Document Title:	Dynamic, Graph-Based Risk Assessments for the Detection of Violent Extremist Radicalization Trajectories Using Large Scale Social and Behavioral Data
Author(s):	Anura P. Jayasumana, Jytte Klausen
Document Number:	303409
Date Received:	November 2021
Award Number:	2017-ZA-CX-0002

Final Research Report

Federal Award Number	2017-ZA-CX-0002
Project Title	**Dynamic, Graph-Based Risk Assessments for the Detection of Violent Extremist Radicalization Trajectories Using Large Scale Social and Behavioral Data**
Principal Investigators	Anura P. Jayasumana Professor, Electrical & Computer Engineering Colorado State University Anura.Jayasumana@Colostate.edu 970-491-7855 Jytte Klausen (Co-PI) Professor, Department of Politics, Brandeis University klausen@brandeis.edu 781-736-2762
Project Period	04/01/2018 TO 03/30/2021
Award Amount	$731,000
Award Recipient Organization	Colorado State University Office of Sponsored Programs Fort Collins, CO 80523-2002
Submission Date	June 30, 2021

Dynamic, Graph-Based Risk Assessments for the Detection of Violent Extremist Radicalization Trajectories Using Large Scale Social and Behavioral Data.

Authors: Benjamin Hung, Anura Jayasumana, and Jytte Klausen

1. Summary of the Project

A. Major Goals and Objectives

The research aimed to produce a new integrated computational technology that can mine, monitor, and screen for the occurrence of behaviors associated with dangerously escalating extremism in large heterogeneous databases and provide early warnings of individuals or groups on behavioral trajectories toward extremist violence. The goal for the research was to harness data science methodologies to enable rapid, semi-automated support for law enforcement analysts and social science researchers to produce structured behavioral indicator profiles from text sources (such as investigator notes, information shared via fusion cells, and suspicious activity reports).

The research produced improvements in the analysis methodology of radicalization pathways, and software to support such analysis. The project has resulted in a library of software routines and computer aided tools that aid the compilation of information from documents and for radicalization indicator detection to pattern detection in networks to visualization.

The project integrates new advances in data science techniques with an evidence-based behavioral model of radicalization leading violence and terrorism-related crimes. The project is in this sense a hybrid of data science and sociology. Three important elements were brought together: an understanding of extremist radicalization as a process during which the radicalizing individual undergo overt behavioral changes, the development of a trained computational algorithm capable of detecting structured and unstructured data known to signify such behaviors, and methods for analyzing and visualizing the data.

To support the development of the computational techniques, the social scientists faced four challenges: First, a robust evidence-based social scientific model of the key features of extremist radicalization had to be tested. Second, a method had to be developed for extracting and tagging text segments containing cues to the overt behaviors associated with the radicalization process. A key aspect has been the pursuit of trained machine learning (ML) models with natural language processing (NLP) techniques for the information extraction task of identifying radicalization indicators present in text documents. Third, to scale the analysis process to cover large volumes of information, extracted data has to be organized as knowledge networks, stored in efficient and appropriate databases. And forth, the knowledge network has to be mined based on complex queries driven by investigators resulting in products such as radicalization pathways.

B. Research Questions

Violent extremism is a marginal, low base rate phenomenon. In probability and statistics, the term "base rate" refers to the population probabilities of exhibiting a certain class or category. If a particular class or a category of activities or pathologies is rarely observed in the general population, the ability to study—or treat—the phenomenon using traditional methods is impaired. Terrorist jihadist extremism falls into this category of rare and complex phenomena.

Rather than search for a profile of extremism—a futile endeavor—research is now focused on explaining how people come to embrace terrorist extremism. The basic idea is that radicalization occurs through a process of deepening engagements that can be observed in changing overt behaviors. Psychologists tend to take the long view. Priming a normal person to accept a moral and personal imperative of self-sacrifice takes time.

Instead of trying to profile and identify at-risk populations, the study attempted to generate a dynamic, evidence-based assessment model of radicalization trajectories of homegrown militants inspired by the jihadist ideology.[1] The concept of a radicalization trajectory implies that an arc exists leading the

---

[1] The methodology and preliminary findings were presented in Jytte Klausen, Selene Campion, Nathan Needle, Giang Nguyen, and Rosanne Libretti. "Toward a Behavioral Model of 'Homegrown' Radicalization Trajectories." *Studies in Conflict & Terrorism* 39, no. 1 (January 2, 2016): 67–83. doi:10.1080/1057610X.2015.1099995.

perpetrator from entertaining extremist ideas to action, and that there is a somewhat predictable pathway from a normal, if perhaps angry state, to the perpetration of a violent attack in the name of the ideology. This insight has tested out well. A study released by the FBI in 2019 of 52 ideologically-motivated lone offenders concluded that "they traveled down the same observable and discernable pathways to violence as other attackers."[2] Our research broadly confirms the findings of the FBI study for a larger and diverse group of ideologically-inspired violent extremists.

From the perspective of technology development, the questions were as follows: Can tools that rigorously examine and account for the activities of close associates better predict the likelihood that an individual would engage in violent extremism? 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? Can computationally efficient tools be used to mine these databases for the specific purposes of monitoring and screening for individuals and small groups posing a significant risk for violence? The fields of computer engineering and computer science as well as many disciplines of science are rich with computer aided tools aimed at tasks such as data mining, data analysis, mathematical modeling, simulation, image and voice recognition and natural language processing. Such a level of computer aided support or automation is not yet available for addressing to social and political sciences related problems. The basic premise of the technology development aspect of the project was that such tools can be developed to help social scientists and investigators scale up their tasks. Of notable interest were technologies that may help analysts to mine text sources for items of interest, representation of such data for further analysis, and analytics for extracting information to support decision making.

[2] Lauren Richards, Peter Molinaro, John Wyman, and Sarah Craun. "Lone Offender: A Study of Lone Offender Terrorism in the United States." *FBI Behavioral Threat Assessment Center* (2019): 3.

C. Research Design, Methods, Analytical and Data Analysis Techniques

The research was organized as two tightly integrated thrusts, data collection and analysis led by the Brandeis University team and technology development led by the CSU (Colorado State University) team. The technology development was driven by the data collection and analysis requirements while the computer aided tools and the results were immediately available for enhancing data collection and analysis.

Data Collection

The Brandeis University team undertook to develop the real-world data used for the project. Three types of data were collected: (1) demographic data about known jihadist terrorism offenders from the United States and the United Kingdom and a small case control dataset comprised of the so-called "Involuntary Celibate," also known as Incels; (2) information about observed behavioral changes and the timing of these behaviors, and (3) a large collection of text data indicative of cues to such behaviors.

The behavioral analysis methodology used by the study is demanding of data. Twenty-four different behavioral indicators were used to infer the timing of a person's embrace of violent extremism. Timestamped cues to such behaviors indicative of growing extremism are required to chart the radicalization process. Demographic information and social network connections are also coded for all subjects to support the analysis of the distinctive pathways to violent extremism taken by different social, racial, and ethnic subgroups within the population of offenders and the influence of peers.

Data were collected from public text documents ranging from news articles to court documents (e.g., indictments and affidavits files in support of a criminal complaint). The type of documents used was chosen on the assumption that the text style resembles case notes filed by law enforcement agents. Coders were trained to follow a detailed codebook that itemizes how to draw inferences from texts and how to translate the information into numerical values.

The Brandeis team provided a secure web-based infrastructure support for the research collaboration. The servers support secure web-based experimentation, exchanges of data and analyses between the collaborators. The data collection portal was specifically designed using a standard relational database management system and query language (SQL) that supports the coordination of structured (numeric) and unstructured (text) data. To be included in the database collection, the study subjects have either (1) been charged and convicted of a terrorism-related crime, (2) died committing a terrorist act, or (3) self-identified as a foreign fighter on social media and subject to public debate. Using sources drawn from prosecutions and statements made by the subjects themselves, and bystanders (family, friends, co-conspirators) detailed forensic biographies of 299 American and British "homegrown" jihadists were developed.

Training data to support the development of a machine learning model of the radicalization process—and the many variations observed in the paths taken to commit a terrorism-related crime—were extracted from public text documents ranging from news articles to court documents such as indictments and affidavits files. The reliance on public documents has the benefit that the research does not conflict with privacy protection standards and may be freely shared and used for experimental research. The downside is that the research is exceptionally time consuming. The culling of relevant texts and the labelling of text segments indicative of cues to a specific extremist behavior that can be used as training data are done by research assistants. On average, it took ten hours to complete the coding of a particular subject's radicalization profile.

Technology Development

CSU team undertook the development of a comprehensive set of software tools aimed at assisting analysts and researchers by automating major tasks related to data collection and analysis. A proof-of-concept prototype system was completed that integrated multiple software modules for human-in-the-loop analysis. Software was developed using a Python[3] language. To ensure the use of leading-edge advances in data sciences and machine learning for radicalization research, open-source software libraries were integrated

[3] Python is a widely used general-purpose computer programming language with an expansive library of open source data analysis, web frameworks, and testing instruments.

to our system whenever possible[4].

Natural Language Processing (NLP) techniques, including those for recognizing named entities together with pronouns referring to them in text, matching text based on specific linguistic rules, and for classification of text segments, were customized to parse text documents for radicalization indicators. Artificial neural network models were trained using hundreds of labeled sentences from the Brandeis team to find the presence of radicalization indicator classes. Examples of indicator classes specifically related to radicalization are as follows: Seeking Religious Authority, Peer Immersion, Desire for Action, Joins a Foreign Terrorist Organization, Issues a Threat and Steps Towards Violence.

CSU team developed a set of graph database[5] tools, including a graph pattern matching technology, INSiGHT (Investigative Search for Graph Trajectories), to help identify individuals or small groups conforming to a radicalization pattern and to follow the formation of radicalization trajectories over time. INSiGHT functions were prototyped using an open-source version of the graph database Neo4j[6]. INSiGHT, for example, can respond to particular forms of user queries related to a person of interest's radicalization which can identify exact or close matches in the knowledge base. Software also supports radicalization trajectory analysis by generating order of occurrence plots for behavioral indicators. Order of occurrence implies the chronological position of a particular indicator in each trajectory. Among the computer tools developed is software for aggregated radicalization pathway analysis and visualization. We were able to, for example, visualize pathways of radicalization among the expanded set of 299 US and UK terrorist profiles in the Western Jihadism Project database. The profiles contained demographic information and the associated timestamped publicly known markers of their radicalization trajectories leading up the

---

[4] Examples of open-source software libraries used include spaCY (a library for natural language processing available at spacy.io), BERT (Bidirectional Encoder Representations from Transformers) language model from Google AI, and Jupyter notebook (available from jupyter.org).

[5] Graph databases are a recent development that enables the storage and analysis of highly connected heterogeneous data. This is a much more powerful approach compared to traditional relational databases for radicalization problem with focus on networks (i.e. graphs), e.g., jihadist social networks and the emergence of patterns therein.

[6] Neo4j is a graph database management system developed by Neo4j, Inc.

criminal act (or acts) that caused their arrest, their death, or their successful departure from their home country to join a foreign terrorist organization.

User-friendly human interfaces were developed for some of the tools which allowed the team at Brandeis to use the tools for their analysis. In most instances, the two teams worked together to use the tools under development immediately for data analysis which also resulted in timely feedback to ensure the accuracy and utility of tools under development.

Research Design

The demographic data were analyzed using traditional social science quantitative methods to identify frequency patterns and correlations between demographic variables—specifically, age and sex. The timestamped behavioral data were analyzed using novel graph techniques to assess the order in which offenders' behaviors change and the most common pathways leading to terrorism-related action. The most salient behavioral indicators of growing extremism were used in the ML model. From their research, the Brandeis team provided annotated raw text data from the source records that was used by the CSU team to train the Machine-Learning (ML) programs under development.

The process of applying supervised ML to a real-world problem is demonstrated in Figure 1. Starting from the top in the diagram, the problem refers to the dynamical radicalization trajectories model developed by Dr. Klausen and her team in previous research. Step 1 is the identification of text data that the Brandeis team does routinely as part of the research required to develop forensic biographies with timelines for the radicalization profiles of known offenders. Steps 2, 3, and 4 are carried out by the CSU teams as part of the development and testing of ML models. The last two steps involve assessing the output files produced by the algorithms for "ground truths" and are the responsibility of the Brandeis team.

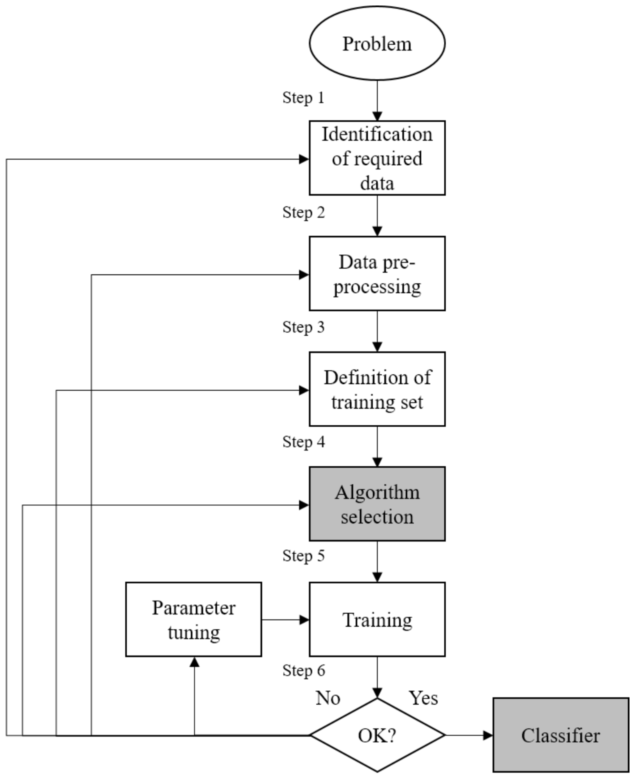

Figure 1. The step-wise process in supervised machine-learning.

Note: Developed from S.B. Kotsiantis, Supervised Machine Learning: A Review of Classification Techniques. *Informatica* 31 (2007): 249-268.

The project relied on a "Human-in-the-loop" (HITL) process to train and develop a machine learning model capable of mapping terrorist radicalization. This effort not only enabled the Brandeis team to more rapidly code biographical radicalization profiles to augment its database, but also builds towards a capability for law enforcement and intelligence agencies to capture key behavioral indicators among analyst and investigator notes and reports. It also enables the transformation of textual data into a graph database of individuals with their social and behavioral cues on which we can apply advanced graph pattern matching techniques to identify persons along a radicalization trajectory. In essence, the HITL process involves a reiterated experimental process in which a computational algorithm is written and rewritten to pick up keywords, phrases or sentences in texts and sort and score those text segments for their relevance to the text segments known to be highly relevant to the social and behavioral model of radicalization trajectories. The HITL process is a highly structured loop between human researchers and the computational logic.

The Brandeis team provided the “human” input in the reiterated experimental process. This entails, first, the development of structured data. The relevant data include, first, the co-PI’s research that has established a feasible and reliable dynamic individual-level psycho-social model of salient behavioral indicators of growing extremism and, second, a list of cues to such behaviors.

The HITL experiments further require a comprehensive annotated library of keywords, phrases, and sentences related to different behavioral indicators of progressive radicalization. To be useful for the “training” of search algorithms, the keywords need to tap distinct and also frequently occurring behavioral indicators of growing radicalization.

The first step in the HITL process was providing annotated (or labeled) text data as training data to a machine learning algorithm, which, once trained, is subsequently applied to a series of texts for the purpose of detecting and scoring text segments in the texts that are supposedly are relevant (or not relevant). Next, the “human” in the experimental loops reviews the output and assesses the accuracy of the computational output. The algorithm is fine-tuned in a second—or third or fourth—trial. New data and variables are added and tested with the aim of improving the capabilities of the algorithmic model to intelligently discern relevant and irrelevant information. This is a time-consuming process of trial and error.

2. Participants and Other Collaborating Organizations

Dr. Jayasumana, who was the lead PI, is a Professor of electrical and computer engineering at Colorado State University, where he also holds a joint appointment as Professor of computer science. His current research interests include detection of emergent patterns such as radicalization profiles in knowledge networks, network analytics, detection of weak distributed patterns in networks, mining of network-based data, networking applications of machine learning, and Internet of Things.

Dr. Klausen, who was the co-PI, is the Lawrence A. Wien Professor of International Cooperation at Brandeis University. She is an affiliate of the Center for European Studies at Harvard University and associate fellow of the International Centre for the Study of Radicalization (ICSR) at the Department of War Studies, King’s College London. In 2006, she founded the Western Jihadism Project, a data collection and archive focused on supporting comparative and historical research of Islamist extremism integrating network studies with demographic analysis of terrorism offenders from North America, Western Europe and Australia.

Senior personnel included Dr. Benjamin Hung is a Post-Doctoral Associate at CSU and an active-duty US Army officer and Operations Research and Systems Analyst. He has operational and practitioner experience in military law enforcement, counter terrorism operations, and defense intelligence analysis. His research interests include pattern detection in social systems, social network analysis, and computational social systems.

Ms. Priyanka Renugopalakrishnan (MA '18, Brandeis University) was the senior staff research assistant until she left the project to take up a position with the United States Government. Ms. Rosanne Libretti (MA '19, John Jay College) took over the position and stayed with the project to its conclusion. Teams of students assisted the principal investigators.

Shashika Muramudalige was a Ph.D. candidate in Electrical & Computer Engineering Department and the Graduate Research Assistant for the project. He is finalizing his Ph.D. dissertation based on this research and is expected to defend it early in fall 2021.

3. Outcomes

A. Accomplishments

A major accomplishment of the CSU team in this research is the proof-of-concept software-based framework called INSPECT (Investigative Pattern Detection Framework for Counterterrorism). The architecture and the functionality of INSPECT is depicted in Figure 2. INSPECT is a human-in-the-loop system that mitigates challenges due to massive data volumes and the dynamic and complex behaviors of extremists and extremist groups. It assists and empowers investigators, scientists and researchers by automating multiple tasks, filters out unrelated information, and provides tools for extracting higher-level structures in data. INSPECT automates multiple tasks for large-scale mining of detailed forensic biographies, formation of knowledge networks, and querying such data for behavioral indicators and radicalization trajectories. It has been developed for human-in-the-loop mode of investigative search and evaluated on data related to known domestic jihadists extensively studied by social scientists. It consists of the following major functional components:

1) NLP to identify radicalization indicators in text sources: Use of NLP dramatically improves the ability to handle large volumes of news articles, court documents, and reports, etc. to identify different behavioral indicators corresponding to different stages of radicalization.
2) Graph databases: The extracted data from text sources are innately captured in the form of a knowledge network that consists of individuals together with their behavioral indicators, as well as links connecting individuals, organizations and behavioral indicators. This dataset is highly linked and storing and processing such data is a challenging task. Graph databases, designed for pattern-based querying over huge volumes, contains many features for mining such networks.
3) Query graph formation: With years of experience observing the extremists' behavior, social scientists have studied the diverse patterns of radicalization. We use their empirical knowledge to model query graphs representing the behavior patterns of interest.
4) Investigative graph search: We have developed and implemented a set of algorithms to explore potentially risky individuals and groups on knowledge networks. Graph searches are performed as custom queries to the graph databases, which enhances the efficiency and the scalability of data processing while utilizing the database features.
5) Synthetic data generation: We propose a novel synthetic data generation technique to mimic the behavior of extremists. Such traces help improve the amount of information available for training human coders and have wide applicability in domains where the available datasets are small, sparse, or insufficient. It can also be used to generate publicly releasable synthetic radicalization profiles that preserve the privacy of real data.

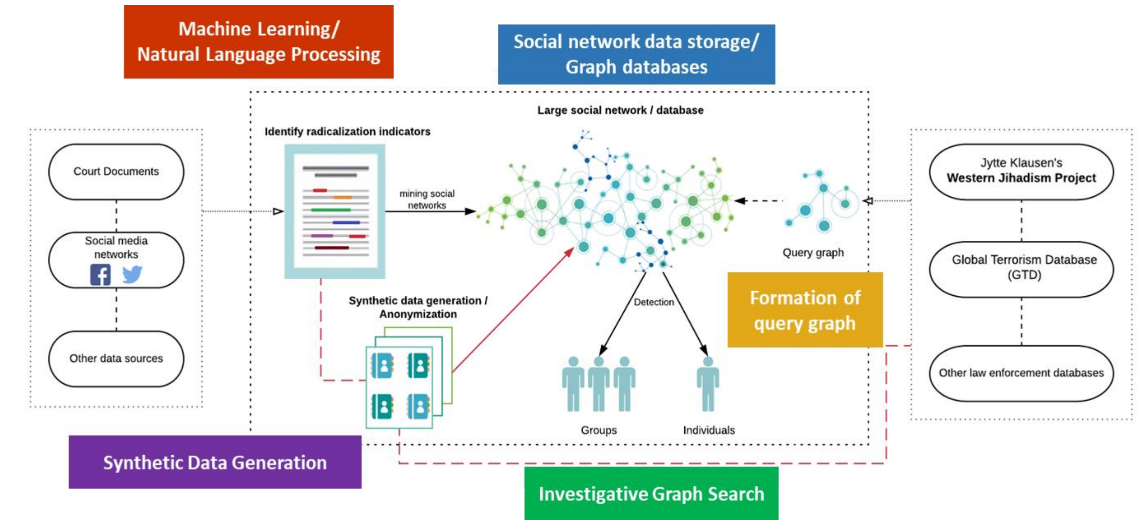

Figure 2. INSPECT computer-aided human-in-the-loop platform for detection of violent extremist radicalization trajectories developed by the project group.

Note: Figure 2 depicts major building blocks (1) Integrated set of tools for Natural Language Processing, (2) Knowledge network database, (3) Query Graph Generation, (4) Investigative graph mining tools, and (5) Synthetic data generation for obfuscating personally identifiable information for sharing data and model testing (CNRL 2021).

The INSPECT software developed is a proof-of-concept platform to aid analysts and researchers in investigative searches. It demonstrates how advances in NLP, ML, graph data bases and data mining can be adapted for use in computational social sciences. Synthetic data generation technique developed for INSPECT has already been used for providing enhanced solutions for phishing website detection and video-trace generation for privacy preservation.

B. Results and Findings

Demography

Demographic variables can only tell us so much. To repeat, there is no demographic profile which can predict who is likely to become a deadly terrorist. Nevertheless, age is highly relevant for who does what. The stereotype is that jihadism is a young person’s crime, like gang membership. In fact, jihadist terrorism is generally an adult crime.[7]

It is axiomatic in criminology that the incidence of crimes among young men, and violent crimes, specifically, peaks in late adolescence and declines swiftly after the age of thirty. The literature on the age-crime relationship is voluminous. It is widely agreed that younger men (and to a lesser extent women) are more likely to commit crimes—and, specifically, violent ones— because they value the thrill and are more susceptible to peer group pressures, whereas older people are more likely to consider the costs, including incarceration to the loss of family and jobs.[8]

Terrorism is a violent crime, which suggests that the age-dependent drivers of the capacity for violence may play a role. Physical strength, absence of familial obligations, and a propensity to discount risk are known risk features of youths. However, terrorism is also a political crime, and a political awakening is a prerequisite. There are, therefore, reasons to think violent terrorist crimes do not conform to the predictable patterns of other types of violent crime. This is what we found. Violent action among terrorist offenders peaks at a later age and occurs across a broader age range than ordinary violent crimes. This may be a consequence of the political motivations behind terrorist action. However, the age-crime relationship distributions varied widely across different types of terrorist offenses. Men (and a few women) at any age may engage in non-violent criminal support for terrorism. It remains the case, nevertheless, that a distinctive age-crime curve may be discerned with young men being more likely to take part in insurgencies abroad and slightly older men engaging in domestic violent terrorist acts.

The relationship between different terrorism-related action types and demographic variables, including age and sex, and whether the offender was a convert to Islam or grew up Muslim were assessed in a correlation model. High incidence rates of diagnosed mental illness have been found among certain types of mass shooters and "lone actor" terrorists. We therefore included diagnosed mental illness as possible variable in determining the outcomes of the radicalization process, specifically the length of the

[7] Jytte Klausen, Tyler Morrill, and Rosanne Libretti. "The Terrorist Age Crime Curve: An Analysis of American Islamist Terrorist Offenders and Age-Specific Propensity for Participation in Violent and Nonviolent Incidents." *Social Science Quarterly* 97, no. 1 (2016): 19–32. doi:10.1111/ssqu.12249.
[8] David P. Farrington. "Age and Crime." *Crime and Justice* 7 (1986): 235. doi:10.1086/449114.

process and the action type.

Data were collected for over twelve hundred American and British jihadists who radicalized after 2001.The group characteristics of the offenders included in the dataset are shown in Table 1 below.

Table 1. Demographic composition of jihadist dataset. Count of each jihadist subgroup (N = 1,241).

Variable	**Value**	**N**
All	All	1241
Sex	FEMALE	152
	MALE	1089
Country	United Kingdom	743
	United States	498
Convert Status	FALSE	889
	TRUE	327
	Missing	25
Mental Illness	Not Present	1158
	Present	83
Action Type	FFNV	390
	NV	162
	VFF	268
	VNV	421

Source: Western Jihadism Project, Brandeis University.

The action types typical for the jihadist were classified in four groups based on the violent nature and the location of the actions undertaken. 421 offenders engaged in domestic terrorist violence (VNV). 390 became foreign fighters (FFNV). 268 became both foreign fighters and domestic violent terrorists (VFF). A minority, 162, committed no violent offenses and were convicted of non-violent terrorism-related offenses (NV).

The coders also assessed, based on the available information, the date at which the offenders committed their first terrorism-related offense (which may not have been the offense for which they were ultimately convicted). The average (mean) age for when they committed their first offense was 25 (mean:

25.31; standard deviation: 7.35). The age-crime curve is displayed in Figure 3 below. The curve has a highly uneven age distribution with 50% of the offenders committing their first offense before or at the age of 23 (median value) and a long tail of older offenders.

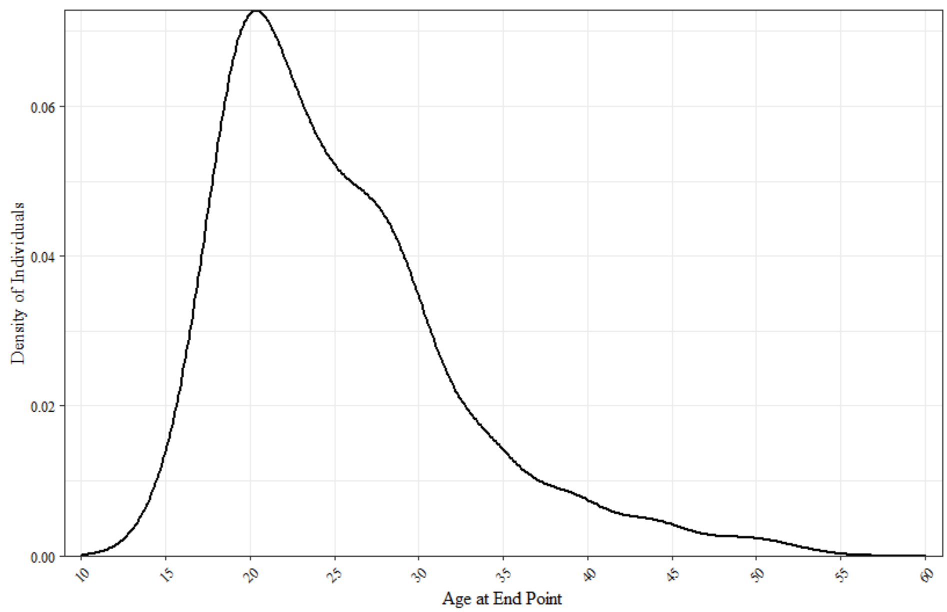

Figure 3. Density plot for age at first jihadist action in months (N = 1239; excludes 2 outliers). Rolling median used for 77 individuals with missing YOB.

The skewed age distribution raises many questions about the relevant counter-radicalization policies and the best means for delivering preventive messaging about the risks of extremist radicalization. Often schools are chosen as the vehicle for public education about extremist recruitment techniques and the warning signs of growing extremism. Given the age distribution with half of the offenders first radicalizing to a point of mobilizing to take criminal action are older and not typically in school means that other types of program delivery are required.

Who then are the most like to become violent offenders? Homegrown jihadists engage in a variety of activities on behalf of the movement. These range from proselytizing to undertaking violent incidents at home. A simple classification scheme was used to sort all actions into three broad categories: becoming a foreign fighter by joining—or trying to join—a foreign jihadist organization; carrying out a violent attack at home or in another Western country or attempting to carry out such an attack; or supporting the global *jihad* through criminal but nonviolent means. The latter includes providing material support to a terrorist organization (e.g., buying and shipping equipment, or purchasing an airline ticket for someone else to travel to join an organization), fund raising, issuing threats (but not acting on such threats), and a range of other supportive acts.

The term "foreign fighter" refers to an individual who, for reasons of ideology or kinship, goes abroad to join a non-state group that is engaged in an armed conflict. The term is also used to include women who travel abroad to join a group in the global jihadist orbit. The United Nations formulation, from which this definition is drawn, is gender neutral. In other words, women who traveled abroad to join ISIS are included as "foreign fighters," even if they perhaps did little actual fighting.

Four action types—here called roles—were treated as dependent in a regression analysis. The demographic variables (convert, native-origin natural-born Muslim, and immigrant-origin Muslim) were used as the independent variables. Because all variables are categorical variables, binary logistic regression tests were used to estimate the odds for the different groups to undertake each of the four types of engagement.

The age variation across different action type was significant. People who carried out the violent attacks—at home or abroad—were significantly younger than those who took on nonviolent support tasks such as fundraising, recruiting, and proselytizing. Nonviolent offenders were on average ten years older than those who became foreign fighters: 32.5 vs. 23 (median values).

Individuals who did "something" violent at home were on average four years older than the foreign fighters. It factors into the interpretation of the data that the standard deviation was large, 6.5 years in the case of the foreign fighters (the youngest was fifteen and the oldest forty). Even if, on average, statistically significant differences exist with respect to the age at which people are likely to choose one role over another, a wide age spread exists generally between the youngest and the oldest perpetrators of violent attacks.

Nevertheless, clearly, age matters greatly for the type of role people play in the service of the movement. Regression analysis is a modeling technique that allows the researcher to isolate the causal effect between two or more variables. Younger people go abroad. Older people stay at home, either to carry out a domestic attack or to undertake support roles. A visual representation of the age-crime curve by type of action taken is shown in Figure 4.

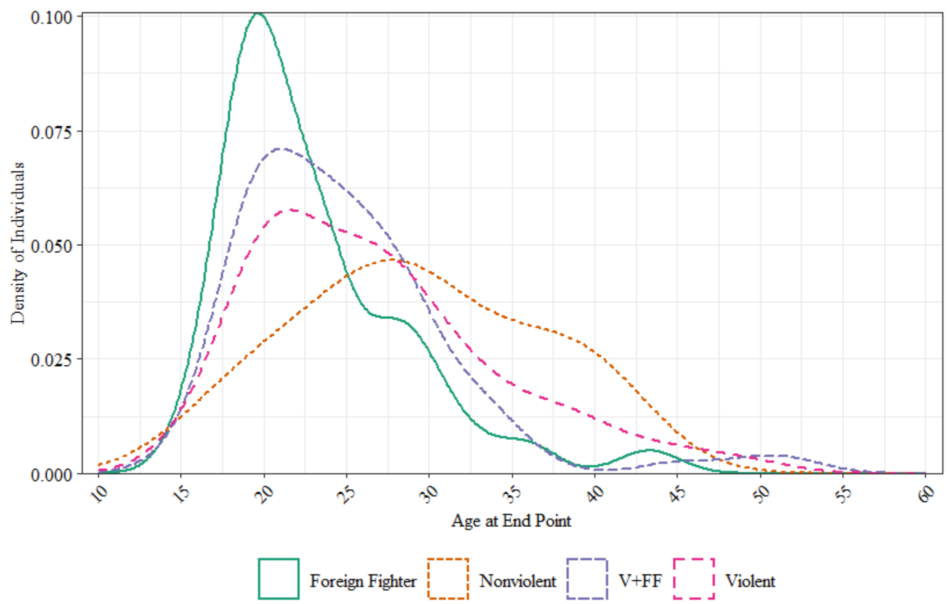

Figure 4. Density graph of age at jihadist action grouped by action type (the "age-crime curve" graph). (N = 1239; excludes 2 outliers). Rolling median used for 77 individuals with missing year of birth.

A Mann-Whitney-Wilcoxon test was carried (See Table 2) out to determine if the observed differences in age by action type were significant differences in age at action for each of the demographic categories. This test was used in preference to a T-test because of the highly skewed distribution and also because this test is preferable in small n situations.

- Non-converts (Med = 23) were significantly younger than converts (Med = 26) at $W = 107219.50$, $z = -6.99$, $p < 0.001$ and had a small effect size ($r = -0.20$). Age differences for comparison among the other categories were not significant.
- Individuals from the United States (Med = 24) were slightly older than that of those from the United Kingdom (Med = 23) at $W = 180284.50$, $z = -0.67$, $p = 0.506$, and had a small effect size ($r = -0.02$).

- Males (Med = 23) were slightly younger than females (Med = 24), at W = 88859.00, z = -1.51, p = 0.130, with a small effect size (r = -0.04).
- Lastly, those without a diagnosed mental illness (Med = 23) were the same age as those with a mental illness (Med = 23), at W = 49352.50, z = -0.44, p = 0.661, and had a small effect size (r = -0.01).

Table 2. Mann-Whitney-Wilcoxon test results for comparing age at first jihadist action by demographic (N = 1239; excludes 2 outliers). Rolling median used for 77 individuals with missing YOB.

Variable	**Value**	**Med.**	***W***	***p***	**CI Low**	**CI High**	***z***	***r***
Country	United Kingdom	23.00	180284.50	0.506	-1.00	0.00	-0.67	-0.02
	United States	24.00						
Sex	FEMALE	24.00	88859.00	0.130	0.00	2.00	-1.51	-0.04
	MALE	23.00						
Convert	FALSE	23.00	107219.50	0.000***	-4.00	-2.00	-6.99	-0.20
	TRUE	26.00						
Mental Illness	Not Present	23.00	49352.50	0.661	-1.00	2.00	-0.44	-0.01
	Present	23.00						

*Key: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$*

Note: Convert excludes 25 individuals with null values.

It has to be assumed that both ideological and social demographics provide scripts for what people do. And then there is the question of motives. A key assumption in discussions of "homegrown" terrorism is that young Muslims are driven to violence by anger over their circumstances. If that were true, we would assume that they above all would choose to attack symbols of their anger here, at home, in the United States. As it turns out, that assumption is incorrect. Going abroad with the intention of becoming a foreign fighter was the most popular choice of action.

Investigating further the relationship between demographic factors and action types chosen on behalf of the foreign terrorist organization, the analysis made use of a type of correlation analysis called binary logistic regression, which is used when the independent variable is highly correlated. Here, the independent variables are classified as "is" or "isn't." The output metric is an odds ratio, which represents the constant effect of the independent variable on the outcome. These estimate the rate at which a particular demographic group engaged in a particular type of action as compared to people who did not belong to that group. In other words, converts are compared with non-converts with respect to the likelihood that they assume X role. Likewise, if the people in a particular subgroup had an odds ratio of 1.5:1 of engaging in a particular activity X, they chose to do X at a 50% higher rate than people who are not in that group. If the odds were 1:1, the particular group was no more likely than the comparison group to do X. (Odds ratios are different from probability. If the odds ratio is 2:1 the probability of this outcome is 1/3.) The results are shown in Table 3.

Table 3. Action type on behalf of a jihadist foreign terrorist organization by demographic factors, American and British jihadists, 2001-2020. Binary logistic regression, grouped by action type.

Variable	**Nonviolent only (Fundraising, recruitment, preaching)**	**Violent domestic plots (No foreign fighting)**	**Foreign fighting (No violent domestic plots)**	**Both violent domestic plots and foreign fighting**
Male	0.81 (0.20)	1.53 (0.31)*	0.58 (0.11)**	1.51 (0.36)
United States	0.53 (0.11)**	0.93 (0.12)	1.39 (0.19)*	1.04 (0.16)
Convert	0.73 (0.17)	1.94 (0.28)***	0.53 (0.09)***	1.06 (0.18)
Mental Illness Present	0.57 (0.28)	2.96 (0.73)***	0.42 (0.13)**	0.62 (0.20)
Age	1.08 (0.01)***	1.02 (0.01)*	0.96 (0.01)***	0.96 (0.01)***
Constant	0.03 (0.01)***	0.18 (0.05)***	2.33 (0.72)**	0.51 (0.18)
Pseudo R^2	0.07	0.04	0.04	0.02

*Key: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$*

Note: N = 1,214; 25 missing values; excludes two outliers older than 60 at the time of their first terrorism-related action). Rolling median used for 77 individuals with missing YOB. Reporting odds ratios, with standard error in parentheses and significance. The constant indicates the baseline odds. Pseudo R^2 measures the total variability of the dependent variable in the equation around the mean. The lower the value, the better the fit.

Close to half of the individuals became foreign fighters or attempted to join a terrorist organization abroad. This underscores how important the aspiration to travel and become a *mujahideen* is to Western adherents. Evidently, what drives them above all is not anger against their circumstances at home but rather an adventure-seeking urge to fight abroad in the name of *jihad*. Converts were nearly three times as likely (2.7:1) to become foreign fighters as non-converts. Men are six times (6.0:1) more likely to commit violent offenses—both domestic and abroad—than are women. The same applies to the variable native-born natural-born Muslim.

Some of the other effects of demography on role types were more limited, although still solid. Native-born offenders were nearly twice as likely (1:7:1) to become foreign fighters as were immigrant offenders. The older people get, the less likely they are to commit violent offenses (although the effect was relatively weak).

Age had a highly significant but moderate effect on the likelihood of becoming a foreign fighter. The odds ratio is 0.9:1. This suggests that as militants age, they are less likely to become foreign fighters. Meanwhile, there was a very significant relationship between being a convert and being a foreign fighter.

It is plausible that older folks might be averse to violent roles, but the relationship between age and the propensity for violent acts is fuzzy. There was no significant relationship between age and becoming a violent domestic operative. Older militants may carry out domestic attacks but are less likely to become foreign fighters.

Moreover, immigrants are nearly twice as likely to become violent domestic operatives (the relationship was highly statistically significant). Another possibility (discussed in what follows) is that the immigrants in the dataset arrive with the intent of becoming recruiters or proselytizers for the global *jihad*. In fact, an immigrant has higher odds of becoming a nonviolent domestic operative, with a ratio of 1.4:1 (highly significant). It is not possible to determine if self-selection is the cause or if the foreign terrorist groups are unwilling to take on older militants as would-be fighters.

Women are most like to assume two roles: providing nonviolent support activities (raising money, buying tickets for people, posting threats online) and going abroad to join the struggle. They generally showed little inclination to carry out domestic attacks.

That men are more violent than women is not a surprise. What may be surprising is the disinclination of natural-born Muslims to participate in domestic attacks (0.37:1). Conversely, immigrants are disinclined to become foreign fighters (0.6:1).

These findings run counter to common assumptions about why people are drawn into violent extremism, which generally have stressed the importance of anger against discrimination and against Western ways of life as motivating the religious militants. Not only are they rarely religious until they embrace the jihadist belief system, militants are also primarily motivated to go abroad to fight rather than taking their anger out on their home surroundings. Those types of sentiments are often expressed as justifications when they do "something," but they are framing their complaints following ideological scripts rather than giving authentic accounts of their passage to militancy.

What then are their basic motivations? Ultimately, nascent individual psychological motives are unknowable. We can only infer behavioral changes from the contexts provided by the data. Young native-born Muslims and converts to Islam may want to travel to jihadi hotspots because they are excited by the dream and opportunities promised by the recruiters to become part of a movement or, simply, to be a big man with a gun. Immigrants, on the other hand, are more focused on homeland activities; raising money, recruiting, etc. They are also more likely to undertake a violent attack in the United States. This may be because they were already inclined to anger against the United States upon arrival in the United States.

Radicalization Trajectories

The study of radicalization trajectories tracks the time it takes for the radicalizing individual to move from exploration to action in the name of the ideology. Table 4 contains a list of the behavioral indicators used to chart an individual's radicalization.

Table 4. List of indicators of radicalization.

Indicator Name	
Date of Conversion	Non-Violent Support
Declaration of Allegiance	Peer Immersion
Desire for Action	Physical and Occupational Training
Disillusionment	Proselytization for the Cause
Educational/Occupational Disengagement	Research and Planning for Traveling Abroad
Ideological Rebellion	Seeking Out New Religious Authority
Information Seeking	Societal Disengagement
Issues Threats	Steps Towards Violence
Lifestyle Changes	Travel Abroad
Marriage Seeking	Violent Action

Recent research indicates that radicalizing individuals of different ideological persuasions generally change their behaviors and in other ways signal that they are about to act on their extremist views. The FBI and other intelligence and law enforcement agencies have embraced the idea that would-be terrorists may be profiled by their behaviors—this approach informed the publication of *Homegrown Violent Extremist Mobilization Indicators* (2019, rev. ed.), which was produced by the FBI, the National Counterterrorism Center, and the Department of Homeland Security. The agencies identified 46 behavioral risk indicators that typically can be observed by bystanders.[9] Numerous concerns have been raised about radicalization being overly inclusive of qualities and running the risk of relying upon stereotypes, particularly for American Muslims.[10] The present research hopes to support a more robust model that goes beyond harmful generalizations and reduce ambiguity. Table 5 contains an example from the codebook of indicator with a description and its cues.

[9] Federal Bureau of Investigation. *Homegrown Violent Extremist Mobilization Indicators.* Washington, D.C: National Counterterrorism Center (2019).
[10] Faiza Patel. *Rethinking Radicalization*. New York, New York: Brennan Center for Justice at New York University School of Law (2011).

Table 5. Example indicator description.

Behavioral indicator name	**Description**	**Cues to the behavior**
Research and Planning for Traveling Abroad	This indicator tracks when an individual begins *actively* planning to travel abroad. All concrete steps taken to facilitate travel abroad are included.	• Buying plane ticket or Applying for passports/visas • Making travel arrangements with individuals abroad • Applying for passports/visas • Obtaining fraudulent travel documents • Selecting a foreign insurgency to join • Asking people online what to bring with during travels • Sought employment in order to travel abroad or to provide support to an extremist organization

We expand the radicalization trajectory analysis by generating order of occurrence plots for each behavioral indicator. The plots for 2 indicators (lifestyle changes [left], physical/domestic training respectively) are depicted in Figure 5. Order of occurrence implies the chronological position of a particular indicator in each trajectory.

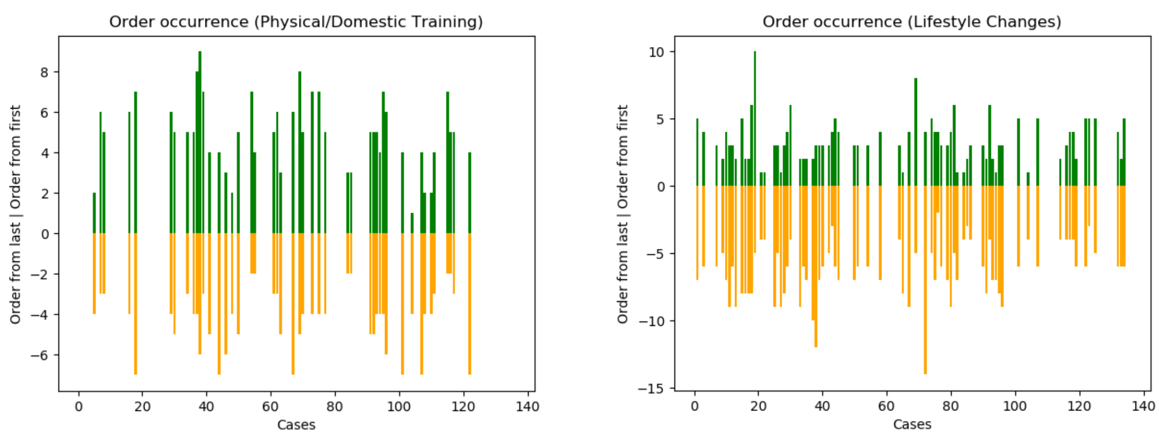

Figure 5. Order of occurrences plots on behavioral indicators: Lifestyle changes [left], Physical/domestic training [right].

Note: In Figure 5, the x-axes represent each trajectory by a number and y-axes depict the order from both the first indicator occurred (green) and from the last indicator occurred (orange). We generated plots for all the indicators, and they provide useful insights on occurrence patterns of each indicator.

A smaller dataset (n = 299) was drawn from the larger dataset of American and British jihadists, who radicalized between 2001 and 2020. Twenty-nine were women. Forty had previously diagnosed mental illness, and 122 were converts to Islam. Fifty-four were British. The British study group was included to test for possible national difference, of which there were a few but none that appears to be of consequence for the overall picture emerging of how jihadists radicalize. The group's demographic characteristics did not fully match the larger dataset of more than 1200 jihadists. The smaller study group had more proportionally converts and slightly older by a few months.

In addition, data were collected about the so-called Incels, to see if what we know about the radicalization process and the signal behaviors that radicalizing individuals inspired by jihadism often exhibit also apply to this group of lethal killers. The name stands for "Involuntarily Celibates." The name was initially used in the late 1990s by a website and an associated mailing list started by a bisexual Canadian woman. It was at that time geared toward making an online community for individuals who wanted to have sex but were unable to find partners. The current Incel movement, known for misogyny and violent rhetoric, has moved worlds away from the original website and spread to forums like Reddit and 4chan. They self-identify as a group, or a movement, but none are known to have met in person, outside the online space. A dataset comprising of self-identified violent Incel offenders (n = 21) was developed with publicly available information about attack-related data, demographics, and time-stamped cues to their behavioral adaptation to the Incel ideology. The data were analyzed using a similar dynamical risk assessment method to track the subjects' radicalization pathways. This research is pertinent to domestic terrorism as over half of the Incels had previous associations with other violent extremist groups, often white supremacist or neo-Nazi groups.

At least six mass shootings have taken place where the perpetrators declared themselves as "Incels." In thirteen successful attacks between 2014 and 2020, Incels killed 69 people and injured 102. In that same time frame, seven successful attacks perpetrated by jihadists inspired by the Islamic State killed 72 and injured 108 individuals in the United States. The jihadist average is greatly increased by the 2016 Pulse

Nightclub attack. They have yet to achieve anything of a similar scale of the Pulse shooting attack but the Incels nevertheless rank as an extremely lethal group. The Incels always act alone and typically engage in shooting attacks (although instances of car ramming are also seen). Surprisingly, they took considerably longer to take action than did the jihadists.

A scatterplot depicted in Figure 6 shows the time (measured in months) it took the 299 jihadists in this small dataset to go through the arc of radicalization from their first observed signature behavior indicative of growing extremism to their moment of final action plotted against their age at the time of their final action. The scatter plot shows broad variation in radicalization trajectory lengths and raises questions also about some standard assumptions about youths being quick to act. The graph also clearly shows that a few people, represented by the dots in the right side of the graphic, take a very long time to act. Trajectory length (in months) for this sample of jihadists had a mean of 33.70 (SD = 30.76) and a median of 23.98, and age at end point had a mean of 25.31 (SD = 6.96) and a median of 23.59.

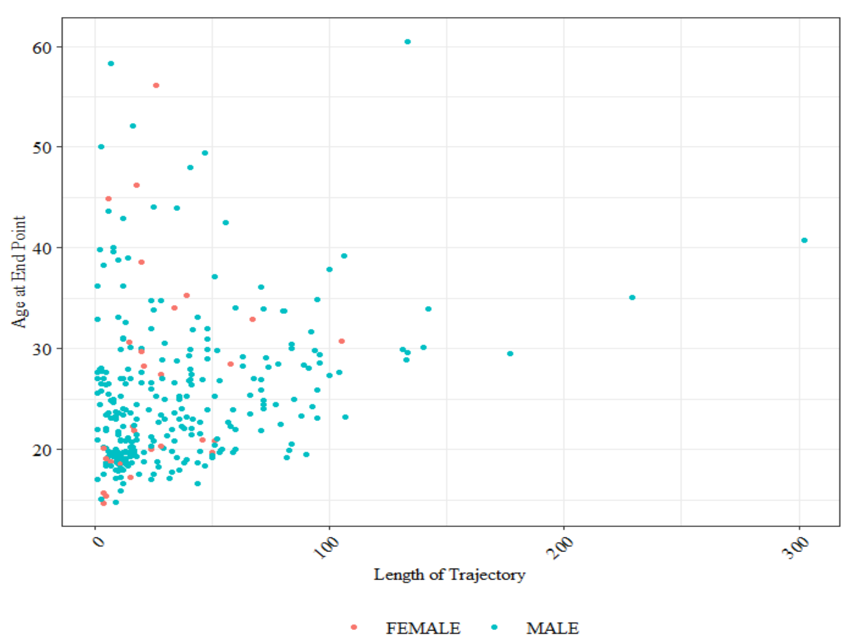

Figure 6. Scatterplot of jihadist trajectory length (months) vs age at end point (N = 299).

The different "tribes" of violent extremists exhibit sometimes radically different demographic and behavioral profiles. The Incels always act alone and typically engage in shooting attacks (although instances of car ramming are also seen). The Incel distribution, using the same measures that were used in the above figure of a scatterplot charting the jihadists' age against the length of time it took them to act, is shown in Figure 7 below. T Surprisingly, they took considerably longer to go through the paces of radicalization than did the jihadists, averaging a whopping 4.2 years from the initial sign of affinity with the ideology to taking—or attempting to taking—violent action. The younger men often acted relatively quickly after embracing the ideology, whereas two older Incel killers took ten years from when they first showed overt signs of having embraced a violently extremist ideology to when they acted.

An important inference is that radicalizing alone and acting alone does not innately correlate with a quick transition to violent action. By way of comparison, the American jihadists' time from initial flirtation with the ideology to criminal action typically *averaged* to 2.8 years.[11] The median, the point at which half the group is at either end of the axis, were 2 years for the jihadists and 3 years for the Incels.

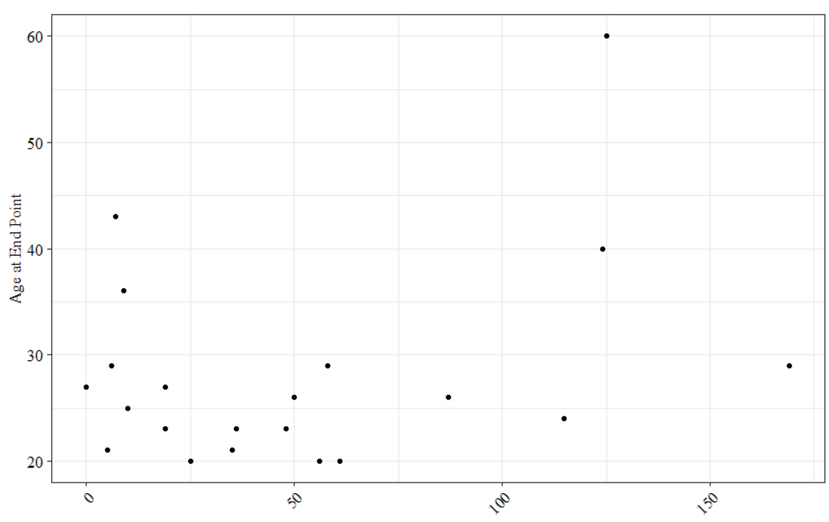

Figure 7. Scatterplot of trajectory length (months) vs age at end point for Incels (N = 21).

[11] Trajectory length (in months) for Incels had a mean of 50.67 (SD = 47.77) and a median of 36.01, and age at end point for Incels had a mean of 28.19 (SD = 9.64) and a median of 26.

Again, using the jihadist as a baseline comparison, a similar pattern of broad variation presents itself from the initial uptake of the ideology to action but even as the standard deviation is large, the jihadists nevertheless generally radicalized into acting more quickly. (Note that the baseline comparison here is made against the smaller subset of the jihadist comprised of nearly three hundred offenders, for which detailed timestamped information was collected.)

In both groups, a small number of individuals take extremely long to mature to violence. The slow "tail" on the time-to-action distribution creates a skewed distribution, which makes the *median* values the more meaningful statistics to use. A standardized radicalization trajectory rate for terrorism offenders may be expressed as the mean time offenders took to progress from evidence of the initial attraction to extremist ideas to criminal action following the ideology. An important inference is that radicalizing alone and acting alone does not, as is often assumed, correlate with a quick transition to violent action.

Although the two groups appeared to present very different patterns with respect to the timetable for the radicalization process, the difference was not actually statistically significant, meaning that the observed variation could be the result of random differences. The broad variation in trajectory lengths among both types of extremists largely explain why the observed difference was not statistically robust. (Using a Mann-Whitney-Wilcoxon test of significance, at $W = 3646.50$, $z = -1.33$, $p = 0.185$, the difference was not statistically significant. Effect size was small ($r = -0.07$).)

An important outcome of the comparison was that it demonstrated it was possible to use analog methods tracking overt cues to radicalizing behaviors on the American jihadists, the British jihadists, and the internationally-mixed group of Incels, a very different ideological framework. But an important caveat was that in the case of the Incels no typical pathways could be identified. It matters also that the Incels had high rates of diagnosed mental illness (9 instances out of 21) and also of previous criminality (13 instances out of 21) unrelated to the ideologically-motivated crimes they eventually carried out, much higher rate on both scores that the jihadists typically exhibit.

The 299 jihadists selected for the detailed study of radicalization trajectories were similarly analyzed to determine if there were significant differences in the lengths of the radicalization process associated with age, sex, nationality or mental illness. The data allowed us to chart the typical distribution curve of radicalization trajectories measured in the number of months it takes an individual to move on to criminal action from the first known instance of having exhibited a growing commitment to engaging with the ideology. The result is shown in Figure 8 below:

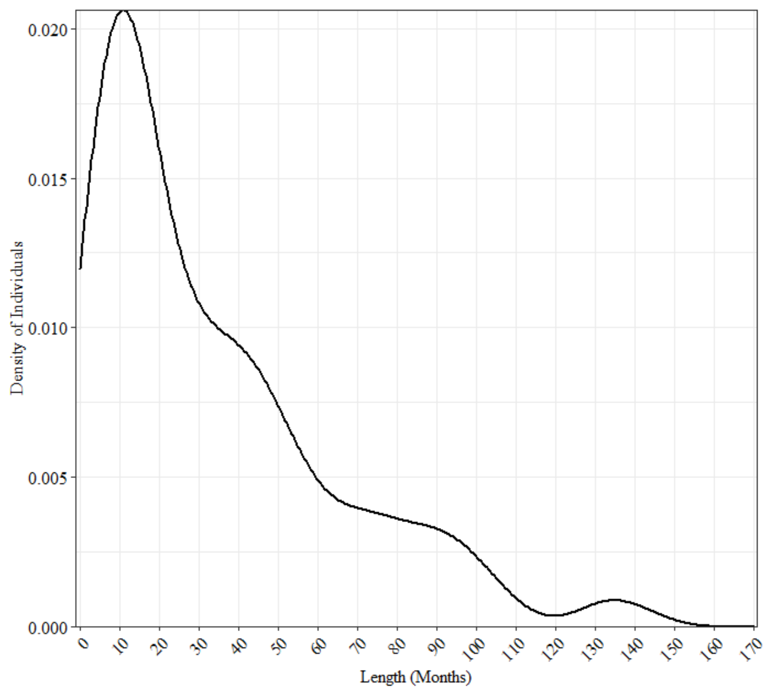

Figure 8. Density plot for jihadist radicalization trajectory length in months (N = 296; excludes 3 outliers).

Some broad regularities may be observed. Typically, most individuals radicalize to a point of action in less than a year—but also not in a "flash." This is important because it means that, if educated about the signifying behaviors, there is room bystanders to intervene—provided they wish to do so.

Tests were performed to determine if there were significant differences in trajectory length for each of the demographic categories. (The Mann-Whitney-Wilcoxon test was used in preference to a T-test because of the skewed distribution of the data.):

- The length (in months) of a radicalization trajectory for individuals from the United States (Med = 24.51) were longer than that of those from the United Kingdom (Med = 16.49), but was not significant ($W = 5666$, $z = -1.53$, $p = 0.127$) and had a small effect size ($r = -0.09$).

- For sex, males (Med = 24.02) had longer trajectories than females (Med = 19.98), but was not significant ($W = 3388$, $z = -1.10$, $p = 0.270$) and had a small effect size ($r = -0.06$).
- In terms of conversion status, non-converts (Med = 23.98) had shorter trajectories than converts (Med = 24.02), but was not significant ($W = 10020$, $z = -0.78$, $p = 0.433$) and had a small effect size ($r = -0.05$).
- Lastly, a comparison of trajectory lengths based on presence of mental illness, those without a diagnosed mental illness (Med = 23.98) had longer trajectories than those with mental illness (Med = 17.00), but was not significant ($W = 5713$, $z = -1.18$, $p = 0.239$) and had a small effect size ($r = -0.07$).

In sum, the British radicalize slightly faster than do Americans, women radicalize faster than do men, and Muslims radicalize faster than do those who convert to Islam. And a diagnosed mental illness does appear to cause people to be less inhibited to act on the violent scripts advocated by the jihadist ideology. Our research did not directly address the possible reasons for these differences but contextual factors ranging from environments to action types may be a factor explaining the observed differences—which in all case were minor and not statistically significant even as the data do indicate underlying group-wide variations. The British, for example, have greater access to offline peer groups sharing the extremist ideology, which is a known accelerant in radicalization. Women may act faster because the threshold for their typical offense—material support for terrorism, becoming a so-called jihadi bride—are lower.

Action types may play a role. The density graph below (Figure 9) shows the timetable for action from first overt sign of radicalization to action, here called trajectory length measures in months, by action type.

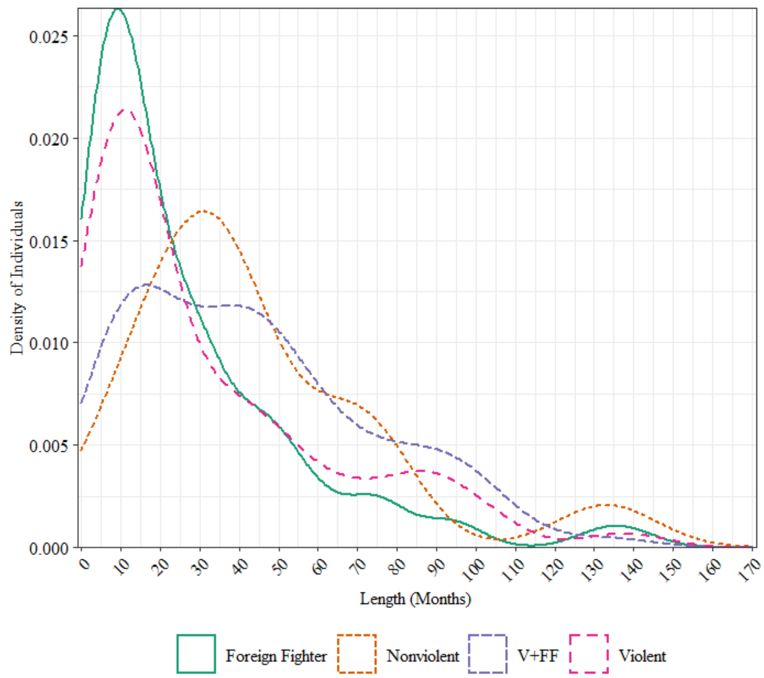

Figure 9. Density graph of jihadist radicalization trajectory length grouped by action type (N = 296; excludes 3 outliers).

To see if the observed difference were statistically robust, a Kruskal-Wallis test was performed to determine if there were significant differences in trajectory lengths when comparing groups with different action types. This test was used in preference to the more commonly used ANOVA test because of the distribution and is preferable in small n situations. The analysis showed that there was a significant difference in the trajectory lengths between action types ($H(3) = 21.72$, $p < 0.001$). A Dunn post-hoc analysis, with Benjamini-Hochberg adjusted p-values to control for the false discovery rate, showed that that:

- Foreign fighters had significantly shorter trajectories than those with only nonviolent action (Med = 34.04; $p = 0.023$) with a small effect size ($r = -0.15$).
- Those who engaged in violent domestic action and also traveled abroad to join a foreign terrorist group (Med = 39.54) took significantly longer to radicalize than those who only engaged in foreign fighting (Med = 14.52; $p < 0.001$) and those who only did domestic violent action (Med = 17.94; $p = 0.003$).

This result is perhaps not so surprising because the group of offenders who travel abroad to join an FTO and also did "something" violent at home often (but not always) travelled abroad before they acted at home. The significant factor here appears to be that this second group slipped between the hands of law enforcement and able to engage in terrorism-related activities for a longer period of time.

To better understand commonalities in radicalization indicators pathways, the team developed an advanced visualization technique and applied it to a set of 299 US and UK terrorist profiles in the Brandeis database. Given that the profiles contain demographic information and the associated timestamped publicly known markers of radicalization trajectories leading up the criminal acts that caused their arrest or death, these aggregate visualizations help analysts discern what risk indicators may be commonly exhibited and in what sequence by those on the path to violent extremism. The resulting visualization is shown in Figure 10a. It summarizes all the occurrences when one radicalization indicator followed another among the violent extremists in the dataset (n=299), where the color of the edge denotes the relative weight or frequency[12] and the size of the node represents the frequency that the indicator was exhibited by the overall population. Figure 10b removes those infrequent transitions and shows only the most frequent ones. It is important to note that Figures 10a and 10b only provide pair-wise (node-to-node) insights and the frequencies of indicators that are subsequently exhibited from a particular indicator. For example, we can discern from 10b that Information Seeking is most frequently exhibited by Date of Conversion (red arrow), Social Disengagement (yellow arrow), and Proselytization for the Cause, Peer Immersion, and Lifestyle changes (green arrows).

Moreover, the researchers also determined a way to depict overall frequently exhibited (i.e., highest weighted) indicator pathways in Figure 10c. The highest weighted pathways were iteratively constructed by chaining together the highest pair-wise paths from End Point and working backwards in order of

---

[12] Due to primary source fidelity, the research team frequently assessed more than one indicator at the same point in time along an individual's radicalization pathway. When considering sequential transitions, there are then combinatorically many 'transitions' between the set of indicators at that point in time and the next indicator. This is the cause for the multitude of indicator transitions (however infrequent) in the aggregate as seen in the thin grey arrows in the figure.

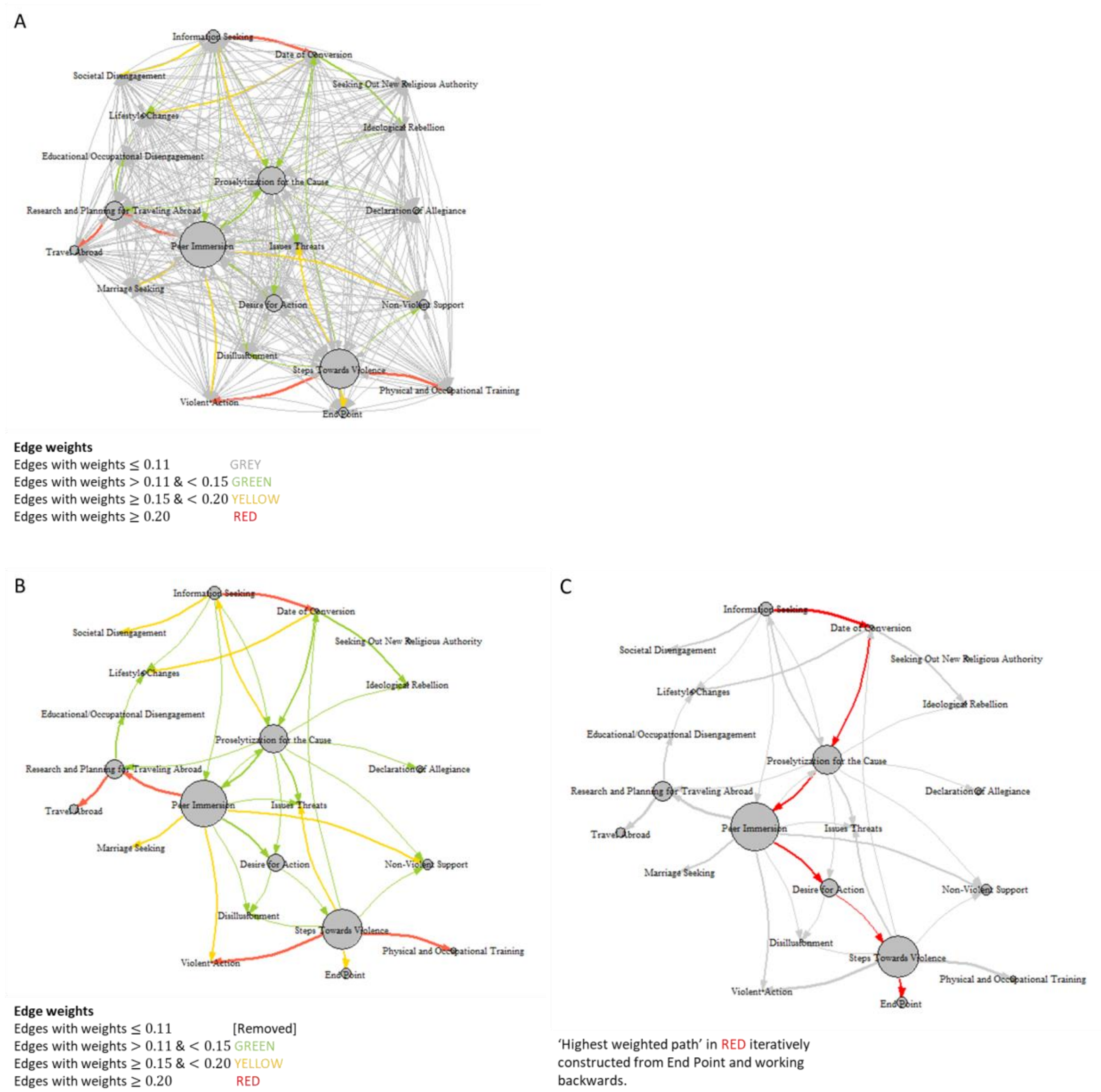

Figure 10. Three visualizations of radicalization pathways among US and UK jihadist extremists showing the frequency of sequential transitions of behavioral indicators (n = 299).

Note: Nodes are features/behaviors and paths represent instances when features/behaviors sequentially followed one another in the WJP empirical data set. Edge weights are proportional to the number of occurrences of such sequential transitions, normalized by the total number of transitions to that indicator. Direction arrows indicate the pairwise sequence of behaviors. Graph A depicts all possible pathways regardless of frequency, graph B depicts only those pathways with an edge weight greater than 0.11, and graph C depicts the "highest weighted path" as constructed from End Point and working backwards.

occurrence. For example, starting with the End Point node, the indicator that most frequently, directly preceded it was Steps Towards Violence. In this way, we continue to inquire at each step what was the most frequently exhibited indicator that preceded it until a full pathway is identified in red. This analysis demonstrates the richness of the insights available to law enforcement, enabling them to make more informed risk decisions and deepens the research community's understanding of radicalization.

This visualization and analysis technique is intended to provide additional insight into radicalization pathways. By analyzing these results further, we aim to discern more about what indicators most commonly precede other indicators. We also produced subsequent visualizations segmented by certain variables (gender, convert, US/UK, mental illness) and continue to analyze for any differences by population.

4. Limitations

The subgroup of jihadists used to track radicalization trajectories was selected based on the availability of sufficiently detailed text sources, which introduced an element of bias. Typically, much of the public information about terrorism offenders derives from prosecutions, and not all terrorism-related offenses are prosecuted in domestic court. Some of the subjects relevant for the study died while fighting for a foreign terrorist organization, in which case often less reliable information is available about the persons' life up until they left their home country. Media reports were also used but often the media reports based on information provided in prosecutions. It can therefore be difficult to diversify the source material used. The reliance on public documents has the benefit that the research does not conflict with privacy protection standard and may be freely shared. The downside is that the research is exceptionally time consuming.

Data about pre-criminal behaviors and psychological states are difficult to assemble, and interview studies with known terrorism offenders and family members are difficult to interpret to cognitive biases introduced by the narrative constructs of the ideology and family dynamics.

The fundamental psychological processes characteristic of the radicalizing individual may have broadly similar features across different extremist sentiment communities, but the specific drivers and behaviors are contingent on the ideology and therefore different. Similarities include detachment from ordinary life and the importance of peer groups and networks in the identity formation process, but the scripts for "what to do" vary greatly between brands of violent extremism. Conducting individualized psychological assessments at a level of abstraction that is willfully blind to the specific action scripts of the extremist belief system inspiring the individual of concern is of marginal practical use for threat assessment.

Finally, another limitation of the research is that it is difficult to know how generalizable the findings are in the absence of data about other types of extremist radicalization processes. The inclusion of the small comparative study of the violent Incel adherents address the issue in a small way but was in part inconclusive. The time-consuming nature of the research and the paucity of publicly available data are obstacles to generalization. Ideally, this problem should be addressed by testing the model on anonymized data developed from analysts' notes and law enforcement investigations.

5. Artifacts

A. Publications

1. S. R. Muramudalige, B.W. K. Hung, A.P. Jayasumana, Indrakshi Ray and J. Klausen, "Enhancing Investigative Pattern Detection via Inexact Matching and Graph Databases," **IEEE Transactions on Services Computing, 2021.** doi : 10.1109/TSC.2021.3073145
2. C.M. Kattadige, S. Muramudalige, K. N. Choi, G. Jourjon, H. Wang, A.P. Jayasumana and K. Thilakarathna, "VideoTrain: A Generative Adversarial Framework for Synthetic Video Traffic Generation," **Proc. 22nd IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM 2021)**, Pisa, Italy, June 2021.
3. The research features in Chapter 10, "Made in the USA" in Dr. Klausen's forthcoming book, Western Jihadism—A Thirty Year History, to be published by Oxford University Press in June 2021 in the U.K. and on September 11, 2021, in the U.S.
4. H. Shirazi, S. Muramudalige, Indrakshi Ray and A. P. Jayasumana, "Improved Phishing Detection Algorithms Using Adversarial Autoencoder Synthesized Data," **Proc. 45th IEEE Conference on Local Computer Networks (LCN2020)**, Nov. 2020.

5. B.W.K. Hung, A. P. Jayasumana and V. W. Bandara, "Finding Emergent Patterns of Behaviors in Dynamic Heterogeneous Social Networks," **IEEE Transactions on Computational Social Systems.** (DOI: 10.1109/TCSS.2019.2938787)
6. B. W. K. Hung, S. Muramudalige, A. P. Jayasumana, J. Klausen, R. Libretti, E. Moloney and P. Renugopalakrishnan, "Recognizing Radicalization Indicators in Text Documents Using Human-in-the-Loop Information Extraction and Natural Language Processing Techniques," **IEEE International Symposium on Technologies for Homeland Security**, Woburn, MA USA, Nov. 2019.
7. S. R. Muramudalige, B.W. K. Hung, A.P. Jayasumana and Indrakshi Ray, "Investigative Graph Search using Graph Databases," **Proc. Graph Computing (GC 2019)**, Laguna Hills, CA, Sept. 2019.
8. J. Klausen, R. Libretti, B.W.K. Hung and A. P. Jayasumana, "Radicalization Trajectories: An Evidence-Based Computational Approach to Dynamic Risk Assessment of 'Homegrown' Jihadists," **Studies in Conflict & Terrorism**, Taylor & Francis, 2018. (https://doi.org/10.1080/1057610X.2018.1492819)
9. B.W.K. Hung, A. P. Jayasumana and V. W. Bandara, "INSiGHT: A System to Detect Violent Extremist Radicalization Trajectories in Dynamic Graphs," **Data and Knowledge Engineering,** Vol, 118, Elsevier, Nov. 2018. (https://doi.org/10.1016/j.datak.2018.09.003).
10. H. Shirazi, S. Muramudalige, Indrakshi Ray, A. P. Jayasumana and H. Wang, "Adversarial Autoencoder Data Synthesis for Enhancing Machine Learning-based Phishing Detection Algorithms," (In preparation).
11. S, Muramudalige, A. P. Jayasumana and H. Wang, "A comparative study of complex dataobject generation with likelihood and deep generative approaches," (In preparation).
12. S. R. Muramudalige, Automating Investigative Pattern Detection using Machine Learning & Graph Pattern Matching Techniques (Tentative). **Ph.D. Dissertation**, Colorado State University, 2021. (In Progress)

B. Software

- Project webpage with links to publications, software and data
  http://www.cnrl.colostate.edu/Projects/RAD/index.html

- PINGS library and four synthetic radicalization datasets are made publicly available at http://www.cnrl.colostate.edu/Projects/RAD/pings.html. Also see GitHub page for PINGS (similarity score-based inexact matching Neo4j to query based on a given graph) https://github.com/cnrl-csu/pings. The program is available as open-source software that may be redistributed and/or modified under the terms of the GNU General Public License as published by the Free Software Foundation, version 3 of the License (GNU GPLv3).

- Rel2Neo (relational databases to Neo4j database conversion) library. The library is capable of generating required data files and Neo4j insert queries based on a given relational database. https://github.com/cnrl-csu/rel2neo Rel2Neo is available as open-source software that may be redistributed and/or modified under the terms of the GNU General Public License as published by the Free Software Foundation, version 3 of the License (GNU GPLv3).

- Significant expansion of the Western Jihadism Database, the transformation of the SQL database into an enhanced graph database, and software products to support graph data mining, pattern detection for radicalization detection, and multi-label text classification to identify radicalization indicators. Additional detail can be found at http://www.cnrl.colostate.edu/Projects/RAD/index.html.

- Four synthetic radicalization datasets are made publicly available at
  http://www.cnrl.colostate.edu/Projects/RAD/pings.html

C. Data Sets Generated

The following data files were developed as part of the research. All of the files will be made available to other researcher through the National Archive of Criminal Justice Data:

2017-ZA-CX-0002_Klausen_Jihadist_Demographic_Data_2001-2019.xlsx

This file contains rows of individual jihadist extremists (n = 1,241), with columns regarding demographic information. Eight demographic variables were coded: county, sex, conversion to Islam (if applicable), year of birth, ethnicity and/or race, year of first terrorism-related action, mental illness, and types of terrorism-related criminality. The information was collected and encoded from publicly available source documents, such as news articles, social media posts, court documents, etc. This file also includes demographic information for the 299 jihadists included in the radicalization trajectory research located in 2017-ZA-CX-0002_Klausen_Jihadist_Trajectory_Data_2001-2020.xlsx.

2017-ZA-CX-0002_Klausen_Jihadist_Trajectory_Data_2001-2019.xlsx

This file contains rows of jihadist radicalization events (n = 7216), with columns describing the events and radicalization indicators that the events are associated with 299 individuals selected from those in 2017-ZA-CX-0002_Klausen_Jihadist_Demographic_Data_2001-2019.xlsx. This information was coded from publicly available source documents such as news articles, social media posts, court documents, etc. The dataset contains timestamped information regarding publicly known markers of events comprising radicalization trajectories leading up to the criminal act (or acts) that caused their arrest, their death, or their successful departure from their home country to join a foreign terrorist organization.

2017-ZA-CX-0002_Klausen_Jihadist_Text_Data_2001-2019.xlsx

This file contains rows of text data (n = 6,476) that provide the basis for coding specific indicators in 2017-ZA-CX-0002_Klausen_Jihadist_Trajectory_Data_2001-2020.xlsx. The columns refer to the raw text segment as copied from the original source, a “cleaned” or parsed down version of that text, and keywords selected from the initial text segment.

2017-ZA-CX-0002_Klausen_Incel_Demographic_Data_2014-2020.xlsx

This file contains rows of individual Involuntary Celibate (Incel) extremists (n = 21), with columns regarding demographic information. Seven demographic variables were coded: county, sex (all are male), year of birth, ethnicity and/or race, year of first terrorism-related action, mental illness, and types of terrorism-related criminality were coded. This information was coded from publicly available source documents such as news articles, social media posts, court documents, etc.

2017-ZA-CX-0002_Klausen_Incel_Trajectory_Data_2014-2020.xlsx

This file contains rows of Incel radicalization events (n = 438), with columns describing the events and radicalization indicators that the events are associated with the 21 individuals found in 2017-ZA-CX-0002_Klausen_Incel_Demographic_Data_2014-2019.xlsx. This information was coded from publicly available source documents such as news articles, social media posts, court documents, etc. The dataset contains timestamped information regarding publicly known markers of events comprising radicalization trajectories leading up to the criminal act (or acts) that caused their arrest, their death, or their successful departure from their home country to join a foreign terrorist organization.

2017-ZA-CX-0002_Jayasumana_Synthetic_Data_no_date.csv

This file includes 1,000 synthesized radicalization trajectories (rows). Each trajectory is defined by 20 behavioral indicators (columns) with a unique identifier. The date of occurrence is coded for applicable indicators, and the value 0 denotes the data unavailability of a particular indicator. Any number of trajectories can be generated based on the requirements. The twenty columns each refer to the behavioral indicators used in 2017-ZA-CX-0002_Klausen_Jihadist_Trajectory_Data_2001-2020.xlsx. This file was generated based upon the parameters set by that data, but does not refer to any person or event from the other files.

Synthetic data can be used to augment or as a substitute for real-world data, which invariably are limited in numbers because of the demanding nature of the research and the small base numbers (it is not possible to identify and encode information for 1,000 subjects). The presented approach allows enhancing such datasets in becoming sufficient to train machine learning models, improve the efficiency of the manual data annotating process by social scientists, and provide a degree of anonymity to the sensitive data.

D. Dissemination Activities

The dissemination activities beyond the above listed publications and software distributions:

- The collaborators are presenting the research at a thematic panel at the 2021 American Society of Criminology Meeting in Chicago, IL, titled Developing and Detecting Extremist Radicalization Trajectories with Machine Learning Techniques. The meeting takes place November 17-20, 2021.
- Klausen and Libretti are presenting a paper titled *Profiling Extremism: Predicting Violent Action Using Behavioral* Indicators at the 2021 conference for the American Society of Evidence-Based Policing, August 28-29, 2021, at the University of South Carolina (virtual and in-person).
- The research also features in a forthcoming book, *Western Jihadism—A Thirty Year History*, by Dr. Jytte Klausen, to be published in September 2021 by Oxford University Press.
- Co-PI Prof. Klausen initiated conversations with the Citizens Crime Commission of New York City, a non-partisan non-profit organization that is engaged in counter-extremism interventions and seeks to develop local-level initiatives to counter domestic terrorism and hate crimes. The Commission and Dr. Klausen's team are committed to develop a partnership to bridge academic research and practical intervention efforts.
- Co-PI Prof. Klausen presented her research on the use of behavioral indicators to track growing extremism and the collaborators' work on using data science to sort evidence and visualize extremist radicalization from behavioral cues to staff at Beyond Conflict, a 28-year old Boston-based non-profit that works internationally to understand and address the causes and consequences of conflict. The organization focuses on delivering research-based solutions to practitioners, policy-makers, and decision-makers globally. Beyond Conflict has worked in Kenya to build resilience against al-Shabaab's recruitment efforts and in Columbia to promote the integration of former rebels.
- A joint research briefing by Brandeis and CSU to FBI Counterterrorism Division (Indicators of Mobilization to Violence Team in the Counterterrorism Division), September 2020, to share our

research efforts to date and to received validation and feedback from important practitioners.

- . The briefing was successful because it constituted an important validation/feedback loop with the key operational stakeholders who perform this analytical work.
- Presentations by CSU team members to others resulted in the adaptation of synthetic data generation method for two specific applications, a) detection of phishing websites and b) synthetic video traces for privacy protection and deception. These collaborations have resulted in peer reviewed publications and on-going collaborations.
- PI Jayasumana made a presentation 'Communication Networks to Terrorist Networks - Exciting Challenges in Networking,' to Industrial Advisory Board, Department of ECE, Colorado State University, Oct. 25, 2019.
- PI Jayasumana, presented a seminar 'Computer Aided Tools for Social Science Research - Experience with Terrorist Risk Assessment Using Large-Scale Behavioral Data,' to a joint audience from Department of Politics & Computer Science Department, Brandeis University, Waltham, MA, Nov. 5, 2019.
- Co-PI Klausen presented an ISTeC Distinguished Lecture on "How 2G Computational Social Science Can Revolutionize the Study of 'Dark' Networks". ISTeC (Information Sciences and Technology Center) caters to a Colorado State University wide audience from different fields. She also presented a lecture entitled "The Call of Duty: The Making of a 'Homegrown' Terrorist" to an audience consisting mainly of students and faculty from social science, political science and computer science. Colorado State University October 2018.