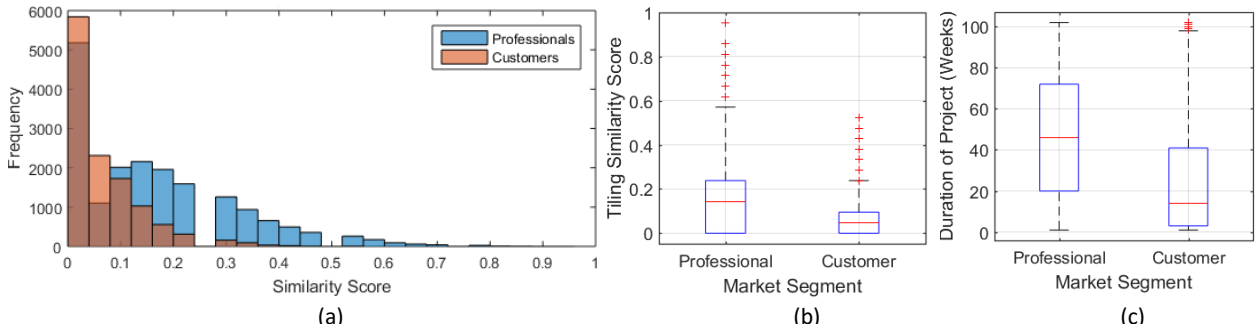


FIGURE 5.23. Histogram of class similarity scores (a) for all 30,443 customers in the subgraph and the corresponding boxplot (b) of the class similarity scores which shows a statistically significant difference in distribution means between professionals and customers. (c) is a box plot of the project durations in weeks for both groups.

may break) more tools needed for a tiling project, and may buy more tile varieties (which are counted as distinct indicators in the query pattern).

Fig. 5.22b also shows that over 5000 professionals and nearly 6000 residential customers had a similarity score of 0 even at the end of the period of analysis (March 2014). This

is due to these customers purchasing only one or more items among the IIRA sub-classes specified in the query (Fig. 5.21c). The use of the labels for investigation indicator types truly helped reduced false positives because it filtered out customers who only purchased items that could have been used for a number of other projects besides tiling.

5.11.3. Discussion on alert criteria development. The objective of all the applications featured in this research is the detection of a latent behavior that may be signaled over time. However, in the detection of both radicalization and evolving blog topics, while analysts would likely want to be aware of those on pathways towards the behavior, they may not necessarily have the urgency to detect the behavior at the earliest opportunity. The occurrence of indicators over time helped analysts follow those who may be of increasing interest, and some red flags may assist in giving some more concrete signals. However, the ultimate goal of this home improvement purchases analysis was to detect the commencement of a specific project at the earliest opportunity for the retailer. Detection achieved at a later time means potentially lost revenue to competing customers or lost opportunities to increase customer satisfaction because of a failure to incentivize project completion. Therefore, more focused analysis is needed on a customer's related purchases *initially* over a short period to determine if the customer was undertaking such a project. Also in this application, red flag indicators may not occur in a specific sequence. For example, if one were engaging in a tiling project, we have not found evidence to suggest that the purchase of a critical item item such as a tiling saw would necessarily occur at the beginning or the end in a period of acquiring materials and tools.

We discuss here the type of analyses needed to develop useful alert criteria for the retailer to identify those most likely to be undertaking a project. We propose two promising means:

association analysis and linear modeling of final class similarity scores based on customer activities.

5.11.3.1. *Association analysis of customer purchases.* We first utilized association analysis, which is a classical technique used to determine purchase patterns in consumer market baskets [2, 293]. This technique is also known as frequent itemset analysis or association rule mining. For this particular application, this method finds among all the customers' itemsets (collection of items purchased) those k-sized sets of items that occurred with some threshold frequency and determines any predictive associations that exist among the items. We discovered that it has the potential to reveal the 'rules' which allow the vendor to predict another single item that may be bought given one or more items already purchased, but seems to be limited to itemsets of size k=3 and only makes prominent those most frequent itemsets while ignoring other indicators. Ultimately, this analysis did provide insight into the quality of our initial query pattern and demonstrated just how relatively infrequently customers bought a significant portion of the items in the query pattern.

Association analysis is based fundamentally on three metrics:

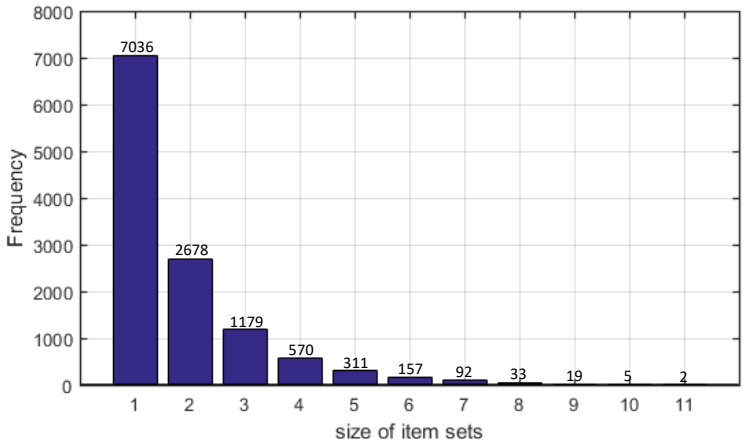
(1) Support: Fraction of all transactions that contain the itemset. This metric describes the relative prevalence of the itemset in the data.

(2) Confidence: Fraction of occurrences when items in a consequent set appear in transactions that contain the antecedent set. This metric describes the relative prevalence of the association rule relating the antecedent and consequent sets of items.

(3) Lift: The probability of the consequent set occurring with the antecedent set, divided by the probability that the consequent set occurs at all. A lift value of 1 means that

the probabilities of the antecedent and consequent itemsets are independent, while a lift value greater than 1 means that there is a dependent relationship [2, 293].

There are now several techniques with varying efficiency to calculate these metrics for a given dataset, but we utilized the popular and well-known a priori algorithm [2] which is a based on breadth-first search. It efficiently generates candidate sets (starting with sets of size 1) using the a priori principle[39] and then prunes those candidate sets that fail to meet a relative frequency threshold. We utilized a MATLAB implementation that is readily available [283].



size of itemset	frequency	percent
1	7036	58.2%
2	2678	22.2%
3	1179	9.8%
4	570	4.7%
5	311	2.6%
6	157	1.3%
7	92	0.8%
8	33	0.3%
9	19	0.2%
10	5	0.0%
11	2	0.0%
Total	**12082**	**100%**

FIGURE 5.24. Histogram and associated table of the size of the itemsets among customers in the dataset

Given the criticality of itemsets in association analysis, we produced the histogram of itemset sizes among customers in Fig. 5.24. This histogram is a variant of the one shown in Fig. 5.23a, without consideration for IIRA items. Based upon this, one can conclude that over 90% of the customers purchase no more than 3 distinct items from the tiling project item list, and that it is relatively rare for customers to purchase a total of 4 or more distinct items. Given the lack of ground truth in the data, it is impossible for us to definitely determine a

[39]Apriori principle (downward closure lemma) says that if an itemset X is not frequent, then any candidate set that contains X is guaranteed to be not frequent [2].

customer's true intentions with respect to the project regardless of the size of their itemsets. Clearly, larger itemsets imply a greater likelihood that the customer was pursuing a tiling project, but a vendor would be interested in identifying those with 2, 3 or 4 size itemsets to purchase more related items.

Selecting the minimum support threshold involves a trade-off between the number of rules mined and their relative frequency. In other words, setting a low frequency would get more association rules but the rules would have a lower relative frequency (and thus in some sense relevancy). Due to the distribution multi-item sets by customer shown in Fig. 5.24, it is clear that a minimum support threshold would have to be set fairly low in order to examine itemsets of size $k \geq 3$, where the purchase of at least 2 items would help infer the third item. Table 5.14 shows the various minimum support thresholds tested with association rule mining and the resulting number of frequent itemsets that met the threshold and the maximum level (largest sized itemset) reached.

TABLE 5.14. Minimum Support and the Resulting Size and Type of Itemsets

Minimum Support	**Frequent Itemsets Found**	**Max Level Reached**
0.20	2	1-itemsets
0.10	8	1-itemsets
0.05	14	2-itemsets
0.03	23	2-itemsets
0.02	42	3-itemsets
0.01	90	4-itemsets

We selected a minimum support of 0.02 in order to examine itemsets of least size 3 while still having some support and a confidence threshold of 0.30. See Fig. 5.25 for a listing of the resulting 22 rules. The results match much of our intuition about products related to the completing of a tiling project. For example, the rules 'CLEANERS AND SEALERS'→'GROUTS', 'DECO AND TRIM TILES'→'GROUTS', and 'MOSAIC'→'GROUTS'

are sensible and have greater than 0.05 support and with life values greater than 2. For 3-itemsets, the rules {'CLEANERS AND SEALERS', 'MOSAIC'}→'GROUTS' and {'LESS THAN 12" TILE/STONE', 'MOSAIC'}→'DECO AND TRIM TILES' each have life values greater than 3 and confidence levels of 78.7% and 63.3%, respectively.

Number	Antecdent	Consequent	Confidence	Lift	Support
1	'12" TO 13" PORCELAIN'	'GROUTS'	0.583	2.414	0.0271
2	'DECO AND TRIM TILES'	'LESS THAN 12" TILE/STONE'	0.472	4.820	0.0453
3	'LESS THAN 12" TILE/STONE'	'DECO AND TRIM TILES'	0.462	4.820	0.0453
4	'LESS THAN 12" TILE/STONE'	'GROUTS'	0.493	2.041	0.0483
5	'MOSAIC'	'LESS THAN 12" TILE/STONE'	0.319	3.261	0.0363
6	'LESS THAN 12" TILE/STONE'	'MOSAIC'	0.371	3.261	0.0363
7	'CLEANERS AND SEALERS'	'GROUTS'	0.495	2.049	0.0685
8	'DECO AND TRIM TILES'	'GROUTS'	0.533	2.209	0.0512
9	'MOSAIC'	'DECO AND TRIM TILES'	0.329	3.434	0.0375
10	'DECO AND TRIM TILES'	'MOSAIC'	0.391	3.434	0.0375
11	'MOSAIC'	'GROUTS'	0.487	2.015	0.0554
12	'DECO AND TRIM TILES' and 'MOSAIC'	'LESS THAN 12" TILE/STONE'	0.614	6.268	0.0230
13	'LESS THAN 12" TILE/STONE' and 'MOSAIC'	'DECO AND TRIM TILES'	0.633	6.601	0.0230
14	'LESS THAN 12" TILE/STONE' and 'DECO AND TRIM TILES'	'MOSAIC'	0.508	4.466	0.0230
15	'GROUTS' and 'MOSAIC'	'LESS THAN 12" TILE/STONE'	0.383	3.908	0.0212
16	'LESS THAN 12" TILE/STONE' and 'MOSAIC'	'GROUTS'	0.583	2.415	0.0212
17	'LESS THAN 12" TILE/STONE' and 'GROUTS'	'MOSAIC'	0.439	3.858	0.0212
18	'GROUTS' and 'MOSAIC'	'CLEANERS AND SEALERS'	0.365	2.632	0.0202
19	'CLEANERS AND SEALERS' and 'MOSAIC'	'GROUTS'	0.787	3.260	0.0202
20	'GROUTS' and 'MOSAIC'	'DECO AND TRIM TILES'	0.396	4.129	0.0219
21	'DECO AND TRIM TILES' and 'MOSAIC'	'GROUTS'	0.585	2.423	0.0219
22	'DECO AND TRIM TILES' and 'GROUTS'	'MOSAIC'	0.429	3.768	0.0219

FIGURE 5.25. Association rules generated for the tiling project using the minimum support of 0.02 and minimum confidence of 0.30.

In summary, we first found that itemsets of size 4 and greater were rare in the customer purchase data. Despite this, and the fact that we had accepted very low support (2% of all transactions) to find 3-itemset association rules, association analysis did provide a set of rules with high lift that provides confidence estimates of a third item with 60-70% probability. There are several areas of future work. First, we would recommend to the vendor to expand its data collection efforts to gain ground truth for project-related purchases. This could be achieved for instance by the vendor asking customers to register their intended project in advance through their store account to receive product discounts. We also intend to incorporate these association rules into INSiGHT to help expand the alert system for emailed

incentives following the matching of antecedent purchases (indicators) for discounts or sale of items in the consequent and other related items. Lastly, we intend to use this technique in a formal process to assist in query pattern refinement. Items that turn out to have very low support over time could be removed from the query, while other new associated items beyond the original query could be added and tracked as important new indicators.

5.11.3.2. *Modeling final class similarity scores based on customer activities.* We first proposed two basic metrics that may be correlated to an individual's ultimate class similarity score, each of which is fast to calculate and utilizes only the system output of time-based similarity scores. The intention is to calculate one or both of these methods in some initial period in order to determine whether an individual ends up purchasing a preponderance of the items from the query for their project. For our purposes, we defined 'initial' time period for each customer i as $\Delta_{z,i} = t_{z,i} - t_{0,i}$, where $t_{0,i}1$ is the week of the first non-IIRA purchase and $t_{z,i}1$ is the z-th week after the first non-IIRA purchase by customer i.

(1) Initial similarity score at time period $t_{z,i}$. Defined as the similarity score achieved by customer i at $t_{z,i}$.

(2) Initial similarity score gradient at time period $t_{z,i}$. Defined as the similarity score achieved by customer i at $t_{z,i}$ divided by $\Delta_{z,i}$.

Both of these methods were tested against the final (ultimate) similarity scores and the resulting correlation values are depicted in Fig. 5.26. The error bars are the 95% confidence intervals around correlation coefficient. At face value, these results show that the correlation between initial and final similarity scores is monotonically increasing for longer lengths of the initial period of consideration. Also, the correlation between initial similarity score gradient and final similarity score steadily decreases for initial periods longer than 1 week

and is always less than the similarity score correlation after the first week. We conclude that while the initial set of purchases made within a week period for a project may is moderately correlated to the final span of items ultimately purchased, this correlation becomes negligible rather quickly. While we allow that this conclusion is not necessarily generalizable to other projects or domains, for the remainder of this analysis we will build models to predict final similarity scores using only the initial scores rather than gradients.

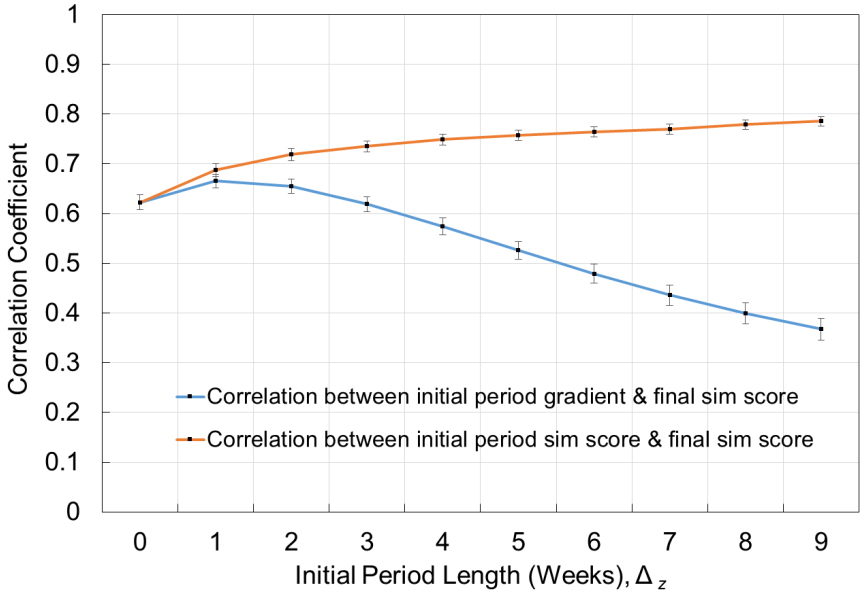
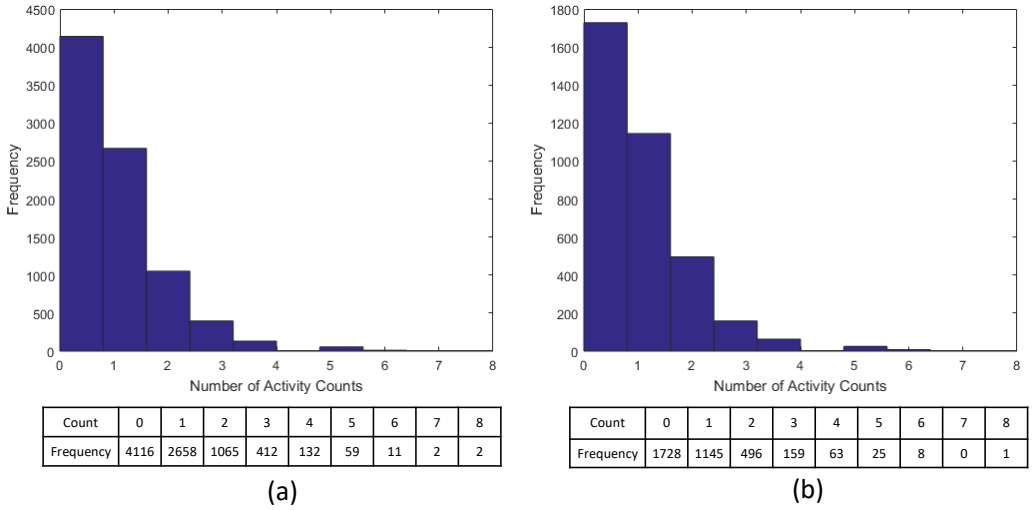


FIGURE 5.26. Correlation coefficients with the 95% CI for both the initial similarity score and initial gradient metrics against the final similarity score.

Having established a positive correlation between the final and initial similarity scores, we continue with an effort to build a fast, simple linear model to dynamically predict final similarity scores (the span of project-related items purchased) in order to enable the vendor detect customers seemingly on the way towards a project so that it could focus marketing efforts on them. The dataset of 12,082 customers was divided into a training (70%) and testing (30%) set consisting of 8,457 and 3,625 individuals, respectively. After trying several variants of the dependent variable related to an initial similarity score, we ultimately

determined a fair performing variable was the similarity score after a specified number of activities, which is defined as the number of times a customer visited the store (binned into weeks) where there was the purchase of an item in a new sub-class in the project query. This variable in some ways mimics the vendor's receipt of information: "Customer X visited our store Y times and has now bought up to Z things in this project pattern." Those who made just one visit to purchase an item that is classified as IIRA would be considered as having zero activities towards the project. The number of customers sorted by their highest activity count in both the training and testing datasets is depicted in the histograms in Fig. 5.27.



Count	0	1	2	3	4	5	6	7	8
Frequency	4116	2658	1065	412	132	59	11	2	2

(a)

Count	0	1	2	3	4	5	6	7	8
Frequency	1728	1145	496	159	63	25	8	0	1

(b)

FIGURE 5.27. The histogram of customers by activity counts in both the training (a) and testing (b) datasets.

Using the training data, we performed the simple linear regression: Final Similarity Score $\sim$ Similarity Score of x Activities, where a separate model was built for each activity count $x = 1...5$. For instance, the first regression was for Final Similarity Score $\sim$ Similarity Score of 1 Activity, and utilized the records in the training set who and 1 or more activities (which as shown in Fig. 5.27a is 4,341 individuals).

TABLE 5.15. Model fit for training set using various activity counts

Item	**Activity 1**	**Activity 2**	**Activity 3**	**Activity 4**	**Activity 5**
n	4341	1683	618	206	74
RMSE	0.0571	0.0503	0.0445	0.0413	0.0379
R^2	0.403	0.559	0.651	0.674	0.752

It is important to note that since we based the dependent variable on activity counts, we no longer used the week from initial purchase as an overt method of selection. For instance, for those individuals who make 2 activities (2 visits binned by week when at least one item from a new sub-class is purchased each visit) could be performing those visits over a period of time 2 or more weeks long depending upon how much time elapse between visits. Secondly, because there are fewer customers at each activity increment, each successive model was built on decreasing sample sizes. The resulting linear models for activity counts of 3, 4, and 5 had greater than 65% R^2 values as shown in Table 5.15. In the same table are the root mean squared errors (RMSE) for each model, which quantifies the spread of the actual final similarity scores around the predicted final similarity score using the same units. The RMSE for models using Activity 3, 4, and 5 were also less than the score increment for buying one product of a new subclass in the query. Since the query had 21 items, each new items contributed a score of $1/21 \approx 0.47619$. This means that the margin of error of prediction of the final similarity score is essentially less than the score change by plus or minus one query item.

The same set of linear models were then run on the test data, which are visually depicted in Fig. 5.28. As before with the training set, the models for each successive activity set were developed using smaller and smaller samples. The R^2 and the RMSE values are in Table 5.16 and show comparable performance. Fig. 5.29 also depicts how the models using activities

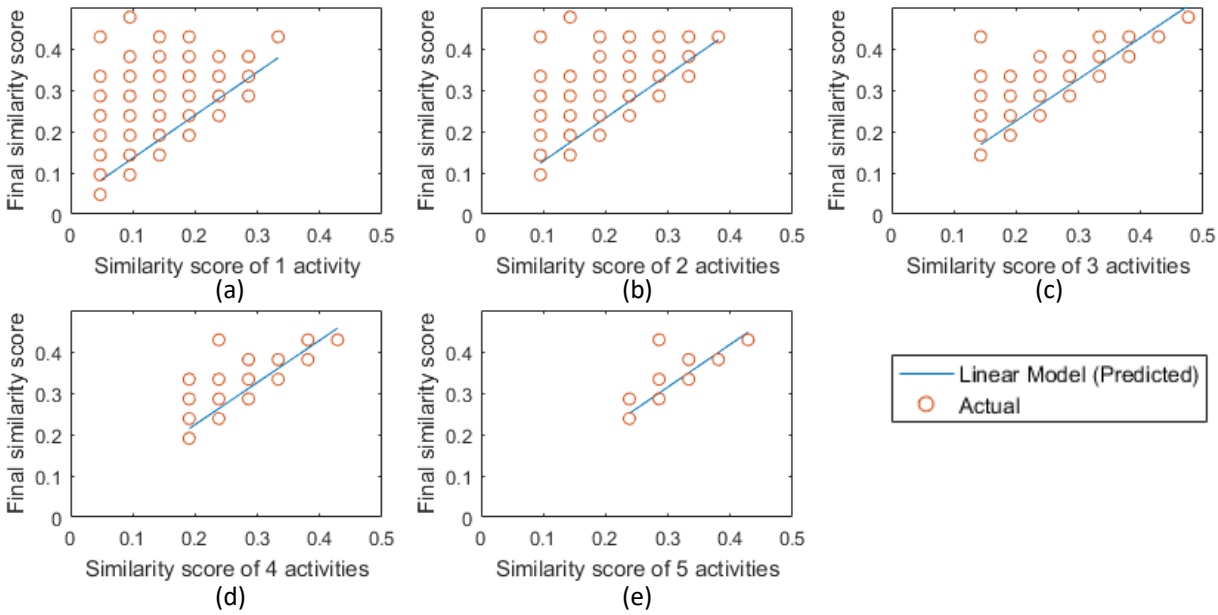


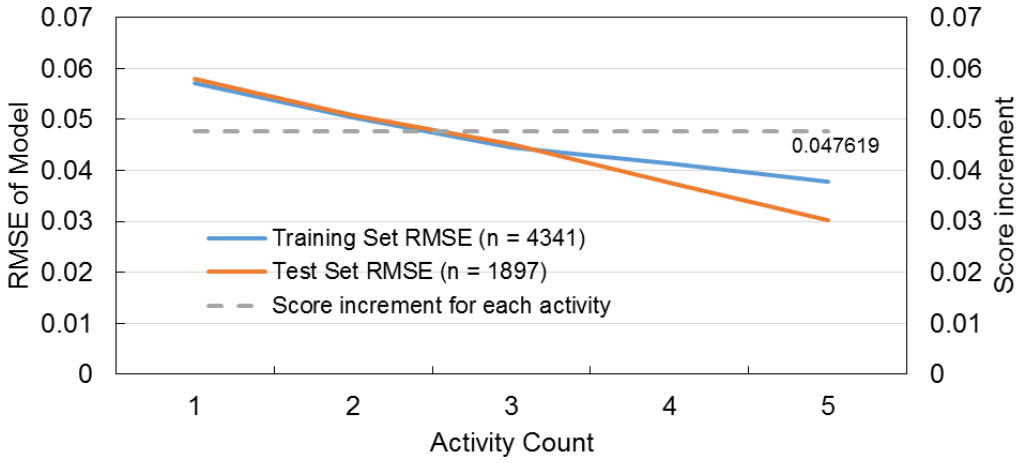
FIGURE 5.28. Plots of the regression models of the final similarity scores against the test data for similarity scores after (a) 1 activity (n=1897), (b) 2 activites (n=752), (c) 3 activities (n=256), (d) 4 activities (n=97), and (e) 5 activities (n=34).

3 or greater produced RMSE that were smaller than the score increment for a single query item.

TABLE 5.16. Model fit for test set using various activity counts

Item	**Activity 1**	**Activity 2**	**Activity 3**	**Activity 4**	**Activity 5**
n	1897	752	256	97	34
RMSE	0.058	0.0509	0.0452	0.0376	0.0303
R^2	0.379	0.545	0.641	0.694	0.698

5.11.3.3. *Conclusion.* The point is that even a simple intuitive model can have some decent results, which can likely be improved upon if one analyzes in closer detail what the individuals actually purchase (i.e., quantity) as well. This is an important first step towards integrating models into the INSiGHT technology in order to move towards the prediction of latent behaviors.



Activity Count	1	2	3	4	5
Training Set RMSE	0.0571	0.0503	0.0445	0.0413	0.0379
Test Set RMSE	0.058	0.0509	0.0452	0.0376	0.0303

FIGURE 5.29. The RMSE of the linear activity models using activity counts for both the training and testing data.

5.12. CONCLUSION

In this important chapter, we formulated INSiGHT, a dynamic inexact graph pattern matching technique that identifies individuals with conforming subgraphs to a query pattern and follows the match trajectories over time. Tailorable to the detection of radicalization indicator patterns, enhancements were also developed to account for the re-occurrence of indicators, the time decay of indicator significance, and the incorporation of red flag and other conditional filters. We demonstrated the performance of our approach on a variety of real-world and synthetic datasets of various sizes and domain applications.

On small synthetic radicalization datasets, we successfully validated the matching mechanics and enhancements, and demonstrated that our technique was useful in producing consistent, informed, and reliable judgments about those stylized individuals who posed a significant risk for violent extremism. On a BlogCatalog dataset of over 470K nodes and 4

million edges, where 98.56% of nodes and 99.65% of edges were filtered out with preprocessing steps, INSiGHT successfully detected the trajectory of the top 1,327 nodes towards a query pattern. INSiGHT also ably determined the radicalization pattern match trajectory of all 135 U.S. violent extremists in the real Klausen time-stamped behavioral dataset. We noted a wide distribution of similarity scores on even just these positive cases of extremists and the difficulty of determining a suitable threshold to distinguish positive and negative cases. Importantly, however, we demonstrated how the inclusion of red flag visualizations and alerts could greatly assist analysts in identifying high risk individuals even when they had relatively lower similarity scores.

Using a real, large proprietary consumer activities dataset from a home improvement retailer with 60K customers and over 11 million time-stamped transactions, INSiGHT was indeed useful in the detection of customers likely undertaking certain home improvement projects based upon the number of project items purchased. However, we were unable to truly validate the utility of INSiGHT to screen for true positives because the data did not contain ground truth. Through secondary analysis for the features which best predict a customer's final similarity score, we found that models using activity sets of three or more produced RMSE values that were less than the score increment for a single query item.

CHAPTER 6

Synthetic Data Generator for Latent Behaviors

6.1. Introduction

The research into the commercial applications such as the detection of project-related purchases revealed a need for large, non-proprietary datasets for training and testing that contain individual-level, time-stamped activities, and, as much as possible, a balanced ground truth on the latent behavior. The deficit is also applicable more broadly to the domain of detecting other behaviors such as radicalization to violent extremism. In this latter case, the challenge is two-fold. First, researchers have difficulty gaining access to sensitive and even classified details of individuals which are held closely by law enforcement and intelligence agencies. See [264, p. 6-7] for a more detailed discussion on the "lack of comprehensive and reliable data" impeding scholarship in terrorism studies. The second challenge is the lack of balanced training datasets because extremist violence is relatively rare (but often times causing horrific mass casualty events). These same or similar challenges in other domains has led to the development of empirically-based synthetic data generators. Notable ones exist in the area of network traffic data [7] and insider threat data [120].

In this section, we take the first step in devising synthetic data generators for the further development and testing of INSiGHT and specifically describe a rule-based synthetic data generator to replicate individual consumer purchases of home improvement projects over time. In rule-based simulation, human behavior is replicated using a set of causal if/then associations to select actions [239, p. 128].

6.2. Related work

There is a quite a large body of literature on modeling customer behaviors. We briefly cover some of the most relevant, and recent research in this section. The probability distributions to model various customer behaviors were discussed by Fader et al in [99], which also includes a proposal for the use of Weibull in purchases times. Leeflang also produced another relevant work which proposed individual demand models related to four customer decision: "whether to buy,", "what to buy," "how much to buy," and "when to buy" [183]. Researchers in [324] also developed an agent-based model of consumer purchase decision-making they specifically used to test for an emergent behavior called the decoy effect. Lastly, we based some the fundamental concepts for our synthetic data generator on [202], who provided the GNU code for a simulator of customers movie purchases based on an assigned genre preference.

6.3. Assumptions

This generator was based upon the following assumptions.

(1) Customers can plan 1 or more projects in 2 year period. Support: 2013 Houzz & Home survey [138] showed that there were more planned projects than respondents, meaning some customers must be planning 2 or more projects.

(2) Customers have a loyalty to a particular retailer. Support: Accenture survey in 2015 said 28% of consumers are loyal to their providers and brands [43].

(3) Customers purchase of the number of items in a home renovation project supply list decreases exponentially. The support for this came from our analysis of real consumer purchase data from major home improvement retailer.

(4) Some customers do not complete a planned project and project completion rates are independent of the customer. Support: Black+Decker 2014 Home Project Survey [28] showed that 52% of respondents said that they had 1 or more unfinished projects at home.

(5) Project start dates are independent and identically distributed over a 2 year period. Support: This is a simplification of a commonly accepted seasonal model for home renovations, which can be improved upon later with additional data.

(6) Customers purchase project items with equal probability from their project list. Support: This is likely a simplification of reality but can modified after more analysis of real purchase data.

(7) Customers purchase unknown project items (noise) through a weighted random sample of an empirical item distribution. The support for this came from our analysis of real consumer purchase data from major home improvement retailer.

(8) Customers item purchase times per project occur at a diminishing rate. The support for this came from our analysis of real customer purchase data from major home improvement retailer.

(9) Order of items purchased per project is determined randomly. This is likely a simplification of reality but can modified after more analysis of real purchase data, particular the sequence of frequent itemsets by the customer.

6.4. Rule-Based Purchase Activity Generation

The steps for the product purchase generation are described below. Table 6.1 summarizes the notation used for the sets, random variables, and parameters.

TABLE 6.1. Defined sets and random variables in the synthetic data generator

Sets and RV Notation	**Description/Meaning**
$A = \{a_1, ..., a_{n_p}\}$	Set of n_p project types (includes 'unknown' project for noise)
Ω	Set of all items related to any project type $\in A$
$B_{a.} \subseteq \Omega$	Set of items for each project type $a. \in A$
$C = \{c_1, ..., c_{n_r}\}$	Set of n_r proportions of planned purchases actually made by customers
X_i	Random variable for the number of projects customer i undertakes within a period of $1, ..., t_{\max}$ days, where $1 \leq s \leq n_p$.
$Y_{i,k}$	Random variable for project type of customer i's k-th project, where $k = 1, 2, .., x_i$
L_i	Random variable for the customer i's loyalty percentage for all projects
$G_{i,k,a.}$	Random variable for percentage of items in project item list $B_{a.}$ that customer i plans to purchase
$H_{i,k}$	Random variable for the proportion of planned purchases actually made for customer i's k-th project, where $k = 1, 2, .., x_i$
J	Random variable for purchase times (in days) for each item purchased in a project counted from the day of the first purchase for that project
Parameter Notation	**Description/Meaning**
n_c	Number of customers to simulate
$t_{\max}$	Time horizon in days
s	Number of different project types that a customer may undertake (possibly simultaneously) within the period of $1, ..., t_{\max}$ days, where $1 \leq s \leq n_p$
$f(X_i)$	Probabilities for each of the s projects in the PMF for X_i (applicable $\forall\ i$)
$f(Y_{i,k})$	Probabilities for each of the n_p projects in the PMF for $Y_{i,k}$ (applicable $\forall\ i$)
λ for $f(G_{i,k}; \lambda)$	Parameter λ (decay) for the exponential PDF of $G_{i,k}$ (applicable $\forall\ i$)
$f(H_{i,k})$	Probabilities for each of the n_r proportions in the PMF for $H_{i,k}$ (applicable $\forall\ i$)
α, β for $f(J; \alpha, \beta)$	Parameters α (shape) and β (scale) for the Weibull distribution PDF of J
a for $f(L_i; a, 1.00)$	Parameter a for the start point in a uniform distribution PDF of L_i between $[a, 1.00]$
t_{noise}	Time horizon in days for the start of purchases for project 'misc' (noise)

(1) *Data requirements.* The user of the synthetic data generator must have apriori a large listing of items Ω, which can be sorted into one or more elements in A, the set of n_p project types. Each of these project types $(a_1, ..., a_{n_p})$ will have a set $B_{a_\cdot}$ of items assigned as a product/materials list; items can belong to one or more item sets.

(2) *Create customer profiles and planned projects.* Profiles for all n_p customers are determined by the outcome of several random variables. First, X_i is the random variable for the number of (possibly simultaneous) projects customer i undertakes over a period of $t_{\max}$ days. The number of projects is randomly assigned to each customer according to the PMF $f(X_i)$. Once each customer has assigned a total project capacity k of one or more projects, the random variable $Y_{i,k}$ determines the project type for each project. The first of these project types is randomly assigned according to the PMF $f(Y_{i,k})$. All subsequent projects (if any) are uniformly at random assigned from the remaining unassigned project types without replacement. Next, each customer i is uniformly at random assigned a purchase loyalty percentage, which is defined as the percent of products bought from the favored store over rival stores. The value is drawn randomly from a uniform distribution L_i between $\big[a, 1.00\big]$.

(3) *Determine planned and actual purchases.* The percentage of project items that customer i plans to purchase for each of the known k projects of type $a_\cdot$ is determined by the random variable $G_{i,k,a_\cdot}$ and is instantiated through a normalized PMF derived from an empirical exponential decay function with parameter λ over the domain of the size of the item list for project $a_\cdot$. For the unknown project ("noise"), the

number of project items planned for purchase is determined by randomly selecting a number between $[0, 1]$ and using the inverse of an empirically derived exponential decay function with λ_{misc}. By rounding the result to the nearest integer, we directly generate the number of unique sub-class purchases. After this, the actual percentage of planned items purchased is determined by the random variable $H_{i,k}$ and draw randomly according to the PMF $f(H_{i,k})$. The resulting number of project items customer i purchased for project $a.$ is determined by the product of $|B_{a_k}|$ (cardinality of the item set B_{a_k}), $g_{i,k}$ (percentage of project items planned for purchase), $h_{i,k}$ (percentage of items actually purchased), and l_i (primary vendor customer loyalty percentage). The actual purchased items for each known project are drawn uniformly without replacement from the full project itemset. For the unknown project ("noise") the purchased items are determined randomly through the empirical weightings of all subclasses project.

(4) *Determine purchase times for all purchases.* The project initiation day for each of customer i's non-miscellaneous projects is selected uniformly at random from $[1, t_{\max}$ For the unknown project, the initiation date is selected uniformly at random over $[1, t_{\text{noise}}]$ days, which effectvely places the "noise" purchases near the beginning of the analysis window. The purchase time for each item actually purchased for a known project is determined by the random variable J and random draws from the Weibull PDF $f(J, \alpha, \beta)$ and added to the project initiation day. For the unkwown project, the time of actual purchases is selected uniformly at random from 1 to $(t_{\max} - t_{\text{noise}})$.

6.5. Example Run of the Synthetic Data Generator

In order to derive a well-founded list of project-related material and tools for an interesting set of projects, we followed the same process as the development of the tiling query described in Section 5.11.1. This included references to home improvement books and selected online resources with tools and material lists such as [78, 191] and the search for relevant item subclass names in the real product list from the home improvement vendor. The resulting project items lists for a minor kitchen remodel, tiling, and attic insulation project are shown in Fig. 6.1. One of the features that we intended to test with this simulation is the ability to analyze how customers undertaking one project could be distinguished from customers undertaking another project when a number of item sub-classes in each of the projects overlap. Specifically, in the project item lists show in Fig. 6.1, one can determine that the Kitchen project has 7 out of 15 item sub-classes which overlap with the Wall and Floor Tiling project (MORTAR/MASONRY/STUCCO, MOSAIC, GROUTS, CAULK GUNS, TILE SURFACE PREPARATION, TILE CUTTERS & WET SAWS, and CLEANERS AND SEALERS). Likewise, the Wall and Floor Tiling project has 7 out of 16 item sub-classes which overlap with the Kitchen project. It is also important to note that the Attic project has no item sub-class overlap with either the Kitchen nor the Wall and Floor Tiling project.

Additionally, we designated a Project 4 as an unknown in order to generate noisy purchases in the data from 1778 sub-classes of items. This is to introduce for analysis who may have bought 'random' items for other projects. False positives can come about when those individuals bought items which by chance happened to be from the list of project items. False negatives come about when they analyst confuses people who were actually

	Minor Kitchen Remodel	Sub-Class	Type
1	Kitchen sink	ACRYLIC SINKS	RF
2	Drain	DRAIN PIPE FITTINGS	IND
3	Undercabinent Lighting	UNDERCABINET LIGHTING	IND
4	Cabinet hardware	S/O CABINET/HARDWARE	IND
5	Faucet	S/O FAUCETS	IND
6	Laminate, Marble countertop	F&D LAMINATE COUNTERTOPS	RF
7	Kitchen lighting	S/O INTERIOR LIGHTING	IND
8		S/O RECESSED LIGHTING	IND
9	Thinset tile mortar	MORTAR/MASONRY/STUCCO	IND
10	Backsplash (mosaic tile)	MOSAIC	IND
11	Tile Grout	GROUTS	IND
12	Caulk	CAULK GUNS	IND
13	Notched trowel	TILE SURFACE PREPARATION	IND
14	Wet saw	TILE CUTTERS & WET SAWS	IND
15	Grout sealer	CLEANERS AND SEALERS	IND

(a)

	Wall and Floor Tile	Sub-Class	Type
1	Tile cutters/Wet saw	TILE CUTTERS & WET SAWS	RF
2	Tape measure	TAPE MEASURES	IND
3	Level	LEVELS	IND
4	Notched trowel	TILE SURFACE PREPARATION	IND
5	Foam brush	ECONOMY/UTILITY BRUSHES	IND
6	Caulk gun	CAULK GUNS	IND
7	Thinset tile mortar	MORTAR/MASONRY/STUCCO	IND
8	Ceramic/Stone wall tiles	12" TO 13" CERAMIC	RF
9		S/O CERAMIC FLOOR TILES	RF
10		LESS THAN 12" TILE/STONE	RF
11		S/O NATURAL STONES/SLATE	RF
12		GREATER THAN 12" STONE	RF
13	Trim tiles	DECO AND TRIM TILES	IND
14		MOSAIC	IND
15	Tile grout	GROUTS	IND
16	Grout sealer	CLEANERS AND SEALERS	IND

(b)

	Attic Insulation	Sub-Class	Type
1	staple gun	STAPLING	IND
2	Protective/safety glasses	PERSONAL SAFETY	IND
3	Gloves	WORK GLOVES	IND
4	Coveralls	PAINT APPAREL	IND
5	loose-fill insulation	INSULATION	RF
6		LOOSE FILL INSULATION	RF
7	pipe insulation	PIPE INSULATION	IND
8	Insulation covers	INSULATION ACCESSORIES	IND

(c)

FIGURE 6.1. The project item lists for three projects (a) minor kitchen remodel, (b) wall and floor tiling, and (c) attic insulation.

undertaking the project, but we set our detection threshold too low to find them (to avoid for instance all the others who are false positives).

Next, we describe the process of determining the two other parameters λ and λ_{misc} for the exponential distribution for the random variable $G_{i,k,a.}$ based on empirical data. Utilizing the data on-hand, we determined the percent of 12,082 customers who purchased one or more of the items in the 21-item tiling project query from Fig. 5.21b. For a given query, Fig. 6.2 shows many customers bought a few items, no one purchased all the items, and only a few purchased as many as 11 items. An exponential function could reasonably approximate this drop-off in item subclasses purchased given a query, but resulting percentages would need to be normalized into a PMF over the number of query items. Given that Q_{total} is the total number of item sub-classes for a project query, $q \in \{1, 2, ..., Q_{\text{total}}\}$ (the number of item sub-classes purchased for a project query), and $S = \sum_{i=1}^{Q_{\text{total}}} e^{-\lambda q}$ (the normalization constant of the sum of the exponential functional values over discrete domain up to the total number of query item sub-classes), Eq. 14 is the function utilized to determine the exponential PMF for project type $a.$.

$$\text{Percent of Customers} = \frac{1}{S} \cdot e^{-\lambda q} \tag{14}$$

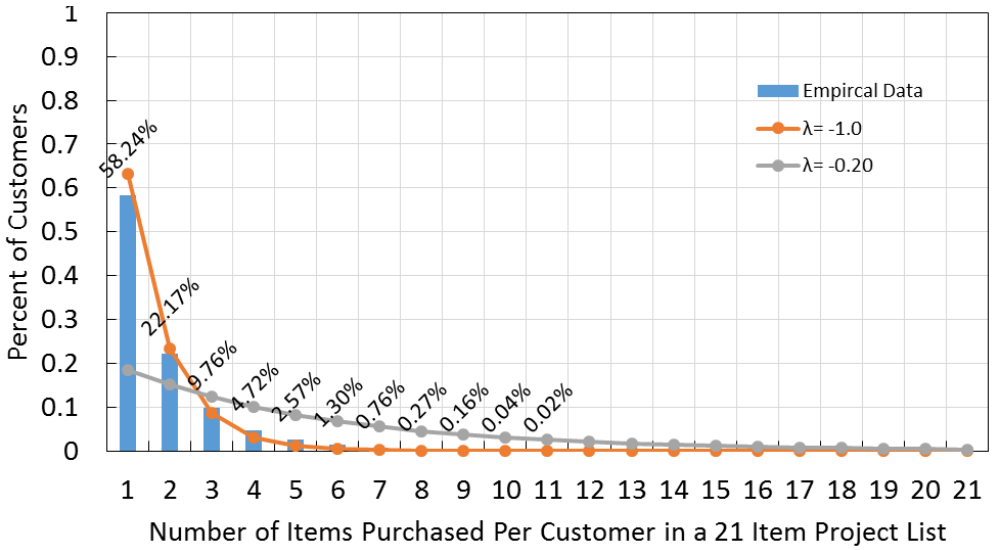


FIGURE 6.2. Percent of customers by the number of unique item sub-classes purchased in a 21-item project materials and tools list. The orange and grey curves depicts a normalized exponential function with $\lambda = -1.0$ and $\lambda = -0.20$, respectively.

As Fig. 6.2 shows, a $\lambda = -1.0$ would approximate fairly well the number of item sub-classes purchased, but in simulation we desired the opportunity to examine the utility of finding customers who purchased more item sub-classes in the query. Therefore, we chose a notional $\lambda = -0.20$ for Eq. 14 that achieves this effect of a heavier tail and utilized it with the specific known project queries of sizes $Q_{\text{total}} = \{15, 16, 8\}$ according to Fig. 6.1. This process generated an exponential PMF for the random variable $G_{i,k,a.}$ for each project of type $a.$.

To select the parameter λ_{misc}, we again drew on the data and determined the empirical distribution of unique sub-classes purchases by each of the customers (irrespective of the project queries). The result is shown in Fig. 6.3. We fitted the exponential function $\text{Frequency} = 1457.1e^{-0.034 \cdot \text{Number of Subclasses}}$ with a $R^2 = 0.9729$ to the data. This function

was then scaled to 1, and utilized as a continuous PDF to generate the number of unique "noisy" item sub-class purchases for customers assigned the unknown Project 4.

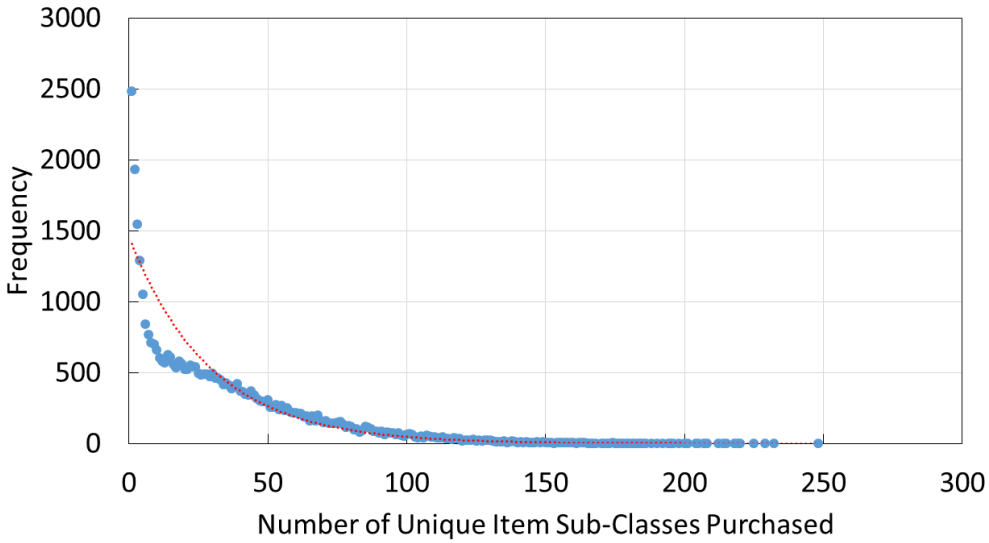


FIGURE 6.3. Histogram of the total number of unique item sub-classes purchases by customers in the real dataset. The fitted exponential curve is shown in dotted red.

Table 6.2 provides the additional details for this particular run of the synthetic data generator. The simulation was implemented in MATLAB using a single random number generator seed `123456`. Additionally, the 2 year period (730 days) was designated to start on January 1, 2013. The portion of the typical output as a simulated purchase table with project ground truth intention is shown in Fig. 6.4.

6.6. Modeling with Ground Truth

Utilizing the parameterization provided for the sample run of the data generator, we obtain a set of simulated customers and their associated purchases over the course of 2 years (730 days). Most importantly, this dataset contained the ground truth of the a priori intentions of each customer as they made their purchases. Table 6.3 provides the statistical summary. The simulated results were validated at face-value. For example, given the PMF of

TABLE 6.2. Sets and parametrization for sample run of the synthetic generator

Notation	Description/Meaning
$A = \{\text{Kitchen, Tiling, Attic, Unknown}\}$	Set of $n_p = 4$ project types
Ω	Set of all items related to any project type $\in A$
$B_{a.} \subseteq \Omega$	Set of items for each project type $a. \in A$
$C = \{1.00, 0.80, 0.60\}$	Set of $n_r = 3$ proportions of planned purchases actually made by customers
$n_c = 1000$	Simulate the purchase activities of 1000 customers
$t_{\max} = 730$	Time horizon of 730 days
$s = 4$	Customers may undertake (possibly simultaneously) up to 4 projects within the 730 day period
$f_{X_i}(x_i) = \begin{cases} 0.67 & \text{if } x_i = 1 \\ 0.20 & \text{if } x_i = 2 \\ 0.10 & \text{if } x_i = 3 \\ 0.03 & \text{if } x_i = 4 \end{cases}$	Probabilities for the number of projects (up to $s = 4$) a customer may undertake in a period of $t_{\max} = 730$ days
$f_{Y_{i,k}}(y_{i,k}) = \begin{cases} 0.12 & \text{if } y_{i,k} = \text{Kitchen} \\ 0.08 & \text{if } y_{i,k} = \text{Tiling} \\ 0.10 & \text{if } y_{i,k} = \text{Attic} \\ 0.70 & \text{if } y_{i,k} = \text{Unknown} \end{cases}$	Probabilities for each of the n_p projects in the PMF for $Y_{i,k}$ (applicable $\forall\ i$)
$\lambda = -0.20$ for $f(G_{i,k}; \lambda)$	Parameter $\lambda = -0.20$ (decay) for the exponential PDF of $G_{i,k}$ (applicable $\forall\ i$)
$f_{H_{i,k}}(h_{i,k}) \begin{cases} 0.75 & \text{if } h_{i,k} = 1.00 \\ 0.15 & \text{if } h_{i,k} = 0.80 \\ 0.10 & \text{if } h_{i,k} = 0.60 \end{cases}$	Probabilities for each of the n_r proportions in the PMF for $H_{i,k}$ (applicable $\forall\ i$)
$\alpha = 0.95, \beta = 30$ for $f(J; \alpha, \beta)$	Parameters $\alpha = 0.95$ (shape) and $\beta = 30$ (scale) for the Weibull distribution PDF of J
$a = .80$ for $f(L_i; a, 1.00)$	Uniform distribution PDF of L_i between $[0.80, 1.00]$
$t_{\text{noise}} = 45$	Time horizon of 45 days as the latest date for the start of purchases for 'Unknown' projects

X_i (number of projects each customer undertakes), the expected value of the number of total number of projects given 1,000 customers is $1,000(0.67 \cdot 1 + 0.20 \cdot 2 + 0.10 \cdot 3 + 0.03 \cdot 4) = 1490$, which is close to the total simulated of $1,488$.

	1 purchase_id	2 customer_id	3 item_code	4 item_name	5 purchase_timestep	6 actual_purchase_date	7 groundtruth_project
1	91	1	945	'PLASTIC'	18	'19-Jan-2013'	4
2	25	1	113	'ADHESIVES'	21	'22-Jan-2013'	4
3	124	1	926	'PERENNIALS'	25	'26-Jan-2013'	4
4	159	1	940	'PIPE CHEMIC...	37	'07-Feb-2013'	4
5	86	1	1583	'VANITY COM...	40	'10-Feb-2013'	4
6	88	1	722	'JIG SAWS'	47	'17-Feb-2013'	4
7	21	1	1493	'TEMPORARY ...	53	'23-Feb-2013'	4
8	54	1	1011	'PT DIMENSIO...	54	'24-Feb-2013'	4
9	3	1	1025	'PVC PIPE'	59	'01-Mar-2013'	4
10	38	1	1011	'PT DIMENSIO...	62	'04-Mar-2013'	4
11	59	1	367	'CUTTING'	66	'08-Mar-2013'	4
12	61	1	818	'MINI ROLLER...	73	'15-Mar-2013'	4
13	126	1	542	'FRAMES / GRI...	76	'18-Mar-2013'	4
14	106	1	759	'LAUNDRY TU...	77	'19-Mar-2013'	4
15	50	1	750	'LANDSCAPE ...	79	'21-Mar-2013'	4
16	69	1	12	'"NEW,NO-LO...	80	'22-Mar-2013'	4
17	135	1	1024	'PVC FITTINGS'	82	'24-Mar-2013'	4
18	52	1	1683	'WORK GLOVES'	84	'26-Mar-2013'	4
19	95	1	1353	'SHOWERHEA...	85	'27-Mar-2013'	4
20	65	1	1389	'SOCKETS/WR...	89	'31-Mar-2013'	4
21	12	1	594	'GENERAL PU...	90	'01-Apr-2013'	4
22	79	1	31	'011-018-003'	92	'03-Apr-2013'	4
23	98	1	471	'F CONNECTO...	92	'03-Apr-2013'	4
24	100	1	194	'BEHR ULTRA'	93	'04-Apr-2013'	4

FIGURE 6.4. Screenshot of the MATLAB output from the synthetic data generator.

TABLE 6.3. Statistics of synthetically generated customers and purchases

Category	**Count**
Number of customers	1,000
Number of purchases	25,444
Number of projects	1,488
-Number of customers with Kitchen (Project 1)	263
-Number of customers with Tiling (Project 2)	223
-Number of customers with Attic (Project 3)	258
-Number of customers with Unknown (Project 4)	744

Table 6.4 shows the number of purchases which contain items from a specific project sub-class list broken down by intended project (ground truth). For example, there were 1,750 purchases that contained items in the Kitchen project sub-class list. However, since there was a query sub-class overlap with the Tiling project, as well as the possibility of individuals undertaking the Unknown project who purchase an item of sub-classes that happen to match the Kitchen project sub-class list, these 1,750 purchases are distributed according to ground truth as shown. There were 1,115 purchases that were the result of an intended Kitchen

project, 406 purchases that were actually for a Tiling project, and 229 purchases that were purchased 'by chance' that were part of the Unknown project. This table reflects the reality that purchases have the potential to be attributed to the incorrect project without additional corroborating purchases.

TABLE 6.4. Table of project purchases by ground truth

		Ground Truth			
Project	**Total purchases that match a sub-class in project query**	**Kitchen (Proj. 1)**	**Tiling (Proj. 2)**	**Attic (Proj. 3)**	**Unknown (Proj. 4)**
Kitchen (Proj. 1)	1750	1115 (63.7%)	406 (23.2%)	0 (0%)	229 (13.1%)
Tiling (Proj. 2)	1762	505 (28.7%)	953 (54.1%)	0 (0%)	304 (17.2%)
Attic (Proj. 3)	1009	0 (0%)	0 (0%)	785 (77.8%)	224 (22.2%)

6.7. Analysis of Results

Like the previous applications, the initialization of INSiGHT required the development of a total of three query patterns (one for each of the three projects). Each of the query patterns followed the same structure as the one in Fig. 5.21b, except each leaf node was limited to the sub-classes for each particular project. The data graph schema was identical to the one in Fig. 5.21a, as well as the parameterizations shown in Table 5.13.

Fig. 6.5 shows the distribution of the final similarity scores as an output of INSiGHT run for each of the project query patterns. In blue are the histogram bars for the number of customers undertaking the respective project by purchasing sub-classes of items in that project query. Within each histogram are light red bars for the number of customers who also purchased items in that project query but were actually undertaking another project (including the Unknown project). These results are expected given the discussion of Table

6.4. By varying the final similarity score thresholds for each of the projects, one can construct the receiver operating characteristic (ROC) curves as shown in Fig. 6.6. The area under the curve (AUC) for Kitchen, Tiling, and Attic are 0.9352, 0.9367, and 0.9642, respectively. We observe that the AUC for the Attic project is higher given the lack of overlap in sub-class items with the other two projects. In fact, our intuition is that a given project's AUC is likely a function of at a minimum 1) the overlap of sub-class items with the other project patterns, 2) the prevalence of the sub-class items among the universal set Ω from which "noisy" purchases are down from, 3) the frequency for each project based upon X_i and $G_{i,k,a.}$, 4) the percentage of planned items actually purchased $H_{i,k}$, and the customer loyalty percentage L_i. In future work, we intend to make this theoretical function more specific.

It is important to note that while these AUC values are reasonably high, one must recall that these performance curves are based strictly on the final similarity scores for each customer and for this particular application. In order to achieve a 100% True Positive Rate (TPR) based on final similarity scores alone, one would need to accept a 32%, 38%, and 18% False Positive Rate (FPR) for the Kitchen, Tiling, and Attic projects, respectively. In the given commercial application, but an analogous number in a radicalization application might be considered insufficiently discerning for law enforcement and lead to an exorbitant number of false leads. In fact, a stepwise logistic regression model can be readily built using 11 product-level features (rather than aggregate matches to a query) to achieve an AUC for the Tiling project at 0.9616. This means that even simplistic machine learning models can outperform the simplistic similarity scoring method.

Another key point of INSiGHT is to perform periodic screening and analysis over time to enable early watching of likely customers for a particular project.

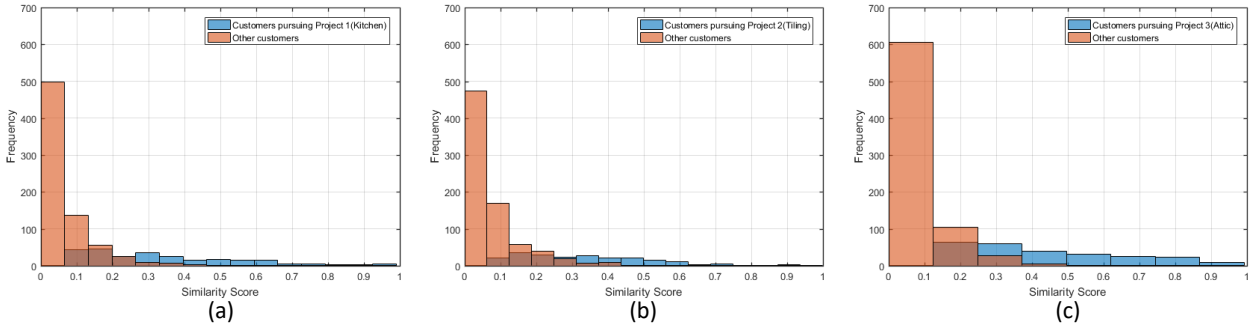


FIGURE 6.5. The final similarity scores for simulated purchase data for those with indicators from (a) Kitchen project, (b) Tiling project, and (c) Attic project. Blue bars are the customers who undertook those projects, while the red bars are the customers who made those purchases for other projects.

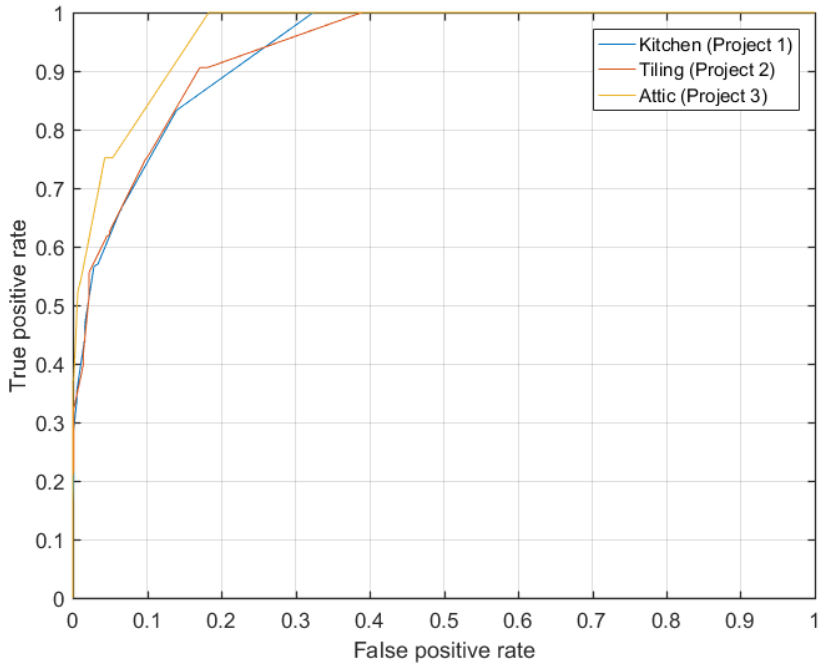


FIGURE 6.6. ROC curve for the classification of three projects based upon the final similarity score. The AUC for Kitchen, Tiling, and Attic are 0.9352, 0.9367, and 0.9642, respectively.

6.8. Combining Investigative Graph Search and Machine Learning

A main focus of the future research is the integration of investigative graph search with machine learning models to make predictions on data stored in dynamic heterogeneous graphs. Graph databases are good for dynamically storing and querying large amounts of interrelated pieces of information. But in order to utilize machine learning classification

or prediction models on such data, it is important first to find the features and associate them with each record (even over multiple hops). The investigative graph search component finds the connected indicators that match a hypothesized pattern of latent behaviors for one or more entities (query focus nodes). Then, once the pattern is detected in whole or in part, the indicators act as features in machine learning models to perform the classification or prediction.

This extension is trivial for 1-hop graph query patterns because presently analysts already keep track of which features are associated with which individuals and the models can be run continuously or periodically for near real-time classification or prediction. However, when latent behaviors may be indicative through more intricate patterns with 2-3 hop connections, the investigative graph search function becomes ever more important.

We propose the following steps to this direction or research and then provide a proof of concept of its potential utility.

(1) Build a machine learning model that can classify latent behaviors based on the presence of indicators (binary for now). Ideally, this model provides a list of features that are statistically significant.

(2) Build graph query with these features as indicators.

(3) Use INSiGHT to find the presence of indicators associated with query focus nodes in heterogeneous graph databases over time.

(4) INSiGHT calculates a) the similarity score to the pattern at each timestep (tells how many indicator matches there are with the query), and b) the classification score at each timestep to provide the classification or prediction.

In order to build the right machine learning model with the interest of providing an early warning to the vendor, we segmented the data into activity counts. For example, Activity 1 refers to using all purchase activities that occurred in the *first* week of a customer's activity, while Activity 2 refers to using all purchase activities that occurred in the *second* week of a customer's activity. We envision that a vendor would desire models which need only a partial set of all purchases in order to make a prediction about whether the customer is pursuing a particular project. We iteratively built 9 logistic regression models through a forward step-wise process, each using a specific activity set (Activity 1, Activity 2, ... Activity 9). We then evaluated each of the 9 models on each of the activity sets and determined the AUC. This result is shown in Fig. 6.7. Models built on more data (i.e., more activities) generally had good performance for more data. On the other hand, decreasing AUC curve for 'Act 1 Items' model shows that models built using very few activities may only be appropriate for classification using a few activities (Activity 1 and 2), but performance will diminish as more activities are considered. The notable exception is 'Act2 Items' model, which has a relatively high AUC at Activity 1 and Activity 2 (0.916 and 0.969, respectively), and remains competitive for all other Activity Counts. Also the dotted blue plot in Fig. 6.7 labeled 'Act Scores' is the resulting AUCs of the best performing a forward step-wise logistic regression model utilizing the just similarity scores and gradients. Notice that these aggregated features have poorer performance than nearly all models which consider specific item granularity.

We, therefore, chose the 'Act 2 Items' model as the machine learning model which would be evaluated using the whole or partial query matches from INSiGHT to dynamically predict whether a customer is undertaking the Kitchen Project. The terms and coefficients and their significance are shown in Table 6.5.

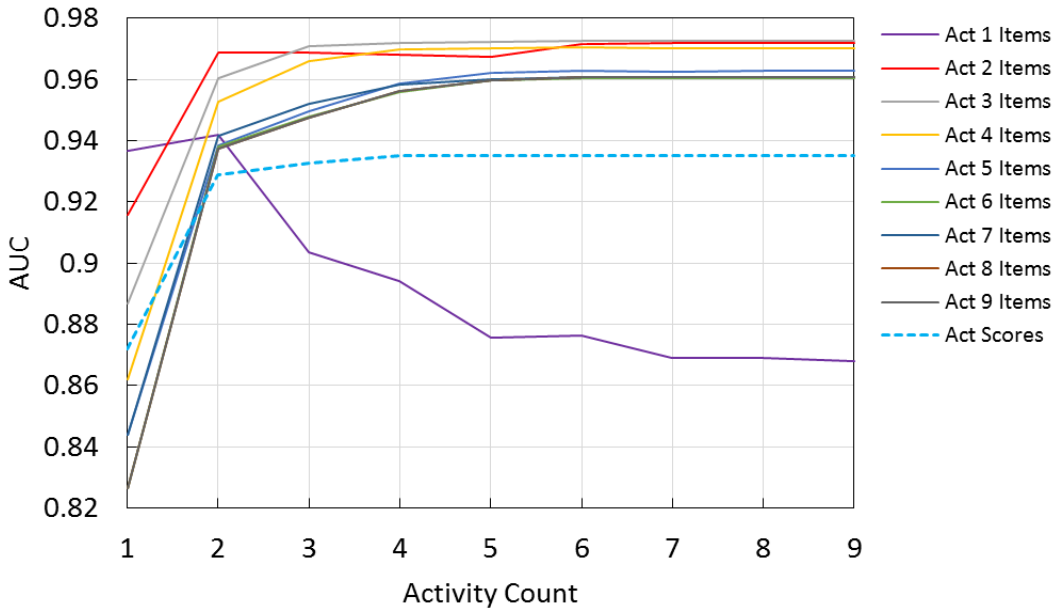


FIGURE 6.7. AUC performance of stepwise logistic regression models on activity sets of various sizes.

TABLE 6.5. Table of terms and coefficients for the 'Act 2 Items' Logistic Regression Model for the Kitchen Project (Project 1)

Term	**Estimate**	**SE**	**tStat**	**p-value**
(Intercept)	-4.108	0.281	-14.633	1.731E-48
SINK	90.424	1.097E+07	8.243E-06	1.000E+00
PIPE	2.806	0.386	7.263	3.795E-13
UCABINET_LIGHTING	4.173	0.615	6.783	1.179E-11
CABINET_HARDWARE	90.707	1.138E+07	7.972E-06	1.000E+00
FAUCET	86.057	1.058E+07	8.133E-06	1.000E+00
COUNTERTOP	5.359	1.115	4.804	1.555E-06
LIGHTING	4.776	0.847	5.639	1.712E-08
RECESSED_LIGHTING	5.419	1.095	4.947	7.550E-07
MORTAR	1.620	0.341	4.755	1.984E-06
GROUTS	1.455	0.332	4.388	1.146E-05
CAULK	1.964	0.415	4.732	2.223E-06
TILE_PREP	2.323	0.370	6.276	3.482E-10
TILE_CUTTER	1.838	0.415	4.432	9.332E-06
CLEANER_SEALER	2.276	0.359	6.337	2.344E-10
CAULK:TILE_PREP	-5.881	1.791	-3.283	1.027E-03
CAULK:TILE_CUTTER	-4.968	3.449	-1.440	1.500E-01

As a proof of concept, the 'Act 2 Items' model was inserted into the INSiGHT implementation, and the classification score was returned for each customer at each timestep. Figs.

6.8a and 6.8b show the plot of classification scores for all 1000 simulated customers and only Kitchen Project customers, respectively.

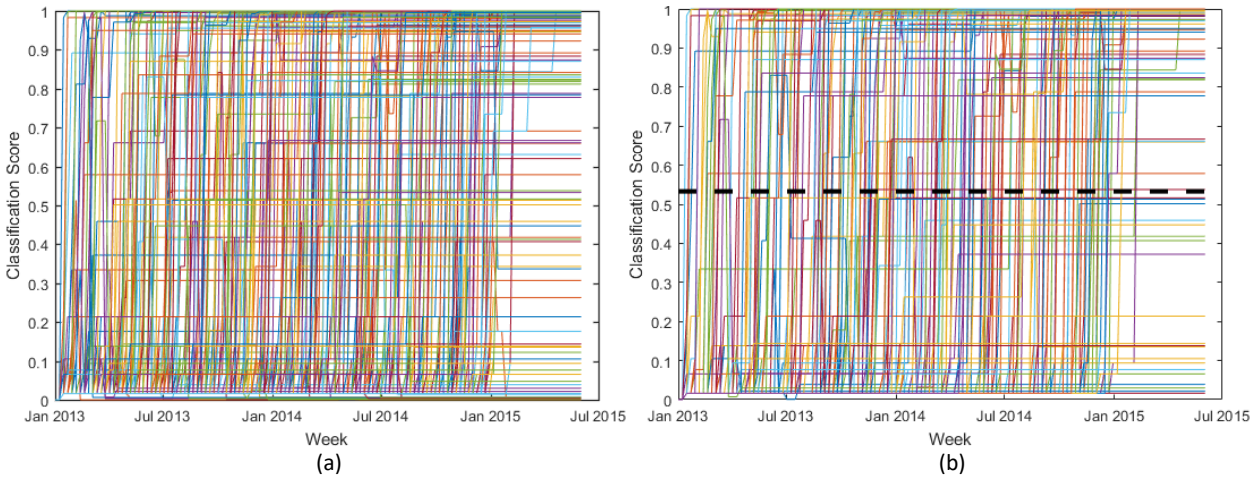


FIGURE 6.8. Classification scores for the Kitchen Project (Project 1) over time for (a) all 1000 simulated customers and (b) only the 263 simulated customers who were actually undertaking the project. The dashed black line in (b) is the threshold classification score of 0.51639.

Recall that logistic regression models the probability of success as a logit function. That is, the dependent variable in a logistic regression model (as well as a general linear model with binomial distribution) is the probability of success $p(\mathbf{x}) = \frac{e^{\beta_0 + \mathbf{x}^T \boldsymbol{\beta}}}{1 + e^{\beta_0 + \mathbf{x}^T \boldsymbol{\beta}}}$, where $\mathbf{x}$ is the binary vector of active terms, β_0 is the intercept, and $\boldsymbol{\beta}$ is the vector of term coefficients. The model implies the log odds of a customer being a positive outcome (undertaking Kitchen Project) increases with every purchase whose coefficient $\beta > 0$. Therefore, with an integration of both a logistic regression classification model with an investigative search over time, increases in the classification time series plot implies a greater log-odds of the entity being 'positive.' Fig. 6.9a shows the dramatic accentuation of the separation between Kitchen Project customers and others, especially when compared to Fig. 6.5a (which was based on a simplistic similarity score). However, we note that an increasing classification score does not necessarily imply the entity is always going to be predicted as a positive case. This is clearly seen in Fig. 6.9b.

We note that two interaction terms CAULK:TILE_PREP and CAULK:TILE_CUTTER have negative coefficients. Thus if either combination was purchased, the log-odds of the project designation decreases, as visible in Fig. 6.8. This is because the exponential function is strictly increasing.

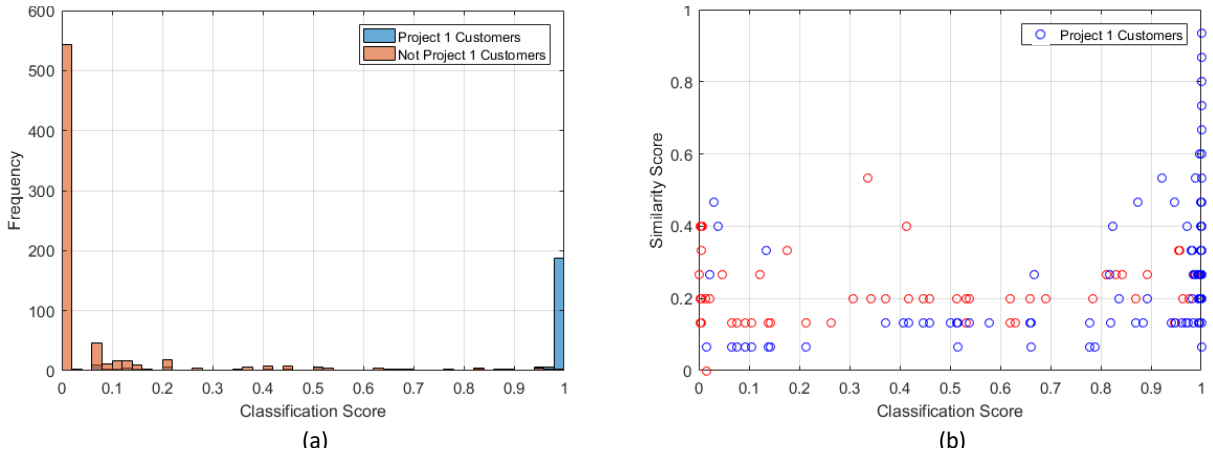


FIGURE 6.9. (a) Distribution of classification scores for all 1000 simulated customers undertaking the Kitchen Project (in blue), and those who were not (in light red).(b) The scatter plot of the final similarity scores the corresponding classification scores for all 1000 simulated customers. Dots in blue are those customers undertaking the Kitchen Project.

Selecting the 'Act 2 Items' model and determining which features for detection by INSiGHT is only one step. For various threshold levels, the logistic regression model will have different specificity and selectivity rates. By analyzing the AUC curves of the 'Act 2 Items' model on various Activity sets (Activity 1 and Activity 9), we determined that a threshold value of 0.51639 achieved suitable selectivity and specificity rates as shown in Fig. 6.10, with a deference towards selectivity especially because the negative cases are nearly three times the number of positives. This threshold became the value that an analyst would utilize to predict those customers who were likely pursuing Kitchen Projects. Fig. 6.8b shows the placement of this threshold on the classification score time series plot.

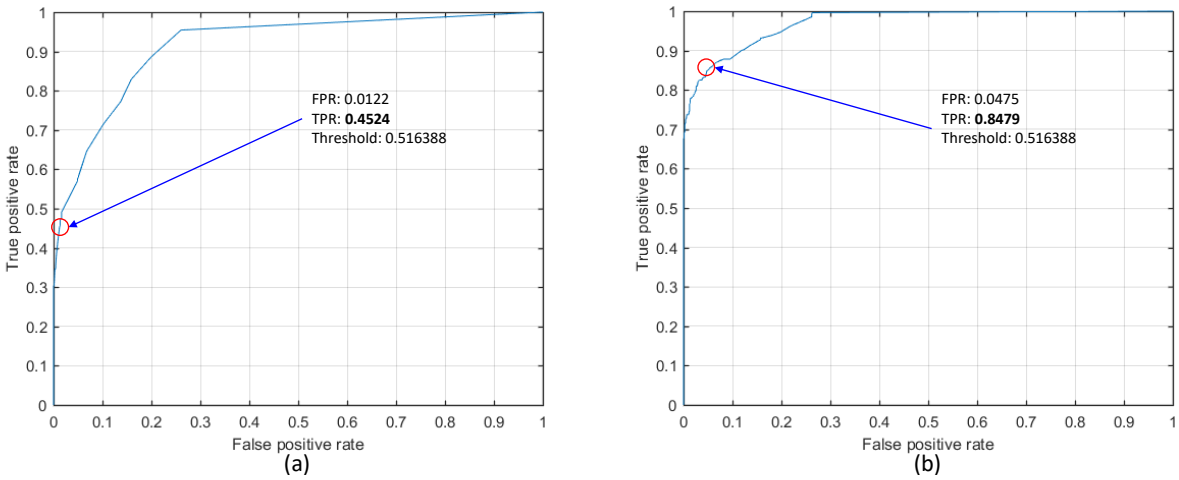


FIGURE 6.10. ROC curves for the 'Act 2 Items' logistic regression model evaluated on (a) Activity 1 data, and (b) Activity 9 data. A common threshold of 0.516388 in both curves achieved suitable selectivity and specificity.

Lastly, we proceeded to evaluate the dynamic performance of this technique by determining the timestep (if any) in which each customer's classification score exceeded the threshold. All those whose score exceeded the threshold (at the time it occurred) was labeled a 'Yes' for Kitchen Project. The confusion matrix against the ground truth is shown in Table 6.6. This dynamic classification achieved a TPR of 0.8555, FNR of 0.1445, and an FPR of 0.0570. Overall the accuracy was 92.0%. This performance exceeded the classification using the final similarity score, which at best had a TPR of 0.8441, FNR of 0.1559, an FPR of 0.1479 and an overall the accuracy was 85.0%.

TABLE 6.6. Classification confusion table upon trigger of exceeding threshold

Actual/Predict	No	Yes
No	695	43
Yes	38	225

6.9. CONCLUSION

Recognizing from the outset that latent behavior data are often difficult to acquire for research, in this section we developed a tunable and empirically grounded synthetic data

generator for customer purchases. Its greatest value is the a priori knowledge of ground truth for the type of project(s) that a simulated customer is pursuing. The synthetic dataset also allowed us to more rigorously validate INSiGHT and examine the final similarity scores with an understanding of ground truth. We determined that AUC is dependent upon several characteristics of the query and underlying data, namely any overlap in the query items and the overall prevalence of query items in the universal set of items.

Significantly, we utilized synthetic data to explore the pairing of investigative graph search and machine learning. The former finds the connected indicators that match a hypothesized pattern of a latent behavior, and the latter performs the classification or prediction (rather than just screening). We demonstrated the utility of this approach with a proof of concept and achieved 92.0% classification accuracy dynamically as the consumers were making purchases.

In the future, we also intend to utilize this synthetic generation framework in order to create large-population, hybrid (synthetic and anonymized) datasets for researchers to test their algorithms related to the detection of radicalization and violent extremism. This contribution has the broader impact of facilitating future development of evidence-based risk assessment technologies.

A proposal for future work is a collaborative effort with the Western Jihadism Project at Brandeis University for the development of a textual database of indicators of a specific radicalization indicator. During the course of data collection, the Brandeis research team will read a text corpus, classify it as indicative of a particular radicalization indicator, and capture the specific text (sentence or paragraph) that led them to make that classification. With enough samples of coded training text and subject-matter expert feature selection,

data scientists will employ natural language processing algorithms to tokenize different texts and utilize machine learning to automatically identify the indicators present for an individual. This achievement will be tremendously helpful for any future efforts to automate the classification of law enforcement data points as specific radicalization indicators

CHAPTER 7

INSiGHT with Neighbor Matching

7.1. Introduction

As discussed previously, recent cases of violent extremism have demonstrated that perpetrators can operate in a conspiracy to commit terrorist acts. While individual actions may not rise up to some threshold of suspicion, the collection of individuals supported by close ties may be able to reveal more obvious threats. In our graph pattern matching approach, this means that conspiratorial graph matches, which we define as match complementarity over more than one query focus node, may further assist law enforcement and intelligence analysts. To our knowledge, such matches are not addressed by current matching notions. Here we formulate a definition for this neighborhood matching technique, describe its implementation in INSiGHT, and demonstrate its functionality on a small example dataset.

7.2. Technique

Definition 9. (l, k) neighbor matching of query Q to data graph G. The graph matching process to find the augmenting sets of distinct conforming subgraphs of G from a binary match relation $S \subseteq V_Q \times V_G$ such that:

- for each of as many nodes $u \in V_Q$ as possible (but at least one), there exists a node $v \in V_G$ such that $(u, v) \in S$, and
- for each pair $(u, v) \in S$, $u \sim v$, and
- for each of as many edges $(u, u') \in E_Q$ as possible (but at least one) there exists at most l edges $\in E_G$ which form a directed path $(v_1, ..., v'_{l+1})$ such that both $(u, v_1) \in S$

and $(u', v'_{l+1}) \in S$ and the directed paths $(v_1, ..., v'_{l+1})$ spans at most k distinct QF-based conforming subgraphs.

Note that this definition allows for each edge in E_G to possibly involve up to $k-1$ different QF nodes other than the ego QF node of the subgraph. Additionally, we observe that $(1, 1)$ neighbor matching reduces to the Definition 4 for Inexact Graph Pattern Matching, where any edge $(u, u') \in E_Q$ must be matched by at most 1 edge $(v, v') \in E_G$ and must occur within the entity's own subgraph. When $k = 2$ we are limiting the matches that result from dyadic relationships (i.e., an edge in the query graph can only be fulfilled by connections involving at most 2 QF nodes). When $k = 3$, we are limiting the matches that result from triadic relationships (i.e., an edge in the query graph can only be fulfilled by connections involving at most 3 QF nodes).

The algorithm for (l, k) Neighbor Matching is shown in **Algorithm 6**, which is executed for each timestep in the window of analysis (line 1). We first construct filtered adjacency matrices $\mathbf{W}'_1(t)_G$ by removing the QF-QF edges which only connect nodes of type QF (line 2). We also construct $\mathbf{W}''_1(t)_G$ by removing the edges $\mathsf{QF}^{(2)}$-$\mathsf{QF}^{(2)}$ (which connect QF forum nodes with each other) as well as edges $\mathsf{QF}^{(2)}$-QF (which connect QF forum nodes back to the QF node originator, as a representation for the receipt of influence for others' messages/content which is communicated online) (line 3).

We then construct modified h-hop adjacency matrices $\tilde{\mathbf{U}}_h(t)^{(k)}_G$ for the specified parameter k, the maximum number of QF nodes that can be involved in the fulfillment of an edge in E_Q. For each hop h up to $l \cdot h_{max}$, each modified h-hop adjacency matrix is a result of the product of the original 1-hop adjacency matrix and/or one or more of the filtered adjacency matrices, depending upon the k (line 4-5). We allow up $l \cdot h_{max}$ hops because each edge in E_Q

Algorithm 6: (l, k) Neighbor Matching algorithm

Input: $\mathbf{W}_1(t)_G$ (1-hop Adjacency Matrices of a graph for a given t), $h_{\max}$ (the desired number of hops), l (maximum path length allowed in G_S to fulfill an edge in E_Q), k (maximum number of QF nodes that can be involved in a fulfilment of an edge E_Q), $\mathbf{A}_G$ (class membership matrix of G), $\mathbf{\Phi}_h(t)$ (binary matrix of ϕ indicator variables, $\mathbf{M}_{G,Q}$ (sparse node class match matrix between query graph Q and data graph G), $\mathbf{C}_h(t)_Q$ (child h-hop class adjacency matrices for all time $t_{\text{start}} \leq t \leq t_{\text{end}}$ for query graph Q)

Output: $\hat{\mathbf{C}}_h(t)_G^{(l,k)}$ (Modified class adjacency matrices of G for (l, k) Neighbor Matching), and $\hat{\mathbf{S}}_h^{(1)}(t)^{(l,k)}$ (parent-to-child class similarity score matrices between G and Q for (l, k) Neighbor Matching), where $t : t_{\text{start}} \leq t \leq t_{\text{end}}$ and $h : 1 \leq h \leq h_{\max}$

1 **foreach** $t = t_{\text{start}}$ to t_{end} **do**
2 Construct $\mathbf{W}'_1(t)_G$ by removing all QF-QF edges in $\mathbf{W}_1(t)_G$
3 Construct $\mathbf{W}''_1(t)_G$ by removing all QF$^{(2)}$-QF$^{(2)}$ and QF$^{(2)}$-QF edges in $\mathbf{W}_1(t)_G$
4 **foreach** $h = 1$ to $(l \cdot h_{\max})$ **do**
5 $\mathbf{U}_h(t)_G^{(k)} = \left(\prod_{i=1}^{k-1} \mathbf{W}_1(t)_G\right) (\mathbf{W}'_1(t)_G) \left(\prod_{i=k+1}^{h} \mathbf{W}''_1(t)_G\right)$
6 $\tilde{\mathbf{U}}_h(t)_G^{(k)}$ =binary matrix of $\mathbf{U}_h(t)_G^{(k)}$, where each entry $\tilde{u}_{h,(i,j)} = 1$ when $u_{h,(i,j)} = 1$, $i \neq j$, and $u_{h\text{-}prev,(i,j)} \neq 1$ for all $h\text{-}prev < h$; and 0 otherwise.
7 $\mathbf{C}_h(t)_G^{(k)} = \tilde{\mathbf{U}}_h(t)_G^{(k)} \cdot \mathbf{A}_G$
8 $\hat{\mathbf{C}}_h(t)_G^{(k)} = \frac{1}{2}\left(\mathbf{1} - e^{-\mathbf{\Phi}_h(t)\mathbf{C}_h(t)_G^{(k)}\mathbf{\Lambda}}\right)\left(\mathbf{1} - \tanh\left(\frac{\boldsymbol{\chi}_h(t) - \boldsymbol{\beta}}{\boldsymbol{\xi}}\right)\right)$
9 **foreach** $h = 1$ to $h_{\max}$ **do**
10 $\hat{\mathbf{C}}_h(t)_G^{(l,k)} = \text{entry-wise } \min\left\{\sum_{i=h}^{l \cdot h} \hat{\mathbf{C}}_i(t)_G^{(k)}, 1\right\}$
11 **foreach** index pair $(i, j) : m_{ij} = 1$ in $\mathbf{M}_{G,Q}$ **do**
12 $\hat{s}_h^{(1)}(t)_{i\cdot}^{(l,k)} = \textbf{similarity}\left(\hat{c}_h(t)_{G,i\cdot}^{(l,k)} \text{ and } c_h(t)_{Q,j\cdot}\right)$
13 **return** $\hat{\mathbf{C}}_h(t)_G^{(l,k)}, \hat{\mathbf{S}}_h^{(1)}(t)^{(l,k)}$

can be fulfilled by paths of up to length l (line 4). For instance, if $k = 1$, then only the filtered adjacency matrix $\mathbf{W}'_1(t)_G$ is used to obtain the h-hop adjacency matrices. However, if $k = 2$ and $h = 3$, the algorithm calls for multiplying the original 1-hop adjacency matrix $\mathbf{W}_1(t)_G$ first with the filtered adjacency matrix $\mathbf{W}'_1(t)_G$ and then second with $\mathbf{W}''_1(t)_G$. This allows the QF-QF edges and (weighted) neighboring attributions due to QF$^{(2)}$-QF$^{(2)}$ and QF$^{(2)}$-QF to be used once in determining the reachability from each node. We then convert these h-hop adjacency matrices into binary matrices, remove any possibility of self-loops, and make each

successive h-hop matrix reflects only new connections (and not previous ones) (line 6). Just as in **Algorithm 2** in [144], we can obtain the class adjacency matrices from the product of the h-hop adjacency matrices with the class membership matrix (line 7). As in 9, we also account for stand-alone 'IIRA' indicators, reoccurring indicators, and indicator recency with the transformation in line 8.

Next for each hop h up to h_{max}, we construct the modified class adjacency matrices $\hat{\mathbf{C}}_h(t)_G^{(l,k)}$ for (l,k) neighbor matching by combining class adjacency matrices for specific hops (lines 9-10). For example, for (2-2)-neighbor matching (where $l = k = 2$), line 10 first sums together the class adjacency matrices $\hat{\mathbf{C}}_1(t)_G^{(k)}$ and $\hat{\mathbf{C}}_2(t)_G^{(k)}$ to account for any matches of an edge in E_Q that is up to $h = 1$ hop away (but can be fulfilled by up to 2 hop lengths in G). Then for $h = 2$ hop matches, line 10 sums together the second, third, and fourth hops of the class adjacency matrices. This is because we are matching for indicators which occurred over two edges in E_Q and thus is allowed to be fulfilled by up to 4 hop lengths in G. Within each h aggregation and for each summed entry $\hat{\mathbf{C}}_i(t)_G^{(k)}$ we utilize the $\min\{\cdot, 1\}$ function to ensure that only new indicators are counted.

Lastly, to calculate the parent-to-child class similarity score matrices $\hat{\mathbf{S}}_h^{(1)}(t)^{(l,k)}$ between each node in G with the class-matching node in Q, we follow the same procedure in [144] using **Algorithm 3** (lines 11-12). Note that in neighbor matching we no longer consider child-to-parent relationships. These score matrices are of size $n \times m$ for each h and t; each row is denoted as $\hat{s}_h^{(1)}(t)_{i\cdot}^{(l,k)}$, where each row i is the vector of weighted class adjacency values which match the class adjacency values in row j of the query graph Q's class adjacency matrix.

It is important to note that our (l, k)-Neighbor Matching procedure is not combinatorial. Specifically, it does not seek to find all groups of QF nodes of size 2, 3, 4,... and select among the greatest resulting scored results. Rather, our procedure captures all resulting QF-based conforming subgraphs which involve up to $k = 2, 3, 4, ...$ other QF nodes for each edge in E_Q. In the simplest sense, our procedure highlights to the analyst all those who would fulfill many, if not all, of the violent radicalization indicators when considering associates' activities. A combinatorial strategy would significantly increase the computational complexity of the search and it is not altogether clear that such results would be more helpful to an analyst.

7.3. Match Goodness Function

Now given the ability to obtain conforming matches by using up to l hops to fulfill each edge $\in E_Q$, it becomes important to measure the goodness of the matches, as well as incorporate the allowances for the reoccurrence or time-decay of occurrence of indicators. In Definition 16 we provide a such a measure.

We recall from line 12 in **Algorithm 6** that $\hat{s}_h^{(1)}(t)_{i\cdot}^{(l,k)}$ is the $1 \times m$ row vector for node $i \in V_G$ of weighted class adjacency values which match the class adjacency values in row j of the query graph Q's class adjacency matrix for each hop h and timestep t. If $i \in V_G$ and $j \in V_Q$ are of the same class (i.e. $L(i) = L(j)$ where $L(\cdot)$ is the labeling function), we know that their class adjacency vectors should have entries in the same positions over time. Since our data model for our connected query pattern graph assumes that a particular class of node occurs once in the query pattern ($|V_Q|$ = number of classes m) and that only the non-QF nodes are connected to with only one edge ($|E_Q| = m - 1$), we can also aggregate across the query graph depth (hops $h_{\max}$) of the node-class adjacency scores for a given allowable neighbor match path length l and timestep t. Each row vector of size $1 \times m$ could

have at most $|E_Q| = m - 1$ nonzero entries, and the column for the QF class is always 0. Thus for node i, we denote the multi-hop aggregate class match score vector with $\gamma(t)_{i\cdot}^{(l,k)}$, where each entry (i, z) for $i = \{1...n\}$ and $z = \{1...(m-1)\}$ is constructed as follows:

$$\gamma(t)_{iz}^{(l,k)} = \begin{cases} \sum_{h=1}^{h_{\max}} \hat{s}_h(t)_{iz}^{(l,k)} & l = 1 \\ \max\left\{ \sum_{h=1}^{h_{\max}} \hat{s}_h(t)_{iz}^{(l,k)} - 2 \sum_{h=1}^{h_{\max}} \hat{s}_h(t)_{iz}^{(l-1,k)}, 0 \right\} & l > 1 \end{cases} \tag{15}$$

For neighbor match path length $l = 1$, this aggregation is simply the sum across hops (15, top). For neighbor match path lengths $l > 1$, we avoid double counting the match scores achieved during the shorter match path length in two ways. This is done by subtracting the twice the sum $\sum_{h=1}^{h_{\max}} \hat{s}_h(t)_{iz}^{(l-1,k)}$ (for match path length of $l - 1$) from $\sum_{h=1}^{h_{\max}} \hat{s}_h(t)_{iz}^{(l,k)}$ (for match path length of $l > 1$) (15, bottom).

Since the summations are from 1 to $h_{\max}$ each time, the first subtracted sum removes from consideration those weighted class adjacency scores that were achieved from the shorter match path lengths. The second subtracted sum leaves only the class adjacency remnant scores which exceed the match scores from the shorter path lengths. In effect, for 2-2 neighbor matching, we use 15 and $l = 1$ to get the class adjacency match scores which result from direct connections and 15 and $l = 2$ to get the marginal class adjacency match score increases from neighboring connections. With this in place, we can now define the match goodness measure.

Definition 10. Match goodness

Consider a query graph Q and a resulting conforming data subgraph G_{S_i} determined through (l, k)-neighbor matching based on QF node i. $|E_Q|$ is the number of edges in Q, α is a decay parameter between $(0, 1]$, and $\gamma(t)_{iz}^{(l,k)}$ for node-class pair (i, z) is the multi-hop class match

score defined in 15. Then the measure of match goodness of conforming data subgraph G_{S_i} to query graph Q at time t is defined by 16.

$$g(Q, G_{S_i}, t; l, k, \alpha) = \frac{\sum_{\text{len}=1}^{l} \sum_{z=1}^{|E_Q|} \alpha^{\text{len}-1} \cdot \gamma(t)_{iz}^{(l,k)}}{|E_Q|} \tag{16}$$

This function obtains a match goodness score for the conforming subgraph for node $i \in V_G$ by utilizing an exponential weighting of a match score over the path lengths l that facilitated the match and normalizes by $|E_Q|$. The numerator in 16 for a particular l multiplies each class match score for node i with a decay factor and sums up the $|E_Q|$ entries. If $l = 1$, then each class match score in $\gamma(t)_{iz}^{(l,k)}$ contributes a value of $\alpha^0 = 1$ to the numerator. However, if $l = 2$, then the contribution is α times the class match score. The outer summation in the numerator then sums up the (weighted) match scores achieved at each allowable path length l. Given our construction of γ as defined in 15, the contributions from $l = 2$ are only those which exceed that which was achieved at $l = 1$.

We note that the value of a multi-hop connection between nodes depends only on the path length between them. For now, we make no distinction of the type or class of any intermediary nodes along the path. However, it follows that one could also designate different α decay parameters for each (QF,QF) pair to model the varying strength of ties between individuals. We intend to investigate the effect of various decay parameters for various pairs of QF nodes in future work.

Our match goodness definition also allows a QF node to achieve indicator matches through $k \cdot |E_Q|$ number of neighbors because there are at most a total of k QF nodes involved in the fulfillment of each edge in E_Q. However, we anticipate that a future improvement to the match goodness function that obtains a score for the κ most impactful

neighbors would better quantify whether a QF node is a high-scoring match due to its connection with a few neighbors, or to a larger number of neighbors who would each contribute a few indicators.

7.4. Experiments

In this section, we detail the experiments we conducted on three datasets to demonstrate the utility of INSiGHT on the detection of radicalization trajectories. First, we continue to utilize the small, synthetic radicalization toy graph we introduced in Fig. 5.1 to specifically test the INSiGHT enhancements. Next, we use an expanded and extended time version of the first dataset, now consisting of 61 nodes and 59 edges, to detect the match trajectories of individuals towards a hypothesized pattern of violent extremism. Lastly, we tested INSiGHT on a large, real world BlogCatalog dataset of over 470K nodes and 4 million edges, which serves as our proxy because it contains structural and behavioral parallels to intelligence networks which could be utilized to investigate radicalization.

7.4.1. Small, Synthetic Radicalization Toy Graph. We return to the motivating problem in Fig. 5.1 and note that the match goodness function now corrects a shortcoming of the earlier class similarity score: scores did not vary linearly with the number of total indicators. The previously presented class similarity score times series plot (from Fig. 5.9) is shown next to the match goodness time series plot in Fig. 7.1 a and b, respectively. Class similarity scoring is the sum of the fraction of node matches at each hop-level from a node. Thus its maximum value is the number of hop-levels in the query Q (which is 2 in this case), and each node match is equally weighted from among all the other nodes at its hop-level (all 4 of the 1-hop nodes are all weighted 0.25, and both of the 2-hop nodes weighted 0.50 in class similarity). The match goodness function, on the other hand, equally weighs all matching

edges in the conforming subgraph (which in this case is $1/6 = 0.167$ for each of the 6 edges in E_Q) and has a maximum value of 1. In Fig. 5.9b, Person 3's straight match goodness score line indicates that this new scoring function gives equal weight to all indicators whether they occur at either the first or second hop.

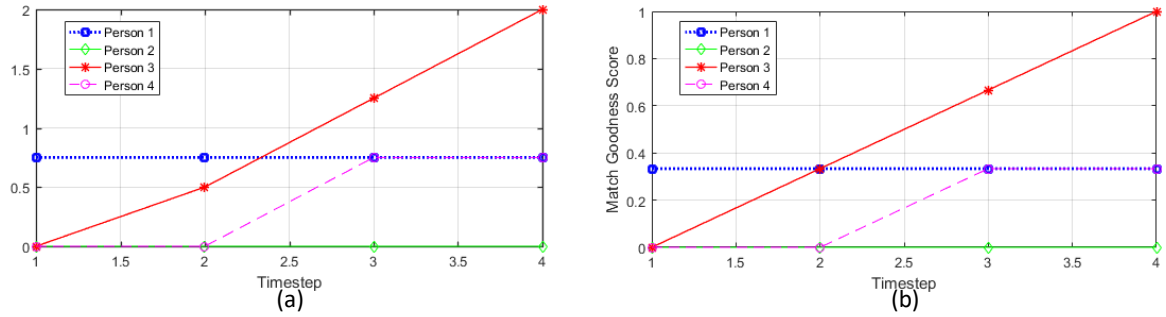


FIGURE 7.1. Multi-hop class similarity (a) and match goodness (b) scores for radicalization example, with investigative indicator type filtering.

7.4.2. SMALL, SYNTHETIC RADICALIZATION TOY GRAPH WITH ADDITIONAL PERSON LINKS. To initially validate neighbor matching, we continue to utilize the small, synthetic radicalization toy graph but now add 3 edges symbolic of person-to-person links shown in Fig. 7.2a. We added a bi-directional edge between Node 1 and Node 6 (representing a close familial relationship, for example), a bi-directional edge between Node 1 and Node 17 (representing a direct phone contact, for example), and a bi-directional edge between Node 18 and Node 10 (representing friendship linkage between social media accounts). As in the motivating example problem, we utilize the same query graph Q shown in 5.1a.

7.4.2.1. *Neighborhood Matching Only, No Effect of Reoccurrence or Time Decay.* We observe that Fig. 7.2b-g show the match goodness score time series for each of the people in the graph for varying α parameters. To confirm the validity of neighbor matching (without incorporating the effect of indicator reoccurrence or time decay), we will first focus on the

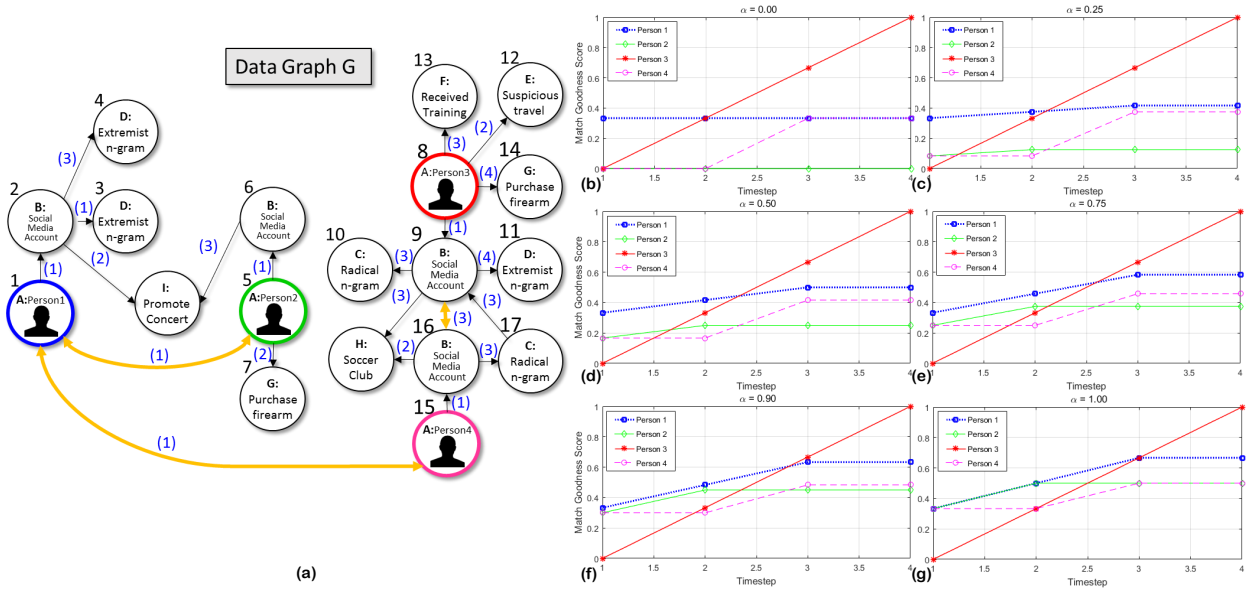


FIGURE 7.2. Expanded motivating example for detecting trajectories of home-grown violent extremists with additional person-to-person links. Beyond the base example from Fig.5.1a, the new data graph Fig. 7.2a above now has additional person-to-person links (in yellow). Fig. 7.2b-g depicts the match goodness scores over time for $\alpha = \{0.00, 0.25, 0.50, 0.75, 0.90, 1.00\}$, respectively.

results show in Fig.7.2g when $\alpha = 1.00$. Recall that at this α, each indicator match at all path lengths l is given the same weight.

Person 1 at timestep 1 has 2 personal indicators (Social Media Account and Extremist n-gram). Note also that at this timestep Person 1 is also connected through person-to-person links to Persons 2 and 4, each of whom also established Social Media Accounts. Because those duplicate social media account matches are reachable at path length $l = 2$ for the 'Person-to-Social Media Account' edge $\in E_Q$, they do not provide a marginal contribution above the edge match for a Social Media Account at $l = 1$. Therefore, Person 1's match goodness score at timestep 1 is simply $2/6 = 0.33$. At timestep 2, Person 1's score increases by $1/6$ due to its neighbor connection to Person 2 who purchased a gun. This connection occurred at $l = 2$ but fully counts because Person 1 did not have a similar indicator on its

own. At timestep 3, Person 1's score increases again by another 1/6 due to its neighbor connection to Person 3 who posted a radical n-gram. Again, because that indicator at $l = 2$ was unique to Person 1, the full score is added. Given inactivity at timestep 4, Person 1's final match goodness score is $4/6 = 0.66$.

We now explain the final match goodness scores for the others. Person 2 has a final match goodness score of $3/6 = 0.50$ because it has 2 personally suspicious indicators (Social Media Account and Purchase Firearm, which are no longer considered 'IIRA' for $\alpha > 0$) and 1 new indicator from associates (Extremist n-gram from Person 1). Person 1's Social Media Account node is not counted because one is already matched at $l = 1$. Also, because we do not yet have scoring for reoccurrence in effect, one of the repeated Extremist n-gram nodes from Person 1 occurring at the same reachability when $l = 2$ is not counted.

Person 3 has a final match goodness score of $6/6 = 1.00$ because it has all 6 personally suspicious indicators. The 2 indicators from other associates (Social Media Account and Radical n-gram both from Person 4) do not provide a marginal contribution above those achieved at $l = 1$.

Finally, Person 4's final match goodness score is $3/6 = 0.50$ because Person 4 has 2 personally suspicious indicators (Social Media Account and Radical n-gram), plus 1 others from associates (Extremist n-gram from both Person 1 and 3). The Radical n-gram from Person 3 and Social Media Accounts from Person 1 and 3 do not provide a marginal contribution above the same matches already achieved at $l = 1$.

Now we discuss the effect of α on the ability of our neighbor matching technique for identifying potentially suspicious conspiracies. While it may not be advisable for analysts to utilize $\alpha = 1.00$ where all neighbors' suspicious behaviors are equally weighted to those that

one personally performed, there may be acceptable parameter values that empower analysts to maintain awareness of one's increasingly suspicious neighbor activities and tempered by law enforcement resource constraints to investigate further. For example, if the connection between Person 1 and Person 2 is strong, it is possible for Person 1 to use Person 2's firearm in some attack (as in the case of San Bernardino). For any level of $\alpha > 0$ and $t \geq 2$, we observe Person 1's match goodness score for a hypothetical violent extremism query pattern is greater than had we not taken neighbors into account. Overall, it is clear that our neighbor matching technique credits activities which occur at varying distances from person nodes and and gives analysts a better sense for how embedded or entrenched a suspect may be among like-minded individuals doing similar or complementary activities on the path towards radicalization. In future work, we intend to explore designating different α decay parameters for each (QF,QF) pair to model the varying strength of ties between individuals.

7.4.2.2. *Effect of Reoccurence and Time Decay.* We now examine changes in the match goodness scores when we consider the effect of both the reoccurrence and time decay of indicators. Based upon the nature of the indicators, we chose the parameters for the time-decay and re-occurrence modules as shown in Parameter Set 1 in Table 5.5. A λ of 10.0 signifies that just one occurrence of that indicator class is necessary to achieve a near maximum score contribution. The classes 'Person' and 'Social Media Account' are basic nodes which need to occur once, while 'Received Training' and 'Purchase Firearm' may be more threatening indicators whose singular occurrence become important. We chose $\lambda = 4.6$ for both the Radical and Extremist n-gram indicators, which equates to 2 or more occurrences of each indicator is needed to achieve a near maximum score contribution.

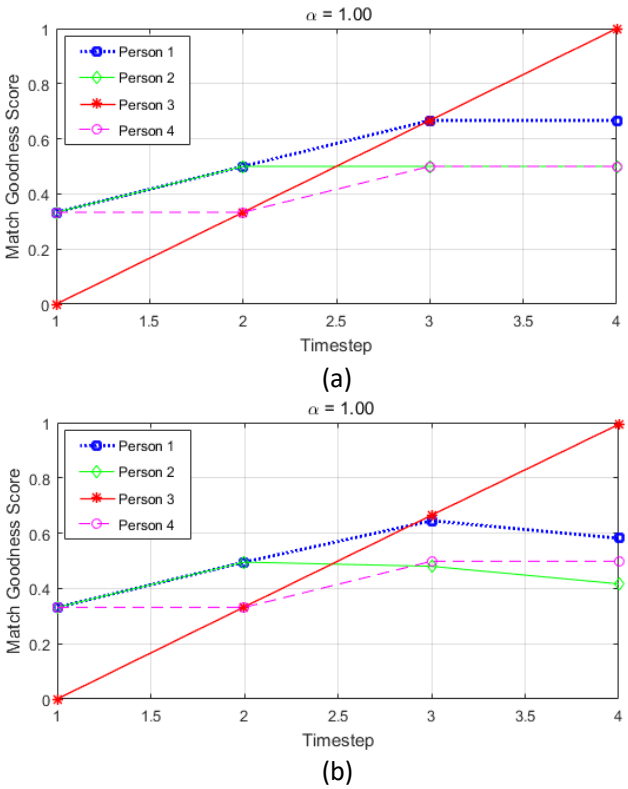


FIGURE 7.3. Multi-hop match goodness scores for radicalization example, without (a) and with (b) the effect of indicator reoccurence and time decay. Using Parameter Set 1 in 5.5, we artificially established the need for more than 2 occurrences each the Radical and Extremist n-gram indicators to achieve an effectively full score for that indicator. We also chose to examine the effect of decaying score contribution of the the Purchase Firearm indicator to half after 2 time steps.

In order to designate the time-decay parameter β, we notionally decided to equate a timestep in our synthetic dataset to represent 3 months (thus the 4 timesteps in the stylized dataset account for indicators which occurred over 1 year). A β of 1000.0 nearly eliminates the decay for the value of an indicator class. We chose this for the 'Person' and 'SM Account' basic classes, but also for the node classes 'Promote Concert' and 'Soccer Club' which are not indicators in the query pattern. We chose $\beta = 2.0$ for the Purchase Firearm indicator, which effectively decays the score contribution from that activity to a half after 2 time steps ($2.0 \cdot 3\text{months} = 6\text{months}$). This latter parameter selection is specific to this toy problem;

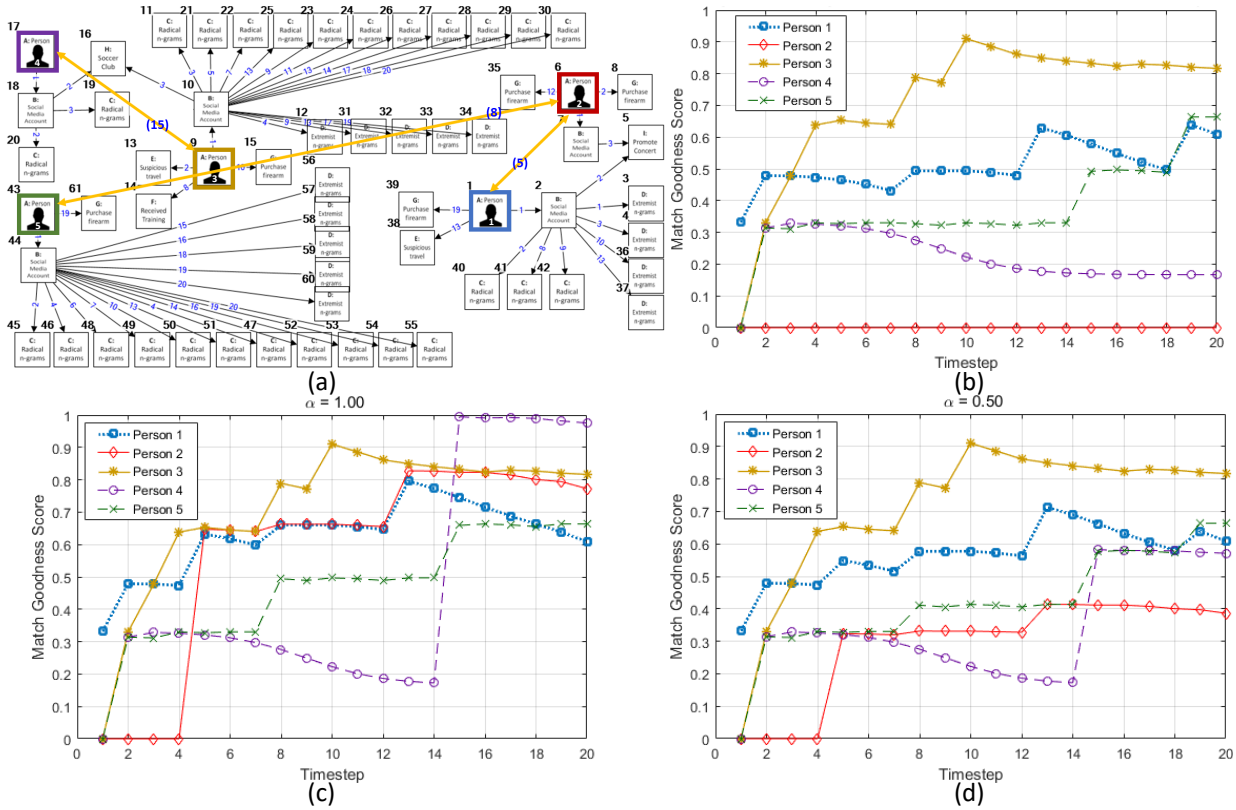


FIGURE 7.4. Expanded motivating example for detecting trajectories of homegrown violent extremists over an extended time. Beyond the base example from Fig.5.1a, the new data graph Fig. 7.4a now has one more individual and depicts additional reoccurring indicators of the on-line behaviors of some homegrown violent extremists. Fig. 7.4b is the match goodness score time series for 1-1 neighbor matching and shows the effect on scores due to the reoccurring indicators and time decay from inactivity. Fig. 7.4c and Fig. 7.4d are the match goodness score times series for 2-2 neighbor matching at $\alpha = 1.00$ and $\alpha = 0.50$, respectively.

in reality, we observe that firearms can be purchased much earlier or shortly before a violent extremist attack.

The time-decay parameter ξ allows the analyst to select the time duration of the decay. In this parameter set, we utilized $\xi = 1.0$ for all node classes, which implies a time duration of approximately 3ξ time units to decay the indicator significance to reach $\frac{1}{2}$.

Fig. 7.3 a and b, we show the match goodness score time series for $\alpha = 1.00$ without and with the reoccurrence and time decay effect, respectively. When comparing the results

to Fig. 7.3a, Person 1 in Fig. 7.3b has diminished scores at both timestep 3 and 4 because the contribution of Person 2's Purchase Firearm indicator decreased as expected with time decay. Also, because it now takes more than 2 occurrences to achieve a full score for an Extremist n-gram, Person 1's score is further slightly diminished at timestep 1 (where the first Extremist n-gram received a value of $0.9899 \cdot \frac{1}{6}$)and increased slightly at timestep 3 due to its repeated Extremist n-gram (which collectively now have a value of $0.9999 \cdot \frac{1}{6}$).

Like Person 1, Person 2 also has diminished match goodness scores at both timestep 3 and 4 because of the time decay of its Purchase Firearm activity. Person 3 has slightly diminished scores because both its own Radical n-gram and Person 4's Radical n-gram from which it receives a connection score is lower for the first occurrence at a given path length l of reachability. Also, Person 4's Purchase Firearm activity (which occurred at timestep 4) has only decayed slightly. Finally, Person 4's score was diminished at timestep 1 due to the single occurrence of Person 1's Extremist n-gram and timestep 3 due to the single occurrence of its own Radical n-gram. However, both Person 1's repeated Extremist n-gram at timestep 3 and Person 3's new Extremist n-gram at timestep 4 give Person 4 a very small score increase due to reoccurrence.

We conclude that the results from this sample test show that the improvements to INSiGHT perform as intended.

7.4.3. Small, Extended Time Synthetic Radicalization Dataset. The new data graph depicted in Fig. 7.4a is an expansion of the one shown in Fig. 5.1b and now has one more individual and depicts additional reoccurring indicators indicative of online behavior of some homegrown violent extremists.

As we developed this stylized, synthetic dataset, we first sought to have each of the 5 individuals fit some profile (without yet considering the Person-to-Person edges in yellow). Person 1 (Node 1) is an extremist with smaller number of social media posts, but later indicators of suspicious travel and firearm purchase. Person 2 (Node 6) is a non-extremist who purchased 2 firearms. Person 3 (Node 9) is an extremist with large number of posts and other indicators early-to-mid in timeline. Person 4 (Node 17) is a former extremist who made a small number of radical posts early in timeline. Finally, Person 5 (Node 43) is an extremist with large number of radical posts throughout, but only extremist posts late in the timeline.

As in the motivating example problem, we utilize the same query graph Q shown in 5.1a. This time, we utilized Parameter Set 2 in 5.5. We designated both Extremist n-gram and Suspicious Travel to have a λ of 4.6, which equates to 2 or more occurrences of an indicator to achieve a near maximum score contribution. We also designated Radical n-gram to have a λ of 2.3, which equates to 5 or more occurrences of an indicator to achieve a near maximum score contribution.

For the time-decay parameter β, we again notionally equated a timestep in our synthetic dataset to represent 3 months (thus the 20 timesteps in the dataset account for indicators which occurred over 5 years). We designated Radical n-gram to have a β of 6, which results in the decay to half of the maximum score contribution of an indicator class after 18 months ($= 6 \cdot 3$ months). The indicators with β parameters 8, 12, and 16 obviously signify that knowledge of its occurrences remains important for longer periods of time. To effectively mask any decay effect for the Person and Social Media Account classes, we set $\beta = 1000.0$.

Lastly, in this parameter set, we selected $\xi = 3$ for all node classes in which time decay was in effect.

Utilizing the improvements to INSiGHT, we produce the match goodness score time series plots in Fig. 7.4b-d for each of the persons of interest in the data graph to quantify the behavioral similarities to a violent extremist profile. We first discuss the results of 1-1 neighbor matching shown in Fig. 7.4b. It is clear that Person 3 has the highest initial gradient towards radicalization, and maintains his score with more instances of indicators over time. However, the trajectories of the other individuals are also worth noting. Person 1 has a relatively high similarity score throughout due to radical and extremists postings, but also has visible spikes when he goes on suspicious travel (timestep 13) and purchases a firearm (timestep 19). Person 5 also clearly has a sustained spike in his score later in the time frame due to the posting of radical and extremists and the purchase of a firearm (timestep 19). The decay of Person 4's earlier radical statements is evident with the shape of its match goodness score curve. Lastly, Person 2 has a match goodness score of 0 throughout the entire time frame of analysis because its activities were limited to the IIRA node-category.

Next, we now consider the Person-to-Person edges shown in yellow in Fig. 7.4a and examine the match goodness score time series for 2-2 neighbor matching. The 3 additional edges are between 1) Person 1 and Person 6 at $t = 5$, 2) Person 6 and Person 43 at $t = 8$, and 3) Person 9 and Person 17 at $t = 15$. Just as in the earlier small radicalization toy graph, these edges are intended to represent evidence for direct contact and/or established relationship ties between each pair of individuals.

Fig. 7.4c shows the times series plot of match goodness scores or neighbor matching when $\alpha = 1.0$, meaning that neighboring activities are given equal weight as those performed by the

individual itself. As previously mentioned, this is likely not the desirable level for an analyst, but we start here because it better illustrates what match changes result from considering neighbor activities.

We make the following observations related to each individual as a result of the visibility obtained from neighbor matching:

- Person 1 now has a significant increase in score at $t = 5$ due to its Person-to-Person connection with Person 2, who had previously purchased a gun. That activity is now counted among the rest of the activities of Person 1. We observe that the score increases again at $t = 13$ (suspicious travel), but not at $t = 19$ despite its firearm purchase. This is because while the attribution of the Purchase Firearm node changes Person 1, the score contribution of this indicator at $\alpha = 1.00$ transfers in its entirety from match scores at $l = 2$ to $l = 1$.
- Person 2, who had personally only done IIRA activities and had a score of 0 throughout the window of analysis, now at $t = 5$ has a high score due to its connection with Person 1. Note that while Person 2 connects to another extremist Person 5 at $t = 8$, Person 2's score does not increase because all of the indicators at that path length were already matched by Person 1.
- Person 3, who had maintained the highest radicalization scores over time at 1-1 neighbor matching, continue to exhibit the same scores at 2-2 neighbor matching and is not marginally affected by its only neighbor Person 4.
- Person 4 was a 'former' extremist but becomes very suspicious again at $t = 15$ when it connects with Person 3. All of Person 3's suspicious activities are attributed to Person 4 initially without the effect of time-decay that has been afforded Person 4.

The treatment of time decayed indicators at the point of a neighbor connection is a subject to different modeling strategies. For now, new indicators acquired through QF-QF connections have their time of last occurrence τ reset, but in future work we intend to investigate the possibility of adopting the existing τ from the QF node in which the indicator occurred.

- Person 5 now has a relative increase in score at $t = 8$ due to its connection with Person 2, who had purchased a gun earlier. Note that Person 4 had personally not purchased a firearm until $t = 19$, but its connection to Person 2 makes it more suspicious and at a higher potential threat. The neighbor match time series captures this. As in 1-1 neighbor matching, Person 5 had a significant increase in match goodness score at $t = 15$ due to the initiation of more Extremist n-grams (as opposed to the Radical n-grams it had previously only made).

Lastly, we show the times series plot of match goodness scores for this dataset when $\alpha = 0.50$ in Fig. 7.4d to illustrate what a different, more realistic parameter would provide. Namely, at this setting, neighbor activities are given some (albeit not full) weight to the individual so that analysts can still keep track of collective threats without making individuals more overly suspicious than they likely are. The plots still show the increases mentioned above as relatively smaller increases, and still differentiates in rank order those who are suspicious as a result of more personal actions rather than the actions of neighbors.

This experiment demonstrated that with a small example, we can indeed detect those who may individually be on a radicalization trajectory towards violent extremism based upon a simplistic query pattern. This may lead to possibilities of testing INSiGHT on real

data, that contains time-based, labeled indicators of bona fide cases of radicalization that both did and did not ultimately lead to violent activity.

TABLE 7.1. BlogCatalog full and subgraph characteristics

Characteristics	$G_{\textbf{full}}$	$G_{\textbf{subgraph}}$
Total Nodes	**382,482**	**2,387**
Number of userids	80,949	1,138
Number of weblogs	127,227	1,245
Number of unique tags	174,306	5
Total Edges	**4,009,467**	**4,918**
Number of userid-userid links	3,223,634	2,329
Number of userid-weblog links	127,226	1,245
Number of weblog-tag links	658,607	1,344

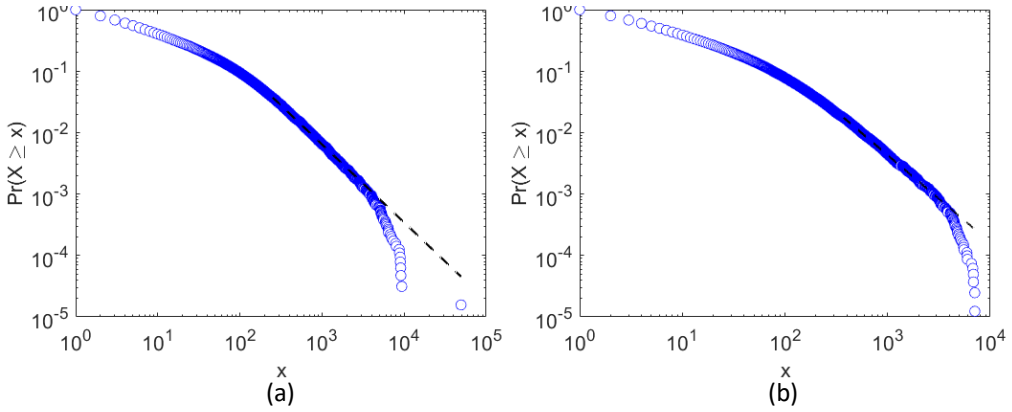


FIGURE 7.5. Log-Log Plot of the In- (a) and Out- (b) degree distributions of the BlogCatalog directed graph. $\mathbf{X}$ is the random variable for the degree distribution. The dotted lines in each plot show the theoretical power law for $\alpha = 2.27$ and $x \geq 251$, and $\alpha = 2.42$ and $x \geq 373$, respectively.

7.4.4. LARGE REAL DATA SET. Finally, we test our approach on the BlogCatalog dataset,[40] a large, real corpus of user activities and the social network in between them [320]. We are using this as a proxy for the type of intelligence or law enforcement data networks that could be used to track the radicalization of violent extremists. This dataset was first described and utilized in our earlier work [144], but we summarize it here for our tests in the neighborhood match setting.

[40]Available at http://dmml.asu.edu/users/xufei/datasets.html

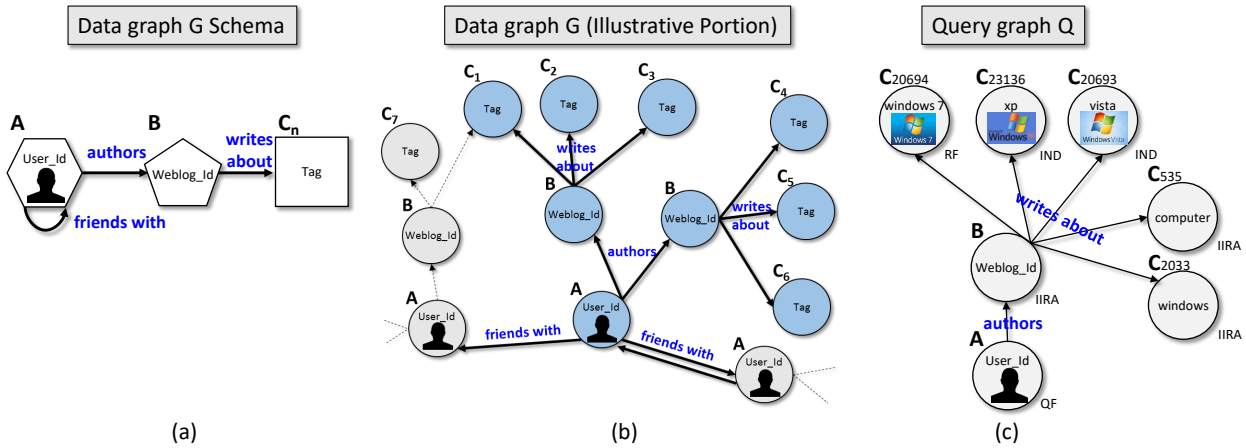


FIGURE 7.6. The schema and partial datagraph of the BlogCatalog graph and the query graph Q. Fig. 7.6a depicts the node types and connections present in the network, while Fig. 7.6b is a partial graph that is illustrative of the larger graph. Fig. 7.6c shows the 7-node query graph Q.

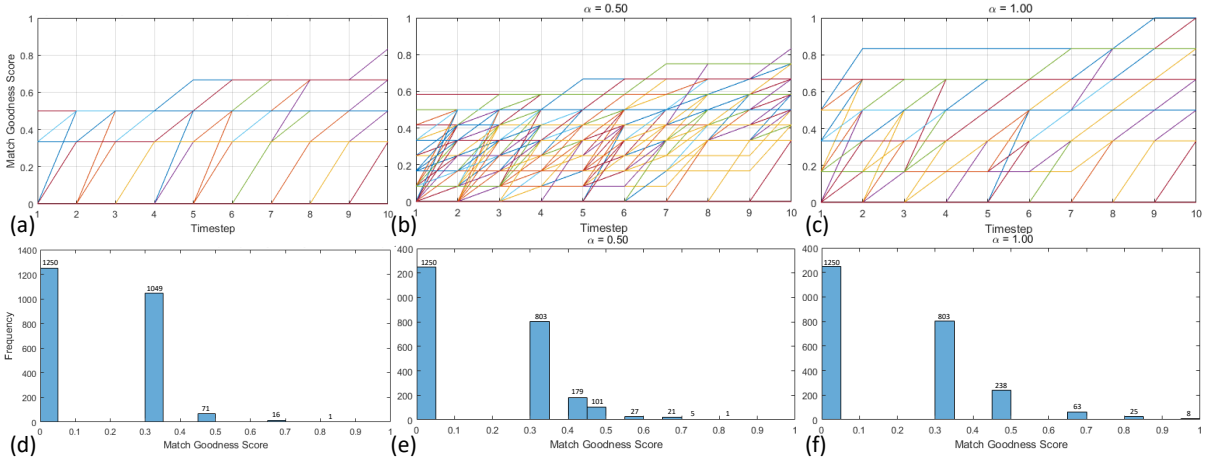
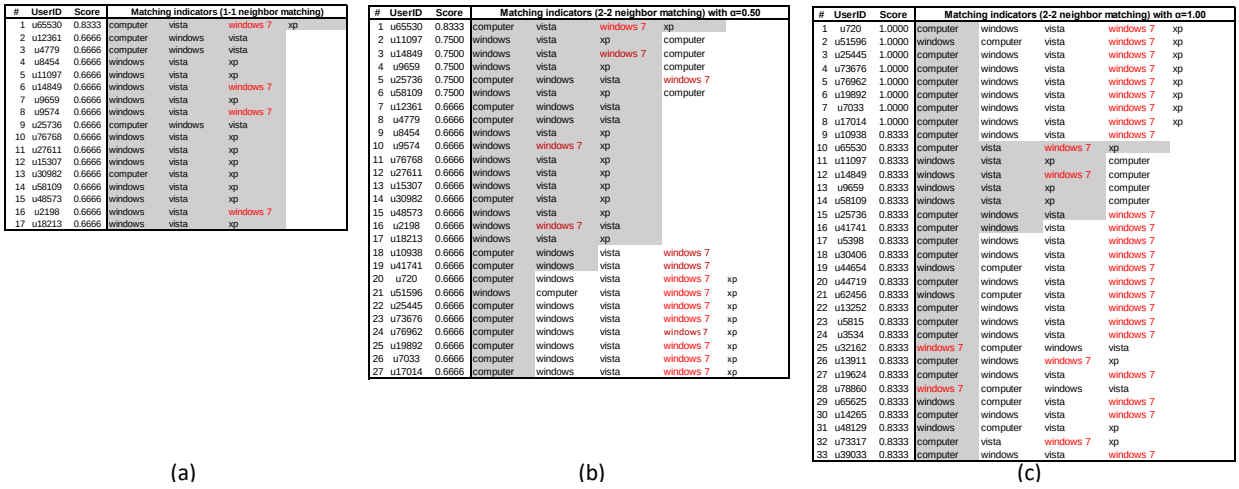


FIGURE 7.7. INSiGHT results on the BlogCatalog dataset. The match goodness time series plots for all nodes in G_{subgraph} for 1-1 neighbor matching (a) and 2-2 neighbor matching at $\alpha = 0.50$ (b) and $\alpha = 1.00$ (c). The corresponding histograms of the final match goodness scores (at $t = 10$) are shown in Fig. 7.7 d-f. The number above each bar is the quantity of nodes with that respective match goodness score.

The BlogCatalog dataset is a scrape taken in July 2009 of a social media site that allows users to register and promote their own blogs, as well as connect with other bloggers. The network schema shown in Fig. 7.6a depicts the node types and connections present in the network, and Fig. 7.6b is an illustrative example of a portion of the data graph. The original graph had over 380K nodes and over 4 million edges (see Table 7.1). We find that the degree



#	UserID	Score	Matching indicators (1-1 neighbor matching)			
1	u65530	0.8333	computer	vista	windows 7	xp
2	u12361	0.6666	computer	windows	vista	
3	u4779	0.6666	computer	windows	vista	
4	u8454	0.6666	windows	vista	xp	
5	u11097	0.6666	windows	vista	xp	
6	u14849	0.6666	windows	vista	windows 7	
7	u9659	0.6666	windows	vista	xp	
8	u9574	0.6666	windows	vista	windows 7	
9	u25736	0.6666	computer	windows	vista	
10	u76768	0.6666	windows	vista	xp	
11	u27611	0.6666	windows	vista	xp	
12	u15307	0.6666	windows	vista	xp	
13	u30982	0.6666	computer	vista	xp	
14	u58109	0.6666	windows	vista	xp	
15	u48573	0.6666	windows	vista	xp	
16	u2198	0.6666	windows	vista	windows 7	
17	u18213	0.6666	windows	vista	xp	

(a)

#	UserID	Score	Matching indicators (2-2 neighbor matching) with α=0.50			
1	u65530	0.8333	computer	vista	windows 7	xp
2	u11097	0.7500	windows	vista	xp	computer
3	u14849	0.7500	windows	vista	windows 7	computer
4	u9659	0.7500	windows	vista	xp	computer
5	u25736	0.7500	computer	windows	vista	windows 7
6	u58109	0.7500	windows	vista	xp	computer
7	u12361	0.6666	computer	windows	vista	
8	u4779	0.6666	computer	windows	vista	
9	u8454	0.6666	windows	vista	xp	
10	u9574	0.6666	windows	windows 7	xp	
11	u76768	0.6666	windows	vista	xp	
12	u27611	0.6666	windows	vista	xp	
13	u15307	0.6666	windows	vista	xp	
14	u30982	0.6666	computer	vista	xp	
15	u48573	0.6666	windows	vista	xp	
16	u2198	0.6666	windows	windows 7	vista	
17	u18213	0.6666	windows	vista	xp	
18	u10938	0.6666	computer	windows	vista	windows 7
19	u41741	0.6666	computer	windows	vista	windows 7
20	u720	0.6666	computer	windows	vista	windows 7 xp
21	u51596	0.6666	windows	computer	vista	windows 7 xp
22	u25445	0.6666	computer	windows	vista	windows 7 xp
23	u73676	0.6666	computer	windows	vista	windows 7 xp
24	u76962	0.6666	computer	windows	vista	windows 7 xp
25	u19892	0.6666	computer	windows	vista	windows 7 xp
26	u7033	0.6666	computer	windows	vista	windows 7 xp
27	u17014	0.6666	computer	windows	vista	windows 7 xp

(b)

#	UserID	Score	Matching indicators (2-2 neighbor matching) with α=1.00				
1	u720	1.0000	computer	windows	vista	windows 7	xp
2	u51596	1.0000	windows	computer	vista	windows 7	xp
3	u25445	1.0000	computer	windows	vista	windows 7	xp
4	u73676	1.0000	computer	windows	vista	windows 7	xp
5	u76962	1.0000	computer	windows	vista	windows 7	xp
6	u19892	1.0000	computer	windows	vista	windows 7	xp
7	u7033	1.0000	computer	windows	vista	windows 7	xp
8	u17014	1.0000	computer	windows	vista	windows 7	xp
9	u10938	0.8333	computer	windows	vista	windows 7	
10	u65530	0.8333	computer	vista	windows 7	xp	
11	u11097	0.8333	windows	vista	xp	computer	
12	u14849	0.8333	windows	vista	windows 7	computer	
13	u9659	0.8333	windows	vista	xp	computer	
14	u58109	0.8333	windows	vista	xp	computer	
15	u25736	0.8333	computer	windows	vista	windows 7	
16	u41741	0.8333	computer	windows	vista	windows 7	
17	u5398	0.8333	computer	windows	vista	windows 7	
18	u30406	0.8333	computer	windows	vista	windows 7	
19	u44654	0.8333	windows	computer	vista	windows 7	
20	u44719	0.8333	computer	windows	vista	windows 7	
21	u62456	0.8333	windows	computer	vista	windows 7	
22	u13252	0.8333	computer	windows	vista	windows 7	
23	u5815	0.8333	computer	windows	vista	windows 7	
24	u3534	0.8333	computer	windows	vista	windows 7	
25	u32162	0.8333	windows 7	computer	windows	vista	
26	u13911	0.8333	computer	windows	windows 7	xp	
27	u19624	0.8333	computer	windows	vista	windows 7	
28	u78860	0.8333	windows 7	computer	windows	vista	
29	u65625	0.8333	windows	computer	vista	windows 7	
30	u14265	0.8333	computer	windows	vista	windows 7	
31	u48129	0.8333	windows	computer	vista	xp	
32	u73317	0.8333	computer	vista	windows 7	xp	
33	u39033	0.8333	computer	windows	vista	windows 7	

(c)

FIGURE 7.8. Top scoring UserIDs in BlogCatalog Experiment. Fig. 7.8a depicts the top 17 scoring UserIDs in 1-1 Neighbor Matching. Fig. 7.8b depicts the top 27 scoring UserIDs in 2-2 Neighbor Matching at $\alpha = 0.50$. Fig. 7.8c depicts the top 33 scoring UserIDs in 2-2 Neighbor Matching at $\alpha = 1.00$. The grey colored indicators denote those matches which were tags established by the corresponding User_ID, while those not shaded denote tags that were established by the User_ID's immediate neighbors. The 'windows 7' red flag indicators are highlighted in red.

distributions in this directed social network follow the power law of a scale free network, where power α for the in- and out-degree distributions are estimated to be 2.27 and 2.42, respectively (see Fig. 7.5). To simplify the dataset first presented in [142], we retained only one person identifier ('User_ID') for each individual and ignored the 'ID' node class.

In keeping with both the consistency and minimization principles established in surveillance ethics research (see Marx in [195]), we developed a preprocessing screening procedure to filter out nodes and edges which have no connection to the indicators in our query[41]. Specifically, we employed a routine that produced a subgraph that included only user IDs which had tags that consisted of the 5 words in the query, any interconnections between the user IDs, as well as their respective weblog IDs. The resulting graph has only 2,387 nodes

[41]Consistency refers to whether individuals have an equal chance of being subjected to a query. Minimization is the principle to "minimize the invasiveness" and "extent of personal and personably identifiable information collected" [195]. While Marx cites numerous other important considerations in the justification and implementation of any proposed surveillance-based approach, those considerations are beyond the scope of this work.

and 4,918 edges (0.624% of original nodes and 0.123% original edges) and is further detailed in Table 7.1.

It is important to mention that we modeled each tag as its own node, regardless of how many blogs utilized it. In **Algorithm 6**, every subsequent hop's adjacency matrix contained only new connections. Any nodes that were connected to previously were not included for that particular hop. So, in this modeling setting, 2-2 neighbor matching does not score for both what the individual and neighbors did, but only credits neighbor activities beyond what the individual did itself.

7.4.4.1. *Query Description.* We utilize the same query pattern developed in [144], which was a proxy query on a benign subject matter with structural parallels to an investigation for a latent behavior (Fig. 7.6c). The query's focus is for user IDs who had been writing blogs related to Microsoft Windows operating systems (XP and/or Vista) and subsequently also began to write about Windows 7 when it was released in July 2009 (which is the month in which the data was collected). Node class A is the query focus user ID, and node class B is the weblog with certain tags. All C class nodes are self-identified tags of the blog, which were meant to be analogous to the behavioral indicators or n-gram topics determined through machine-classification and semantic text analysis in a radicalization application. The labels 'computer' (C535) and 'windows' (C2033) are relatively frequent labels which help provide context or additional clarity on the true topic set, and labels 'xp' (C23136) and 'vista' (C20693) are indicators that the blog is about Windows operating systems. These latter nodes are necessary but not sufficient for the latent behavior of interest. Finally, label 'windows 7' (C20684) is considered a latest-occurring red flag indicator.

As in [144] we are trying to detect individuals who undertake this 'latent behavior.' However, in this work, we also desire to use the INSiGHT's advancements, particularly neighbor matching, to gain visibility on each neighbors' tags and identify any full or partial complement matches. This test on a benign application helps confirm that our approach can work on radicalization-specific data where the detection of conspiratorial clusters is important.

7.4.4.2. *Adding Edge Timestamps.* The BlogCatalog dataset did not contain the timestamps when each of the links occurred, information that is critical to detecting latent behavior trajectories. As detailed in [144], we developed a procedure to generate artificial timesteps over a short time window of analysis (10 timesteps) as a proof of principle.

7.4.4.3. *Analysis of BlogCatalog Results.* INSiGHT successfully identified all the BlogCatalog accounts that match the query in full or in part, as well as produced the match trajectory for each account. In Fig. 7.7 we provide the match goodness time series plots for all nodes in G_{subgraph} for 1-1 neighbor matching (a) and 2-2 neighbor matching at $\alpha = 0.50$ (b) and $\alpha = 1.00$ (c) over the 10 timesteps. The corresponding histograms of the final match goodness scores (at $t = 10$) are shown in Fig. 7.7 b-d. The number above each bar is the quantity of nodes with that respective match goodness score. Note that in each histogram, there are 1250 nodes (all nodes which are not User_IDs) which have a match goodness score of 0. In Fig. 7.8 we also provide the User_Ids, match goodness scores, and matching indicators for the top 17 accounts due to 1-1 neighbor matching (a), top 27 accounts due to 2-2 neighbor matching with $\alpha = 0.50$ (b), and the top 33 accounts due to 2-2 neighbor matching with $\alpha = 1.00$ (c). The grey colored indicators denote those matches which were

tags established by the corresponding User_ID, while those not shaded denote tags that were established by the User_ID's immediate neighbors.

We see the merits of our approach in investigative graph search by segmenting out only a fraction of the entities who are on pathways of partially or completely matching a query pattern. As the histograms show, it is possible to focus investigative efforts on the top-k or top bins of accounts with the highest match goodness scores.

For 1-1 neighbor matching, the top scoring account was User_Id 'u65530' who utilized the 4 of the 5 indicator tags ('computer,' 'vista,' 'xp,' and 'windows 7') over the course of time. After this match, there were 16 others identified in Fig. 7.8a that utilized 3 of the 5 indicator tags. These 1-1 neighbor match results are exactly those found based on our early development of INSiGHT in [144].

For 2-2 neighbor matching, we utilized two parameter settings for α to examine the resulting differences in match results. At $\alpha = 0.50$, neighbors' tags are being considered at half weight for each User_ID. Accordingly, the match goodness score trajectory in Fig. 7.7b shows more various scores throughout the window of analysis, and Fig. 7.7e shows that the distribution of final match goodness scores is more positively skewed due to User_IDs receiving marginal score increases due to neighbors' tags. For instance, although the top scoring match in this setting was still User_ID 'u65530,' the next 5 other high scorers were those who personally selected 3 of the 5 indicator tags and had at least one neighbor who made one additional tag.

At $\alpha = 1.00$, neighbors' tags are being considered at the full weight for each User_ID. The match goodness score trajectory in Fig. 7.7c shows that accounts are achieving higher scores earlier due to connections with neighboring indicators. While the ranking of neighbor

matches seems to make sense for $\alpha = 0.50$, we observe that the results for $\alpha = 1.00$ puts too much allowance on neighbors' matches and the resulting rankings less sensible. Specifically, Fig. 7.8f shows that the top 9 scorers in this setting are those User_IDs who personally exhibited only one indicator, but relied on one or more neighbors for their indicators to make up a partial complement to the set. Recalling the fact that the dual goal of this problem was to detect both individuals and groups performing this latent behavior, the use of the parameter $\alpha = 0.50$ is more sensible because it first brings forth to analysts those accounts which directly perform a preponderance of the indicators and shows the possibility of neighbors completing the query pattern. Overall, these results demonstrate the importance of α in controlling the visibility and impact on match scores for neighbor activities.

7.5. Conclusion

In this chapter, we extended INSiGHT to find full or partial matches against a radicalization query pattern from a single node to a node cluster, while still quantifying the pace of the appearance of the indicators or neighbors' indicators. Its performance was demonstrated on small synthetic radicalization data sets as well as a real data graph of 470K nodes and 4 million edges. It holds promise for assisting law enforcement and intelligence agencies in the radicalization detection problem especially in the presence of linked co-conspirators who are on the trajectory towards violent extremism. In future work, we intend to investigate the effectiveness of varying the α exponential weighting parameter for various pairs of QF nodes in order to model any strong or weak ties between neighbors. Also, we intend to conduct additional experiments on other real-world datasets and are presently seeking real data that contain time-based, labeled indicators of bona fide cases of radicalization that both did and did not ultimately lead to violent activity. Lastly, we intend to devise an incremental

graph pattern matching approach like those found in [49, 101] in order to dynamically and efficiently update the match scores in the presence of new data points.

CHAPTER 8

Analyses of Dynamic Radicalization Indicators

8.1. Introduction

In the field of dynamic risk/threat assessment of violent extremists, much of the extant literature provides itemized behaviors (in checklist-type lists) gathered from empirical evidence and case studies. There is growing recognition that such lists are inadequate for law enforcement personnel and tend to make many more people seem suspicious beyond the relatively small number of those who would commit terrorist acts [67]. In this chapter, we contribute to the body of research on radicalization pathways in the modeling radicalization as a discrete dynamical process and the use of state transition analysis. Our objective here to find more discerning patterns among the behavioral indicators of radicalization of violent extremists that could help law enforcement officials more effectively recognize risks and screen for those more likely to be on paths to violence. These patterns could be set-based (which sets of indicators frequently occur together in a case) or sequence-based (which indicators follow or immediately follow another indicator).

Our efforts are centered on analyzing the unique, restricted-use dataset "A Behavioral Study of the Radicalization Trajectories of American 'Homegrown' Al Qaeda-Inspired Terrorist Offenders, 2001-2015," assembled by Professor Jytte Klausen at Brandeis University and offered by the National Institutes of Justice through the National Archive of Criminal Justice Data (NACJD) [167]. We discussed the originating study that produced this dataset in Section 2.4.4.1.

8.2. Dataset description

The anonymized dataset contains 331 individuals (Group A) whom the researchers identified as meeting three study conditions: "1) he or she must have spent some or all of their formative years in the United States, 2) the radicalization process must have taken place primarily within the United States, and 3) the first instance of verifiable illicit activity leading to charges related to terrorism took place after September 11, 2001" [168]. There are 9 features/variables in this portion of the dataset: ID, Year of Birth, Sex, Ethno-National Origin, Religious Conversion Status, Educational Status, Criminality Before Radicalization, Year of Radicalization (start), and Year of Criminal Action [168].

From this group, 135 individuals (Group B) were selected for more detailed study. Detailed forensic biographies were collected for each of the subjects in Group B, and were compiled from publicly available court documents and investigations into their activities conducted by the United States government and news media, which may include online communications posted by the terrorist offenders [168]. For each of the 135 U.S. violent extremists, this dataset contains the actual or inferred dates (to the closest month, if applicable) when any of the 27 behavioral indicators were exhibited. The dataset also contains 4 researcher-coded binary features (Online/Real Radicalization, Foreign Fighter, Undercover Agent, and Mental Illness), which we intend to utilize as conditioning factors in future analysis. The codebook for these indicators is available in Appendix C, and is directly excerpted from [167] and provides the research team's definitions. Table 8.1 shows the listing of the 27 date features.

The original research reports based upon this data [165, 168] already provided some insights into the duration of the radicalization process. For example, Klausen concluded,

TABLE 8.1. Data features in Klausen Radicalization Dataset. Source: [167].

Date of Birth	Epiphany Date
Religious Conversion Date	Peer-Immersion Date
Disillusionment Date	Domestic Physical Training Date
Trauma Date	Marriage Seeking Date
Personal Crisis Date	Societal Disengagement Date
Information Seeking Date	Desire for Action Date
New Religious Authority Date	Non-Violent Support Date
Ideological Rebellion Date	Joins Foreign Insurgency Date
Lifestyle Changes Date	Issues Threats Date
Occupational/Educational Disengagement Date	Steps Towards Violence Date
School Dropout Date	Date of Criminal Action
Underemployment Date	Arrest Date
Dawah- Virtual Date	Sentencing Date
Dawah- Real Life Date	

(1) The average time period from the initial exploration of extremist ideas to criminal action was over 3 years (38 months) (excluding a few outliers).

(2) Conditioned on those committing criminal action in 2015, the average time period of radicalization was 2 years or less.

(3) Excluding pre-radicalization activities (indicators signaling an initial exploration of belief systems), the average time to criminal action was 6.25 months for those individuals radicalized after 2010 [165].

As mentioned earlier, these statistics ought to aid both practitioners and policy-makers. Klausen also provided summary statistics on the general prevalence of individual indicators and commented on either the consistency or deviation in their order of appearance in the case studies [165, 168].

8.3. Methodology: Modeling Radicalization as a Discrete Dynamic Process

Beyond the analyses in the original research, we modeled an individual's radicalization as a discrete time dynamical process that produces indicator outcomes at timesteps $t =$

$1, 2, 3, \ldots$. We based our preliminary analysis framework on the path-dependence theory in [236] and utilized state transition analysis and network analysis techniques to determine the relative frequencies of transitions between certain states (indicators) and which ones are more or less exhibited. The 26 radicalization date features are possible outcomes in the state space.[42] Following [151], a state refers to an outcome at any moment in time, which in our case can be a single behavior or a set of behaviors. While the rigorously coded and timestamped dataset is a major contribution in the field, we acknowledge upfront that the 26 radicalization features may not completely characterize radicalization processes. We also do not make the assertion that the indicators recorded empirically are in fact the *only* indicators that occurred. However, research suggests that radicalization is a complex, path-dependent process where ones behavioral history matters in subsequent behavior. As Horgan observed, "Overall, for any given individual, becoming involved in terrorism will reflect a dynamic, though highly personalized, process of incremental assimilation and accommodation" [133, p. 85]. With the current state of knowledge, we clearly cannot make the path independence assumption in order to cast the problem as Markovian. This subsequently precludes us from utilizing traditional Markov chain and Hidden Markov Models analysis methods.

However, it is still possible to utilize state transition diagrams to graphically represent the radicalization phenomenon as a means to glean critical insights into detecting those on such pathways. The structure and detailed contained in the Klausen's dataset specifically enables this level of analysis [167, 168]. This is state transition diagram, each of the 26 observable behaviors appears as its own node. In this model, the discernible radicalization behavioral sequence for each of the 135 U.S. radicalized violent extremists would appear as a pathway

[42]All date features were considered, except for sentencing date which we deemed as more a characteristic of the criminal justice system response.

from a 'Year of Birth' node to the 'Arrest Date' node. Because of the available of information (either by fact or by the granularity of the data sources), dates for various behaviors were inferred to the same date, thus creating an outcome (state) which may include multiple behaviors. We first attempted to model each unique outcome as its own state, whether it was a single behavior or a unique set of behaviors. This resulted in state space of 149 unique states (26 single behavior states and 123 combined behavior states) that was less interpretable.

For example, one of the violent extremists studied was assessed to have exhibited disillusionment on July 1, 1993, and then next was assessed to have exhibited a personal crisis, a converted to Islam, and engaged in *dawa* (proselytizing) in person all on July 1, 1995. The first-pass model involved a transition arrow from the state of 'disillusionment' to a combined state of 'personal crisis, conversion, and *dawa* in person' signified that the individual first exhibited one behavior and then next the other three at (approximately) the same time. When considering the multitude of combined states that may vary just slightly or greatly from each other based on possible incomplete information among all 135 cases, this approach proved unwieldy and offered less meaningful insights.

For simplicity and clarity, we devised an alternative method of analysis that focused on which behaviors may immediately follow one another in time sequence while considering that multiple behaviors could be exhibited at the same time. We defined a modified state transition matrix $\mathbf{P} \in \mathbb{W}^{26\times 26}$ where each entry p_{ij} is the number of times a state (outcome) which contains behavior i leads to a state (outcome) which contains behavior j among all 135 individuals in the dataset. This specifically allows for state transitions in which one behavior can lead directly to combined set of behaviors, as well as a combined set of behaviors can lead

directly to a single behavior or another combined set of behaviors. This matrix is depicted in Fig. 8.1, where it is also shaded in green according to the number of transitions. There are 388 distinct sequences of behaviors (transitions where $p_{ij} > 0$) and 1651 total transitions (sum of all entries in $\mathbf{P}$).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
1: Year.of.Birth	0	13	31	16	17	29	27	1	8	5	12	4	6	1	1	10	1	0	0	5	0	0	0	1	0	0
2: Convert.Date	0	0	3	3	1	5	3	1	7	0	2	1	3	1	1	7	1	1	1	5	0	0	0	1	0	0
3: Disillusionment	0	4	0	3	4	16	9	3	9	2	2	0	5	5	0	4	2	1	3	5	1	1	0	0	1	0
4: Trauma	0	2	2	0	5	5	1	0	2	1	1	0	2	0	0	3	2	0	3	2	0	0	0	0	0	0
5: Personal.Crisis	0	3	4	0	0	12	6	0	6	1	1	2	1	2	0	3	2	0	0	3	0	1	0	0	0	0
6: Seeking.Information	0	12	5	0	1	0	21	1	15	2	4	2	5	2	0	12	2	2	3	4	0	2	0	2	1	0
7: New.Authority.Figures	0	2	4	0	1	3	0	0	10	1	4	0	10	4	3	26	4	1	3	17	1	3	0	4	0	0
8: Rebellion	0	0	0	0	0	1	1	0	2	1	0	0	0	0	0	2	2	1	1	5	0	1	1	1	3	0
9: Lifestyle.Changes	0	1	4	1	4	2	6	0	0	0	2	4	8	11	1	17	3	2	3	10	1	3	2	1	4	0
10: Educational.Occupational.Disengagement	0	1	2	1	1	1	2	1	1	0	0	0	1	1	0	1	1	1	1	3	0	0	2	1	5	4
11: Drop.Out.Date	0	2	3	2	2	5	3	1	4	0	0	4	2	2	0	2	1	3	1	5	0	1	0	2	6	5
12: Underemployment	0	0	2	0	1	1	1	2	1	0	0	0	0	1	2	2	2	0	0	8	1	2	0	2	3	0
13: Dawa..Virtual	0	0	0	0	0	0	4	0	2	1	4	0	0	1	2	7	3	2	2	14	3	3	1	6	4	0
14: Dawa.Real.Life	0	1	1	0	0	1	2	0	2	1	1	1	0	0	1	6	4	3	0	8	1	3	0	5	4	0
15: Epiphany	0	1	0	0	0	1	0	0	2	0	1	0	0	0	0	0	2	1	0	3	0	2	1	0	1	0
16: Peer.Immersion	0	1	1	1	4	2	0	4	5	1	4	3	7	2	1	0	17	4	3	30	3	6	3	12	4	0
17: Physical.Domestic.Training	0	0	0	0	0	0	2	0	6	1	1	0	4	2	0	3	0	2	0	7	1	9	0	12	14	4
18: Marriage.Seeking	0	0	0	0	0	0	1	1	1	0	0	1	1	0	0	7	0	0	1	3	1	3	1	5	6	3
19: Societal.Disengagement	0	0	0	0	0	0	0	1	1	2	1	0	3	1	0	2	1	1	0	2	1	3	0	3	0	1
20: Desire.for.Action	0	1	1	0	1	2	0	2	4	5	5	2	6	4	0	6	8	7	2	0	6	28	5	31	24	5
21: Passive.Support	0	0	1	0	0	1	0	0	0	1	1	0	0	1	1	0	1	1	0	0	0	2	0	4	9	4
22: Joins.Foreign.Insurgency.Org.	0	0	0	0	0	1	0	0	1	1	1	1	1	0	2	0	0	0	0	0	3	0	1	8	19	18
23: Issues.Threats	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1	1	0	0	2	11	2
24: Steps.towards.Violence	0	0	1	0	0	1	0	4	1	1	2	0	3	0	1	0	5	1	0	3	2	5	2	0	46	18
25: Date.of.Criminal.Action	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	2	0	2	1	3	0	103
26: Arrest.Date	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

FIGURE 8.1. Radicalization feature/behavior transition matrix $\mathbf{P} \in \mathbb{W}^{26\times 26}$ where each entry p_{ij} is the number of times a state (outcome) which contains behavior i leads to a state (outcome) which contains behavior j.

8.3.1. IDENTIFICATION OF FREQUENT BEHAVIORAL TRANSITIONS. With this in place, we now first discuss how we identified the most frequent behavioral *transitions*. Beyond inventory lists of static behavioral indicators, we intend to extend the level of analysis first performed by Klausen and identify the most frequent radicalization indicators which follow certain indicators in confirmed cases of violent extremism. We seek *dynamic* behavioral indicators.

The green shading on the transition matrix $\mathbf{P}$ aides the reader in identifying those indicators which most frequently follow a specific indicator. However, we can also represent the data as a state transition diagram.

Our first thought was to obtain a right stochastic matrix $\mathbf{P}'$ by row-normalizing $\mathbf{P}$. This allows us to examine the proportion of transitions from each behavior to another behavior. We note, however, that our row normalization of $\mathbf{P}$ by dividing each entry by the total number of transitions by row results in an inaccurate view of the prevalence of a transition. Again, this is based on the fact that transitions from a behavior to one or more behaviors occurred simultaneously. For example: suppose that a particular person only exhibited 3 behaviors X, Y, and Z on the pathway towards radicalization where behavior X is subsequently followed by behaviors Y and Z inferred to have occurred on the same day. Standard row normalization would weight the transition $X \to Y$ and $X \to Z$ would be weighted 0.50 and imply that the transition occurred only half the time. In fact, in this one transition from behavior X, both behavior Y and Z followed.

To correct this accounting, we determined the total number of 'transition occurrences' for each behavior (row), and normalized each row of entries of $\mathbf{P}$ with this total. This resulted in a modified (no longer right stochastic) behavior transition matrix $\mathbf{P}''$. In this previous example, the modified row normalization would result in weights of 1.00 for both behaviors Y and Z since each occurrence would be divided by 1 (the total number of transition occurrences for this person) rather than 2 (the total number of subsequent behaviors involved in the single transition).

All the transitions aggregated from among the 135 violent extremists in this way are depicted in Fig. 8.2a, where the 26 nodes are features/behaviors and weighted paths represent instances when a feature/behavior sequentially followed another. The varied behavioral radicalization paths towards criminal action clearly supports the conclusion by many researchers that there is no single path towards violent extremism [133, 165, 167, 224, 241].

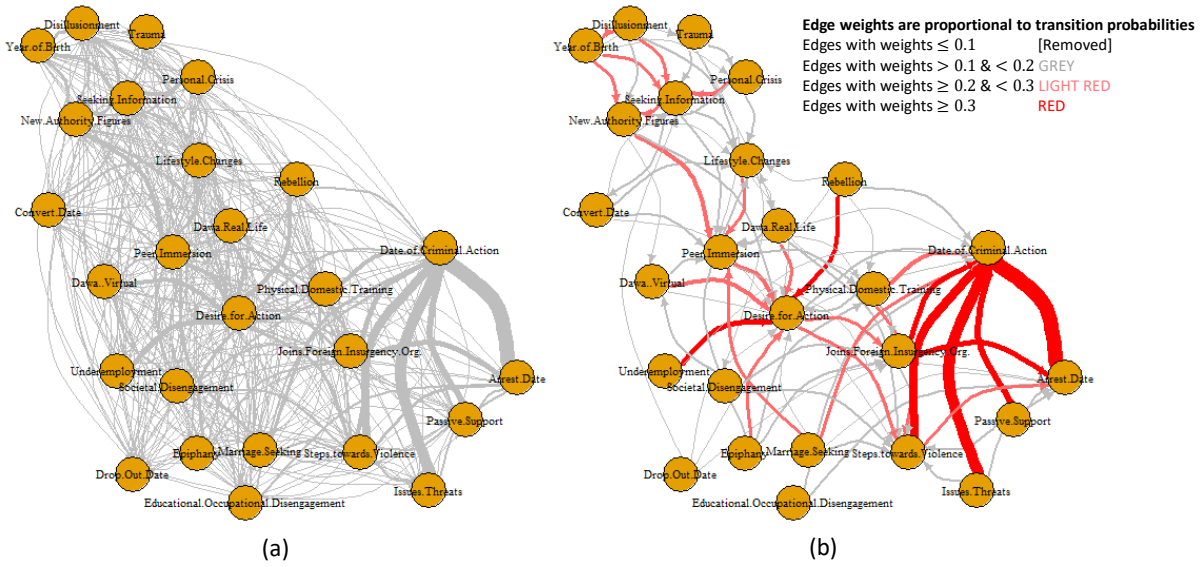


FIGURE 8.2. Radicalization transition diagrams. Nodes are features/behaviors and paths represent instances when features/behaviors sequentially followed one another. (a) Contains all instances of paths from one feature/behavior to another. Edge weights are also proportional to the transition probabilities in $\mathbf{P}''$ for that edge. (b) Filtered and color-coded version of (a). Direction arrows indicate the pairwise sequence of behaviors.

However, we can nevertheless derive some meaningful insights from this analysis. By removing all transitions that comprise less than 10% from each behavior, and color coding the remaining transitions according to their weight, we are able to identify the *common* or *frequented* sequences of behavioral indicators. See Fig. 8.2b. It is important to note that 'Arrest.Date' is considered an absorbing state because nearly all the violent extremists examined in the Klausen dataset [168] ends with an arrest.[43]

In Fig. 8.3, we provide the weights of the light red and red edges in the Fig. 8.2b, which correspond to the entries $p''_{ij} \geq 20\%$ in the radicalization behavior transition matrix $\mathbf{P}''$. We note the high proportion transitions as particularly informative of *what behaviors may immediately come next* for those confirmed cases of violent extremism. For example, out of the 135 perpetrators, 73.3% (11 out of 15) who had issued threats subsequently followed

[43]Only a few permanently relocate abroad or die to a terrorist-related activity.

Start Node	End Node	Proportion of this transition among all transition occurrences from start node
Date.of.Criminal.Action	Arrest.Date	94.5%
Issues.Threats	Date.of.Criminal.Action	73.3%
Steps.towards.Violence	Date.of.Criminal.Action	53.5%
Passive.Support	Date.of.Criminal.Action	40.9%
Joins.Foreign.Insurgency.Org.	Date.of.Criminal.Action	35.8%
Joins.Foreign.Insurgency.Org.	Arrest.Date	34.0%
Underemployment	Desire.for.Action	33.3%
Rebellion	Desire.for.Action	31.3%
Personal.Crisis	Seeking.Information	29.3%
Peer.Immersion	Desire.for.Action	27.8%
Dawa..Virtual	Desire.for.Action	27.5%
Physical.Domestic.Training	Date.of.Criminal.Action	27.5%
New.Authority.Figures	Peer.Immersion	27.4%
Disillusionment	Seeking.Information	25.4%
Seeking.Information	New.Authority.Figures	25.0%
Desire.for.Action	Steps.towards.Violence	24.2%
Marriage.Seeking	Peer.Immersion	24.1%
Physical.Domestic.Training	Steps.towards.Violence	23.5%
Year.of.Birth	Disillusionment	23.0%
Dawa.Real.Life	Desire.for.Action	22.2%
Desire.for.Action	Joins.Foreign.Insurgency.Org.	21.9%
Year.of.Birth	Seeking.Information	21.5%
Epiphany	Desire.for.Action	21.4%
Steps.towards.Violence	Arrest.Date	20.9%
Marriage.Seeking	Date.of.Criminal.Action	20.7%
Lifestyle.Changes	Peer.Immersion	20.5%
Year.of.Birth	New.Authority.Figures	20.0%

FIGURE 8.3. Edge weights for the modified radicalization behavior transition matrix $\mathbf{P}''$ for edges where $p''_{ij} \geq 20\%$.

with some form of criminal action. Other notable transitions that would be informative to intelligence and law enforcement analysts include:

- 53.5% (46 out of 86) who took discernible steps towards violence subsequently followed with some form of criminal action.
- 27.8% (30 out of 108) who immersed themselves with like-minded peers subsequently communicated some desire for violent action.
- 27.5% (14 out of 51) who undertook *dawa* (proselytized) online subsequently communicated some desire for violent action.
- 24.2% (31 out of 128) who communicated some desire for action subsequently took steps towards violence.

There are several important usability considerations to discuss. First, our dynamic insights into the radicalization process go beyond the prevalence statistics that have been traditionally proffered by researchers (see for example [116, 117, 220]. An example would be that 86 out of the sample of 135 perpetrators (63.7%) exhibited some discernible steps towards violence sometime along their path towards violent radicalization. The additional analysis that our methodology allows answers the question, "What may come next?" In this case, empirical evidence shows that the very next step for around 53% of these individuals would be some form of criminal action. In future work, when these positive cases of violent extremism are examined along with a complementary analysis of behaviors of non-violent radicals (see for example, [11, 12]), we can then better discern if these dynamic radicalization indicators do in fact signal a possible *approach* towards violence.

Knowledge about these transitions would then subsequently be incorporated in a system to detect radicalization trajectories through two possible methods. First, each of the paired-state transitions could be included into a new query pattern as their own uniquely weighted indicator and a Red Flag (RF) investigative node type designation. Second, the insights from these state transitions could be incorporated into a conditional scoring scheme to highlight the increased significance of a transition that commonly occurred among positive cases of extremist violence. We intend to develop these two alternative methods in future versions of INSiGHT.

This is first-pass analysis, but much more could be done. For instance, a simultaneous strength and shortfall of determining the proportions of behaviors that immediately follow others is that we do not provide second, third, and subsequent chains of implications for exhibiting a specific behavior.

8.3.2. Analyses of the individual radicalization pathways. We now delve into a more micro-level analysis of each individual's radicalization process. By extending the state transition model, we can view the radicalization process of each individual as a 'path' of discrete steps from birth to arrest or death in the commission of an act of extremist violence and discernible behavioral indicators along the way. We recall that that certain 'path segments' may lead to combined states of more than one behavioral indicator (due to one or more indicators occurring in fact on the same day, or just due to the granularity of the data source or reporting). In the previous section, we also developed a unique transition matrix $\mathbf{P}''$ to record the relative frequency of a subsequent behavioral indicator, whether they occur singularly at that time step or concurrently with other indicators. By normalizing the number of transitions involved from each indicator, we then determined the proportion of specific transitions among all transition occurrence from that indicator.

The histogram of the non-zero transition probabilities in $\mathbf{P}''$ in Fig. 8.4. Importantly, this shows us that over 77% (301 out of 388) of the non-zero transitions have less than 0.10 probability. There are 61 transitions with probabilities between 0.10 and 0.20, and 26 transitions with probabilities 0.20 or greater.

The prevalence of low probability transitions supports the conclusion of the multitude of radicalization paths towards extremist violence and that no single pathway exists [133, 165, 167, 224, 241]. While true empirically, researchers heretofore have failed to make any more specific, quantifiable characterizations of individual-level commonalities along the entire radicalization path. At the individual level, the most rigorous research is from Gill [116–118], but it still did not include an analysis of each individual's set of characteristics or indicative behaviors and comparing it with another individuals' set. Our interest here is to characterize

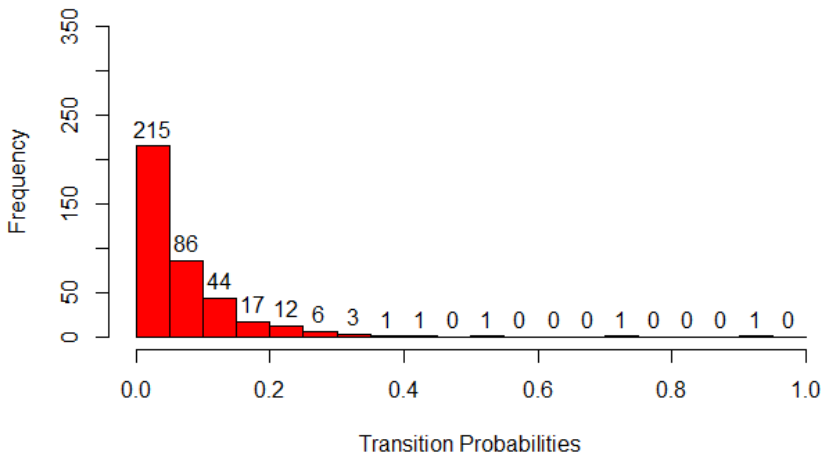


FIGURE 8.4. Distribution of the non-zero transition probabilities (n = 388 unique non-zero transitions in $\mathbf{P}''$).

each individual's radicalization pathway utilizing these frequency-based proportions and gain insights into the commonalities among any portions of the pathway.

We began by first visualizing the distribution of transition probabilities from each individual along radicalization paths towards extremist violence. See Fig. 8.5. Note that transitions from 'Date of Criminal Action' to 'Date of Arrest' because it occurred over 94% of the time and are more reflective of law enforcement response than an individual's radicalization path. A visual inspection of the lengths of each boxplot reveals that while many individuals exhibited both low and high frequency behavior transitions.

To better characterize the occurrence of frequented behavioral transitions, we introduce the notion of a normalized path probability in Definition 11.

DEFINITION 11. Normalized Path Probability

Given a state transition matrix $\boldsymbol{\mathcal{P}}$ where each entry p_{s_1,s_2} is the probability of transition from states s_1 and s_2, and a path π of k states $\{s_1, s_2, ..., s_k\}$, then normalized path probability of π is the $(\prod_{i=1}^{k} p_{s_i,s_{i+1}})^{1/k}$.

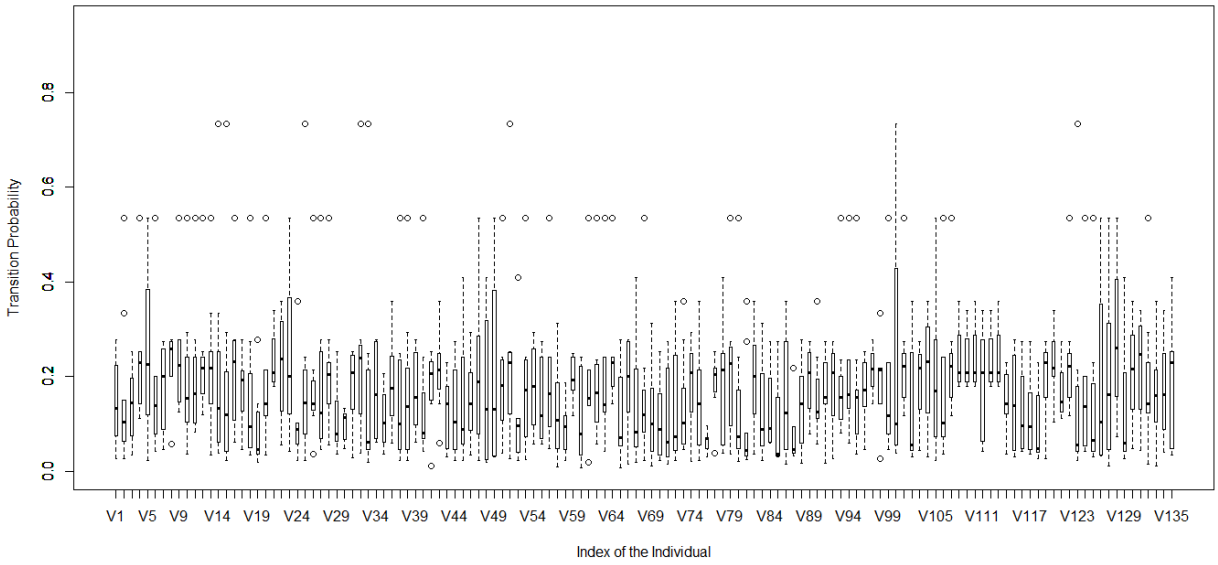


FIGURE 8.5. Box plot of each individual's set of transition probabilities along the radicalization path (n = 135 individuals).

Performing the calculation is an extension of the classical method calculating probabilities of independent events using tree diagrams and conditional probabilities. Specifically, in that procedure, one would multiply each of the probabilities of each path segment along an entire path of a tree, where each level represents another independent event. Normalization of a path with k segments multiplied together involves taking the k-th root of the product in order to quantify the 'mean' probability for each segment.

In the case of when the state space when transitions from a single indicator to a set of indicators occurring concurrently, a meaningful greedy heuristic to determine a normalized path probability is to utilize the largest transition probability among the indicators. We call the result of this heuristic the *maximum* normalized path probability. When a single indicator transitions to more than one indicator each with their associated transition proportions, the largest proportion is used as the multiplicative factor in the path probability. Based upon the data model where every path first involves a birth state leading to another single indicator state or a multiple indicator state, we know that the indicator resulting in the maximum

transition probability is always selected before there is a transition from a combined indicator state to a single indicator state. Thus for such transitions, we simply utilize the transition probability from the max achieving start indicator to the single end indicator. Note, we could have re-calculated which start indicator and end indicator had the largest transition probability at each occurrence, which would have resulted in maximum normalized path probabilities greater than or equal to the ones that resulted from our heuristic.

However, we also need to consider the maximum normalized path probabilities for each of the 135 perpetrators. The histogram of these path probabilities is shown in Fig. 8.6a. Like earlier, we disregarded any of the transitions from 'Date of Criminal Action to Date of Arrest' because it occurred over 94% of the time and would have easily skewed the normalized path probabilities higher without providing much insight on frequented indicators. We notice that only 11% (15 out of 135) of the individuals had 'mean' path probabilities less than 0.10, which could only be achieved if predominantly all path segments were among the rare transitions. For all the others (89%, 120 out of 135), to achieve maximum normalized path probabilities greater than 0.10, there must have been at least several frequented transition path segments whose probabilities were greater than 0.10.

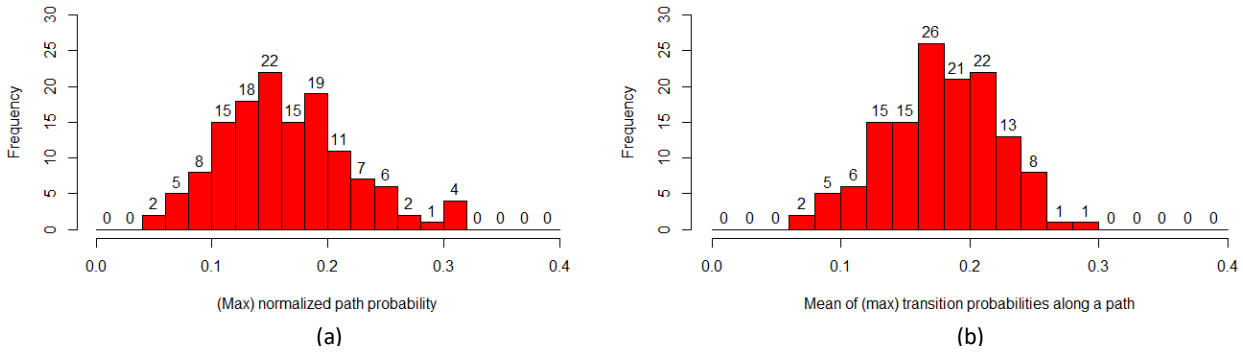


FIGURE 8.6. (a) Distribution of the max normalized path probabilities for each individual (n=135) (b) Distribution of the mean of the max path probabilities for each individual (n = 135).

For completeness, we also calculated the mean transition probability among each individual's radicalization pathway, which we calculated by averaging all the transition probabilities along an individual's path. The distribution of these means is shown in Fig. 8.6b.

Lastly, we determined the maximum transition probability utilized in each pathway. It turns out that among 135 perpetrators, only 1 individual had a pathway that utilized only behavioral transitions with probability less than 0.10, and 3 individuals had only utilized transitions with probability less than 0.20. These individuals' sequence of behaviors could truly be seen as unique and without any of the commonalities with other violent extremists. Fig. 8.7 shows that many others had utilized at least one path segment that was more frequented.

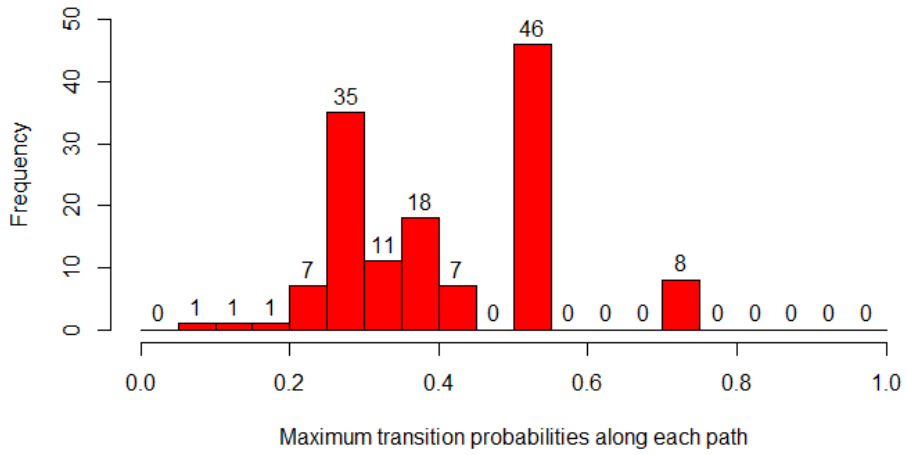


FIGURE 8.7. Distribution of the maximum transition probability in each individual's radicalization path (n = 135).

8.4. CONCLUSION

This chapter presented an approach for modeling radicalization as a discrete dynamical process of individuals exhibiting indicators on pathways towards extremist violence. Significantly, we determined from real-data the presence of highly frequented indicator transitions,

which could then subsequently be incorporated in a system to detect radicalization trajectories.

Additionally, our analysis showed that while indeed perpetrators took widely various paths in total, an overwhelming majority of them followed at least some highly common *segments* of paths. In other words, the radicalization pathway for each of the perpetrators exhibited infrequent pair-wise sequences of behaviors, but almost invariably included a few highly frequent pair-wise sequences that could prove useful for law enforcement and intelligence analysts. In future work, we also intend to perform association (frequent itemset) analysis to find sets of indicators that occur frequently in the data [2]. Utilizing a technique traditionally for determining purchase patterns in market baskets, we propose casting a person's pathway as a single basket and the indicators he/she exhibits as the purchases. This would then allow us to determine those sets of indicators which occur frequently in radicalization paths, and which indicators strongly suggest other indicators that may follow. Such information would be useful to law enforcement analysts to better distinguish those may truly be radicalizing.

CHAPTER 9

Conclusions and Future Work

In this work, we undertook a systems approach to evaluating both the need and feasibility of a radicalization detection system. We introduced an overarching analyst-in-the-loop framework and specific technology called INSiGHT to assist law enforcement and intelligence agencies in mining and screening for the individuals with a risk of extremist violence. Our vectorized, dynamic graph pattern matching approach, provided analysts with the ability to find full or partial matches against a query pattern as well as a means to quantify the pace of the appearance of the indicators. Tracking partial match trajectories provides another dimension of analysis in investigative graph searches to highlight entities on a pathway towards a pattern for a latent behavior such as violent radicalization.

We demonstrated the performance of INSiGHT on small, synthetic radicalization data sets, the real Klausen radicalization behavioral dataset, and a large, real-world BlogCatalog dataset serving as a proxy for the type of intelligence or law enforcement data networks that could be utilized to track the radicalization of violent extremists. We successfully validated the matching mechanics using the small synthetic datasets and identified all the BlogCatalog accounts that match the query in full or in part, as well as produced the match trajectory for each account. INSiGHT also ably determined the radicalization pattern match trajectory of all 135 U.S. violent extremists in the real Klausen time-stamped behavioral dataset. We noted a wide distribution of similarity scores on even just these positive cases of extremists and the difficulty of determining a suitable threshold to distinguish positive and negative cases. Importantly, however, we demonstrated how the inclusion of red flag visualizations

and alerts could greatly assist analysts in identifying high risk individuals even when they had relatively lower similarity scores.

We also extended INSiGHT by developing a non-combinatorial neighbor matching technique as an important first step in enabling analysts to maintain visibility of potential collective threats and conspiracies and account for the role close social ties have in an individual's radicalization. This enhancement was validated first on two small, synthetic radicalization-specific datasets, where we successfully detected over time those who may individually or collectively be on a trajectory towards violent extremism based upon a query pattern. Neighbor matching was also validated on the BlogCatalog dataset through the use of the real social network connections and tagging behaviors for over 80K accounts. The results showed that our algorithm returned whole and partial subgraph matches that enabled the analyst to gain and maintain visibility on each neighbors' tags. Our results also demonstrated the importance of α parameter in controlling the extent of this visibility and impact on match scores for neighbor activities. These tests on a benign application help confirm that our approach can work on radicalization-specific data if real social ties between persons of interest were available and when the detection of conspiratorial clusters is important.

Overall, INSiGHT led to consistent, informed, and reliable assessments about those who pose a significant risk for violent extremism in a variety of settings. Based on these results, we maintain that it is a feasible and useful supporting tool with the potential to optimize law enforcement investigative efforts and ultimately to enable the prevention of individuals from carrying out extremist violence.

The applicability of INSiGHT to detect latent behaviors in other domains such as on-line student assessment and consumer analytics was also demonstrated through experiments with

real data. For on-line student assessment, we tested INSiGHT on a portion of the Knowledge Discovery and Data Mining (KDD) Cup 2015 competition dataset [160] of approximately 1K students and 19K on-line activities to predict those students who persisted in the MOOC course. INSiGHT readily found the whole or partial matches of each student to the query of course materials over time. While it was clear that the preponderance of the students who ultimately 'continued' in the course had some of the highest similarity scores, there was also a significant number with high scores who did not 'continue.' In the end, we found that including just two additional features of 'start day of activity' and 'last day of activity' enabled us to achieve an AUC of 77% with 5-fold cross validation.

Using a real, large proprietary consumer activities dataset from a home improvement retailer with 60K customers and over 11 million time-stamped transactions, INSiGHT was indeed useful in the detection of customers likely undertaking certain home improvement projects based upon the number of project items purchased. However, we were unable to truly validate the utility of INSiGHT to screen for true positives because the data did not contain ground truth. Through a secondary analysis for the features which best predict a customer's final similarity score, we found that models using activity sets of three or more produced RMSE values that were less than the score increment for a single query item. In the end, the challenges with evaluating performance on this real dataset without ground truth motivated us to generate our own synthetic datasets with ground truth.

We, therefore, developed a synthetic data generator of large population, time-stamped, individual-level consumer activities data consistent with an a priori project set designation (latent behavior). We confirmed that the relatively high AUCs (93-96%) that INSiGHT can achieve at various thresholds was largely determined by the overlap of the query indicators

with other latent behaviors of interest, as well as the prevalence of the indicators that can occur by chance or with intention among a universal set of indicative and benign (unrelated) behaviors. These results also spurred us to explore the improved ability at prediction if one integrates investigative graph search with machine learning. While prediction is not necessarily the goal in the radicalization application, it may very well be in other commercial applications. We proposed and tested a modification to INSiGHT that utilized graph pattern matching to find statistically significant indicators in a large heterogeneous graph database and then simultaneously produced a classification score and match similarity score. Since predictive performance had markedly improved, we consider this integrated approach an important extension to formalize and test more thoroughly in future work. This formulation of the synthetic data generator also sets the stage for future work in developing an analogous synthetic data generator for radicalization indicators to serve as a testbed for INSiGHT and other data mining algorithms.

We also intend to press forward in several areas of future work which we describe below.

9.1. Further Enhancements to Radicalization Detection

Besides affirmative indicators of radicalization in the literature, there is some discussion of "mitigating" factors that would dissuade individuals to continue along a radicalization process [302, p. 19]. Pressman, who developed VERA2 calls these indicator items "protective." We envision future work where these key mitigating/protective factors are either added to the original query pattern with negative weights (thus lessening an individual's risk score), or are put into a different query pattern altogether and tracked as a different behavior that is then reconciled with the radicalization scoring.

The discussion may be benefited by a more concrete example. While recent violent extremists motivated by Salafi-jihadism have been shown to predominantly be married with with children (when information was available) [16, 134],[44] this characteristic does not make for a very discerning or informative indicator. However, an example of a discerning, related mitigating indicator is whether a spouse or significant other was supportive of non-violence [249]. This is ably representable in a graph database.

9.2. Incremental Graph Pattern Matching and Distributed Approaches

We intend to devise an incremental graph pattern matching approach like those found in [49, 101] in order to dynamically and efficiently update the multi-hop class similarity scores for new data in the form of additional or deleted edges or nodes. We also plan to deploy INSiGHT on distributed systems in order to handle streaming Big Data and capitalize on multi-core processing.

9.3. Expanding the Scope to Other Forms of Targeted Violence

We also demonstrated that INSiGHT can be applied to other domains, to include the prediction of MOOC persistence and whether a customer is undertaking a home improvement project requiring the purchase of part or all of a set of related items. We are interested in expanding the use of INSiGHT on targeted violence in general and seeing if the early warning behaviors in those cases can also be tracked and the threat detected. This includes the study of violent extremism motivated by other ideologies besides Salafi-jihadism, as well as other targeted violence such as mass murders [203], school shootings [275] .

[44]We recall at least three high profile cases of homegrown violent extremism in which the perpetrators were married with a child (e.g., Tamerlan Tsarnaev, Sayed Farook, and Omar Mateen).

9.4. Suicide Risk Assessment in the Military

There is potential for our work to be applicable to the monitoring the suicide risk among current military members and veterans in the U.S. For several years, U.S. military veterans have been experiencing higher risk for suicide; the research now suggests that at as many 20 veterans are committing suicide per day [80]. Suicide also continues to be a significant problem in the active military. According to the latest report by the Defense Suicide Prevention Office (DPSO), there had been 1,392 suicides within the active component and 1,009 within the reserve component [110].

The problem area is also large, but analytic efforts are greatly benefited here primarily because of the public awareness of and general concern about veteran suicide (which is much more prevalent than deaths from violent extremist), as well as the availability of data because both current and former military members receiving health care through the VA are part of the same integrated medical system.[45]

The U.S. Army has a large research effort called the Army Study to Assess Risk and Reslience in Servicemembers Longitudinal Study (STARR-LS)[46], with a multitude of interrelated data collection and analysis studies. We are currently seeking collaborative opportunities with STARR-LS researchers and explore the utility of our work on the Soldier Health Outcomes Study A and B datasets (for suicide attempts and suicides, respectively). Additionally, the data from longitudinal studies are also very promising for our work, especially considering if any of the survey respondents ended up attempting or committing suicide.

[45]Recently in April 2017, the Department of Veterans Affairs (VA) launched a innovative predictive analytics tool nation-wide called Recovery Engagement and Coordination for Health Veterans Enhanced Treatment (REACH-VET) [80]. The details are currently unavailable, but the public announcement states that the statistical tool identities veterans at risk for suicide, hospitalization, and illness based on already-available medical records.

[46]`http://starrs-ls.org/`

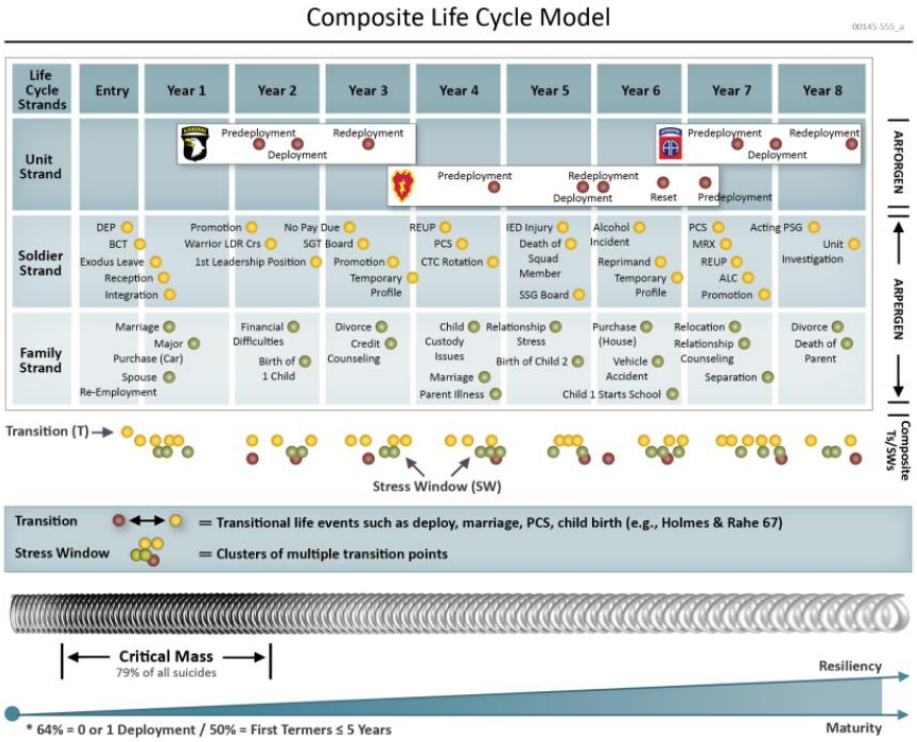


FIGURE 9.1. Army Composite Life Cycle Model. Source: [139, p. 36].

We envision INSiGHT as a supplementary system to current efforts, with a focus on current service members and the integration of graph pattern matching to identify latent behavioral indicators and predictive machine learning models by utilizing on- and off-line longitudinal activities data.

The services have seemed to adopt a "whole of life" approach to understanding the risk factors and develop prevention strategies [110]. The U.S. Army, in particular, utilizes the Composite Life Cycle Model shown in Fig. 9.1 as a means to understand the stressors that appear a service member's individual life, professional life, and the life of his or her family.[47]

[47] "The Composite Life Cycle Model...was designed to provide an aggregate view of the unique 'transitions' that occur in each of the three separate military life cycle strands of Unit, Soldier and Family...The model provides two ways to view the impact of the innumerable transitions and subsequent stressors impacting Soldiers and Families: (1) horizontally across time within a particular strand, and (2) vertically across all three life cycle strands at a particular point in time. The first view illustrates the potential acute and recurring stressors associated within each strand, while the second illustrates the potential for cumulative stressors from all three strands" [139, p. 36-37].

Ultimately, our future research relates to exploring the utility of operationalizing the Armys Composite Life Cycle Model [139, p. 36] for each Soldier to dynamically identify for commanders those most at risk for suicide given the totality of his/her indicator history and current conditions. In addition to tracking unit, soldier, and family strands of indicators/stressors, one could add any diagnosed pre-enlistment mental disorders and possibly even social media signature data over time. Any additional thoughts or suggestions towards this end would be greatly appreciated.

9.5. Cybersecurity Insider Threat Detection

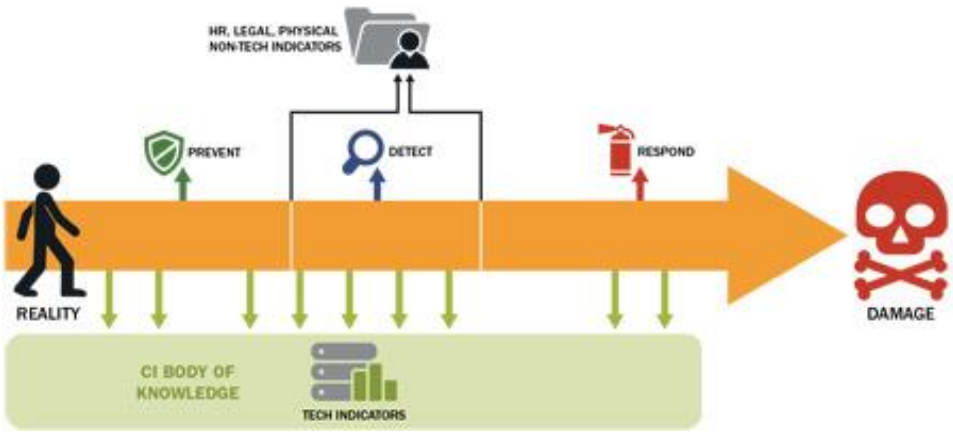


Figure 9.2. CERT insider threat graphic. Source: [45].

Lastly, we see the potential applicability of INSiGHT on detecting insider threat for cybersecurity. Current research suggests that threat actors may exhibit behavioral cues (both physical and psychological) in addition to electronic signatures in the course of their use of workplace computer systems and networks [45, 120]. Fig. 9.2 is a conceptualization from the Computer Emergency Response Team (CERT) division in the Software Engineering Institute (SEI) on how these behavioral cues and technical indicators manifest and how

organizations can deal with them at the appropriate time. According to the Department of Homeland Security, the behavioral indicators of malicious threat activity includes:

(1) "Remotely accesses the network while on vacation, sick or at odd times.

(2) Works odd hours without authorization.

(3) Notable enthusiasm for overtime, weekend or unusual work schedules.

(4) Unnecessarily copies material, especially if it is proprietary or classified.

(5) Interest in matters outside of the scope of their duties.

(6) Signs of vulnerability, such as drug or alcohol abuse, financial difficulties, gambling, illegal activities, poor mental health or hostile behavior, should trigger concern. Be on the lookout for warning signs among employees such as the acquisition of unexpected wealth, unusual foreign travel, irregular work hours or unexpected absences" [73].

In proposed future work, we intend to explore the utility of INSiGHT in detecting insider threats through the testing various threat behavioral patterns on large heterogeneous dataset of an organization's computer network activity augmented with person-centric, performance-related information that may be available in human resource departments or supervisors.

BIBLIOGRAPHY

[1] A. Abad-Santos, "'I Will Die Young': The Eerie Subtext of Dzhokhar Tsarnaev on Social Media," *The Wire*, April 19, 2013, accessed March 24, 2014 at `http://www.thewire.com/national/2013/04/dzhokhar-tsarnaev-social-media-accounts/64400/`.

[2] R. Agrawal, and R. Srikant, "Fast Algorithms for Mining Association Rules," *Proceedings of the 20th Very Large Databases (VLDB) Conference*, Santiago, Chile, 1994.

[3] Arizona Daily Star Staff Reporters, "Newly released FBI files on the Jared Lee Loughner investigation," *Arizona Daily Star*, January 8, 2011, available at `http://tucson.com/news/local/crime/newly-released-fbi-files-on-the-jared-lee-loughner-investigation/collection_6b69747c-c0e9-11e3-8ff1-0019bb2963f4.html`.

[4] S. Asur, and B. Huberman, "Predicting the future with social media." *Web Intelligence and Intelligent Agent Technology (WI-IAT)*, 2010 IEEE/WIC/ACM International Conference on. Vol. 1. IEEE, 2010.

[5] S. Atran, "Pathways to and from Violent Extremism: The Case for Science-Based Field Research," Statement before the Senate Armed Services Subcommittee on Emerging Threats and Capabilities, March 10, 2010.

[6] B. Ball, A. Brooks, A.Langville, "The Nonnegative Matrix Factorization: A Tutorial," National Institute of Statistical Sciences (NISS) NMF Workshop, North Carolina, 2007.

[7] H. Bandara and A. Jayasuamana, "On Characteristics and Modeling of P2P Resources with Correlated Static and Dynamic Attributes," *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM)*, 2011.

[8] P.G. de Barba, G.E. Kennedy, and M.D. Ainley, "The role of students' motivation and participation in predicting performance in a MOOC," *Journal of Computer Assisted Learning*, Vol. 32, p. 218-231, 2016.

[9] M. Bar-Hillel, "The Base-Rate Fallacy in Probability Judgments," *Acta Psychologica*, Vol. 44, p. 211-233, 1980.

[10] B. Barnes,"Confronting the One-Man Wolf Pack: Adapting Law Enforcement and Prosecution Responses to the Threat of Lone Wolf Terrorism," *Boston University Law Review*, Vol. 92, p. 1613-1662, 2012.

[11] J. Bartlett, J. Birdwell, and M. King, *The Edge of Violence: A Radical Approach to Extremism*, London, UK: Demos, 2012.

[12] J. Bartlett and C. Miller, "The Edge of Violence: Towards Telling the Difference Between Violent and Non-Violent Radicalization," *Terrorism and Political Violence*, Vol. 24, p. 1-21, 2012.

[13] S. Basu-Roy, T. Eliassi-Rad, and S. Papadimitriou, "Fast and Effective Pattern Matching on Weighted Attributed Graphs," *ACM Knowledge Discovery and Data Mining*, 2013.

[14] BBC Staff Editor, "FBI 'foils IS-inspired plot to attack US Capitol,'" *BBC*, January 15, 2015, available at `http://www.bbc.com/news/world-us-canada-30824375`.

[15] R. Bender, "Germany Turns to Technology to Assess Threat from Radicals," *The Wall Street Journal*, February 2, 2017, accessed March 8, 2017, available at `https://www.wsj.com/articles/germany-looks-to-tighten-security-net-with-antiterror-tool-1486042836`.

[16] P. Bergen, "Can We Stop Homegrown Terrorists?" *The Wall Stree Journal*, January 22, 2016, available at `http://www.wsj.com/articles/canwestophomegrownterrorists1453491850`, accessed February 4, 2016.

[17] P. Bergen, A. Ford, A. Sims, and D. Sterman, "New America: Terrorism in America After 9/11 Dataset," accessed March 31, 2017, available at `https://www.newamerica.org/in-depth/terrorism-in-america/who-are-terrorists/`.

[18] J.M. Berger, "Visualizing CVE Audiences," published March 3, 2012, available at `http://news.intelwire.com/2012/03/visualizing-cve-audiences.html`, accessed March 30, 2017.

[19] J.M. Berger and J. Morgan, "The ISIS Twitter Census," Brookings Institution, Washington, DC., 2015.

[20] M. Berry, M. Browne, A. Langville, P. Pauca, R. Plemmons, "Algorithms and applications for approximate nonnegative matrix factorization," Computational Statistics and Data Analysis, Vol. 52, pp. 155-173, 2007.

[21] M. Berry and M. Browne, "Email Surveillance Using Non-negative Matrix Factorization," *Computational and Mathematical Organization Theory*, Vol. 11, pp.249-264, 2005.

[22] A. Berzon, J. Emshwiller, and R. Guth, "Postings of a Troubled Mind," *The Wall Street Journal*, January 12, 2011, available at `https://www.wsj.com/articles/SB10001424052748703791904576075851892478080`.

[23] B. Bhargavi and K.P. Supreethi, "Graph pattern mining: A survey of issues and approaches," *International Journal of Information Technology and Knowledge Management*, Vol 5, No. 2, July-December 2012.

[24] T. Bilazarian, "Countering Violent Extremism: Lessons on Early Intervention from the United Kingdom's Channel Program," George Washington University Program on Extremism, October 2016, available at https://cchs.gwu.edu/sites/cchs.gwu.edu/files/downloads/Channel%20UK%20Final.pdf, accessed April 4, 2017.

[25] M. Birnbaum, "France boosts counterterrorism force after deadly Paris attacks," *The Washington Post*, January 21, 2015, available at https://www.washingtonpost.com/world/europe/france-boosts-counterterrorism-force-after-deadly-paris-attacks/2015/01/21/273f7924-a15b-11e4-903f-9f2faf7cd9fe_story.html?utm_term=.683588b19b08, accessed April 4, 2017.

[26] J. Bjelopera, "American Jihadist Terrorism: Combating a Complex Threat," *Congressional Research Service*, January 23, 2013, available at https://www.fas.org/sgp/crs/terror/R41416.pdf.

[27] J. Bjelopera, "The Islamic State's Acolytes and the Challenges They Pose to U.S. Law Enforcement," *Congressional Research Service*, June 13, 2016, available at https://fas.org/sgp/crs/terror/R44110.pdf.

[28] Black and Decker, "How to Avoid Do-It-Yourself Fails," November 21, 2014, available at http://www.blackanddecker.com/ideas-and-inspiration/articles/diy-stumbling-blocks, accessed January 11, 2017.

[29] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *Journal of Computational Science* Volume 2.1, pp 1-8, 2011.

[30] R. Borum, "Radicalization into Violent Extremism I: A Review of Social Science Theories," *Journal of Strategic Security*, Vol 4, No. 4, Winter 2011.

[31] R. Borum, "Radicalization into Violent Extremism II: A Review of Conceptual Models and Empirical Research," *Journal of Strategic Security*, Vol 4, No. 4, Winter 2011.

[32] C. Brinton and M. Chiang, "MOOC Performance Prediction via Clickstream Data and Social Learning Networks," *Proceedings of the IEEE Conference on Computer Communications (INFOCOM)*, 2015.

[33] J. Brown,"Federal Bureau of Investigation Privacy Impact Assessment for the eGuardian System," January 4, 2013, available at `https://www.fbi.gov/services/records-management/foipa/privacy-impact-assessments/eguardian-threat`, accessed April 9, 2017.

[34] J. Brunty and K. Helenek, "Social Media and the Law: Legal Considerations," chapter in *Social Media Investigations for Law Enforcement*, p. 71-87, New York: Routledge, 2013.

[35] J. Brynielsson, A. Horndahl, F. Johansson, L. Kaati, C. Martenson, and P. Svenson, "Harvesting and analysis of weak signals for detecting lone wolf terrorists," *Security Informatics*, volume 2, pp 11-26, 2013.

[36] O. Bureš, "Intelligence sharing and the fight against terrorism in the EU: lessons learned from Europol," *European View*, Vol. 15, p. 57-66, 2016.

[37] F. Burton and S. Stewart, "The 'Lone Wolf' Disconnect, Stratfor Global Intelligence Security Weekly, January 30, 2008, accessed March 29, 2014, at `http://www.stratfor.com/weekly/lone_wolf_disconnect`.

[38] V. Callison-Burch, J. Guadagno, and A. Davis, "Building a Safer Community With New Suicide Prevention Tools," Facebook Newsroom, March 1, 2017, available at https://newsroom.fb.com/news/2017/03/building-a-safer-community-with-new-suicide-prevention-tools/, accessed April 10, 2017.

[39] D. Cameron, "Dozens of police-spying tools remain after Facebook, Twitter crack down on Geofeedia," *The Daily Dot*, October 11, 2016, available at https://www.dailydot.com/layer8/geofeedia-twitter-facebook-instagram-social-media-surveillance/, accessed April 3, 2017.

[40] D. Cameron, "Twitter cuts ties with second firm police use to spy on social media," *The Daily Dot*, October 20, 2016, available at https://www.dailydot.com/layer8/twitter-snaptrends-geofeedia-social-media-monitoring-facebook/, accessed April 3, 2017.

[41] D. Cameron and D. Gilmour, "Twitter cuts off third surveillance firm for encouraging police to spy on activists," *The Daily Dot*, December 9, 2016, available at https://www.dailydot.com/layer8/media-sonar-twitter-social-media-monitoring/, accessed April 3, 2017.

[42] D. Carney, E. Morris, and P. Place, "Identifying Commerical Off-the-Shelf (COTS) Product Risks: The COTS Usage Risk Evaluation," Software Engineering Institute, Carnegie Mellon University, Pittsburgh, PA, 2003.

[43] B. Carter, "Customer Loyalty Statistics: 2015 Edition," December 31, 2015, available at http://blog.accessdevelopment.com/customer-loyalty-statistics-2015-edition, accessed January 11, 2017.

[44] D. Carter, S. Chermak, J. Carter, and J. Drew, "Understanding Law Enforcement Intelligence Processes," Report to the Office of University Programs, Science and Technology Directorate, U.S. Department of Homeland Security, College Park, MD: START, 2014.

[45] CERT Division, Software Engineering Institute, "Insider Threat Best Practices," webpage, available at `http://www.cert.org/insider-threat/best-practices/index.cfm`, accessed on March 29, 2017.

[46] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly Detection: A Survey," *ACM Computing Surveys*, Vol. 41(3), 2009.

[47] L. Chen and C. Wang, "Continuous Subgraph Pattern Search over Certain and Uncertain Graph Streams," *IEEE Transactions on Knowledge and Data Engineering*, Vol 22, No. 8, August 2010.

[48] R. Cheng, J. Hong, A. Kyrola, Y. Miao, X. Weng, M. Wu, F. Yang, L. Zhou, F. Zhao, and E. Chen, "Kineograph: Taking the Pulse of a Fast-Changing and Connected World," *Proceedings of the 7th ACM European Conference on Computer Systems* 2012.

[49] S. Choudhury, L. Holder, J. Feo, and G. Chin, "Fast Search for Dynamic Multi-Relational Graphs," *DyNetMM 2013*, Association for Computing Machinery, June 2013.

[50] T. Cleary, "Garland Shooting: Tweets From Terrorists You Need to See," *Heavy*, May 4, 2015, available at `http://heavy.com/news/2015/05/garland-texas-shooting-terror-attack-elton-simpson-junaid-hussain-isis-tweets-twitter-muhammad-art-contest-exhibit-suspects-claim-responsibility-social-media/10/`.

[51] K. Cohen, F. Johansson, L. Kaati, and J. Mork, "Detecting Linguistic Markers for Radical Violence in Social Media," *Terrorism and Political Violence*, 26: 246–256, 2014.

[52] K. Cohen, F. Johansson, L. Kaati, and J. Mork, "Detecting Linguistic Markers for Radical Violence in Social Media," *Terrorism and Political Violence*, volume 26, pp. 246-256, 2014.

[53] R. Cohn and A. Liao, "Mapping Reveals Rising Use of Social Media Monitoring Tools by Cities Nationwide," Brennan Center for Justice, Internet: `https://www.brennancenter.org/blog/mapping-reveals-rising-use-social-media-monitoring-tools-cities-nationwide`, November 16, 2016, accessed April 1, 2017.

[54] J. Cole, E. Alison, B. Cole, and L. Alison, "Guidance for Identifying People Vulnerable to Recruitment into Violent Extremism." Liverpool, UK: University of Liverpool, School of Psychology, 2010.

[55] J.P. Cole, "Terrorist Databases and the No Fly List: Procedural Due Process and Hurdles to Litigation," *Congressional Research Service*, April 2, 2015, available at `https://www.fas.org/sgp/crs/homesec/R43730.pdf`.

[56] J. Comey, "Update on Orlando Terrorism Investigation," Press Briefing on Orlando Mass Shooting, FBI Headquarters, Washington, D.C., June 13, 2016, available at `https://www.fbi.gov/news/speeches/update-on-orlando-terrorism-investigation`.

[57] J. Comey, "Fifteen Years After 9/11: Threats to the Homeland," Statement Before the Senate Committee on Homeland Security and Governmental Affairs, Washington, D.C., September 27, 2016, available at `https://www.fbi.gov/news/testimony/fifteen-years-after-911-threats-to-the-homeland`.

[58] J. Comey, "Expectations of Privacy: Balancing Liberty, Security, and Public Safety," Remarks delivered at the Center for the Study of American

Democracy Biennial Conference, Kenyon College, Gambier, Ohio, April 6, 2016, available at https://www.fbi.gov/news/speeches/expectations-of-privacy-balancing-liberty-security-and-public-safety.

[59] J. Condliffe, "Facebook Won't Let Insurer Probe Your Profile," *MIT Technology Review*, November 2, 2016, available at https://www.technologyreview.com/s/602772/facebook-wont-let-insurers-probe-your-profile/?utm_campaign=internal&utm_medium=homepage&utm_source=top-stories_2&set=602774, accessed April 3, 2017.

[60] J. Condliffe, "Facebook Forbids the Use of User Data for Surveillance," *MIT Technology Review*, March 14, 2017, available at https://www.technologyreview.com/s/603855/facebook-forbids-the-use-of-user-data-for-surveillance/?set=603859, accessed April 3, 2017.

[61] J. Condliffe, "How Machine Learning May Help Tackle Depression," *MIT Technology Review*, April 5, 2017, available at https://www.technologyreview.com/s/604075/how-machine-learning-may-help-tackle-depression/?set=604089, accessed April 10, 2017.

[62] Cool Springs Press, *Black and Decker: The Complete Guide to Bathrooms*, Minneapolis Minnesota: Creative Publishing International, 2014

[63] R. Covey, "Pervasive Surveillance and the Future of the Fourth Amendment," *Mississippi Law Journal*, Vol. 80(4), p. 1289-1318, 2012.

[64] Criminal Court of the City of New York, "The People of the State of New York against Jose Pimentel, AKA Muhammad Yusuf," (Criminal Complaint), November 19, 2011, available at http://www.nyc.gov/html/om/pdf/2011/jose_pimentel_complaint.pdf.

[65] J. Currall and P. McKinney, "Investing in Value: A Perspective on Digital Preservation," *D-Lib Magazine*, Vol. 12, No. 4, April 2006.

[66] P. Davis, W. Perry, R. Brown, D.Yeung, P. Roshan, and P. Voorhies, "Using Behavioral Indicators to Help Detect Potential Violent Acts: A Review of the Science Base," Santa Monica, CA: RAND Corporation, 2013.

[67] G. Dean, *Neurocognitive Risk Assessment for the Early Detection of Violent Extremist*, SpringerBriefs in Criminology-Policing, Heidelberg, Germany, Springer, 2014.

[68] G. Dean, "SAVE System" [webpage], accessed March 8, 2017, available at `http://geoffdean.com.au/training-expertise/`.

[69] G. Dean, "Structured Assessment of Violent Extremism: Introduction and Overview of the SAVE-30 System," e-email correspondence, December 2016.

[70] L. Dearden, "Khalid Masood: Suspected ISIS supporter used WhatsApp two minutes before London attack," *The Independent*, March 24, 2017, available at `http://www.independent.co.uk/news/uk/home-news/khalid-masood-whatsapp-westminster-london-attack-parliament-message-isis-terror-network-contacts-a7649206.html`, accessed March 27, 2017.

[71] M. DeChoudhury, S. Counts, E. Horvitz, "Social Media as a Measurement Tool of Depression in Populations," *ACM Web Science 2013 Conference Proceedings*, Paris, France, May 2-4, 2013.

[72] Defense Acquisition University, "Federally Funded Research and Development Centers (FFRDC)," available at `https://dap.dau.mil/acquipedia/Pages/ArticleDetails.aspx?aid=5e3079b8-44f2-43df-a0e7-9f379e8c48ed`, accessed April 29, 2017.

[73] Department of Homeland Security, National Cybersecurity and Communications Integration Center, "Combating the Insider Threat, 2 May 2014" available at https://www.us-cert.gov/sites/default/files/publications/Combating%20the%20Insider%20Threat_0.pdf, accessed March 1, 2017.

[74] Department of Homeland Security, "Department of Homeland Security Strategy for Countering Violent Extremism," October 28, 2016, available at https://www.dhs.gov/sites/default/files/publications/16_1028_S1_CVE_strategy.pdf, accessed April 15, 2017.

[75] Department of Homeland Security, Homeland Security Information Network, "Features You Need- Security You Can Trust," available at https://www.dhs.gov/sites/default/files/publications/HSIN-Fact%20Sheet-Features.pdf, accessed April 7, 2017.

[76] Department of Homeland Security, Homeland Security Science and Technology, "S&T Innovation Strategy 2017," available at https://www.dhs.gov/sites/default/files/publications/ST%20Innovation%20Strategy%202017.pdf, accessed April 30, 2017.

[77] C. Derespina, "U.S. placed Berlin terror suspect on no-fly list months ago report says," FoxNews, December 22, 2016, available at http://www.foxnews.com/world/2016/12/22/usplacedberlinterrorsuspectonnoflylistmonthsagoreportsays.print.html, accessed December 22, 2016.

[78] DIYNetwork, "Insulating Attics and Roofs," available at http://www.diynetwork.com/how-to/rooms-and-spaces/storage-space/insulating-attics-and-roofs, accessed January 3, 2017.

[79] Department of Justice, "Fact Sheet: Shifting from Prosecution to Prevention, Redesigning the Justice Department to Prevent Future Acts of Terrorism," May 29, 2002, available at https://fas.org/irp/news/2002/05/fbireorganizationfactsheet.pdf, accessed April 17, 2017.

[80] Department of Veterans Affairs, Office of Public and Intergovernmental Affairs, "VA REACH VET Initiative Helps Save Veterans Lives," April 3, 2017, available at https://www.va.gov/opa/pressrel/pressrelease.cfm?id=2878, accessed April 12, 2017.

[81] R. Depue, "Red flags, warning signs and indicators," Appendix in "Report of the Review Panel on the Mass Shootings at Virginia Tech," 2007.

[82] C. Dewey, "98 personal data points that Facebook uses to target ads to you," *The Washington Post*, August 19, 2016, available at https://www.washingtonpost.com/news/the-intersect/wp/2016/08/19/98-personal-data-points-that-facebook-uses-to-target-ads-to-you/?utm_term=.d3d3e70e5a70, accessed April 10, 2017.

[83] D. Doan, "Commercial off the Shelf (COTS) Security Issues and Approaches," Thesis, Naval Postgraduate School, Monterey, California, 2006.

[84] K. Dolan, "Boston Marathon Bomber Suspect Dzhokhar Tsarnaev's Twitter Account Shows Discontent," *Forbes*, April 19, 2013, accessed March 27, 2014 at http://www.forbes.com/sites/kerryadolan/2013/04/19/boston-marathon-bomber-suspect-dzhokar-tsarnaevs-twitter-account-shows-discontent/.

[85] P. Ducklin, "Why government plans to spy on WhatsApp will fail," March 28, 2017, available at https://nakedsecurity.sophos.com/2017/03/28/heres-why-what-the-government-wants-with-whatsapp-wont-work/, accessed March 29, 2017.

[86] R. Dzhekova, M. Mancheva, N. Stoynova, and D. Anagnostou, *Monitoring Radicalisation: A Framework for Risk Indicators*, Sofia, Bulgaria: Center for the Study of Democracy, 2017.

[87] Economist Data Team, "Terrorist atrocities in western Europe," Internet: `http://www.economist.com/blogs/graphicdetail/2017/03/terrorism-timeline`, accessed April 1, 2017.

[88] M. Eddy,J. Ewing, J. Berendt, and E. Schmitt, "Berlin Attack Sets Off Hunt for a Tunisian in Germany," *The New York Times*, December 21, 2016, available at `https://www.nytimes.com/2016/12/21/world/europe/attack-sets-off-hunt-for-tunisian-who-had-slipped-germanys-grasp.html?_r=1`, accessed April 1, 2017.

[89] D. Edelman and M. Singer, "Competing on Customer Journeys," *Harvard Business Review*, November 2015.

[90] D. Edelman and M. Singer, "The new consumer decision journey," *McKinsey Digital*, October 2015, available at `http://www.mckinsey.com/business-functions/marketing-and-sales/our-insights/the-new-consumer-decision-journey`, accessed February 4, 2016.

[91] V. Egan, J. Cole, B. Cole, L. Alison, E. Alison, S. Waring, and S. Elntib, "Can you identify violent extremists using a screening checklist and open-source intelligence alone?" forthcoming in *Journal of Threat Assessment and Management*, 2016.

[92] B. Egerton, "Imam's emails to Fort Hood suspect Hasan tame compared to online rhetoric," *The Dallas Morning News*, November 29, 2009, available at http://www.dallasnews.com/news/state/headlines/20091129-Imam-s-e-mails-to-Fort-7150.ece.

[93] J. Eisner, S. Funke, A. Herbat, A. Spillner, and S. Strorandt, "Algorithms for matching and predicting trajectories," *Proceedings of the Thirteenth Workshop on Algorithm Engineering and Experiments (ALENEX)*, 2011.

[94] P. Engel, " Here's the ISIS message the female San Bernardino shooter posted on Facebook during the attack," *Business Insider*, December 17, 2015, available at http://www.businessinsider.com/isis-message-tashfeen-malik-posted-on-facebook-during-attack-2015-12.

[95] W. Englund, "In diary, Norwegian 'crusader' details months of preparation for attacks," *The Washington Post*, July 24, 2011, available at https://www.washingtonpost.com/world/europe/in-diary-norwegian-crusader-details-months-of-preparation-for-attacks/2011/07/24/gIQACYnUXI_story.html?tid=a_inl.

[96] T. Escobedo and M. Morgenstein, "Who were suspects in Paris terror attacks?," CNN, November 17, 2015, available at http://www.cnn.com/2015/11/16/world/paris-attacks-suspects-profiles/index.html, accessed May 29, 2017.

[97] Esri, "2017 Terrorist Attacks," available at https://storymaps.esri.com/stories/terrorist-attacks/?year=2017, accessed April 4, 2017, 2017.

[98] J.B. Evans, R.B. Baker, T. Dee, "Persistence Patterns in Massive Open Online Courses (MOOCs)," CEPA Working Paper No.15-09, Stanford Center for Education Policy Analysis, available at http://cepa.stanford.edu/wp15-09, 2015.

[99] P. Fader and B. Hardie, "Probability Models for Customer-Base Analysis," *Journal of Interactive Marketing*, Vol. 25, p. 61-69, 2009.

[100] W. Fan et al, "Graph pattern matching: From intractable to polynomial time," *Proceedings of the VLDB Endowment*, Vol 3, No. 1, 2010.

[101] W. Fan et al, "Incremental graph pattern matching," *Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data*, p. 925-936, 2011.

[102] W. Fan, "Graph pattern matching revised for social network analysis," *International Conference on Database Theory (ICDT)*, 2012.

[103] W. Fan, "Diversified Top-k graph pattern matching," *Proceedings of the VLDB Endowment*, Vol 6, No. 13, 2013.

[104] FBI National Press Office, "2011 Request for Information on Tamerlan Tsarnaev from Foreign Government," FBI, Washington, DC, April 19, 2013, accessed March 28, 2014, at `http://www.fbi.gov/news/pressrel/press-releases/2011-request-for-information-on-tamerlan-tsarnaev-from-foreign-government`.

[105] FBI, "National Data Exchange (N-DEx) System," available at `https://www.fbi.gov/services/cjis/ndex`, accessed April 9, 2017.

[106] FBI National Press Office, "2011 Request for Information on Tameralan Tsarnaev from Foreign Government," Washington DC, April 19, 2013.

[107] FBI National Security Branch, "A New Approach to Countering Violent Extremism: Sharing Expertise and Empowering Local Communities," *FBI Law Enforcement Bulletin*, October 7, 2014, available at `https://leb.fbi.gov/2014/october/a-new-approach-to-countering-violent-extremism-sharing-expertise-and-empowering-local-communities`, accessed April 17, 2017.

[108] M. Fisher, "YouTube account that belongs to a person named Tamerlan Tsarnaev had bookmarked videos on terrorism," *The Washington Post*, April 19, 2013, available at http://www.washingtonpost.com/blogs/worldviews/wp/2013/04/19/youtube-account-that-belongs-to-a-person-named-tamerlan-tsarnaev-had-bookmarked-videos-on-terrorism/.

[109] R. Fisher, T. Breckon, K. Dawson-Howe, A. Fitzgibbon, C. Robertson, E. Trucco, and C. Williams, *Dictionary of Computer Vision and Image Processing*, 2nd edition, John Wiley and Sons, 2014.

[110] K. Franklin, "Department of Defense Quarterly Suicide Report, Calendar Year 2016 4th Quarter," Defense Suicide Prevention Office (DSPO), April 24, 2017, available at http://www.dspo.mil/Portals/113/Documents/DoD%20Quarterly%20Suicide%20Report%20CY2016_Q4.pdf?ver=2017-04-24-112920-020, accessed May 22, 2017.

[111] M. Fredholm, "Jihadists, Al Qaeda, and the Islamic State," in *Understanding Lone Actor Terrorism* M. Fredholm, ed., Routledge Press, New York, 2016.

[112] J. Freilich, R. Belli, and S. Chermak, "United States Extremist Crime Database (EBDB), 1990-2010," available at http://www.start.umd.edu/research-projects/united-states-extremist-crime-database-ecdb-1990-2010, accessed April 15, 2017.

[113] B. Gallagher, "Matching structure and semantics: A survey on graph-based pattern matching," *Journal of the American Association for Artificial Intelligence*, Fall Symposium Technical Report, January 2006.

[114] J. Gansler and W. Lucyshyn, "Commercial-Off-The-Shelf (COTS): Doing it Right," Center for Public Policy and Private Enterprise, University of Maryland, September 2008.

[115] D. Geer, "Why security should monitor social media to prevent violence," *CSO*, February 17, 2014, available at `http://www.csoonline.com/article/2134390/strategic-planning-erm/why-security-should-monitor-social-media-to-prevent-violence.html`.

[116] P. Gill, J. Horgan, P. Deckert, "Bombing Alone: Tracing the Motivations and Antecedent Behaviors of Lone-Actor Terrorists," *Journal of Forensic Science*, Vol. 59, No. 2, March 2014.

[117] P. Gill, *Lone-Actor Terrorists*, New York: Routledge, 2015.

[118] P. Gill and E. Corner, "Lone actor terrorist use of the Internet and behavioural correlates," in *Terrorism Online: Politics, Law and Technology*, L. Jarvis, S. MacDonald, and T. Chen, eds., New York: Routledge, 2015.

[119] N. Gillis, "The Why and How of Nonnegative Matrix Factorization," arXiv, accessed January 21, 2014.

[120] J. Glasser and B. Lindauer, "Bridging the Gap: A Pragmatic Approach to Generating Insider Threat Data," *Proceedings of the IEEE Security and Privacy Workshops*, 2013.

[121] J. Goldstein and W. Rashbaum, "City Bomb Plot Suspect is Called Fan of Qaeda Cleric, November 20, 2011, available at `http://www.nytimes.com/2011/11/21/nyregion/jose-pimentel-is-charged-in-new-york-city-bomb-plot.html?pagewanted=all&_r=0`.

[122] Government Accounting Office, "Countering Violent Extremism: Actions Needed to Define Strategy and Assess Progress of Federal Efforts," GAO-17-300: Report to Congressional Requesters, Washington, D.C., 2017.

[123] G. Graff, "The FBI's Growing Surveillance Gap," *Politico*, June 16, 2016, accessed March 7, 2017, available at `http://www.politico.com/magazine/story/2016/06/orlando-terror-fbi-surveillance-gap-213967`.

[124] Google, "Google Transparency Report: Legal Process," available at `https://www.google.com/transparencyreport/userdatarequests/legalprocess/`, accessed April 25, 2017.

[125] R. Graham, "How Terrorists Use Encryption," *Combating Terrorism Center (CTC) Sentinel*, Vol. 9, Issue 6, p. 20-25, June 2016.

[126] S. Green and K. Proctor, "Turning Point: A New Comprehensive Strategy for Countering Violent Extremism," Center for Strategic and International Studies, available at `https://www.csis.org/features/turning-point`, November 2016, accessed April 4, 2017.

[127] D. Gross, "Why did Colorado shooting suspect avoid social media?" CNN, July 23, 2012, accessed March 30, 2014 at `http://www.cnn.com/2012/07/23/tech/social-media/colorado-suspect-social-media/`.

[128] R.Hämäläinen, T. Lahtinen, "Path dependence in Operational ResearchHow the modeling process can influence the results," *Operations Research Perspectives*, Vol. 3, p. 14-20, 2016.

[129] L. Harding and V. Dodd, " Tamerlan Tsarnaev's YouTube account shows jihadist radicalisation in pictures, *The Guardian*, April 22, 2013, available at

http://www.theguardian.com/world/2013/apr/22/tamerlan-tsarnaev-youtube-jihadist-radicalisation.

[130] C. Haskins, ed., *INCOSE Systems Engineering Handbook*, Version 3, Washington, D.C., June 2006.

[131] P. Helsel, "Terror bomb plotter arrested in NYC," *New York Post*, November 20, 2011, available at http://nypost.com/2011/11/20/terror-bomb-plotter-arrested-in-nyc/.

[132] K. Hill, "The Disturbing Internet Footprint of Santa Barbara Shooter Elliot Rodger," *Forbes*, May 24, 2014, available at http://www.forbes.com/sites/kashmirhill/2014/05/24/the-disturbing-internet-footprint-of-santa-barbara-shooter-elliot-rodger/print/.

[133] J. Horgan, "From Profiles to Pathways and Roots to Routes: Perspectives from Psychology on Radicalization into Terrorism," *The Annals of the American Academy of Political and Social Science*, Vol. 618, July 2008, p. 80-94.

[134] J. Horgan, N. Shortland, S. Abbasciano, and S. Walsh, "Actions Speak Louder than Words: A Behavioral Analysis of 183 Individuals Convicted for Terrorist Offenses in the United States from 1995 to 2012," *Journal of Forensic Sciences*, Vol. 61(5), p. 1228-1237, 2016.

[135] Homeland Security Committee, U.S. House of Representatives, "Going Dark, Going Forward: A Primer on the Encryption Debate," version 2.0, September 2016.

[136] U.S. House of Representatives Permanent Select Committee on Intelligence, "Media Leaks Facts and Context (Long Version)," August 1, 2013, available at http://intelligence.house.gov/sites/intelligence.house.gov/files/documents/talkingpointslong.pdf, accessed April 7, 2017.

[137] P. Hoyer, "Non-negative Matrix Factorization with Sparseness Constraints," *Journal of Machine Learning Research*, volume 5, pp. 1457-1469, 2004.

[138] Houzz & Home, "Renovation in America: Findings from the 2013 Houzz & Home Survey," available at http://info.houzz.com/rs/houzz/images/Houzz%20%26%20Home%202013%20Report.pdf, accessed January 18, 2017.

[139] Headquarters, Department of the Army, "Army 2020: Generating Health and Discipline in the Force Report," Washington, D.C., 2012.

[140] J. Huang, K. Venkatraman, and D. Abadi, "Query Optimization of Distributed Pattern Matching," *Proceedings of the IEEE 30th International Conference on Data Engineering*, 2014.

[141] B. Hung, S. Kolitz, and A. Ozdaglar, "Optimization-Based Influencing of Village Social Networks in a Counterinsurgency," *Association of Computing Machinery Transactions on Intelligent Systems and Technology (ACM- TIST)*, Volume 4, June 2013.

[142] B. Hung and A. Jayasumana, "Investigative Simulation: Towards Utilizing Graph Pattern Matching for Investigative Search," *Proceedings of the Conference on Foundations of Open Source Intelligence and Security Informatics (FOSINT-SI)*, 2016.

[143] B. Hung, A. Jayasumana, and V. Bandara, "Detecting Radicalization Trajectories Using graph Pattern Matching Algorithms," *Proceedings of the IEEE Conference on Intelligence and Security Informatics (ISI)*, 2016.

[144] B. Hung, A. Jayasumana, and V. Bandara, "Pattern Matching Trajectories in Investigative Graph Searches," *Proceedings of the IEEE Conference Data Science and Advance Analytics (DSAA)*, 2016.

[145] B. Hung, A. Jayasumana, and V. Bandara, "INSiGHT: Detecting the Radicalization Trajectories of Homegrown Violent Extremists with Dynamic Graph Pattern Matching," *Proceedings of the IEEE Homeland Security Technologies (HST) Symposium*, 2017.

[146] K. Ilgun, R. Kemmerer, and P. Porras, "State Transition Analysis: A Rule-Based Intrusion Detection Approach," *Proceedings of the IEEE Transactions on Software Engineering*, Vol 21(3), 1995.

[147] INCOSE, "What is Systems Engineering?" Internet: `http://www.incose.org/AboutSE/WhatIsSE`, accessed on March 30, 2017.

[148] Inspectors General of the Intelligence Community, Central Intelligence Agency, Department of Justice, and the Department of Homeland Security, "Unclassified Summary of Information Handling and Sharing Prior to the April 15, 2013 Boston Marathon Bombings," April 2014, available at `https://www.dni.gov/index.php/who-we-are/organizations/ic-ig/ic-ig-news/1604`, accessed May 21, 2017.

[149] "Israel is using social media to prevent terrorist attacks," *The Economist*, April 18, 2016, available at `http://www.economist.com/news/middle-east-and-africa/21697083-new-paradigm-intelligence-israel-using-social-media-prevent-terrorist`, accessed April 3, 2017.

[150] M. Isikoff, "Unaware of Tsarnaev warnings, Boston counterterror unit tracked protestors," NBC News Investigations, May 9, 2013, accessed, March 28, 2014, at

http://investigations.nbcnews.com/_news/2013/05/09/18152849-unaware-of-tsarnaev-warnings-boston-counterterror-unit-tracked-protesters.

[151] J. Jackson and K. Kollman, "Models of path dependence with an empirical application," *Annual Political Methodology Conference*, State College, PA. 2007.

[152] J. Jashinsky, S. Burton, C. Hanson, J. West, C. Giraud-Carrier, M. Barnes, and T. Argyle, "Tracking Suicide Risk Factors Through Twitter in the US," *Crisis*, Vol. 35, p. 51-59, 2014.

[153] S. Jeon, Y. Khosiawan, and B. Hong, "Making a Graph Databse from Unstructured Text," *Proceedings of the 16th International Conference on Computational Science and Engineering*, 2013.

[154] K. Jordan, "Massive open online course completion rates revisited: Assessment, length and attrition," *International Review of Research in Open and Distributed Learning*, Vol. 16(3), p. 341358, 2015.

[155] J. Jouvenal, "The new way police are surveilling you: calculating your threat 'score,' *The Washington Post*, January 10, 2016, available online at https://www.washingtonpost.com/local/public-safety/the-new-way-police-are-surveilling-you-calculating-your-threat-score/2016/01/10/e42bccac-8e15-11e5-baf4-bdf37355da0c_story.html?utm_term=.03b6cb008ba2, accessed April 3, 2017.

[156] J. Kang, "The Online Life of Elliot Rodger," *The New Yorker*, May 28, 2014, available at http://www.newyorker.com/online/blogs/elements/2014/05/the-online-life-of-elliot-rodger.html.

[157] Z. Kastrati, A. Imran, S. Yildirim-Yayilgan, and F. Dalipi, "Analysis of Online Social Network Posts to Investigate Suspects Using SEMCON," *Social Computing and Social Media*, Vol. 9182 of LNCS, p.148–157, 2015.

[158] R. Katz, "Christopher Cornell Expressed Support for Islamic State, Lone Wolf Jihad on Social Media," *SITE* INSITE Blog on Terrorism and Extremism, January 15, 2015, accessed July 7, 2015 at `http://news.siteintelgroup.com/blog/index.php/entry/344-chris-cornell-professed-love-of`.

[159] R. Katz, "Information about the Chattanooga Shooter is Disappearing from the Internet," *SITE* INSITE Blog on Terrorism and Extremism, July 23, 2015, accessed August 13, 2015 at `http://news.siteintelgroup.com/blog/index.php/categories/jihad/entry/390-information-about-the-chattanooga-shooter-is-disappearing-from-the-internet`.

[160] KDD Cup 2015 Competition, May 1, 2015, originally available at `https://biendata.com/competition/kddcup2015/`, accessed December 5, 2016.

[161] M. Key, "Elliot Rodger, Santa Barbara mass shooting suspect, "My Twisted World" manifesto," May 24, 2014, available at `http://www.scribd.com/doc/225960813/Elliot-Rodger-Santa-Barbara-mass-shooting-suspect-My-Twisted-World-manifesto`.

[162] A. Khan, Y. Wu, C. Aggarwal, and X. Yan, "NeMa: Fast Graph Search with Label Similarity," *Proceedings of the VLDB Endowment*, Volume 6, Issue 3, January 2013.

[163] A. Kimery, "Jihad in Texas: In-Depth Look at the Shooters, Islamist Ties and Influences," *Homeland Security Today*, May 5, 2015, available at `http://www.hstoday.us/industry-news/general/single-article/jihad-in-`

texas-in-depth-look-at-the-shooters-islamist-ties-and-influences/09a2b7fc15abb51b28e5449d93bbf852.html.

[164] M. King and D. Taylor, "The Radicalization of Homegrown Jihadists: A Review of Theoretical Models and Social Psychological Evidence," *Terrorism and Political Violence*, 23: 602–622, 2011.

[165] J. Klausen, S. Campion, N. Needle, G. Nguyen, and R. Libretti, "Toward a Behavioral Model of 'Homegrown' Radicalization Trajectories", *Studies in Conflict and Terrorism*, 39:1, 67-83, 2015.

[166] J. Klausen, "Tweeting the Jihad: Social Media Networks of Western Foreign Fighters in Syria and Iraq," *Studies in Conflict and Terrorism*, 38:1, 1-22, 2015.

[167] J. Klausen. A Behavioral Study of the Radicalization Trajectories of American "Homegrown" Al Qaeda-Inspired Terrorist Offenders, 2001-2015 [UNITED STATES]. ICPSR36452-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2016-12-15. http://doi.org/10.3886/ICPSR36452.v1.

[168] J. Klausen. A Behavioral Study of the Radicalization Trajectories of American "Homegrown" Al Qaeda-Inspired Terrorist Offenders," Final Report to the U.S. Department of Justice, Office of Justice Programs/National Institutes of Justice, August 2016, forthcoming in *Studies in Conflict and Terrorism.*

[169] J. Klausen, personal correspondence to the author, March 2016,

[170] J. Klausen, C. Marks, T. Zaman, "Finding Online Extremists in Social Networks," arXiv, available at https://arxiv.org/pdf/1610.06242.pdf, 2016.

[171] J. Klausen, "The Myth of Homegrown Terrorism," *The Georgetown Security Studies Review*, Special Issue: What the New Administration Needs to Know About Terrorism and Counterterrorism, p. 50-60, 2017.

[172] J. Koehler, "The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges," *Behavioral and Brain Sciences*, Vol. 19, p. 1-53, 1996.

[173] D. Koller, A. Ng., C. Do, and Z. Chen, "Retention and Intention in Massive Open Online Courses: In Depth," *Educause Review*, June 3, 2013, available at `http://er.educause.edu/articles/2013/6/retention-and-intention-in-massive-open-online-courses-in-depth`, accessed June 1, 2017.

[174] A. Kossiakoff, W.N. Sweet, S. Seymour, and S.M. Biemer, *Systems Engineering Principles and Practice*, 2nd edition, Hoboken, New Jersey: John Wiley & Sons, 2011.

[175] D. Koutra, A. Parikh, A. Ramdas, and J. Xiang, "Algorithms for graph similarity and subgraph matching," Technical report, Carnegie-Mellon-University, 2011.

[176] C. Kurzman, D. Schanzer, and E. Moosa, "Muslim American Terrorism Since 9/11: Why So Rare?" *The Muslim World*, Vol. 101(3), p. 464-483, 2011.

[177] G. LaFree, "Lone-Offender Terrorists," editorial introduction, *Criminology and Public Policy*, Vol. 12(1), p. 59-62, 2013.

[178] G. LaFree and L. Dugan, "Global Terrorism Database," National Consortium for the Study of Terrorism and Responses to Terrorism (START), available at `https://www.start.umd.edu/gtd/about/`, 2017.

[179] G. LaFree, M. Jensen, and P. James,"Profiles of Individual Radicalization in the United States (PIRUS)," available at `http://www.start.umd.edu/data-tools/profiles-individual-radicalization-united-states-pirus`, accessed April 9, 2017.

[180] V. Lampos, T. Bie, and N. Cristianini, "Flu detector- Tracking Epidemics on Twitter." *Machine Learning and Knowledge Discovery in Databases, LNCS* 6323, pp. 599-602, 2010.

[181] A. Langville, C. Meyer, and R. Albright, "Initializations for the Nonnegative Matrix Factorization," ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Philadelphia, PA, 2006.

[182] J. Lavigne, M. Feldman, K. Meyers, "Screening for Mental Health Problems: Addressing the Base Rate Fallacy for a Sustainable Screening Program in Integrated Primary Care," *Journal of Pediatric Psychology*, Vol. 41(10), p. 1081-1090, 2016.

[183] P.S.H. Leeflang, J.E. Wieringa, T.H.A. Bijmolt, K.H. Pauwels, "Individual Demand Models," chapter in *Modeling Markets*, p. 261-305, New York: Springer Science + Business Media, 2015.

[184] M. Leiter, "U.S. Policy to Counter Violent Extremism is Incoherent," *The Cipher Brief*, April 27, 2017, available at `https://www.thecipherbrief.com/article/north-america/us-policy-counter-violent-extremism-incoherent-1089?utm_source=Join+the+Community+Subscribers&utm_campaign=5171b1b0bd-EMAIL_CAMPAIGN_2017_04_27&utm_medium=email&utm_term=0_02cbee778d-5171b1b0bd-122512261`, accessed April 27, 2017.

[185] S. Lee, Z. Liu, C. Kim, "An Agent Using Matrix for Backward Path Search on MANET," *Agent and Multi-Agent Systems: Technologies and Applications*, *Lecture Nodes in Computer Science* Vol. 4953, p. 203-211, 2008.

[186] D. Lee and H. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, volume 401, pp.788791, 1999.

[187] K. Leggiero, "Countering ISIS Recruitment in Western Nations," *Journal of Political Risk*, Vol. 3(1), available at http://www.jpolrisk.com/countering-western-recruitment-of-isis-fighters/, 2015.

[188] J. Li and A. Wang, "A framework of identity resolution: evaluating identity attributes and matching algorithms," *Security Informatics*, Vol. 4(6), 2015.

[189] Y. Liao and V. Rao Vemuri, "Using Text Categorization Techniques for Intrusion Detection," *Proceedings of the 11th USENIX Security Symposium*, USENIX Association, 2002.

[190] B. Llenas, "Fort Hood Shooter Ivan Lopez's Chilling Facebook Post: 'The Devil Will Take Me...Green Light And Finger Ready,' " *Fox News Latino*, April 3, 2014, available at http://latino.foxnews.com/latino/news/2014/04/03/fort-hood-shooter-ivan-lopez-facebook-status-devil-will-take-megreen-light-and/.

[191] Lowe's, "How to Install a Tile Backsplash," available at https://www.lowes.com/projects/kitchen-and-dining/how-to-install-a-tile-backsplash/project#noop, accessed January 3, 2017.

[192] S. Ma, Y. Cao, W. Fan, J. Huai, and T. Wo, "Strong Simulation: Capturing Topology in Graph Pattern Matching," *Proceedings of the VLDB Endowment*, Vol 5, No. 4, 2012.

[193] S. Ma, J. Li, C. Hu, X. Lin, and J. Huai, "Big graph search: challenges and techniques," *Frontiers of Computer Science*, May 5, 2015, Higher Education Press and Springer-Verlag Berlin Heidelberg, 2015.

[194] S. Malthaner and L. Lindekilde, "Analyzing Pathways of Lone-Actor Radicalization: A Relational Approach," unpublished manuscript presented at Constructions of Terrorism Conference at University of California in Santa Barbara, December 2015.

[195] G. Marx, "An Ethics for the New (and Old) Surveillance," in *Effective Surveillance for Homeland Security*, F. Flammini, R. Setola, and G. Fanceschetti, eds., CRC Press, New York, 2013.

[196] K. Mather, R. Winton, and A. Flores, "Deputies didn't view Elliot Rodger's videos in welfare check," *Los Angeles Times*, May 29, 2014, available at `http://www.latimes.com/local/la-me-rodger-welfare-20140530-story.html`.

[197] M. McCaul, "A National Strategy to Win the War Against Islamist Terror," U.S. House of Representatives Homeland Security Committee, September 2016, available at `https://homeland.house.gov/wp-content/uploads/2016/09/A-National-Strategy-to-Win-the-War.pdf`, accessed April 17, 2017.

[198] C. McCauley and S. Moskalenko, "Individual and group mechanism of radicalization," in "Protecting the Homeland from International and Domestic Terrorism Threats" White Papers, L Fenstermacher, L. Kuznar, T. Rieger, and A. Speckhard, eds., US Department of Defense and the Air Force Research Laboratory, available at `http://www.start.umd.edu/sites/default/files/files/publications/U_Counter_Terrorism_White_Paper_Final_January_2010.pdf`, 2010.

[199] C. McCauley and S. Moskalenko, "Toward a Profile of Lone Wolf Terrorists: What Moves an Individual From Radical Opinion to Radical Action," *Terrorism and Political Violence*, Vol. 26, p. 69-85, 2014.

[200] C. McCauley and S. Moskalenko, "Understanding Political Radicalization: The Two-Pyramids Model," *American Psychologist*, Vol. 72(3), p. 205-216, 2017.

[201] A. McGough, et al, "Insider Threats: Identifying Anomalous Human Behavior in Heterogeneous Systems Using Beneficial Intelligence Software (Ben-ware), *Proceedings of the*

7th ACM Computer and Communications Society International Workshop on Managing Insider Security Threats, p. 1-12, 2015.

[202] J.F. McGowan, "How to Build a Recommendation Engine," September 24, 2012, available at `https://mathblog.com/how-to-build-a-recommendation-engine/#comments`, accessed January 11, 2017.

[203] J. Meloy, J. Hoffmann, A. Guldimann, D. James, "The Role of Warning Behaviors in Threat Assessment: An Exploration and Suggested Typology," *Behavioral Sciences and the Law*, volume 30, pp. 256-279, 2012.

[204] J. Meloy, "Identifying Warning Behaviors of the Individual Terrorist," *FBI Law Enforcement Bulletin*, April 2016, accessed March 8, 2017, available at `https://leb.fbi.gov/2016/april/perspective-identifying-warning-behaviors-of-the-individual-terrorist`.

[205] R. Mendick, G. Rayner, M. Evans, and H. Dixon, "Security services missed five opportunities to stop the Manchester bomber," *The Telegraph*, May 25, 2017, available at `http://www.telegraph.co.uk/news/2017/05/24/securityservicesmissedfiveopportunitiesstopmanchester/`, accessed May 27, 2017.

[206] R. Metz, "Big Questions Around Facebook's Suicide Prevention Tools," *MIT Technology Review*, March 1, 2017, available at `https://www.technologyreview.com/s/603772/big-questions-around-facebooks-suicide-prevention-tools/`, accessed on April 10, 2017.

[207] Y. Miao, W. Han, K. Li, M. Wu, F. Yang, L. Zhou, V. Prabhakaran, E. Chen, and W. Chen, "ImmortalGraph: A System for Storage and Analysis of Temporal Graphs," *ACM Transactions on Storage*, Vol. 11(3), July 2015.

[208] L.E. Miller, "Multihop connectivity of arbitrary networks," `http://w3.antd.nist.gov/wctg/netanal/ConCalc.pdf`, 2001.

[209] C. Mindock, " Who Is Mohammod Youssef Abdulazeez? Chattanooga Shooter Identified; Dead; High School Peers React In Tennessee," *International Business Times*, July 16, 2015, available at `http://www.ibtimes.com/who-mohammod-youssef-abdulazeez-chattanooga-shooter-identified-dead-high-school-peers-2012830`.

[210] MITRE, "FFRDCs- A Primer: Federally Funded Research and Development Centers in the 21st Century," available at `https://www.mitre.org/publications/all/ffrdcs-a-primer`, 2015.

[211] MITRE, "Person-Centric Identity Management- Rapidly Assimilating Data About a Person of Interest," technical paper, January 2017, available at `https://www.mitre.org/publications/technical-papers/person-centric-identity-management-rapidly-assimilating-data-about-a`, accessed March 1, 2017.

[212] M. Molteni, "Artificial Intelligence is Learning to Predict and Prevent Suicide," *Wired*, March 17, 2017, available at `https://www.wired.com/2017/03/artificial-intelligence-learning-predict-prevent-suicide/`, accessed on April 10, 2017.

[213] C. Moody, "Developer Policies to Protect Peoples Voices on Twitter," November 22, 2016, available at `https://blog.twitter.com/2016/developer-policies-to-protect-people-s-voices-on-twitter`, accessed April 3, 2017.

[214] J. Monahan, "The Individual Risk Assessment of Terrorism," *Psychology, Public Policy, and Law*, Vol. 18(2), May 2012, p. 167-205.

[215] M. Mongiovi, R. Di Natale, R. Guigno, A. Pulvirenti, and A. Ferro, "SIGMA: A Set-Cover-Based Inexact Graph Matching Algorithm," *Journal of Bioinformatics and Computational Biology*, Vol 8, No. 2, p. 199—218, 2010.

[216] J. Murphy, V. Berk, and I. Gregorio-de Souza, "Decision Support Procedure in the Insider Threat Domain," *Proceedings of the IEEE Computer Science Security and Privacy Workshops*, 2012.

[217] E. Nakashima, M. Zapotosky, and M. Berman, "The FBI looked into suspected bomber Ahmad Rahami in 2014 and found no 'ties to terrorism'," *The Washington Post*, September 20, 2016.

[218] National Security Agency, "Section 702" fact sheet of Section 702 of FISA, available from `https://www.scribd.com/document/149791922/National-Security-Agency-Section-702-of-FISA-and-Section-215-of-PATRIOT-Act-Fact-Sheets`, accessed April 7, 2017.

[219] National Security Agency, "Section 215" fact sheet of Section 215 of PATRIOT Act, available from `https://www.scribd.com/document/149791922/National-Security-Agency-Section-702-of-FISA-and-Section-215-of-PATRIOT-Act-Fact-Sheets`, accessed April 7, 2017.

[220] National Counterterrorism Center, "Behavioral Indicators Offer Insights for Spotting Extremists Mobilizing for Violence," 2011, accessed August 1, 2015 at `https://publicintelligence.net/ufouo-national-counterterrorism-center-mobilizing-homegrown-violent-extremists-hves-behavioral-indicators/`.

[221] NPR, "German Program Helps Families De-Radicalize Members Prone To Extremism," an interview with Daniel Koehler, March 13, 2015, available at http://www.npr.org/2015/03/13/392845800/german-program-helps-families-de-radicalize-members-prone-to-extremism, accessed May 29, 2017.

[222] Nationwide SAR Initiative (NSI), "NSI Resources," available at https://nsi.ncirc.gov/resources.aspx, accessed April 8, 2017.

[223] L.S. Neo, M. Khader, J. Ang, G. Ong, and E. Tan, "Developing an early screening guide for jihadi terrorism: A behavioural analysis of 30 terror attacks," *Security Journal*, November 2014.

[224] L.S. Neo, "An Internet-Mediated Pathway for Online Radicalization: RECRO," in *Combating Violent Extremism and Radicalization in the Digital Era*, Hersey, PA: IGI Global, 2016.

[225] National Institutes of Justice, Funding Solicitation: "Research and Evaluation on Domestic Radicalization to Violent Extremism," 2017.

[226] Office of the Executive of the United States, "National Strategy for Counterterrorism," June 2011, available at https://obamawhitehouse.archives.gov/sites/default/files/counterterrorism_strategy.pdf, accessed April 17, 2017.

[227] Office of the Executive of the United States, "Strategic Implementation Plan for Empowering Local Partners to Prevent Violent Extremism in the United States," October 2016, available at https://www.brennancenter.org/sites/default/files/2016_strategic_implementation_plan_empowering_local_partners_prev%20(2).pdf, accessed April 17, 2017.

[228] Offices of the Inspector General of the Intelligence Community, Department of Homeland Security, and Department of Justice, "Review of Domestic Sharing of Counterterrorism Information," Department of Justice, Washington, D.C., 2017.

[229] Office of the Program Manager, Information Sharing Environment, "Information Interoperability Framework (I^2F)," version 0.5, March 2014, available at https://www.ise.gov/resources/document-library/ise-information-interoperability-framework, accessed May 21, 2017.

[230] Office of the State's Attorney Judicial District of Danbury, "Report of the State's Attorney for the Judicial District of Danbury on the Shootings at Sandy Hook Elementary School and 36 Yogananda Street, Newtown, Connecticut on December 14, 2012," November 25, 2013, available at http://www.ct.gov/csao/lib/csao/Sandy_Hook_Final_Report.pdf.

[231] M. Olama, G. Allgood, K. Davenport, and J. Schryver, "A Bayesian Belief Network of Threat Anticipation and Terrorist Motivations," *Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense IX*, edited by E.M. Carapezza, and *Proceedings of of SPIE*, Vol. 7666, 2010.

[232] R. Olson, "Suicide Threats on Social Network Sites" Centre for Suicide Prevention, 2011, available at http://www.sprc.org/resources-programs/suicide-threats-social-networking-sites, accessed December 28, 2013.

[233] M. Orcutt, "Why Congress Can't Seem to Fix This 30-Year-Old Law Governing Your Electronic Data," *MIT Technology Review*, February 17, 2017, available at https://www.technologyreview.com/s/603636/why-congress-cant-seem-to-fix-

this-30-year-old-law-governing-your-electronic-data/?set=603667, accessed April 4, 2017.

[234] P. Paatero and U. Tapper, "Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values," *Environmetrics*, volume 5, pp 111–126, 1994.

[235] R. Paffenroth, P. du Toit, R. Nong, L. Scharf, A. Jayasumana, "Space-time signal processing for distributed pattern detection in sensor networks," *Journal of Selected Topics in Signal Processing*, vol. 6, 2013.

[236] S. Page, "Path Dependence," *Quarterly Journal of Political Science*, Vol. 1, p. 87-115, 2006.

[237] M. Palin, "The other 'imminent' terror attacks Australia narrowly escaped," December 23, 2016, http://www.news.com.au/national/crime/the-11-imminent-terror-attacks-australia-narrowly-escaped/news-story/86fc734df0963e21fe038c0eecce7d80, accessed 3 April 2017.

[238] H. Park, "Nonnegative Matrix Factorization: Algorithms and Applications," SIAM International Conference on Data Mining, 2011.

[239] G. Parnell, P. Driscoll, and D. Henderson, eds., *Decision Making in Systems Engineering and Management*, 2nd edition, Hoboken, New Jersey: John Wiley & Sons, 2011.

[240] P. Parveen, J. Evans, B. Thuraisingham, K. Hamlen, and L. Khan, "Insider Threat Detection using Stream Mining and Graph Mining," *Privacy, Security, Risk and Trust (PASSAT)* and *Proceedings of the IEEE Third International Conference on Social Computing (SocialCom)*, 2011.

[241] F. Patel, "Rethinking Radicalization," Brennan Center for Justice, New York University School of Law, 2011.

[242] N. Patki, R. Wedge, and K. Veeramachaneni, "The Synthetic Data Vault," *Proceedings of the IEEE International Conference on Data Science and Advanced Analytics*, 2016.

[243] J. Pearson, "Facebook Banned This Canadian Surveillance Company From Accessing Its Data," *Motherboard*, available at `https://motherboard.vice.com/en_us/article/instagram-banned-this-canadian-surveillance-company-from-accessing-its-data-media-sonar`, accessed April 3, 2017.

[244] E. Perez and S. Prokupecz,"FBI struggling with surge in homegrown terror cases," *CNN*, May 30, 2015, accessed August 10, 2015 at `http://www.cnn.com/2015/05/28/politics/fbi-isis-local-law-enforcement`.

[245] E. Perez and S. Prokupecz,"Paris attackers likely used encrypted apps, officials say," *CNN*, December 17, 2015, accessed May 30, 2017 at `http://www.cnn.com/2015/12/17/politics/parisattacksterroristsencryption/`.

[246] Pew Research Center, "Views of Islam and extremism in the U.S. and abroad," February 16, 2017, available at `http://www.people-press.org/2017/02/16/3-views-of-islam-and-extremism-in-the-u-s-and-abroad/`, accessed April 16, 2017.

[247] R. Pienta, A. Tamersoy, H. Tong, and D. Chau, "MAGE: Matching Approximate Patterns in Richly-Attributed Graphs," *Proceedings of the IEEE Conference on Big Data*, October 2014.

[248] C. Poulin, B. Shiner, P. Thompson, L. Vepstas, Y. Young-Xu, B. Goertzel, B. Watts, L. Flashman, and T. McAllister, "Predicting the Risk of Suicide by Analyzing the Text

of Clinicial Notes," *PLoS ONE*, Vol. 9(1), 2014, available at http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0085733, accessed June 1, 2017.

[249] D. Pressman and J. Flockton, "Calibrating risk for violent political extremists and terrorists: the VERA 2 structured assessment," *The British Journal of Forensic Practice*, Vol. 14(4), 2012, p. 237-251.

[250] J. Preston and M. Roston, "A Closer Look at the Bombing Suspect's Twitter Account," *New York Times*, April 20, 2013, accessed March 27, 2014 at http://thelede.blogs.nytimes.com/2013/04/20/dzhokhar-tsarnaevs-jahar-twitter-account-prompts-scrutiny/.

[251] B.K. Pursel, L. Zhang, K.W. Jablokow, G.W. Choi, and D. Velegol, "Understanding MOOC students: motivations and behaviors indicative of MOOC completion," *Journal of Computer Assisted Learning*, Vol. 32, p. 202-217, 2016.

[252] M. Reddy, R. Borum, J. Berglund, B. Vossekuil, R. Fein, and W. Modzeleski, "Evaluating Risk for Targeted Violence in Schools: Comparing Risk Assessment, Threat Assessment, and Other Approaches, *Psychology in Schools*, vol. 38, 2001.

[253] J. Reich, "MOOC Completion and Retention in the Context of Student Intent," *Educause Review*, December 8, 2014, available at http://er.educause.edu/articles/2014/12/mooc-completion-and-retention-in-the-context-of-student-intent, accessed December 31, 2016.

[254] M. Reilly, "How Facebook Learns About Your Offline Life," *MIT Technology Review*, December 28, 2016, available at https://www.technologyreview.com/s/603283/how-facebook-learns-about-your-offline-life/?utm_campaign=internal&utm_medium=homepage&utm_source=top-stories_2&set=603276, accessed April 3, 2017.

[255] P. Robertson, "How 'glaring' intelligence failures allowed a second bout of terror in Paris," *CNN*, November 18, 2015, available at http://www.cnn.com/2015/11/18/europe/paris-terror-attacks-intelligence-failures-robertson/, accessed July 27, 2016.

[256] J. Robinson, J. Webber, and E. Eifrem, *Graph Databases*, Sebastopol, CA: O'Reilly Media, Inc., 2015.

[257] P. Rosenzweig, C. Stimson, and D. Shedd, "Maintaining America's Ability to Collect Foreign Intelligence: The Section 702 Program," The Heritage Foundation, May 13, 2016, available at http://www.heritage.org/defense/report/maintaining-americas-ability-collect-foreign-intelligence-the-section-702-program, accessed April 7, 2017.

[258] B. Ross and R. Schwartz, "Major Hasans E-mail: 'I Cant Wait to Join You in Afterlife, *ABC News*, November 19, 2009, available at http://abcnews.go.com/Blotter/major-hasans-mail-wait-join-afterlife/story?id=9130339.

[259] R. Rossi, J. Neville, B. Gallagher, and K. Henderson, "Modeling Dynamic Behavior in Large Evolving Graphs," *WSDM 2013*, Association for Computing Machinery, February 2013.

[260] P. Rucker and R. Costa, "In Elliot Rodger, authorities in Calif. Saw warning signs but didn't see a tipping point, *The Washington Post*, May 25, 2014, available at http://www.washingtonpost.com/national/sheriff-calif-shooter-rodger-flew-under-the-radar-when-deputies-visited-him-in-april/2014/05/25/88123026-e3b4-11e3-8dcc-d6b7fede081a_story.html.

[261] W. Ruderman, "Court Prompts Twitter to Give Data to Police in Threat Case," *New York Times*, 7 August 2012, accessed March 29, 2014, at http://www.nytimes.com/2012/08/08/nyregion/after-court-order-twitter-sends-data-on-user-issuing-threats.html?_r=1&.

[262] M. Sageman, *Understanding Terror Networks*, Philadelphia: University of Pennsylvania Press, 2004.

[263] M. Sageman, *Leaderless Jihad: Terror networks in the twenty-first century*, Philadelphia: University of Pennsylvania Press, 2008.

[264] M. Sageman, "The Stagnation in Terrorism Research," *Terrorism and Political Violence*, Vol. 26(4), p. 1-16, 2014.

[265] A. Sanzgiri and D. Dasgupta, "Classification of Insider Threat Detection Techniques," *Proceedings of the 11th Annual Cyber and Information Security Research Conference*, 2016.

[266] K. Sarma, "Risk Assessment and the Prevention of Radicalization from Nonviolence Into Terrorism," *American Psychologist*, Vol. 72(3), p. 278-288, 2017.

[267] D. Schanzer, C. Kurzman, J. Toliver, and E. Miller, "The Challenge and Promise of Using Community Policing Strategies to Prevent Violent Extremism: A Call for Community Partnerships with Law Enforcement to Enhance Public Safety, Final Report," U.S. Department of Justice report, January 2016, available at https://www.ncjrs.gov/pdffiles1/nij/grants/249674.pdf, accessed April 4, 2017.

[268] B. Schneier, "Why Data Mining Won't Stop Terror," *Wired*, March 9, 2005, available at https://www.schneier.com/essays/archives/2005/03/why_data_mining_wont.html, accessed April 9, 2017.

[269] B. Schneier, "Terrorists, Data Mining, and the Base Rate Fallacy," July 10, 2006, available at `https://www.schneier.com/blog/archives/2006/07/terrorists_data.html`, accessed April 9, 2017.

[270] J. Schram, Y. Steinbuch, and D. Fears, "Killer wife pledged allegiance to ISIS on Facebook during attack, *New York Post*, December 4, 2015, available at `http://nypost.com/2015/12/04/killer-wife-in-california-massacre-swore-allegiance-to-isis/`.

[271] B. Schuurman, and Q. Eijkman, "Indicators of Terrorist Intent and Capability," *Dynamics of Asymmetric Conflict*, June 2015.

[272] R. Scrivens, G. Davies, R. Frank, and J. Mei, "Sentiment-based Identification of Radical Authors (SIRA)," *Proceedings of the 2015 IEEE International Conference on Data Mining Workshops*, 2015.

[273] A. Sememov, J. Veijalainen, and J. Kyppo, "Analysing the presence of school-shooting related communities at social media sites," *International Journal of Multimedia Intelligence and Security*, Vol 1, Issue 3, 2010.

[274] A. Semenov, J. Veijalainen, and A. Boukhanovsky, "A Generic Architecture for a Social Network Monitoring and Analysis System," *IEEE International Conference on Network-Based Information Systems*, 2011.

[275] A. Semenov, A. Nikolaev, and J. Veijalainen, "Online Activity Traces Around a "Boston Bomber," " IEEE and ACM Conference on Advances in Social Network Analysis and Mining, *ASONAM* 2013.

[276] A. Semenov, "Principles of Social Media Monitoring and Analysis Software," PhD Dissertation, University of Jyväskylä, Finland, May 31, 2013.

[277] T. Senator, et al, "Detecting Insider Threats in a Real Corporate Database of Computer Usage Activity," *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, p. 1393-1401, 2013.

[278] R. Serrano, "Tashfeen Malik messaged Facebook friends about her support for jihad," *The Los Angeles Times*, December 14, 2015, available at http://www.latimes.com/local/lanow/la-me-ln-malik-facebook-messages-jihad-20151214-story.html.

[279] Z. Seward, L. Mirani, and R. King, "We Know The Boston Bomber's Sleeping Cycles," *Business Insider*, available at http://www.businessinsider.com/we-know-when-the-boston-bomber-sleeps-2013-4, accessed 1 Feb 2013.

[280] D. Shedd, "Technology, New FBI Powers Needed to Combat Terror," The Cipher Brief Expert Commentary, June 4, 2017, accessed June 4, 2017, available at https://www.thecipherbrief.com/article/exclusive/europe/technology-new-fbi-powers-needed-combat-terror-1089?utm_source=Join+the+Community+Subscribers&utm_campaign=ee5e77a258-EMAIL_CAMPAIGN_2017_06_05&utm_medium=email&utm_term=0_02cbee778d-ee5e77a258-122512261.

[281] R. Sherman, "Facebook and Privacy" March 13, 2017, available at https://www.facebook.com/fbprivacy/posts/1624880004207125, accessed April 3, 2017.

[282] C. Shoichet, "Garland, Texas, shooting suspect linked himself to ISIS in tweets," *CNN*, May 4, 2015, accessed July 7, 2015 at http://www.cnn.com/2015/05/04/us/garland-mohammed-drawing-contest-shooting/.

[283] L. Shure, "Introduction to Market Basket Analysis," in "Loren on the Art of MATLAB," available at http://blogs.mathworks.com/loren/2015/01/29/introduction-to-market-basket-analysis/, January 29, 2015, accessed November 5, 2016.

[284] M. Silber and A. Frey, "Detect, Disrupt, and Detain: Local Law Enforcements Critical Roles in Combating Homegrown Extremism and the Evolving Terrorist Threat," *Fordham Urban Law Journal*, Vol 41(1), 2015.

[285] M. Silber and A. Bhatt, "Radicalization in the West: The Homegrown Threat," New York Police Department Intelligence Division, 2007.

[286] J. Silverstein, "FBI arrests 20-year-old Ohio man Christopher Lee Cornell for allegedly plotting ISIS-inspired bombing of U.S. Capitol," *New York Daily News,* January 14, 2015, accessed July 7, 2015 at `http://www.nydailynews.com/news/national/fbi-arrests-ohio-man-plotting-isis-inspired-bombing-article-1.2077959`.

[287] SITE Monitoring Service, "Profiles of Djohar Tsarnaev's Twitter, Tamerlan's Alleged YouTube Channel," *SITE Intelligence*, January 15, 2014, accessed March 27, 2014, at `http://news.siteintelgroup.com/index.php/18-articles-a-analysis/3001-profiles-of-djohar-tsarnaevs-twitter-tamerlans-alleged-youtube-channel`.

[288] G. Smith, " Adam Lanza's Smashed Hard Drive Doesn't Erase His Digital Footprint, Experts Say," December 18, 2012, available at `http://www.huffingtonpost.com/2012/12/18/adam-lanzas-hard-drive_n_2324410.html`.

[289] M. Smith, A. Ceni, N. Milic-Frayling, B. Shneiderman, E. Mendes Rodrigues, J. Leskovec, C. Dunne, NodeXL: a free and open network overview, discovery and exploration add-in for Excel 2007/2010/2013/2016, `http://nodexl.codeplex.com/` from the Social Media Research Foundation, `http://www.smrfoundation.org`.

[290] C. Song, T. Ge, C. Chen, and J. Wang, "Event Pattern Matching over Graph Streams," *Proceedings of the VLDB Endowment*, Vol. 8, No. 4, 2014.

[291] E. Southers, "Op Ed: The U.S. government's program to counter violent extremism needs an overhaul," *Los Angeles Times*, March 21, 2017, available at `http://www.latimes.com/opinion/la-fg-global-erroll-southers-oped-20170321-story.html`, accessed April 15, 2017.

[292] R. Spaaij and M. Hamm, "Key Issues and Research Agendas in Lone Wolf Terrorism," *Studies in Conflict and Terrorism*, Vol. 38, p. 167-178, 2015.

[293] P. Tan, M. Steinbach, V. Kumar, "Association Analysis: Basic Concepts and Algorithms," chapter in *Introduction to Data Mining*, New York: Pearson Education Limited, 2006.

[294] B. Thomee, D. Shamma, G. Friedland, B. Elizalde, K. Ni, D. Poland, D. Borth, and L. Li. "The New Data and New Challenges in Multimedia Research", arXiv:1503.01817, 2015.

[295] Y. Tian and J. Patel, "TALE: A Tool for Approximate Large Graph Matching," *IEEE 24th International Conference on Data Engineering*, p. 963—972, 2008.

[296] D. Zeimpekis and E. Gallopoulos, "TMG: A MATLAB toolbox for generating term-document matrices from text collections". In "Grouping Multidimensional Data: Recent Advances in Clustering", J. Kogan, C. Nicholas and M. Teboulle, eds., pp. 187-210, Springer, 2006. Also Technical Report HPCLAB-SCG 1/01-05, Computer Engineering and Informatics Dept., University of Patras, Greece, Jan. 2005

[297] H. Tong, B. Gallagher, C. Faloutsos, and T. Eliassi-Rad, "Fast Best-Effort Pattern Matching in Large Attributed Graphs," *KDD*, August 2007.

[298] Unknown, "Threads involving Jared Loughner (aka Dare)" *Earth Empires*, available at `http://www.earthempires.com/jared-loughner-arizona-shooter-posts`.

[299] United States District Court for the Central District of California, "Criminal Complaint- United States of American v. Enrique Marquez, Jr," December 17, 2015, available at http://www.justice.gov/opa/file/800606/download.

[300] United States District Court for the Southern District of Ohio, "Criminal Complaint-United States of America v. Christopher Lee Cornell," January 14, 2015.

[301] United States Internet Service Provider Association, "Electronic Evidence Compliance-A Guide for Internet Service Providers," *Berkeley Technology Law Journal*, Vol. 18(4), 2003.

[302] T. Veldhuis and J. Staun, *Islamist Radicalization: A Root Cause Model*, The Hague, Netherlands Institute of International Relations Clingendael, 2009.

[303] K. Verma, M. Jadon, and A. Pujari, "Clustering Short-Text Using Non-negative Matrix Factorization of Hadamard Product of Similarities," Proceedings of the 2013 Asia Information Retrieval Societies Conference, in *Information Retrieval Technology* LNCS 8281, pp.145-155, 2013.

[304] H. Wang, *Innovative Techniques and Applications of Entity Resolution*, Hershey, PA: IGI Publishing, 2014.

[305] J. Welsh, "Psychologist Analyzes Dzhokhar Tsarnaev's Tweets, Says He Committed To Violence As Early As October 2012," *Business Insider*, August 5, 2013, accessed March 28, 2014, at http://www.businessinsider.com/dzhokhar-tsarnaev-committed-to-violent-acts-as-early-as-october-2012-2013-8.

[306] W. Webster, "Final Report of the William H. Webster Commission on the Federal Bureau of Investigation, Counterterrorism Intelligence, and the Events at Fort Hood, Texas, on November 5, 2009," July 19, 2012, available at

https://www.fbi.gov/news/pressrel/press-releases/judge-webster-delivers-webster-commission-report-on-fort-hood.

[307] G. Weimann, "Lone Wolves in Cyberspace," *Journal of Terrorism Research*, vol 3, issue 2, 2012.

[308] West Corporation, West Safety Services, "Beware Fact Sheet," 2016.

[309] Wharton School of the University of Pennsylvania, "Customer Journey Mapping Is at the Heart of Digital Transformation," Knowledge@Wharton, November 2015, available at http://knowledge.wharton.upenn.edu/article/customer-journey-mapping-is-at-the-heart-of-digital-transformation/, accessed February 4, 2016.

[310] S. White, "Workplace Targeted Violence: Threat Assessment Incorporating a Structured Professional Judgment Guide," *International Handbook of Threat Assessment*, J. Meloy and J. Hoffmann, eds. New York: Oxford University Press, 2014, p. 83-106.

[311] Wikipedia, "2014 Fort Hood shooting," last modified February 22, 2016, available at http://en.wikipedia.org/wiki/2014_Fort_Hood_shooting.

[312] Wikipedia, "Sandy Hook Elementary School shooting," last modified March 8, 2016, available at http://en.wikipedia.org/wiki/Sandy_Hook_Elementary_School_shooting.

[313] M. Williams, J. Horgan, and W. Evans, "The critical role of friends in networks for countering violent extremism: toward a theory of vicarious help-seeking," *Behavioral Sciences of Terrorism and Political Aggression*, Vol. 8(1), p. 45-65, 2016.

[314] Z. Winn, "Countering Potential Campus Threats with Social Media Monitoring," *Campus Safety*, Internet: available at http://www.campussafetymagazine.com/article/

countering_potential_threats_with_social_media_monitoring, January 20, 2016, accessed April 1, 2016.

[315] D. Wroe, "Most Australians happy with government social media to stop terror: poll," *The Sydney Morning Herald*, July 2, 2015, available at http://www.smh.com.au/federal-politics/political-news/most-australians-happy-with-government-watching-social-media-to-stop-terror-poll-20150701-gi2uf9.html, accessed April 3, 2017.

[316] X. Yan, J. Guo, S.Liu, X. Cheng, Y. Wang, "Clustering Short Text Using Ncut-Weighted Non-Negative Matrix Factorization," ACM CIKM International Conference on Information and Knowledge Management, Maui, HI, 2012.

[317] X. Yan, J. Guo, S.Liu, X. Cheng, Y. Wang, "Learning Topics in Short Texts by Non-negative Matrix Factorization on Term Correlation Matrix," Proceedings of the 13th SIAM International Conference on Data Mining, SDM'13, Texas, 2013.

[318] X. Yang, H. Qiao, and Z. Liu, "A Weighted Common Subgraph Matching Algorithm," arxiv available at http://arxiv.org/abs/1411.0763, 2014.

[319] R. Yuster and U. Zwick, "Fast Sparse Matrix Multiplication," *ACM Transactions on Algorithms*, Vol. 1., No. 1, p. 2—13, July 2005.

[320] R. Zafarani and H. Liu. Social Computing Data Repository at ASU [http://socialcomputing.asu.edu]. Tempe, AZ: Arizona State University, School of Computing, Informatics and Decision Systems Engineering, 2009.

[321] K. Zavadski, "FBI Knew About 'Draw Muhammad' Gunman Elton Simpson," *The Daily Beast*, May 4, 2015, available at http://www.thedailybeast.com/articles/2015/05/04/fbi-knew-about-draw-muhammad-gunman-eltonsimpson.html.

[322] K. Zavadski, "Read Chattanooga Shooter's Blog," *The Daily Beast*, July 16, 2015, available at `http://www.thedailybeast.com/cheats/2015/07/16/read-chattanooga-shooter-s-blog.html`.

[323] D. Zeimpekis, E. Kontopoulou, and E. Gallopoulos, Text to Matrix Generator software, 2017, available at `http://scgroup20.ceid.upatras.gr:8000/tmg/`.

[324] T. Zhang and D. Zhang, "Agent-based simulation of consumer purchase decision-making and the decoy effect," *Journal of Business Research*, Vol. 60, p. 912-922, 2007.

APPENDIX A

Reference Table for Recent Violent Extremist Attacks

A.1. Background

The pervasiveness of social media posts in the form of microblogs such as Twitter has led researchers to the potential of detecting latent signals of human behavior. Some, in particular, have begun to study the social media signals of 'lone wolf' terrorism[48] and radical violence and proposed analytical methods to detect such signals [35] [52]. In recent history, some domestic terrorists left social media footprints, including Jared Loughner in Arizona in 2011 and Anders Breivik in Norway in 2011 [127]. Brynielsson recently proposed the use of a subset threat behaviors from [203] as indicators of a lone wolf's intent to commit an act of terror: 1) activity on a radical webpage, 2) radical expressions in postings, 3) leakage of intent to do harm, 4) identification of oneself with a previous attacker or as an agent of a particular cause, and 5) fixation on issue, idea, or person [35]. He also went on to identify potential online detection methods for each of these indicators, including Bayesian and non-Bayesian classifiers, supervised and unsupervised machine learning, and semantic text analysis, but stopped short of providing any results [35].

To develop a basis for understanding the Tsarnaev threat signals and many others, we compiled an open-source case study database of the digital signals of targeted violence. It consists of 12 prominent and recent cases of targeted violence (a majority of which are incidents of homegrown violent extremism) where online communications in some form may had contained some latent signal of the threat. See Tables A.1-A.3. This compilation is unique in the literature with its focus on bringing to light those available digital indicators,

[48] A 'lone wolf terrorist' is defined as "a person who acts on his or her own without orders from– or even connections to– an organization" [37].

with an emphasis on recent cases of homegrown violent extremism (both those carried out as well as foiled). There are three other collections worth mentioning. First, the website "New America" has a periodically updated dataset of U.S. terrorist plots (to include those inspired by the jihadist ideology) [17]. A recent compilation of jihadist lone actor terrorists can also be found in [111], which summarizes the warning signs or indicators but does not include specifics. Additionally, [273] detailed the digital signals in 12 cases of only school shootings from 2005 to 2009 mostly in North America and Europe. In forthcoming research, we intend to provide a further analysis of the signals in each of the cases we covered and their implication for future study in social media targeted violence detection.

A.2. Related Work

Many have been applying the fields of both machine learning and text analysis (or text mining) to social media data. Twitter, a popular micro-blog, is a widely utilized data source because of its accessibility and availability[49]. In the last few years, researchers have found macro-level, crowd-sourcing correlations of Twitter data to some real world phenomenon, including movie box office results [4], stock market performance [29], the incidence of influenza [180], and even the prevalence of depression [71]. To date, however, we have been unable to find any research that tries to make micro-level (individual) correlations to real world activity, which is of prime importance in lone-wolf indicator detection.

[49]Twitter has also made content and user information available to the police [66] [261].

TABLE A.1. Recent Incidents of Targeted Violence with Social Media Signals (1 of 3)

Name	Plot Name	Date	Location	Completed?	Description	Digital Footprint Summary	Electronic Signal Detected?	Data Available and Sources
Hasan, Nidal Malik	Fort Hood shooting	5-Nov-09	Fort Hood, TX	Yes	Target: US Army Soldiers at an inprocessing center on post. Description: Fatally shot and killed 13 people and injured more than 30 others	1) Exchanged 18 emails with Awlaki [306] [258]. 2) Developed powerpoint presentation entitled ("The Koranic World View as it Relates to Muslims in the US Military") [307]; unknown if it was distributed.	Yes, 2 FBI task forces intercepted the messages, but they and the Army deemed them "innocent" [306] [258].	Yes, emails in text form from the published Webster Commission [306].
Loughner, Jared Lee	Tucson shooting	8-Jan-11	Tucson, AZ	Yes	Target: US Representative Gabrielle Giffords. Description: Shot and severely injured Congresswoman Giffords, and killed 6 people.	1) MySpace account with 216 followers. Posted numerous disturbing messages [22] [3]. 2) Online private gaming forum postings since 2010 at the Earth Empires Massive Multiplayer Online game site [298] [115]. 3) YouTube videos [22] [3].	No.	Yes, some. MySpace account was terminated, but FBI released documents which show the most troubling posts and messages as well as described some of his YouTube videos [3]. Online forum posts were private but released a few under his "Dare" username [298].
Breivik, Andres	Bombing and mass shooting in Norway	22-Jul-11	Oslo and Utoya Island, Norway	Yes	Target: Government building bombing and teen camp of the Workers' Youth League (AUF). Description: Bombed government buildings in Oslo killing 8 people; Mass shooting on the island of Utoya killing 69 people.	1) Manifesto [307] [95].	Yes.	Yes [95].
Pimentel, Jose	Pipe bombs to blow up police and postal facilities in NY	19-Nov-11	Manhattan, NYC, NY	No	Target: NYC Police and Postal Offices. Description: One-man terror plot in retaliation for the US killing of Anwar al-Awlaki in SEP 11.	1) Noticed by police by website postings associated with al-Awlaki in 2009 [131] [121]. 2) Maintained his own website www.trueislam1.com (formerly www.trueislam12@blogspot.com), which encouraged violence against America and posted bomb-making instructions [64].	Yes, police knew about the website postings and website maintenance [64].	None found.
Lanza, Adam	Connecticut school shooting	14-Dec-12	Newtown, CT	Yes	Target: Sandy Hook Elementary School. Description: Lone gunman killed 27 others at school and home.	1) Lanza destroyed his hard drive before the shooting and he reportedly did not have a Facebook or Twitter account [288]. 2) However, investigators recovered other digital evidence of a fascination with mass shootings and entries in a unspecified blog focused on mass shootings [230].	No.	None found.
Tsarnaev, Tamerlan and **Tsarnaev**, Dzhokhar	Boston Marathon Bombing	15-Apr-13	Boston, MA	Yes	Target: Boston Marathon finishline. Pressure cooker bombs placed at the finish line. Description: Killed 3 and wounded 264 others. Also killed MIT Police Officer Sean Collier.	1) Tamerlain's suspected YouTube account http://www.youtube.com/user muazseyfullah [287] [129] [108]. 2) Dzokhar's Twitter account https://twitter.com/J_tsar [279] [1] [84] [250].	No.	1) Tamerlan's YouTube videos [287] [129] [108]. 2) Dzokhar's tweets from Twitter account in text (scraped on 19 April 2013)– Note: there have been many subsequent tweets deleted since then [279].

TABLE A.2. Recent Incidents of Targeted Violence with Social Media Signals (2 of 3)

Name	Plot Name	Date	Location	Completed?	Description	Digital Footprint Summary	Electronic Signal Detected?	Data Available and Sources
Lopez, Ivan	Fort Hood shooting	2-Apr-14	Fort Hood, TX	Yes	Target: Various unit areas on Fort Hood. Description: Mass shooting that killed 4, and injured 16.	1) Facebook status under the name Ivan Slipknot [190].	No.	Yes, only the posts highlighted in the media; actual Facebook account was pulled [190]. Posted on 1 MAR 14: "I have just lost my inner peace, full of hatred, I think this time the devil will take me...I was robbed last night and I am sure it was 2 "flacos" (guys). Green light and finger ready. As easy as that." [190]
Rodger, Elliot	Mass shooting in California	23-May-14	Santa Barbara, CA	Yes	Mass stabbing and shooting in the college town. 6 people killed, 13 injured.	1) YouTube videos discussing murder and suicide posted as early as April 2014. 2) YouTube video, "Elliot Rodger's Retribution" posted day of the shooting [260] 3) Manifesto "My Twisted World" was emailed to mother and therapist minutes before shooting on 23 May, and possibly a day before to others [196].	Yes, parents alerted police in April 2014 due to his videos. Police questioned him on 30 APR, but did not search his house [260] [196]. Also on 22 MAY, someone posted one of his disturbing YouTube videos on reddit.com. Some were alarmed but did not warn anyone [156].	1) Manifesto (137 pages) in text [161]. 2) YouTube videos https://www.youtube.com/user/ElliotRodger and transcript in text [260] [132]. 3) Remants of converations on Bodybuilding.com [132]. 4) Remnants of threads on Puahate.com [132].
Cornwell, Christopher	Pipe bombs to attack the US Capitol	14-Jan-15	Green Township, OH	No	Target: US Capitol Building. Description: Planned to plant pipe bombs in US Capitol and shoot people inside.	1) Twitter account(s) using the online persona "Raheel Mahrus Ubaydah" @ISBlackFlags [158] [300] [286]. 2) Links to prominent pro-radical jihadists including "Israfil Yilmaz," a prominent Dutch jihadist fighting in Syria, and "Muslim-Al-Britani," the pro-IS British fighter who had made headlines after flooding Twitter with weapons-making manuals in late 2014 [158]. 3) Communication with confidential informant eventually went to a messaging service (BBC) and typed "I believe we should meet up and make our own group in alliance with the Islamic State here and plan operations ourselves" [300].	Confidential informant who was being investigated by the FBI on a different charge offered to reveal information about Cornell in exchange for clemency. Cornwell was arrested on 14 Jan 15 outside a gun shop after purchasing two semi-automatic rifles and about 600 rounds of ammunition. He had long been tracked by an undercover agent [300].	1) Yes. Some tweets published in Katz online blog [158].
Simpson, Elton and **Soofi**, Nadir	Attack on Curtis Culwell Center in Texas	4-May-15	Garland, TX	Yes	Target: People in the Curtis Culwell Center viewing a Mohammed cartoon contest. Description: They wore body armor and carried assault rifles to attack those in the building.	1) Simpson Tweet: "May Allah accept us as mujahideen," and "given bay'ah to Amirul Mu'mineen" (pledge of allegiance the leader of the faithful) (likely ISIS leader Abu Bakr al Baghdadi) [282] [50]. 2) Nadir Soofi reportedly had Facebook page (unknown) that showed his pro-Palestine views and devotion to Islam [321] and his critique of US Middle East policy [163].	Elton Simpson was investigated by the FBI as early as 2010 before resurfacing on social media with a "renewed interest in jihad" [282] [321]. He subsequently posted references about the Prophet Muhammad cartoon contest in Garland Texas in 2013, which prompted the FBI send a notice to the Joint Terrorism Task Force that was monitoring the event [282] [321].	1) Yes. Some tweets posted under the name Simpson's Twitter account "Shariah is Light" [50].

TABLE A.3. Recent Incidents of Targeted Violence with Social Media Signals (3 of 3)

Name	Plot Name	Date	Location	Completed?	Description	Digital Footprint Summary	Electronic Signal Detected?	Data Available and Sources
Abdulazeez, Muhammad Youssef	Attack on two military recruiting centers in Tennessee	16-Jul-15	Chattanooga, TN	Yes	Target: Marines and Sailor at a Naval Reserve Center and military recruiting station. Description: Assailant was armed with an AK-47 style weapon at the time of the attack. Shot at military recruiting center in strip mall and drove 7 miles to Naval reserve center. 5 service members killed, 2 wounded. Assailant was eventually killed by police.	1) He maintained blog and wrote 2 entries about Islam on 13 July [322] [159]. 2) It is suspected that Abdulazeez tried to sanitize much of his social media accounts before the attack.	No.	1) Yes, some portions of the blog were posted. However, it is suspected that his Facebook, WordPress, Photobucket, Daily Motion, YouTube accounts were all erased [159].
Farook, Syed and **Malik**, Tashfeen (and accomplice Marquez, Enrique)	San Bernardino attack	2-Dec-15	San Bernardino, CA	Yes	Target: Department of Public Health at Inland Regional Center Description: Mass shooting and attempted bombing that killed 14 people and injured 22.	1) Malik is suspected of sending at least 2 private Facebook messages to friends in Pakistan in 2012 and 2014 that were described to be "pledging her support for Islamic jihad and saying she hoped to join the fight one day" [278]. 2) On 2 Dec 2015, the day of the attack, an account suspected to be associated with Malik posted on Facebook, "We pledge allegiance to Khalifa bu bkr al bhaghdadi al quraishi" [likely referring to Abu Bakhr Al Baghdadi, leader of ISIL) [299] [270] [94]. 3) Marquez, now charged with conspiring to provide material support to terrorists, posted on Facebook on 5 Nov 2015: "No one really knows me. I lead multiple lives and I'm wondering when its all going to collapse on M[e]...Involved in terrorist plots, drugs, antisocial behavior, marriage, might go to prison for fraud, etc" [299]	No.	1) Yes, small portions of posts in a FBI affadavit [299] No other knowledge about account names etc is publically available

APPENDIX B

Case Studies of Homegrown Violent Extremism

B.1. Introduction

In this section, we present three short case studies of recent homegrown violent extremism to provide real-world context to the modeling of on- and off-line behaviors as heterogeneous data graphs, as well as to our investigative search approach. While these graph-based connections were established after the plot or attack in a subsequent law enforcement investigation, we also aspire to demonstrate the analysis possible if there were a better fusion of law enforcement and public security databases with open-source social media. In the last case study, we also present a preliminary analysis of the social media account of a violent extremist as a basis for illuminating the complexities of social media analysis and the importance of examining indicators and connections beyond the text.

B.2. Case Study #1: Christopher Cornell

Christopher Lee Cornell, an example of a recent homegrown violent extremist in the United States, was arrested by the FBI in January 2015 for allegedly planning to employ pipe bombs at the U.S. Capitol and then open fire on nearby people. From the criminal complaint [300] and other open sources [158], we employed a methodology called process-tracing to identify increasing indicators of radicalization that ultimately led up Lee's purchase of weapons to use in a planned attack. The indicators include Lee's activity on Twitter with references to jihadist recruiter Al Awlaki and other homegrown terrorists, posting of Islamic State propaganda videos, and his attack planning with an FBI confidential human source. When the signals and indicators are combined into a heterogeneous data graph, there are a

total of 43 nodes and 16 discernible classes in the class graph as shown in Fig. B.1 (produced in NodeXL [289]). These classes are devised from the political science literature on potential indicators of homegrown terrorists [116] [165] [203] [271] [285].

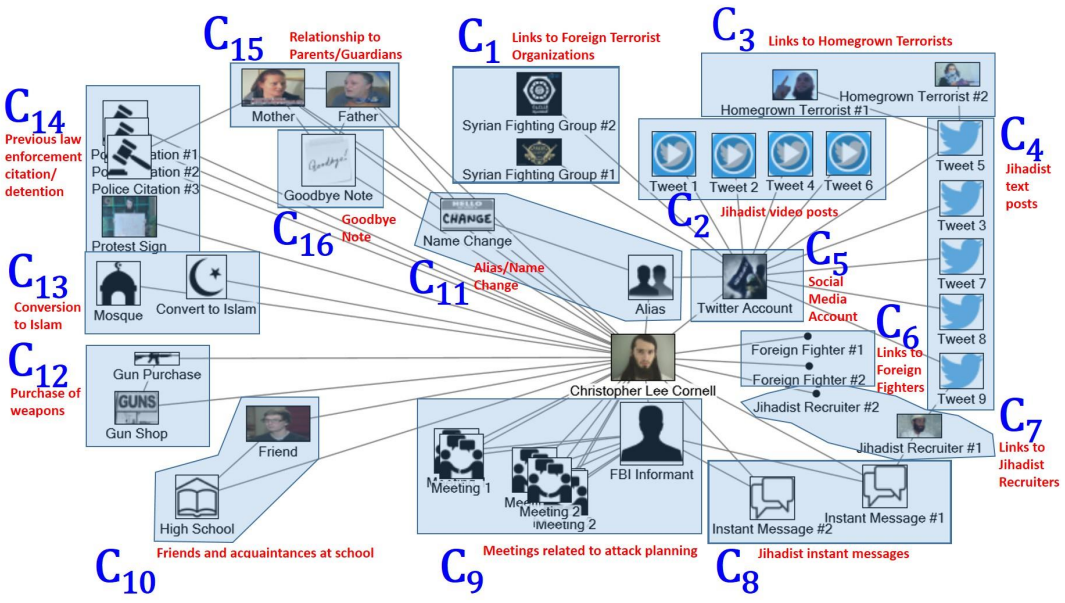


FIGURE B.1. US Capitol Attack Plot, 2015. Example class graph of Christopher Lee Cornell showing the indicators and signals of his radicalization and progress towards an attack.

B.3. CASE STUDY #2: SAYED FAROOK AND TASHFEEN MALIK

The next case study is the San Bernardino, California terrorist attack on December 2, 2015. The perpetrators Syed Farook and wife Tashfeen Malik conducted a mass shooting and attempted bombing that killed 14 people and injured 22 at the Inland Regional Center. Enrique Marquez has also been charged with conspiring to provide material support to terrorists [299]. Just as in the Lee case, critical signals here were embedded in the perpetrators' social media posts. For example, nearly a month before the attack, Marquez posted this exchange on Facebook with another user: "No one really knows me. I lead multiple lives and I'm wondering when its all going to collapse on M[e]...Involved in terrorist plots, drugs,

antisocial behavior, marriage, might go to prison for fraud, etc." [299]. In future work, we propose to include n-grams or key words as nodes in the data graph. Connections to those nodes from the social media posts serve to link the most suspicious phrases and words that may warrant further investigation from law enforcement officials.

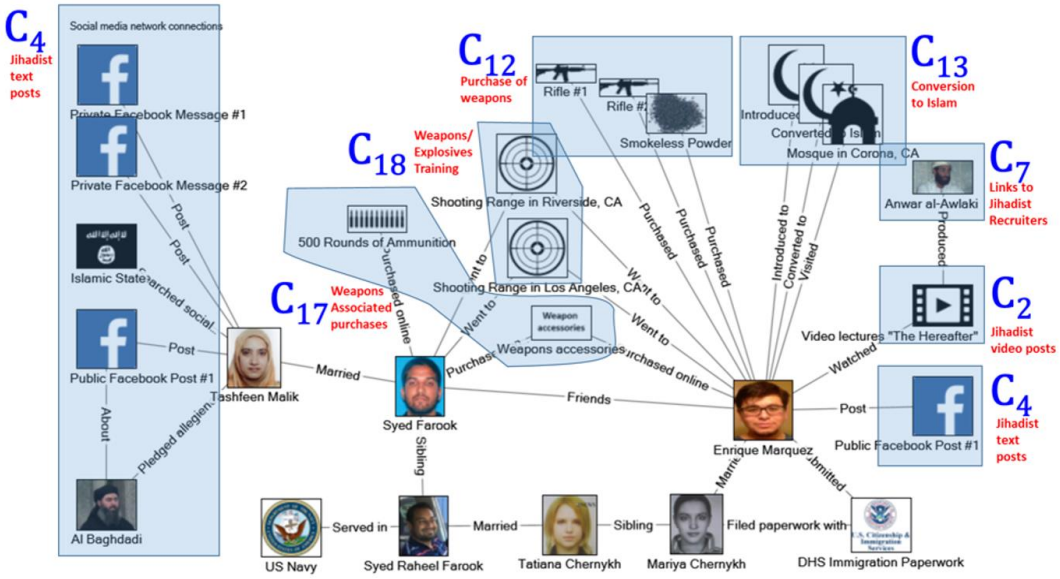


FIGURE B.2. San Bernardino Terrorist Attack, 2015. Example class graph of Syed Farook, Tashfeen Malik, and Enrique Marquez showing the indicators and signals of their collective radicalization and preparations for the attack.

It is also interesting to note that the indicators of radicalization and attack preparations were not present in just a single individual, but in all *three*. This case study points outs that conspiratorial graph patterns (match complementarity over more than one query focus node) are not addressable by current matching notions. As part of our on-going work in investigative graph simulation, we propose to identify these types of conspiracy cells through query-focus node clusters matching.

B.4. Case Study#3: Tamerlan and Dzhokhar Tsarnaev

Tamerlan and Dzhokhar Tsarnaev were the prime suspects in the explosions at the Boston Marathon on April 15, 2013 that killed three people and wounded around 250 others. The former was killed by Boston Police four days later, and the latter was taken into custody. It was later revealed that Russian intelligence had made vague warnings to both the FBI and the CIA in 2011 about Tamerlan becoming increasingly radical with connections to jihadists in Dagestan [104] [150]. The FBI subsequently interviewed Tamerlan and conducted an assessment by "check[ing] US government databases and other information to look for such things as derogatory telephone communications, possible use of online sites associated with the promotion of radical activity, associations with other persons of interest, travel history and plans, and education history," but found no evidence of terrorism [104].

In the investigation that followed the bombing, authorities discovered that both suspects had a presence in social media: Tamerlan had his own YouTube page, and Dzhokhar was active on his Twitter account [287] and less so on the Russian personal profile site called vk.com [275]. Since then, many people have tried to manually analyze these data sources[50], specifically Dzhokhar's Tweets, in order to determine whether they contained any hints of their motivations or intentions [1] [84] [250]. However, despite the recent detailed examination of the Tsarnaev brothers' social media footprint that exposed suspicious tweets and linkages to jihadist media, there does not seem to be a consensus that one would have been able to single them out in advance as an imminent danger. Of interest, then, is whether NMF has the potential to reveal the hidden topics which could indicate the seriousness of the threat.

[50]Additionally, the psychologist James Pennebaker attempted to apply automated linguistic tools and the detection of pronouns to determine when Dzhokhar may have started plotting the attack [305].

B.4.1. Preliminary social media analysis.

B.4.1.1. *Related Work.* This work is a preliminary effort to introduce NMF on Twitter data as an anomaly detection method of key indicators of lone wolf terrorism. It is part of a larger research effort to develop a semi-automated anomaly detection system of these threat indicators using multiple social media sources.

Many have been applying the fields of both machine learning and text analysis (or text mining) to social media data. Twitter, a popular micro-blog, is a widely utilized data source because of its accessibility and availability[51]. In the last few years, researchers have found macro-level, crowd-sourcing correlations of Twitter data to some real world phenomenon, including movie box office results [4], stock market performance [29], the incidence of influenza [180], and even the prevalence of depression [71]. To date, however, we have been unable to find any research that tries to make micro-level (individual) correlations to real world activity, which is of prime importance in lone-wolf indicator detection.

NMF was first introduced by Paatero in 1994 [234] as an alternative decomposition technique for producing low-rank approximations, and was made more prominent by Lee and Seung in 1999 [186]. One of the appeals of this method is the natural interpretability of the purely additive results [181], which researchers have found many applications for: image processing, hyperspectral imaging, signal processing, and even bioinformatics [119] [238]. Others have since made considerable progress in using NMF for text mining, to include topic detection and document clustering [238] on large volumes of text including medical abstracts[181], newswire posts [181], corporate emails [21].

[51]Twitter has also made content and user information available to the police [66] [261].

More recently, researchers have been examining modified NMF methods specifically to improve topic detection or document cluster for microblogs (short texts), which tend to have very sparse term-document matrices [317] [316] [303].

B.4.1.2. *Approach. Term-Document Matrix.* The term-document matrix is a method to turn text data into a representation suitable for learning algorithms and classification tasks; it is often referred to as the vector space model [189]. Each document (or tweet) is represented by a vector of terms (or words). The collection of documents forms a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, where m is the number of distinct terms and n is the number of documents. Each entry of the matrix a_{ij} is the weight of word i in document j. We simply use $a_{ij} = f_{ij}$, where f_{ij} is the frequency of word i in document j. However, there are other schemes such as Boolean weighting (where a_{ij}=1 if the word occurs in the document, 0 otherwise), or *tf-idf* (term frequency-inverse document frequency).

Nonnegative Matrix Factorization. NMF is the method of creating low-rank approximations of a matrix $\mathbf{A}$ by finding nonnegative factors $\mathbf{W} \in \mathbb{R}^{m \times k}$ and $\mathbf{H} \in \mathbb{R}^{k \times n}$. The desired rank of the approximation, $k \ll \min(m, n)$, is determined by the user and also serves as the number of (hidden) topics in the corpus [181]. The factorization is done by solving the following nonlinear optimization problem [181].

$$\begin{aligned} \min \quad & \|\mathbf{A} - \mathbf{WH}\|_F^2 \\ \text{s.t.} \quad & \mathbf{W} \geq 0 \\ & \mathbf{H} \geq 0 \end{aligned} \tag{17}$$

Unfortunately, because this formulation is convex in $\mathbf{W}$ and $\mathbf{H}$ but not both, it is difficult to find the global minimum. The matrix $\mathbf{W} \in \mathbb{R}^{m \times k}$ is particularly informative. Each column

of $\mathbf{W}$ contains the set of words found simultaneously in several documents [119] and acts as a basis vector of words for the dataset.

We also wanted to examine the effectiveness of a modified NMF procedure for short texts such as Twitter entries in order to overcome the sparsity of the original $\mathbf{A}$ matrix.

This involves first constructing the term-correlation matrix $\mathbf{S} \in \mathbb{R}^{m \times m}$ and then factoring this matrix into $\mathbf{U}$ and $\mathbf{U}^T$ [317]. In this case, the optimization problem is the following [317]

$$\begin{aligned} \min \quad & \|\mathbf{S} - \mathbf{U}\mathbf{U}^T\|_F^2 \\ \text{s.t.} \quad & \mathbf{U} \geq 0 \end{aligned} \tag{18}$$

Algorithms. There are a variety of algorithms to solve (17), including multiplicative update algorithms, gradient descent algorithms, and the alternating least squares algorithms (ALS) [20] [6]. Initially, we chose the latter due to its speed and simplicity.

It is important to point out that the NMF is not unique, meaning that both resulting factors $\mathbf{W}$ and $\mathbf{H}$ can vary depending upon the particular local minima the algorithm arrives at [181]. Furthermore, the method of initialization of $\mathbf{W}$ and $\mathbf{H}$ in NMF algorithms also has impacts on its speed and accuracy. We initially chose the simple random initialization of $\mathbf{W}$ from [186] (all that is required for ALS), but intend to test others in future work.

Algorithm 7: ALS NMF Algorithm

Input: Term-Document Matrix $\mathbf{A}$ and *max_iter*
Output: Term-Topic Matrix $\mathbf{W}$ and Topic-Document Matrix $\mathbf{H}$

Initialize $\mathbf{W}$ as a random dense matrix;
for $i \leftarrow 1$ **to** *max_iter* **do**
    Solve matrix equation $\mathbf{W}^T\mathbf{W}\mathbf{H} = \mathbf{W}^T\mathbf{A}$ for $\mathbf{H}$;
    Set all negative elements of $\mathbf{(H)}$ to 0;
    Solve matrix equation $\mathbf{H}\mathbf{H}^T\mathbf{W}^T = \mathbf{H}\mathbf{A}^T$ for $\mathbf{W}$;
    Set all negative elements of $\mathbf{(W)}$ to 0;

B.4.1.3. *Experiments. Experimental Set-up.* We accessed the publicly available Twitter dataset of "J_star," which has been confirmed to be the account of Dzhokhar Tsarnaev, one of the two alleged Boston Marathon bombers [279]. This dataset contains 1055 entries made from 10/25/2011 to 4/17/2013. Most tweets were in English, but 6 entries were written in Russian. We proceeded to convert the data into a text-document matrix using the Text to Matrix Generator (TMG), a MATLAB toolbox [296] [323]. TMG allows the removal of stop words (common, short function words), but cannot process the 6 non-English entries. The resulting matrix $\mathbf{A} \in \mathbb{R}^{2689 \times 1049}$ is a collection of 2689 terms in 1049 tweets as shown in Figure B.3. The matrix has a density of 0.00197.

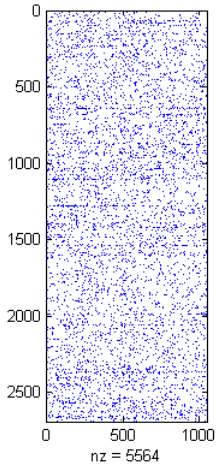


FIGURE B.3. Term document matrix of "J_star" Twitter account.

Furthermore, we also constructed the term correlation matrix $\mathbf{S} \in \mathbb{R}^{2689 \times 2689}$. However, because there were more terms than documents in the original matrix ($m > n$), the term-correlation matrix $\mathbf{S}$ resulted in a matrix that had only a slightly higher density of 0.00464 as shown in Figure B.4. While $m < n$ was the unspecified assumption in [317] (which may not be valid in many social media datasets for a single individual), we nevertheless decided to proceed with this modified NMF approach in order to examine the result.

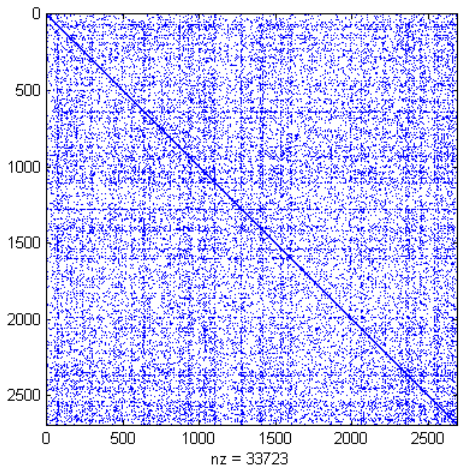


FIGURE B.4. Term correlation matrix of "J_star" Twitter account.

We performed nonnegative matrix factorizations of $\mathbf{A}$ into $\mathbf{W}$ and $\mathbf{H}$, and $\mathbf{S}$ into $\mathbf{U}$ and $\mathbf{U}^T$ using the alternating least squares algorithm [20] [6]. We initially chose rank $k = 15$, which also serves as the number of (hidden) topics in the corpus [181]. In the future, we intend to vary this parameter and analyze the effect of masking or revealing more topics.

Results. First, Table B.1 contains many of the Twitter usernames of those whom Dzohkhar directed his message (with an "@" symbol), or those whose messages he re-tweeted ("RT"). For example, "montana," "copdawholethang," "alanhungover," "drjohnyblaze," "therealabdul," "sotirop_evi," "xxjungaxx," "troycrossley," "wonkatweets," "crispylips212" are all such usernames. We had intentionally left these usernames in the term-document matrix in order to gain insight on the predominant words associated with these individuals, but after seeing the results it might have been better to remove these 'mentions' because it limits the topics to substantive words. For example, usernames alone take up 2-3 of the top 6 highest weighter terms in some topics.

Second, we focus our analysis on topic $\mathbf{W}_6$ in Table B.1 not only because it is a good example to explain what the NMF does to produce each of the k topics, but because it

also addresses the police. The expletive "fu**" appeared many times in the corpus, but it co-occurred several times with "wit" (sic for "with"), "tho" (sic for "though"), as well as "police." This demonstrates that stopword removal in the pre-processing step must be improved because it did not work on misspelled words. It also reveals that Dzhokhar used the expletive with regard to the police several times in the corpus, which we verified semantically in examining the actual messages.

Third, we are still examining the results in Table B.2 (using the modified NMF procedure for short texts on the term-correlation matrix). We notice that there are less usernames listed, but also that the terms no longer seem related to the frequency of occurrence in the data. For example, the terms in $\mathbf{W}_{14}$ in Table B.2 all came from just a single tweet. We will continue to evaluate the usefulness of the short text procedures or how they can be improved to give a more intuitive result.

Overall, the rest of the topics in both Tables seem to be typical of young millennials and do not show linkages to any of the warning behaviors for lone wolf terrorism as identified in [35].

The first set of results is from the NMF of the term-document matrix. Table B.1 shows the 6 highest weighted terms in each topic (basis) vector. The second set of results is from the NMF of the term-correlation matrix. Table B.2 shows the 6 highest weighted terms in each topic (basis) vector.

On the surface, it appears that there is nothing about the topics detected that would lead one to suspect that Dzhokhar Tsarnaev would participate in radical violence as a lone wolf terrorist. However, there are a few points of interest worth mentioning specifically related to how the NMF worked on the dataset and the insights they give to us for future work.

TABLE B.1. NMF results of term-document matrix

$\mathbf{W}_1$	$\mathbf{W}_2$	$\mathbf{W}_3$	$\mathbf{W}_4$	$\mathbf{W}_5$
baby	say	don	doesn	shit
montana	hate	amp	matter	real
nigga	things	care	tell	hate
copda-wholethang	turn	sleep	hell	drjohnny-blaze
today	guess	alanhung-over	met	fake
food	idiot	think	feel	niggas
$\mathbf{W}_6$	$\mathbf{W}_7$	$\mathbf{W}_8$	$\mathbf{W}_9$	$\mathbf{W}_{10}$
fu**	kid	money	therealabdul	xxjungaxx
wit	wavyy	dont	evi	lol
tho	tho	need	sotirop	troycrossley
police	tsar	sex	fast	bro
wonkatweets	basiklee	ahaha	yea	bad
ahaha	yea	honey	aha	show
$\mathbf{W}_{11}$	$\mathbf{W}_{12}$	$\mathbf{W}_{13}$	$\mathbf{W}_{14}$	$\mathbf{W}_{15}$
wait	dudes	beautiful	love	http
twitter	dam	girl	heart	earthpix
text	hit	guy	dont	amp
complain	think	crispylips-212	meant	find
texting	guy2	thing	twins	amazing
guy2	heart	woman	cats	close

B.5. Conclusions and Insights on Real-World Investigative Search

. These two case studies provide a real-world context to the challenges of applying graph pattern matching to aid in the search for homegrown violent extremists. In particular, constructing the query is problematic because almost all existing graph pattern matching approaches rely on certainty in the query and do not have a categorical node labeling structure for indicators. Between the two case studies, one can discern the variability in the presence of indicators as well as the significant number of nodes which might be classified as 'individually innocuous but related activities.' Investigative simulation seeks to address these issues by allowing for partial matches of a comprehensive indicator query, suggesting node categories for investigative searches, and pruning or augmenting the match relation to produce sensible matches and less false positives.

TABLE B.2. NMF results of term-correlation matrix

$\mathbf{W}_1$	$\mathbf{W}_2$	$\mathbf{W}_3$	$\mathbf{W}_4$	$\mathbf{W}_5$
stop	baby	chicken	don	fake
street	montana	egg	money	eyes
cleaning	today	letting	need	isn
noparking-spacesleft	guy	decided	head	mouth
rest	future	hatch	themike-derby	amp
streets	black	surface	thing	distinguish
$\mathbf{W}_6$	$\mathbf{W}_7$	$\mathbf{W}_8$	$\mathbf{W}_9$	$\mathbf{W}_{10}$
kids	haven	game	coming	phones
tell	room	ohjeyy	jjr	stopped
entertain-ment	shits	cheater	undeniable	guys
faces	seen	french	school	hit
finish	wear	request	crucial	attention
permit	didn	retard	goingham	paying
$\mathbf{W}_{11}$	$\mathbf{W}_{12}$	$\mathbf{W}_{13}$	$\mathbf{W}_{14}$	$\mathbf{W}_{15}$
video	person	zombie	den	kid
games	1year	apocalypse	aggy	wavyy
jordans	aging	dream	andthatsfast	tsar
meekmili	appears	1st	clicked	smh
sports	brooke	base	spanish	basiklee
wanting	greenberg	triggered	stds	dat

There are several areas we intend to pursue in the future. First, we propose the functional decomposition of a semi-supervised threat topic detection system in texts as shown in Fig. B.5. The threat lexicon filter (provided through supervised machine learning) is designed to separate out the possibly relevant documents from the irrelevant by examining the presence of suspicious keywords. An analyst can visually inspect for any temporal trends in keyword frequency, and then select the subset of documents for topic detection. The resultant topics are again inspected by an analyst for suspiciousness and identified for further investigation. Within this system, we will continue to experiment with more short text NMF techniques to this dataset to determine if there is an improvement in the results. This includes testing a previously-proposed Ncut weighting scheme for the term-document matrix [317] and the Hadamard Product of Similarities [303]. We will also apply different algorithms as well as

explore the affect of the deliberate imposition of sparsity on the clarity of the detected topics [137].

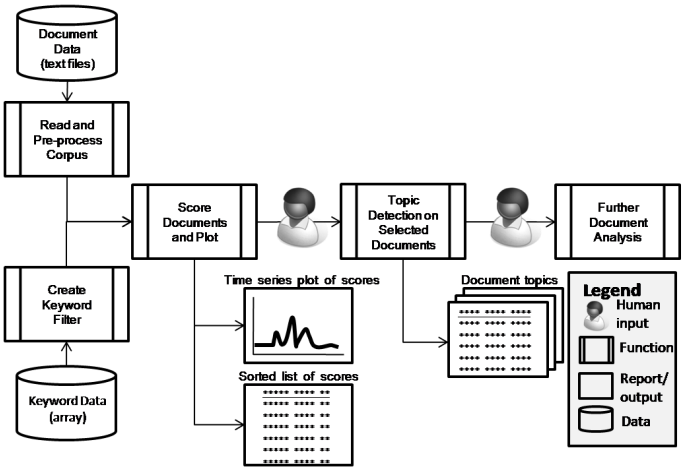


FIGURE B.5. Functional decomposition of the topic detection system.

Second, we intend to examine the possibility of systematically incorporating *special texts and context* into the anomaly detection. While we observe that the topics derived from just the tweeted words may not contain a detectable signal for the lone wolf terrorism, we suspect that the signals could be hidden in special text and non-text context of the Twitter corpus. This is evidenced by a few examples of suspicious tweets that our topic detection methods did not identify. 1) In March 2013, Tsarnaev wrote "September 10th baby, you know what tomorrow is. Party at my house!," which seems to imply a celebration about the September 11th terrorist attacks [84]. 2) In April 2012 Tsarnaev wrote in Russian "I will die young" [1]. 3) Finally, Tsarnaev had been following a Twitter user named "Al_firdausiA" (translated: "the highest level of Paradise, Allah willing"), who had encouraged followers to listen to Anwar Awlaki, an American-born al Qaeda terrorist [287]. If an anomaly detection system could pick-up on these special contexts in addition to the words in the Twitter corpus, perhaps it could better signal the potential threat of violence.

Beyond this, our broader research effort seeks to apply anomaly detection analysis through social media of all forms targeted violence, of which 'lone wolf' terrorism is only one type. Targeted violence is a "violent incident where both the perpetrator and target(s) are identified or identifiable prior to the incident" [252]. As in [203], we hope to also consider detection of potential work place violence, campus and university violence, school shootings, adolescent and adult mass murder. We also seek to develop a larger, semi-automated anomaly detection system of these threat indicators using multiple social media sources (such as Facebook, MySpace, YouTube) and mediums (such as text, videos, and network connections).

APPENDIX C

Codebook Excerpt from Klausen's Radicalization Trajectories Dataset

The following is a direct excerpt from [168] of the definitions of the 27 features we utilized in our data analysis.

Year of Birth This field is a numeric variable that tracks the year in which an individual was born. If the exact year is known and available, enter the full year. If a source document only provides an individuals age, subtract the age from the year in which the source was published. If age and year of birth are both unknown, leave the field blank.

Convert Date This field tracks the approximate date at which an individual marked converted to Islam. If information is not available, enter the year in which an individual claimed their conversion to Islam. If there is no available information pertaining to conversion, but the individual did convert to Islam enter Unknown. This should be left blank if the individual is not a convert to Islam. Dates of conversion to other religions are not included here. Year of conversion and year of radicalization should not be assumed to be identical.

Pre-Radicalization: The initial stage in the radicalization trajectories is Pre-Radicalization. This is hallmarked by searching behavior indicative of cognitive opening.

Disillusionment This field tracks the first verifiable date at which an individual began to overtly express disillusionment with world affairs, religion, or Western society. Expression of disillusionment could be made evident in various ways, either virtually or in real-life. Possible cues include disdain towards mainstream societal trends, opposition to Democracy

or political ideals, expressed anger in regards to the oppression and behaviors of other Muslims, etc. Cues used to code this field are not clearly ideological. Record the first instance if an individual takes part in a political protest on a certain date, but it is evident that the individual has taken part in prior protest activity, investigate further to ascertain when the behavior began taking place.

Trauma This field tracks a discrete occurrence of adverse personal circumstances yielding feelings of self-dissatisfaction or introspection. To be coded in this field, the event must take place at a set point in time (ex., death of a loved one, personal injury or illness, separation from spouse or of parents, etc.) The event must have an impact on the individual.

Personal Crisis This field tracks the date range during which an individual experienced adverse personal circumstances yielding self-dissatisfaction or introspection. Personal crisis must be catalyzed by continuous, prolonged phenomenon (ex., incarceration, drug addiction, unemployment, homelessness, etc.). Personal crisis might also entail a more covert pattern of "searching behavior," such as seeking out new a new ideology (though not necessarily focused on Islam).

Information Seeking This field tracks the earliest known date on which an individual began actively seeking out sources of Jihadist information (ex., downloading or procuring Jihadist literature, opening dialogue with extremist figures, seeking out new friendship on a doctrinal basis, seeking out a more radical place of worship, etc.). Information seeking might take place virtually through an online platform, in a real-life community setting, or both.

New Religious Authority This field tracks the most exact date on which an individual began actively seeking out and adhering to new figures of religious authority. This might be indicated by overt signs (ex., direct support via personal communication), or by more passive

measures (ex., attending, reading, or listening to material produced by spiritual authority figures). Unlike "information seeking" as described in the previous column, seeking out new authority focuses on a specific figure and entails ideological motivation on part of the individual. Nevertheless, the two are not mutually exclusive.

Stage 1: The first stage of radicalization is denoted by detachment from previous life and combined with changes to daily life, combined with experiencing revelations regarding the ideology.

Ideological Rebellion This field tracks the most exact date on which an individual began acting out against and detaching from formerly central life figures. To be coded in this field, the action must be idealized, ex. picking fights with Imams, mosque members, or parents in regard to increased piety.

Lifestyle Changes This field tracks the first outward indication of an individuals movement towards radical ideology, such as through change in clothing (ex., beginning to wear kaftan or niqab; grows a beard or starts wearing trousers cut above the ankle (males only)), abstention from food/substances deemed haram, or sudden overt changes to expressions of religious piety. Please note that these examples are examples and overt behavioral life styles changes vary widely based on an individual's prior activities and interests.

Note: the following three (3) fields are not mutually exclusive of one another, but track different versions of disengagement from mainstream employment or study. As such, each field applies to different subsets of individuals. "Dropout" is only applicable for individuals who are enrolled in an educational institution, while "underemployment" only applies to people who are employed.

Occupational/Educational Disengagement This field tracks the most exact date on which an individual began to disengage from responsibilities at school or work. This field requires discretion and inference in coding, and will often not be immediately apparent. Possible cues include, but are not limited to failure to adhere to occupational or educational obligations, suddenly falling grades, being placed on academic or occupational probation, and excessive hostility towards colleagues or superiors in professional environment.

School Dropout This field tracks the most exact date at which an individual officially disassociated from their educational responsibilities. Possible indicators include, but are not limited to withdrawal from classes or educational program and dismissal from occupation or education program.

Underemployment This field tracks the most exact date on which an individual commenced employment at a job lacking skill or trade knowledge requirements, and sought such employment on an ideological basis. Not all service or manual workers are coded in this field. Only enter a value if the individual sought underemployment under the premise that such work would not interfere with religious beliefs or obligations, or did so in order to travel abroad or to provide support to an extremist organization.

Da'wah Virtual This field tracks the most exact date on which an individual began disseminating extremist material in an online setting (ex., publishing or recirculating material on social media). To be considered in this field, the dissemination must be active. For example, an individual who reads material published by extremist hubs but does not republish that information would be coded under "information seeking," whereas an individual who republishes that information would be considered in this field.

Da'wah Real Life This field tracks the approximate date at which an individual began actively taking part in the dissemination of extremist material. This might include encouraging friends and family to adopt more orthodox beliefs, handing out literature in public places, espousing radical beliefs to a group of people, etc.

Stage 2: The second stage of radicalization is the time during which an individual leaves home to become closer to a peer group of like-minded individuals. They seek out ways to demonstrate their commitment to the new ideological community and its mission.

Epiphany This field tracks the most accurate date at which an individual starts broadcasting a personal revelation or proclaims a revelation that participation in violent Jihad is a necessary and imperative individual obligation.

Peer-Immersion This field tracks the most exact date on which an individual began seeking out and associating with a group of like-minded individuals, such as through cohabitation or by spending increased amounts of time with the group. Peer-immersion can take place either in a virtual setting or in real-life. Peer-immersion is often preceded by a disassociation from former social settings, sometimes made evident through "lifestyle changes" or "rebellion." The group with which a radicalizing individual associates is often perceived by the individual as more ideologically knowledgeable than the subject.

Domestic Physical Training This field tracks the date at which an individual first took part in domestic training for militant action, either by participating in training exercises with a group of like-minded individuals (ex. shooting practice, weight training, etc.), or by otherwise cultivating useful battlefield skills (ex., enrollment in nursing/EMT classes). Training in this category is typically general in nature, rather than being tied to a specific organization, location, or group.

Marriage Seeking This field tracks the most exact date on which an individual began actively seeking, or expressed interest in seeking, a like-minded spouse following the precepts of the Jihadist ideology.

Societal Disengagement This field tracks the most exact date on which an individual began actively dissociating from political or societal obligations. Unlike "disillusionment," this field denotes active abstention from societal processes, generally subsequent to expressed disdain. Societal disengagement occurs due to either belief that such engagement goes against religious ideals (ex., not voting under the premise that democracy is forbidden to Muslims), or associated rules of the extremist belief system such as that such engagement supports action against like-minded persons or groups (ex., paying taxes is "forbidden").

Desire for Action This field tracks the most exact date on which and individual first expresses desire to take part in extremist action (ex., foreign fighting, financial support, domestic plot, etc.), but occurs prior to the development of concrete plans.

Stage 3: The final stage of radicalization is when the individual attempts or enacts violent action, or joins a terrorist group abroad or attempts to join a group. They actively support another person carrying out violent action on behalf of the ideology.

Non-Violent Support This field tracks the date at which an individual first lent tangible non-violent support to a terrorist group or organization (ex., by fundraising or smuggling materiel).

Joins Foreign Insurgency This field tracks the most exact date on which an individual travels abroad, or is known to have traveled abroad with the intention of taking part in an overseas extremist insurgency.

Issues Threats This field tracks the most exact date on which an individual issues violent threats, either broad or specific, against another individual or group of individuals. Threats considered in this category must occur subsequent to radicalization and must be made in connection with an extremist organization or ideology.

Steps Towards Violence This field tracks the most exact date on which an individual began actively preparing to carry out action on behalf of an extremist organization or ideology.

Date of Criminal Action This field tracks the specific date in which an individual partook in terrorism-related activity. Terrorism related activity includes, but is not limited to, religiously motivated violent acts, violent intent, fraud, smuggling, robbery, proselytization, incitement, or travel occurring after the individual has been radicalized. Enter the last chronologically occurring date from the following:

- Year of plot. In the case that an individual has partaken in multiple terrorist activities, use the final date.
- Permanent relocation abroad, but only when that relocation is related to foreign fighting or joining an overseas insurgency.
- Death. Do not include date of death when not related to terrorist activity (ex., natural causes).

Arrest Date This field tracks the specific date in which an individual was arrested and taken into custody by law enforcement officials for involvement in terrorism-related activity. Terrorism related activity includes, but is not limited to, religiously motivated violent acts, violent intent, fraud, smuggling, robbery, incitement, or travel occurring after the individual

has been radicalized. This may also include being charged in absentia in the case of foreign fighters.

Sentencing Date This field tracks the year in which an individual received a penal sentence relating to involvement in terrorist activity. This is only applicable when the individuals arrest and trial related to terrorist activity resulted in sentencing. Sentences to deportation or being sentenced in absentia are reflected in this field as well. This variable should be left blank if the sentencing has not yet occurred, but it is scheduled for a future date.