UNIVERSITÀ CATTOLICA DEL SACRO CUORE DI MILANO

DOTTORATO DI RICERCA IN CRIMINOLOGIA

Ciclo XXXII

S.S.D.: SPS/12

ON META-NETWORKS, DEEP LEARNING, TIME AND JIHADISM

Tesi di Dottorato di:

Gian Maria Campedelli

Advisor:

Prof. Kathleen M. Carley

Coordinatore di Dottorato:

Prof. Francesco Calderoni

ANNO ACCADEMICO 2018/2019

On Meta-Networks, Deep Learning, Time and Jihadism

Cycle XXXII

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Criminology

Advisor:
Prof. Kathleen M. Carley

November 2019

UNIVERSITÀ CATTOLICA del Sacro Cuore

International PHD in Criminology

transcrime

This page intentionally left blank

1 The world is everything that is the case.
1.1 The world is the totality of facts, not of things.
1.11 The world is determined by the facts, and by these being *all* the facts.
1.12 For the totality of facts determines both what is the case, and also that is not the case.
1.13 The facts in logical space are the world.

Ludwig Wittgenstein, *Tractatus Logico-Philosophicus*, 1921

This dissertation is dedicated to
my mom, my dad, my sister, my grandparents, and to Marta.

This page intentionally left blank

Acknowledgments

Right after the beginning of my Ph.D. I was told that a doctorate is intrinsically a long individual walk. Indeed it has been. I have lived - and enjoyed - this feeling of loneliness many times throughout the path. The topics and the methodological choices that shaped this work have probably reinforced the sense of solitude of this journey. Moving out from the classic analytic approaches of criminology and the traditional research areas covered at Transcrime, the research center where I have been working in the last three years, has been thrilling and, at the same time, challenging. For this reason, I first want to thank the two anonymous reviewers of this dissertation for their comments and their appreciation of my work. Their words meant a lot to me.

In spite of the inherent individual nature of my path, I owe thanks to many other people that shared with me professional and life experiences during my doctorate.

I am very grateful to my advisor, Kathleen M. Carley. She believed in my work from the very first time we spoke together, and allowing me to stay at Carnegie Mellon University for a semester has been decisive in getting closer to realize my ideas. Her constant support and her inclusive philosophy of research helped me in navigating safely towards my objectives. This work would have not been possible without her.

Francesco Calderoni, who has been my supervisor while working as a researcher on the Horizon2020 project PROTON, has been a never-ending stimulus for my intellectual activity. We energetically debated about a myriad of research problems and concepts in the past years, and those long discussions have often become fertile seeds for my scientific work. Although our ideas were and still are divergent on most matters, working with him has certainly made me a better scientist, strengthening my critical thinking and fostering my desire for rigor.

I would also like to thank Serena Favarin, Antonio Bosisio, Cecilia Meneghini, Martina Marchesi, and Alberto Aziani, my friends and colleagues at Transcrime. The moments we lived together lighted up many dark steps of this path. I will one day look back and remember only the sweetest things. A special mention goes to

Tommaso Comunale, who hosted me on his couch for one year and a half and shared with me almost every second of this Ph.D. The vibrant, endless, and deep chats we had will never be forgotten.

The inspiring recommendations and suggestions I have received from other colleagues around the world have been precious to me, not only with regard to this work but in general in relation to my broader research experience. Thanks to Iain Cruickshank, Mihovil Bartulovic, Alfred Blumstein, Daniel Nagin, Bruce Desmarais, Stephen Tench, Paul Gill and to all the fellows at Carnegie Mellon University who made me feel like I was one of them.

I am immensely indebted to my family for the unconditional support provided to me since the beginning of my undergraduate studies. I hope that every sacrifice you did for me will one day pay you back, wherever I will be, wherever you will be. The pride in your eyes is to me the biggest reward for all my efforts. I love you.

My deepest gratitude to Marta, who has been a cornerstone of my life and has always demonstrated her interest in my work and my ideas. Thank you for all the times you have put up with me. Thank you for your patience, for your billions of questions, for the love you gave me and for having pushed me to always do my best.

Although they have not yet understood what I do, and will probably never read these lines, I also wish to thank Matteo, Lorenzo, Nicolò, Riccardo, Michele, and Enrico for being my friends. I feel blessed to have you all in my life: you may not be able to comprehend what my work is about, but you certainly cannot understand what you represent to me.

Finally, I want to thank Wilson, my cat, for having been a loyal and honest friend even when I was far away from home. His company during the nights while writing this dissertation has been a delicate and heartwarming sign of friendship.

Abstract

Jihadist terrorism represents a global threat for societies and a challenge for scientists interested in understanding its complexity. This complexity continuously calls for developments in terrorism research. Enhancing the empirical knowledge on the phenomenon can potentially contribute to developing concrete real-world applications and, ultimately, to the prevention of societal damages. In light of these aspects, this work presents a novel methodological framework that integrates network science, mathematical modeling, and deep learning to shed light on jihadism, both at the explanatory and predictive levels. Specifically, this dissertation will compare and analyze the world's most active jihadist terrorist organizations (i.e. The Islamic State, the Taliban, Al Qaeda, Boko Haram, and Al Shabaab) to investigate their behavioral patterns and forecast their future actions. Building upon a theoretical framework that relies on the spatial concentration of terrorist violence and the strategic perspective of terrorist behavior, this dissertation will pursue three linked tasks, employing as many hybrid techniques. Firstly, explore the operational complexity of jihadist organizations using stochastic transition matrices and present Normalized Transition Similarity, a novel coefficient of pairwise similarity in terms of strategic behavior. Secondly, investigate the presence of time-dependent dynamics in attack sequences using Hawkes point processes. Thirdly, integrate complex meta-networks and deep learning to rank and forecast most probable future targets attacked by the jihadist groups. Concerning the results, stochastic transition matrices show that terrorist groups possess a complex repertoire of combinations in the use of weapons and targets. Furthermore, Hawkes models indicate the diffused presence of self-excitability in attack sequences. Finally, forecasting models that exploit the flexibility of graph-derived time series and Long Short-Term Memory networks provide promising results in terms of correct predictions of most likely terrorist targets. Overall, this research seeks to reveal how hidden abstract connections between events can be exploited to unveil jihadist mechanics and how memory-like processes (i.e. multiple non-random, parallel and interconnected recurrent behaviors) might illuminate the way in which these groups act.

This page intentionally left blank

Foreword

This work has bloomed during my semester spent as a visiting research scholar at the School of Computer Science of Carnegie Mellon University, under the supervision of Professor Kathleen M. Carley. During those five months in Pittsburgh, I had the chance to brainstorm and discuss about terrorism, political violence and other criminology topics with professors and young researchers that came from a wide variety of backgrounds and fields.

This process of freely sharing ideas with other scientists and scholars helped me in the process of developing my own work, shining a light on the path to take. Besides my thesis, this experience led me to work on several publications related to adjacent topics. Three research articles have been already published in the area of the computational analysis of terrorism. The first one, entitled *Complex Networks for Terrorist Target Prediction* has been published in the 2018 Proceedings of the "International Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction and Behavior Representation in Modeling and Simulation" (SBP-BRiMS). The second one, *Detecting Latent Terrorist Communities Testing a Gower's Similarity-Based Clustering Algorithm for Multi-partite Networks* is part of the proceedings of the "International Conference on Complex Networks and Their Applications", held at the University of Cambridge in December 2018. The third one, *Pairwise similarity of jihadist groups in target and weapon transitions* has been recently published in the Journal of Computational Social Science (the fourth chapter of this thesis is an extended version of it). Finally, the fourth one, entitled *A complex network approach to find latent clusters of terrorist groups*, has appeared in "Applied Network Science".

Collaborating with Professor Kathleen M. Carley and doctoral students Iain Cruickshank and Mihovil Bartulovic has been fruitful and scientifically thrilling. I am hopeful that the outputs that have emerged and will eventually derive from these and future works may represent a proof of the effectiveness of cross-disciplinary approaches to the study of terrorism. With this belief, I start from here.

This page intentionally left blank

Contents

This page intentionally left blank

List of Tables

This page intentionally left blank

List of Figures

This page intentionally left blank

List of Abbreviations

ADF	Augmented Dickey-Fuller
AI	Artificial Intelligence
AIC	Akaike Information Criterion
AMISOM	African Union Mission in Somalia
AQAP	Al Qaeda in the Arabian Peninsula
AQI	Al Qaeda in Iraq
AR(1)	Autoregressive Process of Order 1
AR	Autoregressive Model
ARFIMA	Autoregressive Fractionally Integrated Moving Average Model
ARIMA	Autoregreesive Integrated Moving Average Model
AR(k)	Autoregressive Process of Order k
ARMA	Autoregressive Moving Average Model
BIC	Bayesian Information Criterion
Bu	Business
Cov	Covariance
DL	Deep Learning
E/B/D	Explosives, Bombings, Dynamite
Fi	Firearms
GG	Government (General)
GTD	Global Terrorism Database

HMM	Hidden Markov Model
IEP	Institute for Economics and Peace
i.i.d.	Independent and Identically Distributed
In	Incendiary
IS	Islamic State (synonym of ISIS)
ISI	Islamic State of Iraq
ISIS	Islamic State of Iraq and Syria (synonym of IS)
JTJ	Jamaat al-Tawhid wal-Jihad
KS	Kolmogorov-Smirnov
ln	Natural Logarithm
LSTM	Long Short-Term Memory
MA	Moving Average Model
MAE	Mean Absolute Error
Me	Melee
Mi	Military
ML	Machine Learning
MLE	Maximum Likelihood Estimation
MSE	Means Squared Error
NGO	Non-governative Organization
NN	Neural Network
NTS	Normalized Trail Similarity
OECD	The Organisation for Economic Co-operation and Development
PC&P	Private Citizens and Property
PGIS	Pinkerton Global Intelligence Service
Po	Police
ReLU	Rectified Linear Unit

RF/I	Religious Figures/Institutions
RNN	Recurrent Neural Network
RNV	Rank-version of Von Neumann's Ratio Test for Randomness
SRNN	Simple Recurrent Neural Network
STRNN	Spatial-Temporal Recurrent Neural Network
T/NSM	Terrorists and Non-State Militia
Un	Unknown
US	United States
VAR	Vector Autoregressive Model

This page intentionally left blank

List of Symbols

$\Phi_{g_k}(s_i \to s_j)$	Number of transitions between state i and state j
$1\mathrm{d}_r$	1-dimensional receiver
$1\mathrm{d}_s$	1-dimensional sender
$2\mathrm{d}_r$	2-dimensional receiver
$2\mathrm{d}_s$	2-dimensional sender
α_i	In the context of rare and weak signals: impact function of a given event
$\Delta\omega_t$	Update of model weights within the Adam optimization algorithm
γ	Shrinkage parameter
$\Gamma_{T_{n\neq 0}}$	Set-wise accuracy
$\lambda(t)$	Intensity functin of a Hawkes self-exciting process
$\{t_i\}$	A set of event times
$\mathbf{A}(\mathfrak{W})_{ij}$	Adjacency matrix of transitions with 1-dimensional states
$\mathbf{A}_N(\mathfrak{W})_{ij}$	Adjacency matrix of transitions with N-dimensional super-states
$\mathbf{G}^{\mathrm{T}}$	Transpose of matrix $\mathbf{G}$
$\mathbf{M}^{n\times n}$	Square symmetric matrix of dimension n $\times$ n
$\mathbf{P}(\mathfrak{W})$	1-step stochastic transition matrix with 1-dimensional states
$\mathbf{P}^t(\mathfrak{W})_{ij}$	t-step stochastic transition matrix with 1-dimensional states

$\mathbf{P}_N(\mathfrak{W})_{ij}$	1-step stochastic transition matrix of N-dimensional super-states
wM	Weighted square matrix
$\mathcal{H}(t)$	The history of events up to time t
$\mathfrak{G}^N$	Multi-partite graph (manifold) that contains N partitions
$\mathfrak{M}_T$	Unified meta-network associated to the entire vector T
$\mathfrak{T}$	Set of possible targets
$\mathfrak{W}$	Set of possible weapons
$\mathbf{G}^{m\times n}$	Weighted adjacency matrix of dimension m $\times$ n
D_n	Komogorov-Smirnov test statistic
$\ln\mathcal{L}$	Loglikelihood function
M_{ti}	Meta-network associated to time t_i
norm $C_D^W\,(i)_t$	Normalized total degree centrality of node i at time t
Y_i	Distribution of the residuals
μ	Constant rate of a Homogeneous Poisson process
ω	Decay rate after a rise in the event rate
ω^{-1}	Average length of the period with persisting increased rate of events
Ω_i	Rarity/impact indicator
$\Phi_{T_{n\neq 0}}$	Event-wise accuracy
$\psi_{g_i}\,(d,t)$	A time-ordered trail of targets
$\psi_{g_i}\,(d,t,w)$	A time-ordered trail of targets and weapons
$\psi_{g_i}\,(d,w)$	A time-ordered trail of weapon
σ	In the context of neural networks: a logistic sigmoid function

σ	In the context of time-series properties: variance of the RVN test for randomness
τ_i	Residual of an element i of a set of event times $\{t_i\}$
$\Theta_{i \to j}$	A 2-dimensional super-state connecting i and j
A_{gi}	Sequence of attacks of group gi
$a_{gi}(d, t, w)$	Tuple defining a terrorist event
c	Centroid of all the points excluded the worst one in the Nelder-Mead algorithm
$C_D^W(i)_t$	Total degree centrality of node i at time t
c_t	Cell gate
f_t	Forget gate
$g(t)$	Exponential kernel to model the decay of λ
$G = \langle N, E \rangle$	Mono-partite graph comprising a set of nodes N and a set of edges E
$G = \langle U, N, E \rangle$	Bi-partite graph comprising two set of nodes U and N
$g_1, g_2, \cdots, g_T$	In the context of neural networks: a list of gradients for each time unit
$G_{m,n}$	A single partition of the manifold $\mathfrak{G}^N$
h, s, l	Indices for worst, second-worst and best points in the Nelder-Mead algorithm
h_t	Recurrent hidden state
i_t	Input target
k	In the context of Hawkes modelling: jump factor
N	In the context of rare and weak signals: total number of sampled events

n	In the context of rare and weak signals: total number of times in which the event i has occurred
$N(t)$	A Counting Process
o_t	Output gate
$S(\mathfrak{T}_{gi})$	State space for the set of targets for group i
$S(\mathfrak{W})_2$	State space of 2d super-states
$S(\mathfrak{W}_{gi})$	State space for the set of weapons for group i
T	A vector of discrete time-stamps
$T(M)_i$	A memory-keeper for event i
U	In the context of neural networks: a hidden state-to-hidden state matrix
W	In the context of neural networks: a weight matrix
$W(v_i \to v_j)$	Number of transitions between node i and j
x	In the context of rare and weak signals: number of classes to which an event belongs
x_c	Contraction point
x_e	Expanded point
x_j	Shrink-contracted point
x_r	Reflected point
X_{d_k}	Random variable mapping the position of a node X at time d_k

Introduction

From 2000 to 2015, the number of civilians killed due to terrorist attacks dramatically increased by 550%, ranging from 2,000 to 12,500 deaths. In 2015, more than 60% of the member states of the Organisation for Economic Co-operation and Development (OECD) experienced at least one terrorist attack, accounting for a total of 577 deaths (Institute for Economics and Peace, 2016). This represented the highest peak since 2004, the year of the Madrid bombings. In 2016, 79 countries reported at least one death caused by terrorists (Institute for Economics and Peace, 2017). The estimated global economic impact of terrorism (in U.S. Dollars) was 52 billion in 2017 (Institute for Economics and Peace, 2018). These few data contribute to picture terrorism as a global threat that has caused damages to civilians, governments, and economic systems. Terrorism is indeed an actual and very complex issue that continuously affects many countries in the world. While keeping in mind the existence of different types of terrorism, jihadism has played a leading role in recent decades in terms of its geographical scope of action, lethality, and impact. Overall data, when disaggregated, demonstrate that jihadists are the main characters of the dramatic terrorist scenario. The rise of jihadism as the most relevant form of terrorism worldwide has contributed to the spread of attention over Islamist organizations in research, and the progressive availability of data attracted the attention of scientists from different fields. Indeed, government and institutional funding for scientific projects related to terrorism in the last years has been increasingly directed towards applied research, calling for technical expertise and skills that can only be achieved through the collaboration among distinct research domains.

In parallel, in the years following the 9/11 attacks, studies tried to describe and assess the state of the terrorism research domain, in order to detect critical issues and positive trends. Reid and Chen (2007) were among the first to identify a shift of perspective in research: indeed, they reported how the focus on terrorism as a low-intensity conflict was substituted by the widespread idea that the phenomenon assumed the connotation of an actual global threat. Soon after this study, Silke (2008) pointed out how the attacks led to a substantial increase in the number of individuals working in the field. However, the author highlighted how, albeit statistical analyses started to become more popular (work applying inferential statistics on terrorism data more than trebled since then), still the shift from mere literature review-based studies was insufficient to guarantee solid and reliable findings and conclusions on the phenomenon.

In spite of the cautious optimism that was circulating among researchers and funding agencies, Sageman (2014) harshly criticized and contrasted the narrative of a successful research path towards the aim of unfolding terrorism. In a largely commented and debated paper, Sageman blamed the government strategies of funding projects on terrorism without sharing primary source information with academics for the problem of "stagnant" research on terrorism. According to the author, these strategies avoided solving the gap between the lack of expertise of intelligence agencies in terms of scientific methods and technical skills and the scarcity of relevant and reliable data at disposal of academic researchers and scientists. His solution was to make non-sensitive data available to academia and to inaugurate a new process of dialogue and collaboration between the two sides. Sageman developed his thesis starting from the fact that research is still too far away from answering the fundamental question "why a person should turn to terrorism?". Critiques and comments about his opinions were addressed to Sageman by other scholars in the field that either contested his diagnoses or his evaluation of the results achieved by terrorism research, besides existing failures (Taylor, 2014; Schmid, 2014).

In the last (in chronological order) attempt to assess the situation of terrorism research, Schuurman (2018) identified encouraging improvements in the way terrorism is investigated and studied, starting from increasing use of primary and original data and the (although slow) diffusion of quantitative works. Notably, Schuurman also recognized existing issues, such as the predominance of qualitative works and the nature of many authors who are one-time contributors, failing thus to provide a constant and significant contribution to the field. Besides these matters of concern, however, the author in the conclusions of the works claims that research on terrorism is flourishing, rather than stagnating. Each of the different points regarding the state of the field contains ideas and considerations that may deserve attention. Indeed, the lack of agreement has to be coupled to the difficulty to obtain estimates on real-world impacts and consequences of terrorism research, therefore forcing the debate to build upon views and opinions, rather than actual facts. In such context, where data sources are few and still government funding are reticent in information sharing, a researcher should try to extract the best from the available. In spite of the popularity and great diffusion of the existing open access databases on terrorism, my intuition is that there is still a lot of knowledge that can be exploited for research and policy purposes, and the only way to gather this knowledge is to invest on innovative methodologies that go beyond the state-of-the-art. Investing in novel techniques may be risky in the first phase: pitfalls might arise, and results may not be accurate or relevant enough to encourage government and institutions to fund new research on terrorism. However, my strong belief is that terrorism research needs a boost in order to attain success not only in terms of journal articles but also in terms of real-world initiatives, and this boost can only go through innovation and training of young researchers that shall be highly skilled in analytical terms and have a strong knowledge of the phenomenon. Some considerations regarding this point will be made later in the work: this belief, however, represents the main motivation that led me to write this thesis, on this precise topic, with this precise shape and this precise focus.

That considered, this dissertation will focus on jihadist terrorism (also called Islamist) and will specifically seek to detect jihadist organizations' behavioral patterns and assess the predictability of their actions over time through the attempt to merge network science and artificial intelligence. The analytic part will rely on a three-fold structure. The main conceptual intuition behind the work is that treating multi-dimensional event data as meta-networks that are connected through time allows detecting hidden relations that traditional methodological frameworks and techniques fail to capture. These relations, which in most cases are abstract, connect together multiple entities that can be observed and monitored over time: this complex typology of networks is effective in capturing recurring behaviors and detecting potential anomalies, making it possible to conduct analyses at explanatory and predictive levels. Results are encouraging and indicate that network-derived models and analysis are able to capture inter-dependencies and to use them to enhance the knowledge of how these organizations act in the global scenario. Moreover, time-series analysis using Neural Networks indicate how terrorist organizations follows two parallel behavioral pathways in terms of attacked targets: while some patterned dynamics emerge, other mechanics seem to be random or extremely challenging to capture and will require additional data or further analyses to be better depicted and understood.

This thesis develops as follows. The first chapter will illustrate the conceptual background of the work. Specifically, it will first focus on the main contributions on the issues of defining terrorism, also proposing a four-dimensional focus on its most relevant dimensions. Additionally, the theoretical framework of the work is introduced and described. The second chapter will then explain the aims and motivations of the work, with an additional note on the state of research in social sciences and criminology.

The third chapter will first review the origins, history and main features of the jihadist groups that will be analyzed throughout the dissertation, namely the Islamic State, the Taliban, Al Qaeda, Boko Haram, and Al Shabaab. Data will be then described, providing a general overview of the Global Terrorism Database, the source from which all the data used in this dissertation are retrieved.

The fourth, fifth and sixth chapters will represent the core of the work. The fourth chapter (*Stochastic Matrices of Terrorism: Complexity and Heterogeneity of Jihadist Behavior*) will present the work on N-dimensional super-state transition networks and trails of terrorist events, highlighting the complexity of terrorist patterns in their operational choices and presenting a novel pairwise coefficient for assessing similarity among groups.

The fifth chapter (*Hawkes Processes of Jihadism*) will focus on point processes of jihadist attacks, showing the spatio-temporal clustering of events and the presence of memory-like dynamics in attacks against specific targets.

The sixth chapter (*Deep Learning and Terrorism: Long Short-Term Memory Networks for Jihadist Target Forecasting)* will then present the framework that combines dynamic network science and machine learning for the prediction of future terrorist targets.

Finally, a conclusive chapter will provide a homogeneous and comprehensive overview of the strengths and limitations of the work, highlighting the most relevant results and the future research pathways, also proposing a broader reflection on the state of the study of terrorism.

1 Background

1.1 Conceptualizing Terrorism

Framing terrorism is an arduous task. Over the decades, academics with different backgrounds have tried to develop definitions as precise as possible, nevertheless failing to produce an unambiguous and universally accepted definition of the concept. One of the motivations is that terrorism is deeply embedded in the historical, political and social context in which it is manifested and therefore it becomes difficult to judge it in a totally objective way. In many cases in the past, the border between terrorists and freedom fighters has been labile and largely dependent on the observation point of view (Laqueur, 1987). Terrorism research has exploded after the 9/11 attacks (Hoffman, 2002; Buckley and Fawn, 2003; Enders and Sandler, 2006; Simons and Tucker, 2007), nonetheless scholars have started to attempt to frame the issue long before that date. This paragraph will review some of these attempts.[1]

1.1.1 Defining Terrorism

In order to find a starting point in reviewing the debate around the concept of terrorism, it is helpful to introduce the review with the work of Jongman and Schmid (1988). The authors analyzed 109 definitions of terrorism coming from a questionnaire with the aim of systematizing the most recurring features. The outcome of the analysis, however, did not lead to a satisfactory homogeneity of the results. Definitions analyzed by the two scholars highlighted the presence of twenty-two recurring characteristics. Similarly, Weiberg et al. (2004) sought definitions from three main academic journals in the area of terrorism, comparing the frequencies of the twenty-

[1] For the purposes and nature of this thesis, this section will only refer to definitions produced by researchers and academics, thus excluding from the review legal definitions provided by governments and international institutions.

two features appeared in Jongman & Schmid with the outcome frequencies of their analysis. Table 1.1 summarizes both the outcomes. The results in the table demonstrate the wide heterogeneity of elements that emerge from experts' definitions of terrorism.

N	Element	Jongman & Schmid	Weibert et al.
1	Violence, force	83.5	71
2	Political	65	60
3	Fear, Terror emphasized	51	22
4	Threat	47	41
5	Psychological effects and (anticipated) reactions	41.5	5.5
6	Victim-target differentiation	37.5	25
7	Purposive, Planned, Systematic, Organized action	32	11
8	Method of combat, strategy, tactic	30.5	31.5
9	Extranormality, in breach of accepted rules, without humanitarian constrains	30	0
10	Coercion, extortion, induction of compliance	28	5.5
11	Publicity aspect	21.5	18
12	Arbitrariness, impersonal, random character, indiscrimination	21	0
13	Civilians, noncombatants, neutrals, outsiders as victims	17.5	22
14	Intimidation	17	11
15	Innocence of victims emphasized	15.5	10
16	Group, movement, organization as perpetrator	14	29
17	Symbolic aspect, demonstration to others	13.5	5.5
18	Incalculability, unpredictability, unexpectedness of occurrence of violence	9	1
19	Clandestine, covert nature	9	7
20	Repetitiveness, serial or campaign character of violence	7	0
21	Criminal	6	5.5
22	Demands made on third parties	4	1

Table 1.1: Prevalence (%) of Definitional Elements of Terrorism in Jongman & Schmid (1988) and Weiberg et al. (2004). Source: (Weiberg, Pedahzur, and Hirsch-Hoefler 2004)

Considering the complexity of building a comprehensive definition of terrorism, some scholars have limited their scope to a list of features that distinguish terrorism from other forms of violence. Hoffman (1998) claims that terrorism, compared to other types of political violence is:

> "(1) ineluctably political in aims and motives; (2) violent - or, equally important, threatens violence; (3) designed to have far-reaching psychological repercussions beyond the immediate victim or target; (4) conducted either by an organization with an identifiable chain of command or conspiratorial cell structure (whose members wear no uniform or identifying insignia) or by individuals or a small collection of individuals directly influenced, motivated, or inspired by the ideological aims or example of some existent terrorist movement and/or its leaders; (5) and perpetrated by a subnational group or nonstate entity".

Matusitz (2012) - partially mirroring the approaches of Jongman and Schmid and Hoffmann- has collected the most commonly cited definitions in literature, either from academic and institutional sources. The common elements were:

1. The use of violence to create fear for political, religious or ideological reasons;
2. The civilian (and sometimes iconic) targets;
3. The spectacular nature of the actions in order to gain publicity for a cause;
4. A change in the regulatory system;
5. The concept of "asymmetric warfare"[2]

In addition, Ganor (2002), in an effort to separate terrorism from the concepts of revolutionary violence, guerrilla and national liberation, proposes three fundamental discriminating factors that qualify terrorism:

1. Violence as the very essence of terrorist activity;
2. The political aim of the activity;
3. The civilian targets

This brief review underlines the challenges in defining terrorism. To solve this problem, several authors have tried to focus on the definition of specific types of terrorism. Indeed, terrorism has many different faces depending on the object of the analysis. In the last years, the impossibility to build a universal concept of terrorism has shifted the spotlights to the different shades of the phenomenon.

[2]Asymmetric warfare is intended as "the use of random/unpredictable violence by a weak group (i.e., one with a smaller force) against a stronger power (i.e., military, government, or even society in general) to gain advantage"(Matusitz, 2012, p.4-5)

1.1.2 A Four-Dimensional Focus on Terrorism

The review presented above demonstrates the difficulties to provide a *tout-court* definition of terrorism. Nevertheless, to disentangle the problem, terrorism can be distinguished by many dimensions (Ganor, 2008). The present sub-section proposes a four-dimensional focus on terrorism motives, strategies, structure and geographical range of action as fundamental aspects for framing it under different perspectives.

1.1.2.1 The Motives and Goals of Terrorist Organizations

In the aftermath of 9/11, Rapoport published an article developing the thesis of the "four waves of terrorism". According to the American scholar, terrorism has evolved in its values and motives across the centuries, stressing the radical shifting from its positive meaning during the French Revolution to the terrible attacks occurred in the United States (US). Even though the word "terrorism" first appeared as a product of the French Revolution, Rapoport claimed that the first wave started around 1880 as a consequence of some economic reforms introduced by the czars in Russia. These reforms caused political assassinations and the wave expanded across borders, conducting also to the anarchist violence that invaded many countries in Europe. The second wave started in the 1920s and lasted until the 1960s: its main feature was national self-determination. The grievance against colonial powers led to atrocities against police forces, making this wave an intersection between terrorism and guerrilla. The third wave was represented by the reaction of the Vietcong against the US military forces. Rapoport notes that this third wave gave birth to a territorial displacement. The violent resistance against the US showed the vulnerability of the Western countries. Many terrorist groups emerged in Europe with the aim to fight capitalism and the power of Western governments (e.g.: The Italian Red Brigades or the German Red Armed Fraction). This third wave ended in the 1980s, leaving room to the final fourth wave. Two events contributed to the starting of this new era: the Iran Revolution in 1979 and the defeat of the Soviet Union in Afghanistan in 1989. The most important element of this fourth wave is the role of religion and the protagonist of this wave is Al Qaeda. In his work, Rapoport called for the possibility of a new wave, relying on the fact that terrorism can renovate its ideology and its motives, re-emerging with new strength. In his theory, Rapoport states that terrorism is deeply rooted in modern culture.

The use of an historical review to describe political and ideological motives of terrorist organizations described in Rapoport's work can be found also in other contributions. Ganor (2008) reviews the most popular terrorist motives categories and the subsequently demonstrates that terrorism is not solely a religious business, even though today jihadism is the protagonist of the terrorist scenario. Table 1.2 shows these categories.

Typology	**Description**
Revolutionary Organizations	Organizations that act to change a nation's regime with the aspiration to bring about a change in the government (e.g.: Nepalese Maoists)
National Liberation Organizations	Organizations that act to vanquish and expel an occupying force and to achieve national independence (e.g.: Palestinian Fatah organization)
Social Organizations	Organizations that act to change a nation's socioeconomic order (e.g.: El Salvadorian FMLN)
Separatist Organizations	Organizations that advocate the territorial separation of an ethnic minority in a multi-ethnic state (e.g.: The Irish IRA)
Radical Ideological Organizations	Organizations that act to advance extremist ideologies. This category includes communist, anarchist, and fascist groups (e.g.: Italian Red Brigade)
Religious Organizations	Organizations that aim to advance religious interests or disseminate a religion via violence while fulfilling what the organization's members believe is "the will of God"). At times, such organizations' activities stem from an aspiration to defend a religion from hostile sources or sources that are interpreted as such (e.g.: Al Qaeda, IS)

Table 1.2: Classification of terrorist organizations based on their motives. Source: author's adaptation of Ganor (2008)

In the effort to describe the objectives of terrorist organizations, Kydd and Walter (2006) detect five ultimate goals of terrorism:

1. Regime change;
2. Territorial change;
3. Policy change;

4. Territorial control;
5. Status quo maintenance

Kydd and Walter point out that many organizations hold more than one goal and can use one as a facilitator for another. They analyzed the Foreign Terrorist Organizations list released by the US State Department, highlighting how most these organizations (31 out of 42) seek regime change. Another significant part (19) seek territorial change, while policy change and maintenance of the status quo are residual goals. Finally, Picco (2004) notes that goals can change over time and that they can be less important than the idea of "perpetual confrontation". This process introduces the confrontation between tactical and strategic terrorism, well described by the behavior of Al Qaeda. According to the author, tactical terrorism is characterized by stable and well-known objectives, while strategic terrorism is dynamic with respect to goals because contraposition itself is the most important feature regardless of political objectives.

1.1.2.2 The Strategies of Terrorism

Terrorism is often described as "senseless" or "mindless". However, many have contested this statement, proving that terrorism is the product of strategic calculations. Strategies of terrorism are intrinsically connected with its aims, and scholars have sought to investigate not only the "what" but also the "how". The attempts to understand the strategies and tactics of terrorism are many and rooted in history (May, 1974; Fromkin, 1975; Price, 1977; Dobson and Payne, 1979). Due to the great number of these attempts and the multifaceted nature of terrorism, there is no universal consensus on a universal set of strategies that comprehensively describes the "how" used for reaching the "what". Nevertheless, many of the conceptualizations of the strategies of terrorism provided by scholars are similar.

Harmon (2001) enumerates five different strategies that he considers as the most common in the behavior of terrorist organizations:

1. Creation of societal dislocation and chaos;
2. Discrediting or destroying a particular government;
3. Rendering economic and property damage;
4. "Bleeding" state security forces and doing other military damage;
5. Spreading fear for international effects

The author indicates that all these strategies involve calculation and that an organization can employ different strategies at different points in time to adapt to its environmental context.

Kydd and Walter (2006) start their explanation of terrorism strategies from the concept of uncertainty. According to the authors, the uncertainty about power, resolve and trustworthiness governs the process that leads a terrorist group to act against another subject (namely a government, a community, etc.) in a certain way at a certain time. To resolve uncertainty, terrorists use costly signaling instead of other legal methods to convince the audience they talk to. These audiences are of two kinds: 1. governments and 2. individuals they hope to positively influence. Combining the three uncertainty dimensions (power, resolve and trustworthiness) with the two audiences, Kydd and Walter develop a theoretical set of five distinct terrorist strategies. Table 1.3 shows this classification.

<table>
<tr><th></th><th></th><th colspan="2">Target of Persuasion</th></tr>
<tr><th></th><th></th><th>Enemy</th><th>Own Population</th></tr>
<tr><td rowspan="3">Subject of Uncertainty</td><td>Power</td><td rowspan="2">Attrition</td><td>Intimidation</td></tr>
<tr><td>Resolve</td><td>Outbidding</td></tr>
<tr><td>Trustworthiness</td><td>Spoiling</td><td>Provocation</td></tr>
</table>

Table 1.3: Strategies of terrorism/political violence. Source: Author's adaptation of Kydd and Walter (2006)

The theoretical development of objectives and strategies carried out by Kydd and Walter started from the assumption that terrorist organizations often achieve their goals. Abrahms (2006) largely contested this assumption, trying to demonstrate that the literature which claims that terrorist works (Dershowitz, 2003; Pape, 2005) is actually confined to game-theoretic models or case studies and therefore it does not reflect the generalized reality. In fact, Abrahms tries to empirically show that terrorism actually does not work. He analyzes the strategic effectiveness of 21 terrorist organizations included in the list released by the US State Department. The results corroborate the inability of terrorists to achieve their goals as posited by Schelling (1991): indeed, Abrahms demonstrates that terrorist organizations included in his sample have achieved their goals less than 10% of the times. The conclusion of Abrahms is that terrorism is not solely ineffective and inefficient as an instrument itself, but that its failures are often determined by the tactics that these terror groups use, especially when these tactics involve deaths of civilians.

Neumann and Smith (2005) also focused on the strategies adopted by terrorist organizations and on their inherent limitations. The authors identify three different modi operandi with related objectives (Table 1.4).

Modus Operandi	**Objectives**
Disorientation	To alienate the authorities from their citizens, reducing the government to impotence in the eyes of the population
Target Response	To induce a target to respond in a manner that is favorable to the insurgent cause
Gaining Legitimacy	To exploit the emotional impact of the violence to insert an alternative political message

Table 1.4: Modus operandi and objectives of terrorism. Source: Author's adaptation of Neumann and Smith (2005)

Besides this classification, the authors argue that strategic terrorism has many limitations since it is based on wrong assumptions related to the psychological behavior of individuals or institutions in critical situations. These assumptions are two:

1. The target's group determination to hold on a particular policy or possession will collapse once it has been exposed to terrorist violence;
2. A terrorist campaign will instill a degree of fear within the target population;

In their work, Neumann and Smith review these assumptions listing cases that prove their weakness, concluding that strategic terrorism is intrinsically limited because it principally "relies on the exploitation of the psychological rather than the destructive effect of armed action" (Neumann and Smith, 2005, 591).

1.1.2.3 The Organizational Structure of Terrorist Organizations

As for the concept of terrorism, many scholars have attempted to provide a comprehensive definition of terrorist organization. Considering the many-sided nature of these organization, this is a hard task. Matusitz (2012) defines a terrorist organization as "an illicit clandestine organization that generally consists of planners, trainers, and actual bombers/killers". For security purposes, institutions and governments are conducting censuses of the existing terrorist organizations all over the world. The United Nations Security Council (UNSC) has adopted two resolutions in 1999 (1267) and 2011 (1989) for the creation of a list of "entities and other groups and undertakings associated with al Qa'ida". This list was intended to counter the financing to terrorism, the transit of terrorist affiliates and the supply of arms. Moreover, in

2001 the European Union, in the framework of the Common Foreign and Security Policy, has published an additional list of terrorist organizations to extend the efforts to prevent terrorist activities. This list was lastly updated in 2016[3]

The investigation of organizational processes and dynamics of terrorist groups has long interested academia. Many approaches have been used to describe how terrorist organize, communicate, employ resources. Notably, Crenshaw (1987) compares the instrumental theory with the organizational perspective, developing the "Organizational Process Theory" (Table 1.5).

Instrumental Perspective	**Organizational Perspective**
The act of terrorism represents a strategic choice	The act of terrorism is the outcome of internal groups dynamics
The organization using terrorism acts as a unit, on the basis of collective values	Individual members of an organization disagree over ends and means
The means of terrorism are logically related to ends and resources, surprise compensate for weakness	The resort to terrorism reflects the incentives leaders provide for followers and competition with rivals
The purpose of terrorism is to bring about change in an actor's environment	The motivations for participation in terrorism include personal needs as much as ideological goals
The pattern of terrorism follows an action-reaction process, terrorism responds to what the government does	Terrorist actions often appear inconsistent,erratic and unpredictable
Increasing the cost of terrorism makes it less likely, decreasing cost or increasing reward makes it more likely	External pressure may strengthen group cohesion; rewards may create incentives to leave the group
Terrorism fails when its practitioners do not obtain their stated political objectives	Terrorism fails when the organization disintegrates; achieving long-term goals may not be desirable

Table 1.5: Assumptions of instrumental and organizational perspectives on terrorist organizations. Source: (Crenshaw 1987, 27)

At that time, the organizational perspective was rarely used in the study of terrorism. After listing the founding assumptions of the two, Crenshaw claims that the final goal of every organization is the maintenance of the organization itself, regardless of the aim the organization had when it was created. This statement clearly divides the individual reasons and interests of the subjects that form the group to the behavior

[3]After the signing of the Colombia peace agreement between the government and the Fuerzas Armadas Revolucionarias de Colombia (FARC), the Council suspended the sanctions against the FARC, excluding the organization from the list.

of the group itself. The theory had the aim to explain the position of the organization related to the use of violence and the process through which an organization decides to shift to it.

After Crenshaw's work, considered the historic events happened during the 1990s and the 2000s, the organizational study of terrorist organizations gained importance in the academic debate. Abrahms (2008) refers to the instrumental perspective calling it "the strategic model", empirically demonstrating its invalidity. Abrahams lists seven "puzzles" that show how the strategic model has many weaknesses despite its policy relevance in the field of counter-terrorism strategies. The "natural systems" model proposed by the author in the final chapter of his work intersects the organizational and the motivational areas. The model states that terrorist organizations "will routinely engage in actions to perpetuate and justify their existence, even when these undermine their political agenda", in a similar fashion to what Crenshaw wrote.

Focusing specifically on the structure of terrorist organizations, Ganor (2008) reviews the most popular typologies emerged from the literature. The main distinction is between hierarchical and network organizations. On one hand, hierarchical organizations have a clear division with authority between the leadership, officials with specific responsibilities, activists and supporters (e.g., Hezbollah, Hamas). On the other hand, network organizations are composed of weakly connected cells and do not present any clear hierarchical authority (e.g., Al Qaeda). Piazza (2009) proposes a further distinction, however focusing only on jihadist groups. Using a case study of post-invasion in Iraq, Piazza develops two categories, merging groups' goals and structure. He divides jihadist organizations acting in Iraq between "strategic groups" and "abstract/universal" groups. The former typology is composed by groups that are similar to secular national-liberation and regime change movements, while the latter is made of groups affiliated with the Al Qaeda network.

The organizational issue also poses relevant questions on whether the organizational structure itself has a significant impact on the lethality of attacks if compared with other types of non-organized terrorism. Indeed, despite the historical prevalence of organized terrorist structures (either hierarchical or network based), during the last years, Western countries have experienced the attacks of lone-wolf terrorists. Even though lone wolf terrorism stems its root long before these days (Malet, 2010, 2013), the increase of their actions has marked a significant turning point. The emergence of the Islamic State (throughout the work also referred as IS)[4] has brought a wave of

[4]Other authors refer to the IS also as ISIS (Islamic State of Iraq and Syria) or ISIL (Islamic State of Iraq and the Levant) or Daesh (an acronym for the Arabic phrase al-Dawla al-Islamiya al-Iraq

violent attacks (especially in Europe) carried out by individual lone wolves (Institute for Economics and Peace, 2016). Comparing a lone wolf with a member of a terrorist organization, the former is a person that "act without group or organizational support" (McCauley et al., 2013). Similarly, Spaaij (2010) defines lone wolves as persons who "(a) operate individually, (b) do not belong to an organized terrorist group or network, and (c) whose modi operandi are conceived and directed by the individual without any direct outside command or hierarchy". Concentrating on more recent dynamics, Feldman (2013) defines lone wolves terrorism as:

> "self-directed political or religious violence undertaken through the "terrorist attack cycle" by individuals—typically perceived by its adherents to be an act of asymmetrical, propagandistic warfare—which derives from a variable amount of external influence and context (notably now online), rather than external command and control"

The definition provided by Feldman foregrounds the use of the internet. Indeed, leaderless terrorism has exploited the role of social media and web propaganda, creating the conditions for easier radicalization processes (Aly et al., 2016).

Among OECD countries, the deadliest recent attacks from lone actors occurred in Turkey, France and United States. Specifically, in the United States attacks from lone actors have been estimated to account for the 98% of the total number of terrorist attacks since 2006 (Institute for Economics and Peace, 2016). To testify the criticality of this dimension, it is worth to note that the San Bernardino and the Nice attacks have originated from the actions of lone wolf terrorists. Moreover, scholars have tried to model recurring patterns among lone actor profiles: several studies emphasize the prevalence of young males, and rapid radicalization processes occurred quite recently before the actions as important features (Spaaij, 2011; Bates, 2012; Basra and Neumann, 2016). Nevertheless, Alakoc (2017) empirically demonstrates that organizationally linked suicide attacks are deadlier than lone wolf attacks, suggesting that the issue of lone wolves must not overshadow the power of proper terrorist organizations.

1.1.2.4 The geographical range of action of terrorist organizations

When looking at the activities of terrorist organizations, it is necessary also to focalize on their geographical range of action. Usually, the main distinction is between domes-

al-Sham (Islamic State of Iraq and the Levant). IS is the english translation of how the terrorist organization calls itself. All these acronyms refer to the same terrorist organization tic and international (also foreign) groups. Many scholars have analyzed these two different dimensions (Dugard, 1973, 1974; Hoffman, 1997; Kurowski and Sussman, 2011).

Simply put, Bergesen and Lizardo (2004) define international terrorism as that kind of terrorism in which the perpetrator, the target groups or the location of the incident involve at least two different countries. Domestic terrorism, on the other hand, is related in its various dimensions only to a single country. Merari (1978) proposed a bi-dimensional point of view to label terrorist organizations, based on their target population and base of operation. The author developed four profiles:

- Domestic-based xenofighters
- Foreign-based xenofighters
- Domestic-based homofighters
- Foreign-based homofighters

Thereby, the base of operation can be either domestic or foreign and the action can be conducted either against a foreign entity (i.e.: xenofighters) or a domestic target (i.e.: homofighters). These four categories lead to three results:

- Xenofighters terrorist groups tend to adopt more indiscriminate tactics than homofighters;
- Foreign-based terrorist groups tend to perpetrate international terrorism;
- Foreign-based terrorist organizations are mostly dependent upon foreign countries' support

Drawing upon more recent patterns, the Institute for Economics and Peace (2016) provides descriptions of the main features of domestic and international terrorist groups. Domestic groups actions are mostly motivated by anti-government sentiment, nationalism, separatism, racism, bigotry or anarchy. In the OECD area, the most prominent domestic groups are the IRA in Northern Ireland, the Euskadi Ta Askatasuna in Spain (ETA) and the PKK in Turkey. All the three organizations are animated by nationalist or independence ideologies. Under the category of domestic terror, the IEP includes also the home-grown terrorism, citing the attackers of the London bombings occurred in 2005. Data analysis included in the same report highlights how PKK is the deadliest domestic terrorist groups, accounting for 529 deaths from 2000 to 2016. The recruitment to domestic groups is mostly influenced by friend and family ties (Institute for Economics and Peace, 2016, 48). Conversely, the recruitment to international groups is driven by education and employment conditions. IS is nowadays the most prominent international terrorist group.

1.2 Theoretical Framework

Social sciences account for a massive number of theories that aim at explaining human nature by means of human behavior in several contexts and domains (Hull, 1943; Klein et al., 1993; Endsley, 1995; Monroe and Maher, 1995; Hechter and Kanazawa, 1997; Naylor et al., 2013). The study of crime and terrorism is no exception. For instance, various theories have been developed, proposed and tested to explain why and how individuals (or, eventually, collective organizations) engage in crime (Burgess and Akers, 1966; Sutherland et al., 1992; Moffitt, 2003). Some of them have been partially or completely falsified by empirical evidence (for instance the General Theory of Crime proposed by Gottfredson and Hirschi (1990)), while others are still debated and have not led to comprehensive agreement. The difficulty of agreeing on specific theories to explain social phenomena is strongly connected to the inherent complexity of human nature. Theories generally provide simplified and restricted explanations of human decisions and actions, failing to take into account the whole set of concurring and intervening factors that may have an impact on an agent's thoughts and actions.

This considered, the use of a theoretical framework as the backbone of this dissertation should not be seen as an attempt to universally explain how jihadism behave and why it behaves in certain given ways. I am fully aware that, as a portion of reality is captured by the selected theories, other components are missing. However, it is useful to frame the present work in relation to established theoretical and empirical explanations so that these theories may offer insights to interpret and read the results of the different analyses.

This work will then rely on two theoretical components: theories aiming at explaining the spatio-temporal clustering of terrorism, and theories aiming at explaining terrorist decision-making.

1.2.1 The Spatio-Temporal Concentration of Terrorism

The first documented application of the idea that crime is unequally distributed across spatial units dates back to the work of Quetelet on crime in France, Belgium and Holland (Quetelet, 1831). Decades after, within the School of Chicago context, Shaw et al. (1929) showed that crime was dramatically unequally distributed across the neighborhoods of the city.

Following these two pioneering and seminal works, criminologists have widely provided further evidence on the spatial nature of crime and, specifically, on its tendency towards clustering across some given areas. This stream of research produced the development of a subfield within criminology, the so-called "criminology of place" (Sherman et al., 1989). Furthermore, a law, named "law of crime concentration" has been also proposed by Weisburd (2015). The "law of crime concentration" indeed states that crime is densely distributed in a small number of micro-places (e.g., street segments) in a city, and that this distribution is generally stable over time. Many studies have then tested this proposition, providing evidence of its actual meaningfulness in different cities and contexts (Weisburd and Amram, 2014; Wheeler et al., 2016; Favarin, 2018).

Besides the actual empirical corroborations of this law, and even before its official formalization, a wide number of works have reasoned around the concept of the non-random distribution of crime across places, applying also a variety of statistical methods that became lately more and more sophisticated (Brown, 1982; Dutt and Venugopal, 1983; Felson, 1987; Evans and Herbert, 1989; Gorr and Olligschlaeger, 2010; Murray et al., 2001; Ackerman and Murray, 2004; Harries, 2006; Nath, 2006). The numerous empirical evidence towards crime concentration across space (and time) went beyond the borders of academia, inspiring a great shift also in policing practices. The diffusion of CompStat in the late 90s first marked a change towards resource allocation of police departments in the United States, with data-informed decisions working on the assumption that crime does not occur the same way in the same areas of the city (Henry, 2002; Weisburd et al., 2003).

CompStat certainly changed the policing scenario, but eventually opened the path towards more sophisticated and algorithmic supporting systems for the law enforcement. The spread of AI companies brought predictive policing software into the market (Perry, 2013; Dunham and Alpert, 2015; Bennett Moses and Chan, 2018). These software simply make use of machine and statistical learning algorithms that are able to identify hot-spot areas (i.e., areas in which crime tends to concentrate heavily) to suggest to officers where to intervene. Besides the technical aspects of these systems, they all rely on the accepted and confirmed fact that crime (although not all crimes follow this law) clusters in certain areas and at certain times of the day. Different explanations have been provided to justify these findings, with scholars

testing well-known criminological theories such as social disorganization, crime opportunity and rational choice.
While criminology generally focused on predatory crimes to evaluate the "spatio-temporal clustering" hypothesis, parallel fields as political science and international studies reached similar consensus on a tightly-related matter: the spatial and temporal dynamics of violent conflicts and terrorism. Although with many differences (regarding unit of analyses and the nature and impact of the social phenomena), scholars have empirically showed that political violence and terrorism tend to cluster in certain areas and to behave in patterned ways. This can happen by means of contagion and diffusion processes (Midlarsky, 1978; Pitcher et al., 1978; Hamilton and Hamilton, 1983; Myers, 2000) bursts and micro-cycles, within the context of near-repeat victimization (Behlendorf et al., 2012). For instance, the seminal work of Midlarsky et al. (1980) highlighted the presence of contagion in international terrorism and the existence of autocorrelation processes within and between regions.
Additionally, terrorism, as crime, not only clusters in space but also exhibit patterned dynamics of concentration in time. Besides the theoretical explanation and description that one should give to its decision making (which will be covered in the next subsection), terrorism is localized in time and behaves through non-random timings (Enders and Sandler, 2006; Medina and Hepner, 2008; Siebeneck et al., 2009; LaFree et al., 2012; White et al., 2013; Tench et al., 2016).
Although criminology (as the social sciences in general) are often far away from providing universally accepted answers to explanations and even descriptions of social phenomena, the spatio-temporal concentration of crime and terrorism has been confirmed and proved by a wide number of studies, spread over a century of research. This finding represent a crucial frame for my work, given the nature of the data and analysis that will be presented later on in the manuscript. All the three analytic chapters will combine either temporal or geographic (or even both) information on terrorist events plotted by a sample of jihadist groups, thus relying on the concentration of terrorism over time and space as a strong component for interpreting real-world dynamics of terrorist groups.

1.2.2 Strategic Terrorism

The definition of terrorism has been long debated and so are its origins. However, there has been an agreed turning point in which terrorism started to become systematically formalized as to attract, train, educate individuals. This turning point is tightly related to the essay "*Mord und Freiheit* - Murder and Freedom", written by

Heinzen (1853). In this essay, the German revolutionary argued that murder must be turned into a science and that revolutionaries have to overcome the asymmetry between the State warfare and resources by means of strategies characterized by high-profile violence (Bessner and Stauch, 2010).
This call inspired many other ideologists of terrorism and anarchists, as Bakunin, Zaichnevski, Recluse and Romanenko. These were all central contributors to the development of the modern doctrine of terrorism, marked by the so-called "propaganda of the deed", introduced by Carlo Pisacane and afterwards made popular by Carlo Cafiero and Errico Malatesta in the declaration to the Anarchist International in 1876 (Fleming, 1980; Linse, 1982; Garrison, 2004; Kassel, 2009).
Since then, and following the advances made by technology and communication, theorists of terrorism have proposed philosophies and developed schools of thought with regard to terrorist actions and behavior. The first philosophy was rationalism. Rationalism posited that violence represented a mean to an end. A second competing philosophy was expressionism. Expressionism, in turn, considered terrorist violence as a form of individual expression, and resorting to terrorism meant an existential choice (McCormick, 2003). These two different approaches towards terrorism developed over time and inspired contemporary theories of decision making. Rationalism contributed to the formalization of the "strategic theories", while expressionism mostly influenced "psychological theories". Finally, a third contemporary frame is represented by "organizational theories".
These three approaches are not mutually exclusive, but instead provide different interpretations from distinct standpoints to read and understand terrorism. Given the nature of my work, I will solely focus on strategic theories. The whole work will rely around data and objectives that deal with the tangible actions of a sample of jihadist groups, namely its terrorist attacks. Attacks comprise a number of different characteristics and features that can reasonably fall under the umbrella of strategic choices. As McCormick (2003) explained, a group's choice of targets, tactics and timing together defines the group's own "operating profile". This is the reason behind the decision to focus solely on strategic theories, as organizational theories focus on the internal structure and the formal symbolism of each group as a way to read their behavior, and psychological theories are more concerned with other non-rational considerations covering motivations, individual drivers and personal traits.
Strategic theories regarding terrorist decision-making originate in the study of conflicts. Schelling (1980) posited that the parties engaging in a conflict are adaptive strategic agents. This means that the parties try to find the most suitable ways to

win, ruling out the opponent, as in a game or a contest. This simple consideration has been widely adopted by terrorism researchers. Scholars have then formalized terrorism as an instrumental type of activity carried out to achieve a given set of long- and short-run objectives (Corsi, 1981).
As noted by McCormick (2003), terrorist groups are then organizations that aim at maximizing their expected political returns or minimizing the expected costs related to a set of objectives.
Besides this adaptive and adversarial characteristic, the strategic frame assumes that terrorist groups act with a collective rationality (Crenshaw, 1987; Sandler and Lapan, 1988; Lake, 2002; Sandler, 2003): a terrorist group is therefore a unique actor, that exist "per se". This is certainly a simplifying assumption, as organizational theories explain, considering that terrorist groups can be structured in very different ways and this would impact decision making processes (Crenshaw, 1987; Ganor, 2008; Piazza, 2009; Heger et al., 2012). However, at a general level, when considering historical events and their multidimensional characteristics, the simplifying assumption of collective rationality can hold and can be useful in interpreting and framing the life-cycle or behaviour of a group (Enders et al., 1992, 2011; Enders and Sandler, 2006; Behlendorf et al., 2012; Campedelli et al., 2019a).
Two alternative views constitute the collective rationality paradigm: a strong and a weak alternative. The strong variant, derived from neoclassical economics, assumes that there is no asymmetry between the real world and the group's view, while the weaker variant (also called "procedural") states that these actors act rationally based on their beliefs that, however, are asymmetric and incomplete (Simon, 1987). This latter view has certainly found much more empirical evidence in literature (Bowen, 2004; Fussey, 2011).
A number of constraints severely limit the strategic decision making of a group, even if there is a rational way of behaving behind an organization's actions. These aspects have an impact on the type of attacks (as the ultimate and visible step of a decision making process) that a terrorist group will plot. Tangible and intangible constraints have the power to deeply influence the frequency, severity and impact of a terrorist events.
Drake (1998b), for instance, listed and described the constraints that lead to the final selection of a target by a terrorist group. These constraints include ideology (group dynamics, capabilities), strategy (protective measures) and tactics, as decision making does not occur in a vacuum. The ideology helps in defining a set of potential targets, as also showed by Asal et al. (2009). Furthermore, the need for support,

their own capabilities and the security environment also play a role in determining the boundaries for strategic actions. Support is fundamental, as terrorist groups in general wish to benefit some portion of a given society, and therefore unintentionally harming or damaging those who might represent future recruits or support providers could have a very negative payoff for the group itself (Schwartz et al., 2009; Byman, 2013; Benigni and Carley, 2016).

Groups capabilities are then fundamental to assess the potential outcomes of the decision making process of a given organization. Material resources and individual skills are crucial in the actual computation of potential strategies: trivially, groups with a higher number of affiliates, with higher material and economic resources and with a wider set of technical capabilities will have much more options compared to smaller and less powerful organizations.

McCormick (2003) reports Schelling words in defining the strategic approach as a "cheap theory" and notes how the simplicity of this approach is both a strength and a weakness. Certainly, such approach completely fails in really understanding the way in which decisions are made within a group and leaves out other important components for fully understanding terrorism. Nevertheless, the strategic frame helps in unfolding some of the visible dynamics that data can reveal. Variations in combinations of tactics, weapons and targets, for instance, could be better understood if assuming (even partially) rational decision making processes. Temporal information on over-time event characteristics may reveal very much regarding the group itself (Martin and Perliger, 2012; Gilli and Gilli, 2014; Polo and Gleditsch, 2016). Long periods of low or no activity might imply an insufficient level of resources. Conversely, a long series of attacks against hard targets with sophisticated weapons would imply exactly the opposite, providing further information on the current state of a group and on its short- and long-term goals.

2 Motivations and Aims

2.1 Aim of the Work

Though terrorism has not been considered a core topic in classic criminology for a long time, scholars have lately tried to apply, tailor or test renowned criminological theories to the problem of political and terrorist violence (LaFree and Freilich, 2016), although its mechanisms are certainly different from most ordinary and organized crimes. Terrorism has the power of affecting thousands of lives with one single action, with its consequences spanning also to the economy and the stability of the political process of targeted countries. That considered, within the study of criminal and deviant processes terrorism certainly plays a role in terms of impact on humanity. As previously said, even though terrorism is a multifold creature that relies on different ideologies and motivations, in the last decades jihadism has become its main protagonist. Jihadism itself is not a unique, homogeneous concept, but yet it possesses some specific features that distinguish it from other types of terror and represents an actual threat to peace, development and security in many regions of the world in spite of its declining trends in many parts of the world.

For this reason, the need for accurate and useful research is of indisputable value. Nevertheless, research in terrorism is often of little help in solving problems. Sageman (2014) highlighted how, despite years of government funding for research in terrorism, the research itself is still too far away from solving concrete problems.
This issue is determined by the lack of empirical data and empirical studies: the author notes how the majority of research is still (poorly) interview-based or historical-focused. Data on terrorism have been scarce for a long time: only recently few institutions started to provide well-built databases for conducting high-level research in the field. Furthermore, governments and intelligence agencies usually avoid sharing information with academia for privacy and political motivations. In this context, researchers should extract the best from the available. Having these aspects in mind,

the primary aim of this work will be to build solid knowledge on the existing operational patterns and strategic choices of jihadism, also trying to forecast the actions of the groups under analysis. In parallel, a second aim is, as previously highlighted, to develop and present a novel methodological framework that integrates network science and artificial intelligence.
This framework additionally seeks to demonstrate how complex but flexible models can benefit research on terrorism. To pursue these aims, the dissertation will try to answer the following research questions:

1. What are the recurrent patterns in the behavior of the considered jihadist groups?

2. What are the relevant similarities and the significant differences in how these terrorist groups act?

3. To what extent jihadist terrorism show memory-like processes when multi-dimensional information on attacks is considered?

4. What can be the contribution of complex networks, mathematical modeling, and artificial intelligence to the study of terrorism?

The research problems and related questions have implications both from the academic and policy standpoints. Identifying patterns and features related to the behavior of different jihadist terrorist organizations and predicting their strategies can help in advancing the knowledge on future scenarios, providing policy-makers with concrete insights for combating terror. Investigating the concept of "memory-like" processes in terror behaviors can enhance the knowledge on how these groups act and plan future attacks, highlighting possible strategies or recurring patterns, contrasting the hypothesis that terror happens randomly and is intrinsically unpredictable.

Finally, the specific attention that will be dedicated to targets throughout the work is of particular relevance considered that analyzing what entities jihadist groups attack can shed light on potential consequences, damages, and impacts of terror events. While certainly predicting tactics or employed weapons might be helpful, forecasting targets is even more important also from an intelligence perspective. Agencies and law enforcement seek to understand what terrorism will hit in the future: they potentially care less about what type of weapon jihadists will use, as if they cannot forecast against whom they will be using it, the effort is almost useless. Generally, weapons and targets are inter-related (to exemplify, we can assume that it is almost impossible that a group will attack a government building using stones), and if resources have to be efficiently allocated, then a good way to understand how to manage them is to define which are the most probable targets that have to be protected.

In light of these aspects, this dissertation will focus on the world's five most active jihadist organizations (namely the Islamic State, Al Qaeda, Boko Haram, Afghan Taliban, and Al Shabaab). The analytic part will be divided into three different dimensions:

1. Presenting a technical framework originating from Markov chains to detect the complex behavioral structure of jihadist groups and developing a coefficient to measure similarity in the dynamic processes of weapon and target selection by jihadist groups;

2. Investigating memory-processes and self-excitability in terrorist events via Hawkes processes modeling;

3. Integrating dynamic meta-networks within Long Short-Term Memory Networks with the aim to predict the most probable targets in the future, developing the basis for a prediction model.

The outcomes may open new pathways towards the implementation of these techniques to evaluate the risk of incidents, illuminating covered patterns and decision-making processes to design effective prevention policies aimed at countering jihadism.

2.2 On the Need for Rethinking Research in Criminology and Terrorism

The focus of this dissertation is not merely criminological. This aspect has to be taken into consideration from the very first pages. The rapid advances in computational methods and the need for scalable and efficient solutions for complex social problems are calling for a revolution in the way social sciences are studied and addressed. A call to which many have recently started to respond all over the world. Many universities have now hybrid departments of quantitative social sciences, computational sociology and research groups in which scientists from a variety of backgrounds work together to solve societal issues. Nevertheless, still resistance from a wide part of the sociological and criminological community exists towards the evolution of the field. This is especially the case concerning the rise of machine learning and Artificial Intelligence (AI), which are going to represent the mainstream approaches in a lot of disciplines at the moment (including health-care, finance, and industry-oriented research), but still lie in a primitive phase of application in social sciences.

Resistance to innovations is typical in human societies, and the research community is, indeed, a mere subset of the greater set of humans living on the planet. We should not be surprised then to realize how slow is the process of accepting and exploiting this massive change in academia, too. Many reasons can lie behind this resistance: the impression that machine learning algorithms are only black boxes that are unable to provide theoretical interpretation for results, and the discontinuity that algorithms mark not only with qualitative social sciences but also in relation to classical quantitative approaches, are among the explanatory factors of this reaction. While social scientists using quantitative methods generally apply these techniques to test, verify or falsify theories (therefore tweaking the data based on the specific problem, without letting the numbers speak), machine learning and AI algorithms explicitly look at the performance, rather than the explanation. Resistances are physiological, then, but unlike the past, we are now facing a key turning point in history and science we shall carefully consider.

My deep belief is that the future of social sciences on one side, as the future of Artificial Intelligence (AI) on the other, have both to rely on a strong interdisciplinary dialogue to make things work. The only difference is that while AI will not disappear if social sciences will not be able to communicate with it, for social sciences the potential cost is radically higher. On one side, AI will continue to grow and influence our lives and if no control will be provided, outcomes may become harmful. On the other side, traditional social sciences as intended in the past, where works were mainly qualitative or ethnographic or, at most, softly quantitative, face the concrete risk of becoming completely useless for policy purposes and real-world applications. The ultimate end of this type of vision of social sciences will be of a futile academic exercise with no concrete impacts. In parallel, as AI starts to become prominent in the public sphere, we - as the criminology scientific community - cannot leave the design and implementation of algorithms in the hands of engineers and computer scientists alone. This especially in consideration of the risk of the unequal distribution of towards more applied research in the hand of the computer science community as a consequence of the hype that surrounds AI worldwide. We – as humanity – will need a deep knowledge of social processes, dynamics and phenomena to avoid the risk of biased AI systems, unintended consequences at scale and potential massive pitfalls in the long run.

In light of this, I have devoted my dissertation to the goal of bringing together network science and AI to advance the knowledge on jihadism and to design possible applications that may be useful in the future out of the pure academic debate, relying on the idea that terrorism research can benefit from advanced computational techniques, and that AI shall devote more resources and efforts to tackle societal problems, such as terrorism. This vision is related to the concept of responsibility. As I intend it, research shall not be confined to departments and laboratories and conferences, but should instead serve as much as possible (especially in this field) to provide tools or answers to questions that deal with the wider human community. Additionally, scholars in areas which are changing more rapidly than in the past - as happens to the vast realm of criminology and social sciences - should be afraid of the consequences of this shift only if they will decide to keep themselves away from it, without getting involved.

The prominent methodological shape of this work does not mean, of course, that research on terrorism and criminology shall only be a matter of data and computational models. Many past attempts of applying algorithms and computational techniques to terrorism and other social and criminal phenomena by computer scientists, statisticians and mathematicians have been completely useless because of the lack of conceptual and theoretical knowledge on the analyzed phenomenon. Integration, interdisciplinarity, and cohesion are three keywords that shall guide the “new deal” of social sciences.

Too many times research in criminology and terrorism has been a mere “replication” of past methodologies, without investing adequate resources in novel approaches. This, of course, avoided to take the risk for possible failures but at the same time prevented scholars from trying to push the border a little bit further. Furthermore, if scholars will continue to use theories and theoretical frameworks as chains to which the scientific community has to pay a perpetual debt, they will only limit the path of science towards knowledge and innovation. Too many times I have seen or heard colleagues and researchers worried about the theoretical framework to use instead of focusing on the right research questions. This, to me, has always appeared as a gigantic problem. Being worried about what legacy of the past should act as the cornerstone of a new research and not investing time and resources to try to create something new in order to respond to meaningful and relevant questions is, in my perspective, a sign of somehow diverted research. For guaranteeing a role to social sciences in the real world of tomorrow, I do believe this process has to be changed. Hence, this humble work explicitly aims at pointing in that direction.

This page intentionally left blank

3 Case Studies and Data

3.1 Jihadist Terrorism: Concepts and Actors

As outlined in the previous chapter, there exist different categories of terrorism. Religious terrorism in one of these categories, and within this category stands Jihadism. There are other forms of religious terrorism (e.g. Christian or Jewish terrorism) but it is indubitable that, in the last decades, jihadism accounts for the large majority of attacks, deaths, damages and tangible consequences. Its existence has deeply influenced the recent history and still have a dramatic impact on the security and safety of many countries and people today. The following subparagraphs briefly overview jihadism as a typology of terrorism, focusing then on the IS, Boko Haram, the Taliban, Al Qaeda, and Al Shabaab to introduce their history, goals and nature.

3.1.1 Defining Jihadism

The prominent role acquired by jihadism is proportional to its complexity: Islamist terrorism is rooted in history and involves a multitude of religious, social, economic and political factors. From a philological point of view, Jihad means "to strive" in Arabic. The word has a different origin compared to what it communicates today. Jihad, indeed, has a primary spiritual meaning, while the idea of the "jihad by the sword" was developed afterward. The concept of "jihad by the sword" was defended by radical groups that used it as the main tool to legitimate Islamist terrorism (Springer et al., 2009). In the last decades, after realizing the prominence of Islamist terrorism, researchers and scholars have started to investigate the origins of the contemporary jihadist ideology to understand how it changed and spread all over the world.

The multifold and interconnected events of the Twentieth century have surely influenced if not fueled the expansion of jihadism. Colonialism, the establishment of Israel in 1948, and the influence of the United States after World War II are among the main external key factors. Besides external factors, internal factors to the Middle East also had an impact on the evolution of the problem. Springer et al. (2009) points out that also the presence of secular regimes (e.g. Egypt), corruption, unemployment, failed economies and the Arab loss against Israel in 1967 have led to the Islamist insurgency. Moreover, influencing personalities like Sayyid Qutb built intellectual and ideological theories to support the concept of "Global Jihad" (Nettler, 1996). Original members of Al Qaeda and many leaders like al Zawahiri relied on his theories and ideas (Perry and Negrin, 2008) and, consequently, Global Jihad became the most relevant doctrine in Jihadist terrorism. In this context, Al Qaeda acted as the leader of a worldwide number of Islamist networks and organizations, encouraging all the others to embrace the idea of "holy war". Since 9/11 and the events of Madrid (2004) and London (2005) a lot has happened. In more than fifteen years, Jihadism has partially changed its face, even becoming more determined and merciless. The rise of the Islamic State and the birth and reinforcement of many other Islamist terrorist groups have shaken the world dozens of times. Attacks in Western countries have increased and Islamist terrorism has shifted towards a more decentralized and loose structure. Lone wolves and self-radicalized individuals have joined the idea of jihad, making attempts to predict and anticipate the attacks almost impossible.

In this scenario, four terrorist organizations have played a key role. In 2015, the actions of the IS, Boko Haram, the Taliban in Afghanistan and Al Qaeda have accounted for 17,721 deaths all over the world (Institute for Economics and Peace, 2016). Two years after, in 2017, Al Shabaab substituted Al Qaeda in the list of the four deadliest groups worldwide provided by the Institute for Economics and Peace (2018): the Somali group, with the Islamic State, the Taliban, and Boko Haram were alone responsible for 10,632 deaths from terrorist events. These single aggregate figures stress the tremendous impact that these five jihadist organizations have on society.

3.1.2 The Islamic State

In the last years, the world has experienced the radical and dramatic actions of the IS. While during the 2000s Al Qaeda was considered the main terrorist actor in the global scenario, since 2007 the IS has slowly but constantly overcome all the rivals and acquired the role of most critical terrorist threat for many countries in the world.

The origins of IS dates back to 2000 in a militant group called Jamaat al-Tawhid wal-Jihad (JTJ), founded and led by Abu Musab al-Zarqawi (Hashim, 2014). JTJ was

one of the groups fighting against the US invasion of Iraq in 2003. In 2004, the JTJ joined Al Qaeda and changed its name in Al Qaeda in Iraq (AQI) (Fishman, 2016). Hashim (2014) reports that the goals of the group were mainly focused on the removal of the aggressor from the territories of Iraq, on waging Jihad to liberate Muslim territories from infidels and establish a caliphate ruled by Sharia. Subsequently, in 2006, some divergences between al-Zarqawi and the central structure of Al Qaeda and the final death of the leader of AQI, induced to the attempt to create the prototype of the current IS. The chosen name was Islamic State of Iraq (ISI), and the organization control was given to Abu Omar al-Baghdadi. The project failed due to the lack of resources and by 2008 Iraq experienced a relatively peaceful period. Nevertheless, between 2010 and 2013, the ISI came again to the attention of the world. Its leader started to explicitly call for the creation of a caliphate, and the group changed again its name in "Islamic State of Iraq and Syria" (ISIS). The violent victories in the territories of Iraq and Syria gave favorable conditions to al-Baghdadi for the establishment of the caliphate. In 2014, ISIS changes definitely its name in "Islamic State" (IS).
The anomalous nature of this terrorist organization has been emphasized by many. Cronin (2015) sustained that the IS cannot be described as a terrorist group, stressing the differences with Al Qaeda. It is indisputable that IS has many features that distinguish it from all the other terrorist groups of the world. Firstly, it has established a caliphate (Jabareen, 2015). Secondly, the IS could count on a number of members, affiliates, and fighters much higher than any other group. Thirdly, the IS can rely on economic, communication and military resources that no one else possesses in the terrorism scenario (Stergiou, 2016). With regard to this, studies show how the IS relies on the power of the internet to recruit members and to spread its propaganda more than any other group (Farwell, 2014; Klausen, 2015; Mahood and Rane, 2017).

Data returns an impressive snapshot of the human flows that go from the Western countries to Iraq and Syria to join the caliphate. The IS benefits from the so-called foreign fighters too. Hegghammer (2013) defines a foreign fighter as an agent that "*a.* has joined, and operates within the confines of an insurgency; *b.* lacks citizenship of the conflict state or kinship links to its warring factions; *c.* lacks affiliation to an official military organization; *d.* is unpaid". Estimates tell that until 2015, between 27,000 and 31,000 individuals have joined the IS, traveling from 86 countries (Soufan Group, 2015). They first become human capital for terrorist groups (Benmelech and Klor, 2018). Battlefield experiences in Syria and Iraq enhance their military capabilities and increase links and relations with other terrorists. Then, a part of these foreign fighters, after an experience in the Middle East, goes back to the European countries.

Sometimes, returning foreign fighters put into practice what they have learned in the war zones, with the specific intent to carry out individual or group terrorist attacks (Hegghammer, 2013). Vidino (2014) points out that foreign fighters that depart from Europe have different ages, origins, social and economic backgrounds: this finding makes it difficult to develop tailored policies since no defined profiles exist.
The IS exploited all these elements to become the world's deadliest terrorist group. In 2015, IS plotted attacks in 252 cities and its members were responsible for a total of 6,141 deaths (Institute for Economics and Peace, 2016). In, 2016 the number of casualties rose up to 9,132 (Institute for Economics and Peace, 2017). The primary goal of the IS is to create an area of Islamic rule. Although in the last two years the group has been almost defeated and has lost many territories, unlike all the other groups it has succeeded in its aim. The leader is still al-Baghdadi and the group is still considered the deadliest worldwide (Institute for Economics and Peace, 2018).

3.1.3 The Taliban

Despite Kleiner (2000) simplistically defines its members as "warriors", the Taliban is an Islamist fundamentalist terrorist group based in Afghanistan composed by Pashtun tribesmen and Mujahedeen. These Mujahedeen participated in the resistance against the Soviet Union invasion of the country in the 1980s. The group ruled Afghanistan from 1996 to 2001. In 2001, the Taliban have been defeated by a NATO coalition led by the US. In 2013, the NATO coalition reduced its presence in Afghanistan and this provoked an increase in the terrorist activity of the Taliban group (Institute for Economics and Peace, 2016). The origins of the Taliban go back to the Soviet invasion and the subsequent fall of the pro-Soviet government in 1989 (Goodson, 2001; Barfield, 2010; Hyman, 2016). The Mujahedeen groups started to fight against each other and the country was fragmented in many regions. In this scenario, some former fighters formed a group led by the Mullah Omar, with the final aim to "restore peace, disarm the population, enforce Sharia law and defend the integrity and Islamic character of Afghanistan" (Rashid, 2002). The group was called "the Taliban" and rapidly gained popularity and local legitimacy. This escalation led the group to conquer Kandahar in 1994 and Kabul in 1996 (Barfield, 2010).
Besides the other historical events that brought the Taliban to re-enforce their terror activity (which is not the very purpose of this work), analyzing their strategies and targets is important to understand their activities. Johnson (2013) stresses how the Taliban proved to be a highly adaptive group. Their tactics have evolved over time: the long tradition of conflicts has helped the members of the Taliban to learn and

employ different strategies. Despite a low technological level, their strategies are rather sophisticated. The control of the territory through different methods (e.g.: the institution of shadow governments in rural areas, as noted by Johnson) and the battlefield tactical behaviors demonstrate their skills. In this scenario, the current first aim of the Taliban is to overthrow the Afghan government. For this reason, in the last years the Taliban have mostly targeted police forces (Institute for Economics and Peace, 2016). Moreover, it is relevant to note how since the US invasion in 2003, the Taliban have highly increased the use of suicide bombings (Johnson, 2013). The Taliban killed 4,502 people in 2015 (deadliest year ever, +29% if compared with 2015) and 3,583 in 2016 Institute for Economics and Peace (2016, 2017). Concerning economy revenues, the Taliban rely on opium and heroin smuggling as the first funding source for their activities (Thruelsen, 2010; Piazza, 2012).

3.1.4 Al Qaeda

Al Qaeda is the Islamist terrorist organization responsible for the 9/11 attacks. The dawn of Al Qaeda dates to the Soviet invasion in Afghanistan when its founding leader Osama bin Laden and Abdullah Azzam were collaborating in the conflict. In 1988, while the conflict was ending, bin Laden and Azzam founded Al Qaeda in Pakistan. After the assassination of Azzam perpetrated by bin Laden himself, the Saudi-Arabian born terrorist became soon one of the leaders of the Global Jihad Network. After the first phase where the organization strived for the internal jihad (the attempt to conquer the countries with Muslim populations in Middle East, Central Asia, the Indian Subcontinent, and Southeast Asia), Al Qaeda turned to the external jihad. This shift happened across 1997 and 1998 (Gunaratna and Oreg, 2010). One of the main motives of this strategic choice was the intent to stop the Western (and specifically the American) influence in that area of the world (Migaux, 2007). Since that moment, US citizens and building became the very target of the terrorist activity. Some successful attacks (among the other, the attacks against US embassies in Nairobi and Dar es Salaam) preceded the 9/11 attacks which are considered the most devastating and lethal terrorist actions of modern history.
After the death of bin Laden in 2013, the organization has renovated itself to better achieve its goals. Nowadays, Al Qaeda is a global organization, decentralized and franchised around a central control group. Currently, Al Qaeda's main affiliates are Al Shabaab, Al Nusrah Front, Al Qaeda in the Arabian Peninsula (AQAP), Al Qaeda in the Islamic Maghreb, Abdullah Azzam Brigades and Al Qaeda in the Indian Continent (Byman, 2014). Zehr (2017) develops the concept of "Al Qaeda phenomenon"

to describe that process that has led to the worldwide proliferation of terrorist organizations similar to Al Qaeda. According to the author, Al Qaeda had the power to inspire the IS and many others to join the jihad. As mentioned, a relevant aspect of Al Qaeda is the connection that the organization has with many other groups, relying on the ideology of the Global Jihad. With the rise of the IS, the terrorist narrative has mainly concentrated on these two entities to understand differences, similarities and possible evolution in the relations between the two. Abu Musab al Zarqawi was helped by Al Qaeda in the foundation of the JTJ, but after his death and with the expansion of the IS, the two organizations started a feud. In fact, there have been clashes between the IS, Al Qaeda and other groups like Al Nusrah. Holbrook (2015) (2015) notes that Al Qaeda sought to present the group as a moderate alternative to the IS, but the events of the last years demonstrated that the IS managed to become the leading force in the Jihadist terrorist scenario. Despite the loss of power, in 2015 Al Qaeda was responsible for 1,620 deaths (-17% if compared to 2014) and 368 incidents, while in 2016 the jihadist group has killed 1,349 individuals (Institute for Economics and Peace, 2016, 2017).

3.1.5 Boko Haram

Boko Haram is a Nigerian terrorist group which has first come to the attention of its country chronicles in the early 2000s (Onuoha, 2010). The name means *"Western education is forbidden"*. Its main goal is to establish an Islamic caliphate in Nigeria. The group has partially succeeded in his goal, imposing the Sharia law in some of the states of Nigeria. Its operations mainly concentrate in Nigeria - the state of Borno is the epicenter of the terror - and in adjoining African countries like Burkina Faso and Cameroon (Institute for Economics and Peace, 2016). Several scholars have tried to analyze the context in which the terrorist organization has developed. The group was born in the north of Nigeria, and it is currently the major security issue for the Nigerian government (de Montclos, 2015; Abubakar, 2017).
Due to its limited geographical range of action, Boko Haram has not been directly considered a priority for Western institutions and academia. Nevertheless, it poses great challenges for the stability of the area threatened by its presence. According to many scholars (Onuoha, 2010; Loimeier, 2012; Akinola, 2015; Iyekekpolo, 2016), the expansion of Boko Haram in Nigeria is caused by multiple factors that are independent of the mere religious aspect. These factors are the extreme Nigerian poverty, the weak efforts of the government in countering the terrorist threat and the grievance of large local areas against the institutions. According to these authors, economic

conditions and political opportunities have fuelled the Boko Haram expansion. Moreover, Abubakar (2017) includes also the corruption of the government and the failure of the northern elites in implementing Sharia as important causes of the rise of Boko Haram.
Despite the first phase in which Boko Haram targeted Nigerian security forces and mainly applied "hit and run" strategies, in the last years the group has started to carry out attacks also against religious and educational institutions and civilians (Regens et al., 2016). The strategies have also changed: the group started to act to occupy and conquer territories, increasing the brutality of the attacks (Weeraratne, 2017). The terrorist group was responsible for 5,478 death in 2015 with an 18% decrease if compared with 2014 when Boko Haram was recorded as the deadliest group of the world (Institute for Economics and Peace, 2016). In 2016 the number of deaths continued to decrease (1,079) due to the defeats inflicted by Nigerian military forces. Institute for Economics and Peace (2017) also reports how these military defeats led to the separation of three factions within Boko Haram in late 2016: a violent, an Islamic State-aligned and an Al Qaeda-aligned one.

3.1.6 Al Shabaab

Harakata al-Shabaab Mujahideen, mostly known as Al Shabaab, is a jihadist terrorist organization that first appeared in the area of Mogadishu, Somalia, in the early 2000s (Hansen, 2013). Initially, the group was an urban militia aiming at defending the Islamic Courts Union in the capital of Somalia. Since then, Al Shabaab has gained importance in the country and started to control an increasing number of rural territories and cities. In 2010, the group attracted the attention of the international community because of the "World Cup bombings" in Kampala (Anderson and McKnight, 2015). In the last years, the group evolved and expanded across different territories. As noted by Mueller (2018) while the group covered a marginal role in Somalia in the early stage of its existence, Al Shabaab is now one of the most relevant players in the process of armed opposition against the nascent Somali government (allies included). This transition has also affected the geographic scope of attacks, which has been expanded in the last years, spanning over different adjoining countries. While the political discourse tried to paint the group as in the middle of a gradual receding process, scholars have demonstrated its resistance and resilience, claiming that Al Shabaab benefits of more legitimization if compared to the federal government of Somalia (Lind et al., 2017). After a longstanding informal relation, since 2012 Al Shabaab is officially considered part of the Al Qaeda network (Joosse et al., 2015).

The African Union Mission in Somalia (AMISOM), a regional peacekeeping mission that is operated by the African Union under the consent of the United Nations, is one of the actors that better contrast the role of Al Shabaab in the region. AMISOM forces have taken over strategic locations and deprived Al Shabaab of resources and physical territories but, as pointed out by Cannon and Ruto Pkalya (2017), the group kept maintaining capabilities and clear strategic goals, also exploiting the absence of a strong and effective statehood.
Cannon and Ruto Pkalya, additionally, argue that the group has evolved towards a universal/abstract organization. This classification is based on the multi-casualty and indiscriminate nature of its attacks. Attempts to classify the group have been done also by other scholars. However, no universal agreement on the organizational model or typology that Al Shabaab represents. Ingiriis (2018) tried to enhance the knowledge on the group exploring the relationship between it and the twentieth-century anti-colonial Somali movement of the Dervishes, highlighting some imitation in the way Al Shabaab operates. Tar and Mustapha (2017) argued that the success of the group is partially explained by the cooperation and alliances it has consolidated with local warlords. The authors even try to categorize Al Shabaab itself as a warlord group, due to the activities it has carried out in Somalia, including racketeering and plundering. The group has presumably killed over 4,000 individuals since it was born in 2006 (Institute for Economics and Peace, 2017).

3.2 Data

3.2.1 The Global Terrorism Database

The analyses in this work rely on data drawn from the Global Terrorism Database (GTD), developed by LaFree and Dugan. The GTD is the most comprehensive and detailed open access dataset on terrorist events at global scale.[1] The GTD originates from data collected by the Pinkerton Global Intelligence Service (PGIS): researchers at PGIS were trained to include information on terrorist events from 1970 to 1997 and in 2006 the START Consortium received funding to continue the data collection and update the dataset (LaFree and Dugan, 2007; LaFree, 2010).

[1] While writing this dissertation, the GTD manager has announced that the funding from the State Department has been cut, as no follow-on contract has been granted after the expiration of the previous one, in May 2018. This means that, at this moment, the START research center has no funding to complete collection of 2018 data, nor are they able to publish data beyond 2018.

Data collection continued to date, and START releases an updated version of the dataset every year. The dataset includes now data on more than 180,000 events. Information is gathered from different open sources, and events have to meet specific criteria to be included in the database. These criteria are divided into two different levels. The first level criteria are three and have all to be verified. These mandatory ones are related to the intentionality and the violence (or immediate threat of violence) of the incident and the subnational nature of terrorist actors. The second level criteria are three and the condition is that at least two of them are respected. Second level criteria relate to (1) the specific political, economic, religious or social goal of each act, (2) the evidence of an intention to coerce, intimidate or convey messages to larger audiences than the immediate victims, (3) the context of action which has to be outside of legitimate warfare activities. Finally, although an event respects these two levels and is included in the dataset, an additional filtering mechanism (variable *doubter*) is introduced to control for conflicting information or acts that may not be of exclusive terrorist nature (START, 2017b).

In this work, I have focused on the five most active jihadist groups in terms of plotted attacks from 1970 to 2016. Since each event in the GTD may have up to three perpetrators cooperating in a single attack, I have calculated the cumulative sum of all the appearances of each group even though the attack was executed by one actor or more. Moreover, I have decided to merge all the attacks perpetrated by all the factions belonging to the Al Qaeda network that in the dataset were labeled as separate, creating a single "Al Qaeda" category. The factions that were combined are: Al-Qaida, Al-Qaida in Iraq,[2] Al-Qaida in Saudi Arabia, Al-Qaida in the Arabian Peninsula (AQAP), Al-Qaida in Yemen, Al Qaida in Lebanon, Al-Qaida in the Islamic Maghreb, Al-Qaida in the Indian Subcontinent, Islambouli Brigades of al-Qaida, Secret Organization of al-Qaida in Europe, Al-Qaida Organization for Jihad in Sweden, Al-Qaida Network for Southwestern Khulna Division, Jadid Al-Qaida Banglades (JAQB), Al-Qaida Kurdish Battalions.

This process identified the Taliban, IS, Boko Haram, Al Shabaab and Al Qaeda as the most active groups present in the dataset. After group selection, I have removed

[2]Al Qaeda in Iraq represents the seed of the Islamic State, given that the group then evolved changing its name in "Islamic State in Iraq". One may thus dispute the decision to fold the group into the broader "Al Qaeda" network. This decision, however, is motivated by two reasons. First, Al Qaeda in Iraq has been a formal affiliate of the Al Qaeda network, and it would have then been conceptually wrong to exclude it based on retrospective knowledge about its history. Second, the GTD reports attacks perpetrated by the Islamic State in Iraq, thus clearly distinguishing between the two different groups.

all the attacks that were of labeled as of doubtful terrorist nature (relying on the *doubter* variable). This led to a reduced amount of attacks for each group (Table 3.1).

Group	Original N	Cleaned N	Attack Freq.	First Attack	Last Attack
Taliban	6,607	5,633	0.71	4/20/1995	12/31/2016
Islamic State	4,343	3,562	2.63	4/18/2013	12/31/2016
Boko Haram	2,090	1,901	0.70	7/27/2009	12/31/2016
Al Shabaab	2,689	1,695	0.50	11/2/2007	12/30/2016
Al Qaeda	2,058	1,506	0.17	12/29/1992	12/25/2016

Table 3.1: Number of attacks (original and cleaned) for each of the selected groups

Similarly to the doubtful event, for what concerns attacks for which the perpetrator was not reported, I decided to exclude them avoiding heuristic techniques based on probabilistic methods to estimate the likely perpetrator of the event. The objective was to rely solely on reliable and verified information. Estimating the likely perpetrator is possible, however in absence of ground truth (i.e., the real perpetrator), it would have been impossible to verify the correctness of the response. This would have constituted a risk of bias for the whole methodological architecture of the work. While it is reasonably possible that some attacks lack their perpetrator, I considered safer to only use events with certified perpetrators.

Given the relevance of the temporal dimension for the aims of the analyses, I have instead treated events with no precise date reported (at daily unit detail) in two different ways. If the additional variable *approxdate* was available, I imputed data relying on the information included there. However, if *approxdate* was not precise enough to derive any type of imputation, I filled missing data using the median date based on each month's distribution. If, for instance, in February 2008 there were 9 attacks on days 1, 3, 3, 5, 9, 14, 23, 26, 26, then the median value for the missing dates corresponding to that month would be 9. When the number of attacks in a specific month is even, then I take the average of the two median ones and input the mean value rounded up if the two dates are different, exact otherwise (e.g. if in another month of another year there were 4 attacks on days 3, 5, 6, 8, the median in this case will be $(5+6)/2=5.5\sim 6$) Additionally, I have employed data on attack types (in this work referred as "tactics"), weapons, targets, and targeted countries. In the GTD, each event may have information on multiple tactics, weapons, and targets: in this analysis I have kept all available information, without dropping any variable or a specific category, not to alter the distribution of events and their specific characteristics (Table 3.2).

Group	**Targeted Countries**	**Tactics**	**Weapons**	**Targets**
Taliban	2 (Afghanistan)	9	9	23
Islamic State	21 (Iraq)	10	8	23
Boko Haram	6 (Nigeria)	8	7	21
Al Shabaab	7 (Somalia)	9	7	22
Al Qaeda	31 (Yemen)	9	9	24

Table 3.2: Descriptive statistics of Attack Variables Per Terrorist Group 8Most targeted countries between parentheses).

It is worth to outline that the GTD provides different levels of details for both weapon and target features. Weapons are labeled on two different levels of detail. Variable *weaptype* records the general type of weapon that terrorists used in the attack (e.g. Firearms), while variable *weapsubtype* gives a more detailed and specific type of information related to the weapons used in the event (e.g.: Automated weapon). Targets are instead labeled on four different levels of detail. Variable *targtype* is the most general one, providing a broad class to which the specific target belongs (e.g. Government), *targsubtype* gives further specification, introducing additional information (e.g. Government building/facility/office). Variable *corp* identifies the corporate entity or government agency that was targeted (e.g.: Spanish government) and variable "target" labels the specific person, building or installation that was victimized.

In the present work, I have used in both cases the most general type of categorization (i.e.: variables *targtype* and *weaptype*): this decision was driven by the fact that using a finer-grained level of detail would have led to over-specification, eventually compromising the generalization of results and models. Nonetheless, in order to build more meaningful models, in the case of events where *targtype* was equal to "other", I have used the variable *targsubtype* instead. The residual label "other" includes heterogeneous targets that become more informative if analyzed as separate objects (examples of *targsubtype* variables are Fire Fighters and Ambulance).

3.2.2 Limitations

The GTD is a fundamental open access resource for researchers interested in the study of terrorist events. Nonetheless, this work comes with several limitations intrinsically related to the data or to the choices made in processing the information contained in the database.

Firstly, the GTD presents a limitation that is unsolvable: in fact, all events that occurred in 1993 are missing. As done in previous studies (Santifort et al., 2013), I did not estimate the number events of 1993: I have instead treated them as missing. Albeit four of the five jihadist groups started to plot attacks long after 1993, this lack of information might have biased the statistical analyses for the Taliban.

Secondly, another structural limitation of the dataset as a whole relates to the potential presence of unmeasured events or attacks with no reported perpetrator that are therefore excluded from the analyses. Given that this dataset is built on open access data it can be that attacks, especially low magnitude ones, are not recorded by press agencies or newspapers. Alternatively, events may not be attributed to a specific group due to a tactical or strategical ambiguity that makes it extremely difficult to create a link to a specific subject. These excluded attacks, that can be labeled as "false negatives" pose the risk of constructing biased models that fail to capture real-world dynamics. For this reason, in my future work I hope to be able to test the goodness, fitness, and reliability of the models via statistical techniques that would eventually overcome this inherent limitation of these data. One potential technique would involve the inclusion of random noise data generated by some stylized probability distributions, in order to assess the magnitude of the effects of potential biases explained by unmeasured events. For what concerns attacks with an unattributed perpetrator that might have been instead plotted by one of the groups under analyses, probabilistic methods derived from unsupervised learning techniques may be a feasible way to estimate likely perpetrators. While keeping this in mind, it is worth to point out three things. First, given the digitalization of information that caused an increasing availability of news coming from all the corners of the planet, it is reasonable to think that these problematic events have been reduced in the last two decades, therefore only marginally influencing the present analyses. Second, the groups that I have included in my sample have been under the spotlight due to their high-frequency and high-magnitude activities worldwide. These activities have often led to high-casualty attacks and worldwide attention. Third, jihadist groups generally publicly claim their attacks, further reducing the risk of false negatives. For this reason, it is hard to think that the number of unmeasured attacks for such organizations is that high to disqualify the results of this thesis. Nonetheless, given my future aim to potentially expand my work to other groups or contexts for which these two last caveats may not hold, the problem of unmeasured attacks represents an issue to be tackled seriously.

Thirdly, my own decision to concentrate on countries for the geographical dimension of terrorist events might pose some serious limitations to the utility of the models from the practical point of view. While from the research standpoint, such a macro perspective has a respectable value, policy-makers and analysts might be interested in finer-grained spatial configurations. In fact, countries may be too general spatial approximations and models relying on this abstraction could not provide useful insights for intelligence purposes. A model that correctly predicts attacks in Afghanistan, for instance, does not help suggest where to precisely allocate resources. Afghanistan is a wide country, and data show that the distribution of terrorist events is not equal across provinces. It is thus necessary to take into account that, from the perspective of applied and policy-oriented research, future work will have to address this issue and that limited inferences should be made relying on the geographical component of the analyses. Hopefully, the database will be able to support the improvement of the models from the geographical point of view, given that events are geocoded and more detailed spatial resolution is available for most of the attacks.

Fourthly, and related to the third point, the choice operated by me to use the most general level of categorization for Weapons and Targets certainly preserves the integrity of the information, avoid excessive sparsity and noise, but again limits the potential application of the work for intelligence aims. Certain categories (e.g., "Firearms" as a weapon or "Private Citizens and Property" as a target) are too general and may hide relevant levels of information that can hide further micro-patterns in terms of specific weapons, tactics or targets. What if, within the general "Private Citizens and Property" category stands a certain patterned distribution of multiple different subcategories? The reader must keep this aspect in mind when reading the description of the methodological setup and the results of the different analyses made throughout the work.

This page intentionally left blank

4 Stochastic Matrices of Terrorism: Complexity and Heterogeneity of Jihadist Behavior

Preliminary Note The present chapter reports part of the analyses of the research article *"Pairwise similarity of jihadist groups in target and weapon transitions"* published in the Journal of Computational Social Science in May 2019, with Mihovil Bartulovic and Kathleen M. Carley as co-authors (DOI: 10.1007/s42001-019-00046-8).

4.1 Introduction

Terrorism and its multi-fold complex dimensions are increasingly studied from different perspectives, attracting scholars from several scientific fields. Advanced quantitative techniques, derived from mathematical and statistical sciences, have been applied to increase the knowledge of how this phenomenon evolves and occurs. Although almost all social phenomena spark interest in the scientific community, terrorism - especially in the last decades - has been capable of fostering unprecedented attention due to how it has shocked recent contemporary history. From 1970 to 2016, the Middle East, North Africa, South Asia and Western Europe were the regions with the highest number of attacks (START, 2017a). However, data reveal the global relevance of the issue, considering that in the last four decades terror events have occurred in more than 200 countries in the world. Indeed, the terrorist threat has pushed scientists to provide help through research to contrast the phenomenon. Complex systems, statistics, security studies are few among the several communities from which major applied contributions have been made to the study of terrorism.

In the attempt to innovate and advance the knowledge on jihadist dynamics from a network science perspective, the present study seeks to explore the behavioral dynamics of the world's most active jihadist groups to shed lights on the existing recurring patterns in terms of operational choices and quantify the extent to which these groups show similar or dissimilar tactical choices in their attack sequences. In light of this, I will introduce a novel framework based on Markov chains that will use super-states and stochastic transition matrices to analyze the complexity of jihadist event sequences and propose a pairwise coefficient that maps the similarity of jihadist groups in terms of transitions between targets, weapons and targets and weapons combined. The study will use open access data and will create transition networks and related network trails treating attacks as ordered state sequences. The relevance of investigating the nature of state sequences is strongly related to the inherent nature of terrorism itself. As a matter of fact, the complex changes of tactical and operational decisions by terrorist groups makes it extremely difficult to predict and eventually prevent attacks and consequent damages. While bounded by limited resources and manpower, terrorists usually have at their disposal many potential scenarios to maximize the utility of their actions. It is thus crucial to improve the knowledge of transitions between different states to detect regularities and irregularities in the behavior of jihadist groups.

This study is an exploration in this direction and opens the path for future research. Concerning results, on one side, the creation of N-dimensional super-state transition networks underlines the complex repertoire of combination and sequential patterns existing for all the five groups in all the three scenarios. On the other side, the stability of several pairwise similarities across different transition networks testifies how certain groups actually share (or do not share at all) very common dynamic behaviors. Furthermore, groups that are similar in terms of weapon transitions but very dissimilar in target transitions demonstrate how there is not always a strong connection between the two dimensions, and that the same terrorist goal can be reached by different means, and vice versa.

The rest of the study is organized as follows: the next section presents the Data and the processing techniques applied to structure the information for the purposes of the chapter. The third section will focus on stochastic transition matrices and their applicability to research on terrorism dynamics. It will specifically introduce the basic concepts behind Markov chains and present the N-dimensional super-states framework, also illustrating the quantitative results. The fourth section will instead present Normalized Trail Similarity. The rationale and mathematical formalization will serve to better interpret the outcomes of the analyses commented in the same section. Finally, the last section will provide final remarks with a focus on policy and intelligence implications and highlight potential future research paths.

4.2 Data Processing

This section aims at presenting the data and the operations carried out to manipulate the data for the aim of the work.

The data used in this chapter are a reduced version of the entire history of attacks present for each group in the GTD. For Al Qaeda and the Taliban, very few attacks were recorded from 1992 to 2001. Hence, to avoid noise and too strong assumptions on the presence of dependencies between events years distant from each other, attacks for Al Qaeda and the Taliban plotted prior to 2001 have been excluded. Table 4.1 shows the descriptive statistics of the reduced data, and Figure 4.1 displays the distribution of events at the monthly unit for all the groups.

Group	**Cleaned N**	**Attack Frequency**	**First**	**Last**
Taliban	5,629	1.04	1/7/01	12/31/16
Islamic State	3,562	2.63	4/18/13	12/31/16
Boko Haram	1,901	0.70	7/27/09	12/31/16
Al Shabaab	1,695	0.47	11/2/07	12/30/16
Al Qaeda	1,502	0.26	09/11/01	12/25/16

Table 4.1: Number of Attacks (Original and Cleaned) for Each of the Selected Groups

Formally, for each group g_i is given a sequence of terror events $A_{gi} = (a_1, ..., a_n)$ and a sequence $D = (d_1, ..., d_n)$, representing temporally ordered discrete days. These two sequences are inherently related because the mapping $f: A \rightarrow D$, which connects every event with a unique time-stamp, is always verified. Elements in A are ordered based on D, therefore all the events are ordered by the time-stamp they are associated to, and this order goes from the most distant to the least distant with reference to the present time.[1] Additionally, we define $\mathfrak{T} = \{t_1, ..., t_k\}$ as the set of possible target

[1] It is worth to specify that in the analyses, events will be ordered temporally but without taking into account the actual delta between attacks. This means that there is no difference between two attacks plotted within a range of four days and other two attacks plotted within a range of five months. Additionally, when two or more attacks are plotted on the same day, we order them by the *eventid* variable included in the original dataset, assuming that the information coded in the

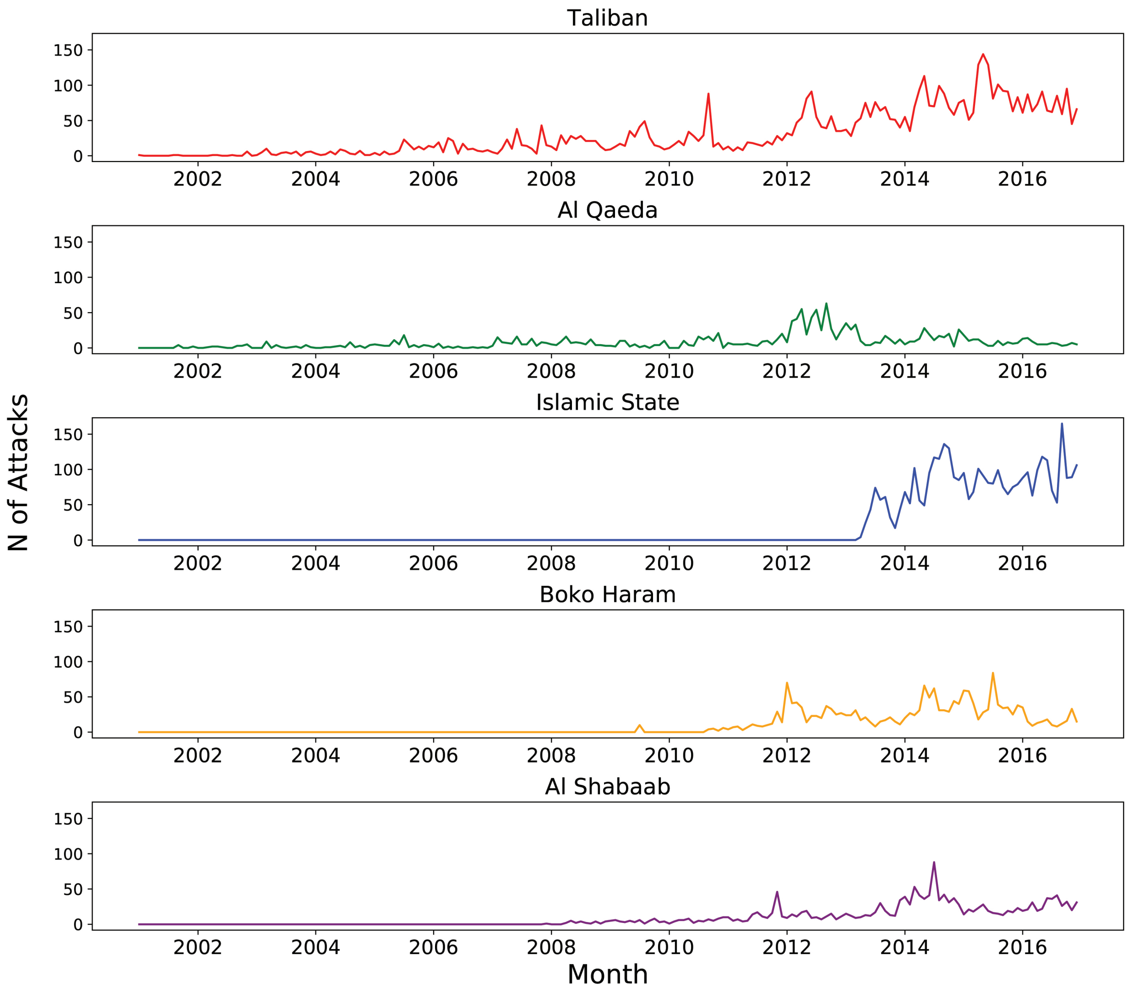

Figure 4.1: Monthly Time Series of Attacks per Each Group (Jan 2001-Dec 2016)

types and $\mathfrak{W} = \{w_1, ..., w_l\}$ as the set potential weapons. Within this frame, we can thus formalize an event in the following compact format:

$$a_{gi}\,(d, t, w) \quad 0 < t \leq 3\,;\; 0 < w \leq 4 \tag{4.1}$$

The format above posits that an event plotted by group g_i is abstractly defined as a combination of three elements: the day it has occurred (temporal element), the targets that have been attacked and the weapons that have been employed. In fact, each event might have been directed to up to three targets simultaneously and might

variable provides a more robust ordering criterion than pure randomly distribution.

have been carried out using up to four weapon types as denoted in Equation 4.1. This further poses the problem of setting up the state spaces for the three classes of analyses I am interested in, namely the analysis of weapons, targets and targets and weapons combined. The state space of a chain, S, is defined as the set of values each element of X_d of the process can take. To exemplify, $S(\mathfrak{T}_{gi})$ is the state space of the existing combination of targets (i.e., states) that g_i can take, and $|S(\mathfrak{T}_{gi})|$ is the cardinality of the set, namely the number of states that it contains.

As mentioned above, the fact that each attack may include multiple targets and weapons dramatically increases the possible combinations of selected targets and weapons to include in the analysis. In fact, in the worst-case scenario, Al Qaeda has attacked twenty-four types of targets during its existence and employed nine types of weapons. To better depict the extremely wide range of possibilities arising from this problem, it is useful to express it by applying basic combinatorics. I first deal with the two single-entity classes, namely weapons and targets, hence excluding for the moment the case of targets and weapons combined. In both cases, combinations of multiple objects (up to three in case of targets and four in case of weapons) are possible, repetitions are plausible (therefore have to be considered), and order matters. This means that in a given combination of three targets, there may be two identical targets and a third different one. Additionally, the order in which these targets or weapons are included in the dataset is important, assuming a hierarchic criterion (descending in terms of importance of the specific feature). Thus, two hypothetical combinations of elements (x, y, z) and (y, z, x) have to be treated as different. Applying these rules, and knowing that our combinations can lie in a finite range, the equations that yield the theoretical cardinality of the two state spaces are the following:

$$\text{Theoretical}|S(\mathfrak{W})_{gi}| = \sum_{r=1}^{4} w^r = 9 + 9^2 + 9^3 + 9^4 = 7,380 \tag{4.2}$$

$$\text{Theoretical}|S(\mathfrak{T})_{gi}| = \sum_{r=1}^{3} t^r = 24 + 24^2 + 24^3 = 14,424 \tag{4.3}$$

These two equations already highlight the huge amount of possible combinations for the two simplest trails and this already showcases that we need a more practical way to tackle such data. Additionally, the number further increases when considering the third type of trails. Indeed, Theoretical$|S(\mathfrak{T}, \mathfrak{W})_{gi}|$ aims at mapping the trajectories of groups behavior when targets and weapons are considered together. In this particular case, we are dealing with a simpler probabilistic problem. As an example, we should think as two sets of finite elements $A = \{a_1, a_2, a_3\}$ and $B = \{b_1, b_2\}$. Our particular problem is finding the number of possible combinations of unique pairs of elements, and each pair must contain one element from set A and one element from set B. Thus, we are not interested in finding unique pairs such as (a_2, a_2) or (b_1, b_2). Given these constraints, it is straightforward to verify that the number of potential unique pairs in our ad hoc example is given by the product of the number of elements in A and the elements in B. Therefore, going back to the main case:

$$\text{Theoretical}|S(\mathfrak{T}, \mathfrak{W})_{gi}| = \left(\sum_{r=1}^{3} t^r\right)\left(\sum_{r=1}^{4} w^r\right) = 14,424 \times 7,380 = 106,449,120$$

Applying the calculation to our third trail leads to multiplying 14,424 by 7,380. The final result is 106,449,120. Considering more than 106 million combinations would have led to a very high expense of computational resources. Additionally, considering all potential combinations might slow down the computation of the Normalized Transition Coefficient. Therefore, to simplify this task, we have coded and considered only the combinations (for all trails) existing in the dataset. The only existing rule, indeed, was that that specific combination of targets or weapons was present in at least one of the events of at least one group. The decision of not considering all potential combinations might be contrasted by one's critique, saying that only considering recorded combinations is a way to artificially bound the extremely wide range of options in the hand of terrorist organizations (especially if considering that the organizations in analyses have – or had – availability of many resources in economic and operational terms).

In spite of this, the analysis focuses on the past, hence concentrating on the universe of existing combinations without paying attention to potential future unexplored combinations. This justifies my choice. That considered, applying this reduction led to sensibly smaller numbers. The first state space $|S(\mathfrak{W})_{gi}|$ is limited to a total of 55 states, $|S(\mathfrak{T})_{gi}|$ is bounded by a total of 200 states, and $|S(\mathfrak{T}, \mathfrak{W})_{gi}|$ has 703 states. In the case of $|S(\mathfrak{W})_{gi}|$ it means that only 0.71% of weapon combinations have been found in the data, while for $|S(\mathfrak{T})_{gi}|$ the percentage is 1.38%. Finally, for $|S(\mathfrak{T}, \mathfrak{W})_{gi}|$ the number reduces further sensibly: data yields the 0.0006% of total potential combinations of targets and weapons. Figure 4.2 visually presents the distribution of most common states (i.e., combinations) for target, weapons and both combined across each group and Table 4.2 guides the reader in decoding the abbreviations used in the plots and the text.

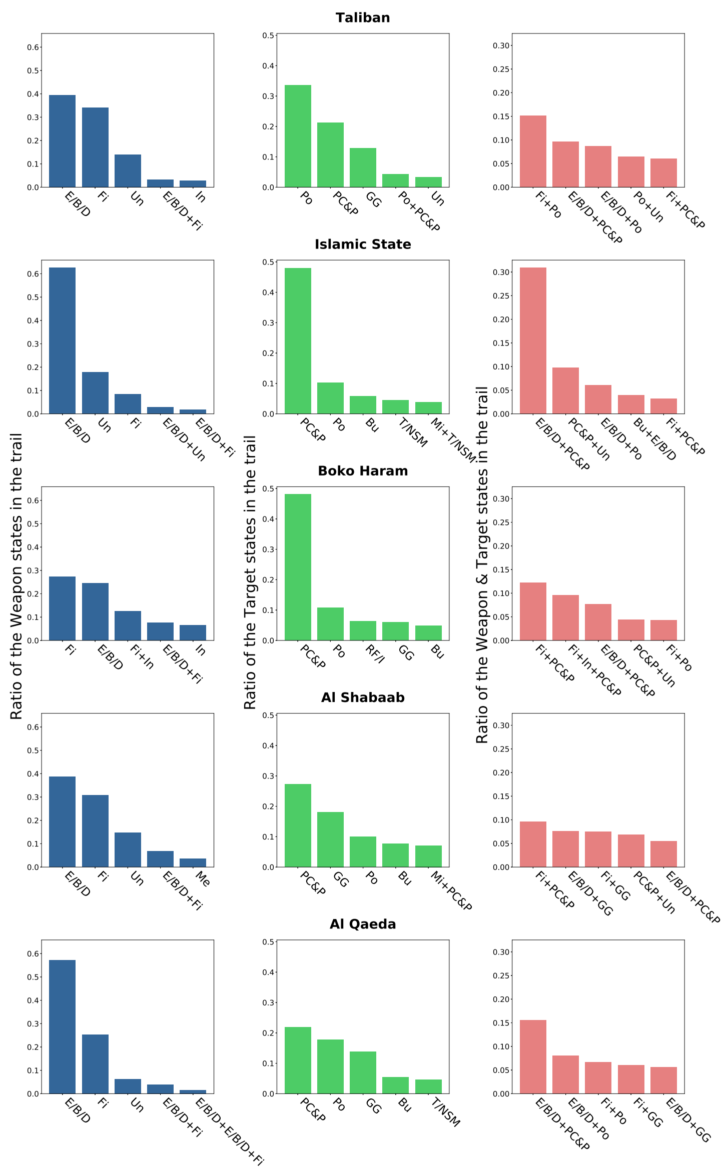

Figure 4.2: Histograms of 5 Most Common States from Each of the Jihadist Groups' Trails

Figure 4.2 shows how in terms of weapons (first column, in blue), the Islamic State has a sensibly higher preference for the use of Explosives, Bombings, Dynamite (E/B/D) in its attacks. E/B/D accounts for almost 60% of the weapons used in each event. A similar finding is displayed for Al Qaeda, while the Taliban, Boko Haram and Al Shabaab tend to diversify more and to have less stronger preferences. Notably,

Abbreviation	**Type**	**Full Name**
E/B/D	Weapon	Explosives/Bombs/Dynamite
Fi	Weapon	Firearm
Un	Weapon & Target	Unknown
In	Weapon	Incendiary
Me	Weapon	Melee
Po	Target	Police
PC&P	Target	Private Citizen & Propriety
GG	Target	Government (General)
Bu	Target	Business
Mi	Target	Military
RF/I	Target	Religious Figures/Institutions
T/NSM	Target	Terrorists/Non-State Militia

Table 4.2: List of abbreviations of targets and weapons used in Figure 4.2

Boko Haram is the only group that uses more Firearms (Fi) than Explosives in its attacks.

For what concerns targets, the Islamic State again shows a very strong preference for a particular type (i.e., Private Citizen and Property, PC&P). This also applies to Boko Haram. Both groups tend to hit PC&P in almost half of their events. Taliban, Al Shabaab, and Al Qaeda exhibit less polarized distributions. It is worth to mention how the Taliban is the only jihadist organization that prefers to target Police rather than PC&P.

Finally, in the combined scenario, distributions of the five most common combinations are more homogeneous for all groups but the Islamic State. In this last case, the group led by al-Baghdadi further demonstrates to have a very clear tendency to hit PC&P using E/B/D ($\sim 30\%$ of attacks). Although with different proportions, this combination is the most common one also for Al Qaeda. Boko Haram and al Shabaab tend instead to target PC&P using Firearms, while the Taliban use Firearms to attack Police.

4.3 Stochastic Transition Matrices

Markov chains are stochastic classes of models named after their inventor, the Russian mathematician Andrey Markov.[2] Markov chains are designed for handling sequence

[2] For a comprehensive review of Markov chains see Norris (1998) and Revuz (2008).

data, where sequences can map objects or entities as events, locations, characteristics, abstract states: they are thus extremely fit for the problems of unfolding jihadist dynamics via the analysis of sequences of terrorist attacks. These models have gained extreme popularity during the Twentieth century, due to their applicability to a variety of scientific and industrial domains. Within the realm of social sciences, for instance, they have been employed in economics (Judge and Swanson, 1962; Bickenbach and Bode, 2003; Le Gallo, 2004), finance (Kijima, 2016), sociology (Sorensen, 1978; Heckathorn, 2002), political science (Martin and Quinn, 2002; Jackman, 2004) and, also, criminology (Holland and McGarvey, 1984; Stander et al., 1989; Pettiway et al., 1994).

To introduce the concepts that will be used for modeling terrorist activity and behavior throughout this part of the chapter, a brief introduction to stochastic transition matrices and Markov chains is required. The following subsection will specifically deal with this task. Following, the results of the analysis will be showcased and commented.

4.3.1 Mathematical Framework

4.3.1.1 A Very Short Introduction to Markov Chains

In this first introduction, I will use the notation used for weapons in the previous section: the reader shall keep in mind that the same equations and formulas have to be applied for targets and targets and weapons combined. Given a sequence of attacks:

$$A_0, A_1, ..., A_k \tag{4.4}$$

and a state space $S(\mathfrak{W})$ of potential values that the various A_k can take, then the sequence is a Markov chain iff the Markov property holds.[3] The Markov property is indeed formulated in mathematical notation as follows:

$$\begin{gathered}\mathbb{P}(A_{t+1} = w | A_t = w_t, A_{t-1} = w_{t-1}, ..., A_0 = w_0) = \\ \mathbb{P}\left(A_{t+1} = w | A_t = w_t\right)\end{gathered} \tag{4.5}$$

Equation 4.5 means that in a Markov process the distribution of A_{t+1} only depends on A_t only, and does not take into account previous time units. This is often referred as

[3] In this chapter I will refer to Markov chains limited to the discrete-time case. The data at my disposal are, in fact, discrete and therefore I will not cover the mathematics behind continuous-time Markov chains.

the "memorylessness" nature of Markov processes. The property has to be valid for all t=1,2,3 and for all states $w_0, w_1, ..., w_t$. To compactly map the list of all possible states in the state space $S(\mathfrak{W})$, a transition matrix $\mathbf{P}(\mathfrak{W}) = p_{i,j}$ is created. $\mathbf{P}(\mathfrak{W})$ is a square matrix of dimension $|S(\mathfrak{W})| \times |S(\mathfrak{W})|$ and each row should sum to 1:

$$\begin{gathered}\sum_{j=1}^{|S(\mathfrak{W})|} p_{ij} = \sum_{j=1}^{|S(\mathfrak{W})|} \mathbb{P}(A_{t+1} = j | A_t = i) = \\ \sum_{j=1}^{|S(\mathfrak{W})|} \mathbb{P}_{\{A_t = i\}}(A_{t+1} = j) = 1\end{gathered} \tag{4.6}$$

To exemplify, a hypothetical transition matrix $\mathbf{P}(\mathfrak{W})_{ij}$ where the state space $|S(\mathfrak{W})|=$ $\{w_1, w_2, w_3, w_4, w_5\} = 5$ is given. Visually, the matrix takes the form:

$$\mathbf{P}(\mathfrak{W})_{ij} = \begin{pmatrix} . & p & q & . & . \\ p & q & . & . & . \\ p & r & . & . & s \\ . & . & p & s & r \\ . & . & . & p & q \end{pmatrix} \tag{4.7}$$

where . denotes zero-entries for simplicity, p is the probability associated to the entry ij, q is a probability equal to $(1-p)$, r is a probability equal to $(1-p-s)$ and, trivially, s is a probability equal to $(1-p-r)$. The case above only regards single-step transitions between i and j, however, Markov chains allow also to formalize t-step transitions. If, for instance, there is a need to quantitatively describe the probability of going from state i to state j in two steps, using Partition Theorem of matrices, the 2-step transition matrix $\mathbf{P}^2(\mathfrak{W})_{ij}$ is obtained via:

$$\begin{gathered}\mathbb{P}(A_2 = j | A_0 = i) = \mathbb{P}(A_2 = j) = \\ \sum_{k=1}^{|S(\mathfrak{W})|} \mathbb{P}_i(A_2 = j | A_1 = k)\mathbb{P}(A_1 = k) = \\ \sum_{k=1}^{|S(\mathfrak{W})|} \mathbb{P}_i(A_2 = j | A_1 = k, A_0 = i)\mathbb{P}(A_1 = k | A_0 = i) = \\ \sum_{k=1}^{|S(\mathfrak{W})|} \mathbb{P}_i(A_2 = j | A_1 = k)\mathbb{P}(A_1 = k | A_0 = i) = \\ \sum_{k=1}^{|S(\mathfrak{W})|} p_{kj}p_{ik} = \sum_{k=1}^{|S(\mathfrak{W})|} p_{ik}p_{kj} = \mathbf{P}^2(\mathfrak{W})_{ij}\end{gathered} \tag{4.8}$$

Following this procedure, the general case to obtain a t-th step Markov chain and create the related $\mathbf{P}^t(\mathfrak{W})_{ij}$ transition matrix is:

$$\mathbb{P}(A_t = j | A_0 = i) = \mathbb{P}(A_{n+t} = j | A_n = i) = \mathbf{P}^t(\mathfrak{W})_{ij} \ \ \forall(t) \tag{4.9}$$

A sample transition network of weapons derived from a toy stochastic transition matrix is depicted in Figure 4.3. It is worth to note how for any state, the sum of the probability of the out-links is equal to 1. This brief primer on the basics of Markov chains served to introduce further machinery behind the analyses.

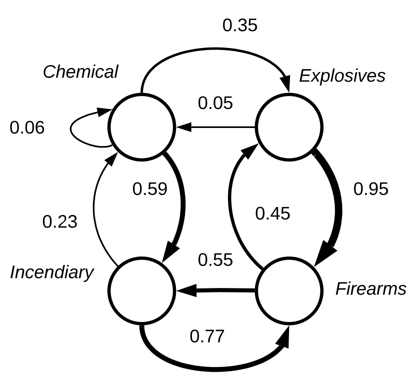

Figure 4.3: Sample Transition Network with Transition Probabilities - Weapons

4.3.1.2 Super-States of N-Dimension

So far, only Markov chains mapping the t-step transitions between single states have been addressed. However, for the purposes of the analysis of terrorism dynamics, it is relevant to understand whether, within the data, typical cycles are present and, if so, what is the knowledge associated with them. Going back to the case of the state space seen in Equation 4.7, I now propose the matrix $\mathbf{A}(\mathfrak{W})_{ij}$ in which the entries are not anymore probabilities but, instead, absolute values mapping the number of times that there has been a transition from a state i to a state j:

$$\mathbf{A}(\mathfrak{W})_{ij} = \begin{pmatrix} . & w_{1,2} & w_{1,3} & . & . \\ w_{2,1} & w_{2,2} & . & . & . \\ w_{3,1} & w_{3,2} & . & . & w_{3,5} \\ . & . & w_{4,3} & w_{4,4} & w_{4,5} \\ . & . & . & w_{5,4} & w_{5,5} \end{pmatrix} \tag{4.10}$$

where, again, . denotes zeros. At this point, to detect recurring patterns occurring in attack sequences is it useful to create a new augmented state space $S(\mathfrak{W})_2$ of dimension 2. Given that $w_{ij} \neq 0$, then I define a 2-dimension super-state as:

$$\Theta_{i \to j} =: (w_i \to w_j) \tag{4.11}$$

in which w_i is called the "1d-sender" 1d_s and w_j is called the "1d-receiver" 1d_r. The new super state now incorporates the single-step transition between the previous original state w_i to w_j. This gives birth to a new squared adjacency matrix $\mathbf{A}_2(\mathfrak{W})_{ij}$ of order two where the state space is equal to:

$$|S(\mathfrak{W})_2| = \sum w_{ij} \ \forall \, w_{ij} \neq 0 \tag{4.12}$$

After the creation of the adjacency matrix, the related stochastic transition matrix $\mathbf{P}_2(\mathfrak{W})_{ij}$ is obtained simply by row-normalization so that the sum of the entries of each row-vector is equal to 1. To help the reader who might not be comfortable with the mathematics shown above, it is useful to think about the following example. Given an existing link in the adjacency matrix $\mathbf{A}(\mathfrak{W})_{ij}$ between "Firearms" and "Explosives", in the augmented matrix $\mathbf{A}_2(\mathfrak{W})_{ij}$, thus there will exist a new state:

$$\Theta_{\text{Firearms} \to \text{Explosives}} \tag{4.13}$$

In the present example, "Firearms" is the sender and "Explosives" is the receiver. I have thus now obtained a new information space with super-states of dimension 2: to further investigate cycles, I have iterated the operation above up to super-states of dimension 5. The procedure is not exactly identical as we increase the dimensionality, and needs a bit of additional explanation. Imagine a hypothetical adjacency matrix of order 2 $\mathbf{A}_2(\mathfrak{W})_{ij}$:

$$\mathbf{A}_2(\mathfrak{W})_{ij} =$$
$$\begin{array}{c} \\ \Theta_{i\to j} \\ \Theta_{i\to k} \\ \Theta_{j\to l} \\ \Theta_{j\to j} \\ \Theta_{l\to i} \end{array} \begin{array}{c} \begin{array}{ccccc} \Theta_{i\to j} & \Theta_{i\to k} & \Theta_{j\to l} & \Theta_{j\to j} & \Theta_{l\to i} \end{array} \\ \begin{pmatrix} \cdot & \cdot & w_{(\Theta i\to j,\Theta j\to l)} & w_{(\Theta i\to j,\Theta j\to j)} & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & w_{(\Theta j\to l,\Theta l\to i)} \\ \cdot & \cdot & w_{(\Theta j\to j,\Theta j\to l)} & w_{(\Theta j\to j,\Theta j\to j)} & \cdot \\ w_{(\Theta l\to i,\Theta i\to j)} & \cdot & \cdot & \cdot & \cdot \end{pmatrix} \end{array} \tag{4.14}$$

Then, it is worth to specify that the new matrix is populated based on some conditions. So, for any given pair of super-nodes $\Theta i \to j$ and $\Theta j \to k$, mapping the super-states of dimension two:

$$\begin{gathered} w_{(\Theta i\to j,\Theta j\to k)} \neq 0 \leftrightarrow 1\mathrm{d}_r(\Theta_{i\to j}) = 1\mathrm{d}_s(\Theta j \to k) \\ \wedge \exists\, w\,[1\mathrm{d}_s(\Theta i \to j) \to [1\mathrm{d}_r(\Theta i \to j) = 1\mathrm{d}_s(\Theta j \to k)] \to 1\mathrm{d}_r(\Theta j \to k)] \end{gathered} \tag{4.15}$$

The equation means that a transition edge w between $\Theta i \to j$ and $\Theta j \to k$ will exist iff (1) the 1d receiver of $\Theta i \to j$ is equal to the 1d sender of $\Theta j \to k$ and (2) there is at least one 3-d chain that connects the 1d sender of $\Theta i \to j$, the two equal states $1\mathrm{d}_r\Theta i \to j$ and $1\mathrm{d}_s\Theta j \to k$ that are collapsed in a single entity, and the 1d receiver of $\Theta j \to k$. This process leads to the creation of a new adjacency matrix containing super-states of dimension 3, $\mathbf{A}_3(\mathfrak{W})_{ij}$, from which it is easily derivable a stochastic matrix $\mathbf{P}_3(\mathfrak{W})_{ij}$. The state space of the new matrix is given by:

$$|S(\mathfrak{W})_3| = \sum w_{(\Theta i\to j,\Theta j\to k)}\ \forall w \neq 0 \tag{4.16}$$

This procedure allows us to create 3-dimensional superstates in the form $\Theta_{i\to j\to k}$ where information regarding the sequence across states i, j, k is incorporated. In this superstate, i is the 1d sender, and j and k are the 2d receivers ($2\mathrm{d}_r$). At the same time, we can consider another 3-dimensional superstate $\Theta_{j\to k\to m}$. If we hypothesize the existence of a link between the two, intending to further obtain a 4-dimensional superstate, we have to check again for the basic condition expressed in Equation 4.15. In the second sequence, j and k are called the 2d senders ($2\mathrm{d}_s$), and m is the

1d receiver. This considered, as the first fundamental condition the transition may exist if:

$$2\mathrm{d}_r(\Theta i \to j \to k) = 2\mathrm{d}_s(\Theta j \to k \to m) \tag{4.17}$$

Furthermore, if this is verified, there must exist at least one transition such that:

$$\begin{gathered} w\,[1\mathrm{d}_s(\Theta i \to j \to k) \to \\ [2\mathrm{d}_r(\Theta i \to j \to k) = 2\mathrm{d}_s(\Theta j \to k \to m)] \to \\ 1\mathrm{d}_r(\Theta j \to k \to m) \end{gathered} \tag{4.18}$$

This condition, if verified, lead to the creation of a new 4-dimensional superstate that, relying on the example, takes the form:

$$\Theta_{i \to j \to k \to m} \tag{4.19}$$

This new super-state, in a real-world case, could be seen as a chain of four weapons taken from our original $S(\mathfrak{W})$, such that it would potentially look like the following:

$$\Theta_{\text{Firearms}\to\text{Explosives}\to\text{Chemical}\to\text{Incendiary}} \tag{4.20}$$

To obtain further augmented super-states, it is sufficient to slightly modify the condition that I have described to create the matrix of order 2, in order to generalize it to other multidimensional cases. I define $\Theta^{'}_{N-1}$ and $\Theta^{''}_{N-1}$ as two arbitrary super states of dimension $|N-1|$. Imagining a N-dimension matrix where N is the number of states that we want to include in the creation of the super-state Θ_N, which is to say that N is the length of the sequence of states that we consider, then the entries of the matrix $\mathbf{A}_N(\mathfrak{W})_{ij}$ can be different from 0 if:

$$|N-1|d_s(\Theta^{'}_{N-1}) = |N-1|d_s(\Theta^{'}_{N-1}) \tag{4.21}$$

and, additionally, there exist at least one transition such that:

$$1\mathrm{d}_s(\Theta^{'}_{N-1}) \to \Big[|N-1|d_s(\Theta^{'}_{N-1}) = |N-1|d_s(\Theta^{'}_{N-1})\Big] \to 1\mathrm{d}_r(\Theta^{''}_{N-1}) \tag{4.22}$$

A visual depiction of the process that lead to the creation of a 4d-super state starting from two separate 2d super-states is provided in Figure 4.4.

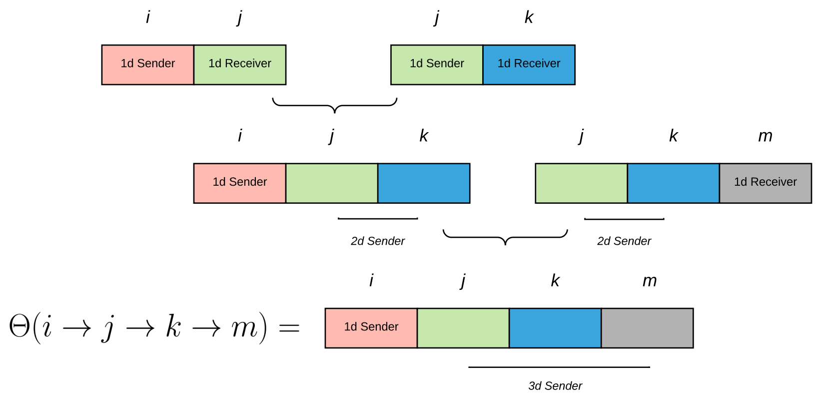

Figure 4.4: Visual Depiction of the Creation of 4-dimensional Super-states

4.3.2 Super-States and the Heterogeneity of Jihadist Behavior

After the introduction of the mathematical concepts that led to the derivation of the super-states stochastic transition matrices for each group with regards to weapons, targets and targets and weapons combined, it is necessary to analyze and interpret what these new sets of information tells us about how jihadist behave. As anticipated throughout the previous subsection, the intent is to capture inherent recurring trends, e.g. cycles, that can better picture the behavior of the groups in the sample. Hoffman (1993) showed that, besides differences in the lethality of attacks between the 1970s and 1980s, the majority of terrorist organizations remain stable in their operational choices, drawing from the same consistent repertoire of weapons and tactics. Merari (1999), comparing terrorist violence with conventional conflicts and wars, noted that that terrorism has not changed in the course of a century in terms of weaponry and modes of operation. This view is shared also by Dolnik (2007).

While many scholars agree on the recurrent and persistent regularity of terrorist behavior, Gill et al. (2013) addressed the problem focusing on the concepts of “creativity” and “innovation”, drawing upon the existing literature on industrial psychology. They have qualitatively focused on specific case studies, demonstrating how the conversion of a new creative idea into an innovation is a process that requires time. Authors have proposed a new framework that, according to the authors, could better indicate to policymakers the specific innovation-drivers sub-units of a terrorist organization.

Data can be treacherous when trying to analyze the extent to which terrorist groups show homo or heterogeneity of operational behavior in their attack patterns. Aggregate descriptive statistics may reveal very strong preferences towards certain types of weapons or targets, while obscuring other more complex patterns that can provide much more useful insights, also from a policy point of view. With this regard, complex networks can be of help. Imagine, for instance, the case of two attacks, A_1 and A_2, where the employed weapons are (Explosives, Explosives, Explosives) in one the former, and (Explosives) in the latter case. A first superficial intuition may lead to consider the two operational characteristics as identical and aggregate statistics would suggest that Explosives account for the 100% of the weapons used in the attacks. However, while in both attacks the type of weapons is the same, the first one shows a much higher logistic capacity, and may be seen as "more complex".

Looking for heterogeneity solely in relation to drastic changes in terms of operational choices (e.g., shift from Explosives to Chemical weapons) is a fallacy that does not take into consideration other underlying logistic mechanisms. These mechanisms might be the expression of the power, strength or more general goals of a group. Furthermore, terrorists might not really need to change their operational choices if certain stylized types of actions are part of long-range strategies or demonstrate to be effective for the organization. The same type of reasoning can be applied to targets, although in the literature there is a more pronounced consensus towards the idea of higher variability of target types over time (Brandt and Sandler, 2010; Santifort et al., 2013).

In light of this debate, what can be drawn from the construction of super-states of weapons, targets, and both combined? What are the insights on the complexity of jihadist behavior, when approaching the question from a complex network perspective, taking into account the sequential nature of attacks?

Figures 4.5, 4.6, 4.7 help in picturing the evolution of the transition networks across the different N-dimensions of super-states.[4] The reader shall keep in mind the figures previously showed in Table 3.2: the number of employed weapons for each group oscillates from 7 to 9, while the number of targets falls in the range 21-24. What is presented below is the product of looking at the combinations of these small sets of features over-time.

Figure 4.5 first displays the trend in the number of states of the state space of weapons when increasing the dimension of the super-states. The trend is positive for all the groups, with some differences. While all groups start with a similar number of

[4] A Super-state of Dimension 1 is simply a single state related to a single event.

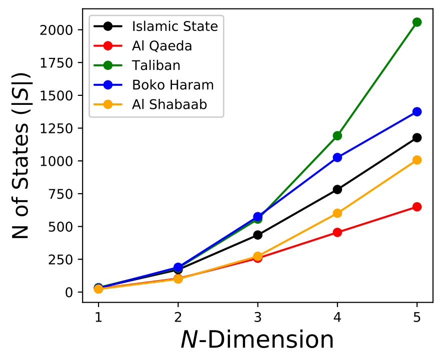

Figure 4.5: N of States in the State Space of Weapons per Group across N-Dimension Stochastic Matrices

states in the 1d-case (ranging from 23 to 34), the differences become more prominent as the dimension is increased. Taliban appears to be the group that, in its history, has experimented overall a higher number of higher-dimensional super-states of weapons. The differences emerge quite clearly from the 4d and 5d scenarios. Conversely, Al Qaeda shows the lowest attitude towards heterogeneity in weaponry. The picture in

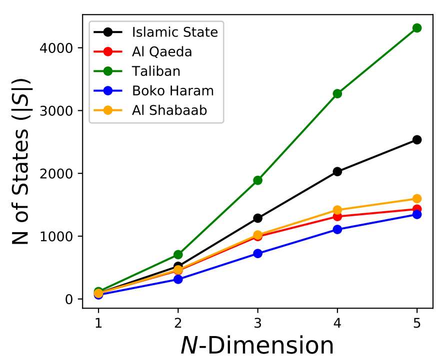

Figure 4.6: N of States in the State Space of Targets per Group across N-Dimension Stochastic Matrices

Figure 4.6 slightly changes only when focusing on individual trends, with Boko Haram, for instance, being the least prone to expand his repertoire of cycle-combinations. Nonetheless, overall, the trend is still positive for all groups, with various levels of steepness in the passage from 4d to 5d super-states, and with the Taliban again showing by far the widest state space.

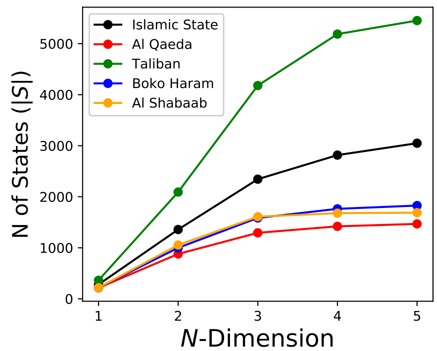

Figure 4.7: N of States in the State Space of Targets and Weapons per Group across N-Dimension Stochastic Matrices

Finally, the most comprehensive scenario, namely the combined case of targets and weapons, is visually depicted in Figure 4.7. The patterns revealed by the previous Figures here assume more defined shapes. Each group again shows a positive trend, while the curves tend to flatten for three groups (Al Qaeda, Boko Haram, Al Shabaab) after the 3d dimension. The increase is less evident from the fourth to the fifth step also for the Islamic State and the Taliban. However, the absolute differences between the two and the rest of the sample are evident, with the Taliban dominating in terms of heterogeneity of combinations. It is worth to note that the flattening pattern means that, increasing the dimension of the super-states, the matrices will start to become less populated and the conditions for the creation of super-states listed in Equations 4.21 and 4.22 become more and more challenging to be respected. In other words, this means that the matrix will slowly become smaller and smaller: indeed, when $|D|= |A|$, the matrix becomes of trivial dimension 1×1.

For what concerns the heterogeneity of terrorist behavior with regard to operational choices, it is worth to note that the numbers of the existing super-states are in the order of thousands. Starting from the very few options of weapons and targets seen in the descriptive statistics of Table 3.2, the jihadist groups demonstrate a complex operational repertoire. How they select combinations of targets and weapons over time gives much more information on their actual nature, compared mere aggregate statistical distributions. While innovation may pertain to other aspects of the operational context (Crenshaw, 2010), heterogeneity is certainly a component of terrorism complexity that captures the not directly observable inherent nature of jihadism.

Another indication can be drawn from the previous plots: the higher the number of attacks, the higher the heterogeneity in operational choices. The reader shall keep this in mind as the difference between the absolute and the relative standpoint for comparing sequences will be better entailed in subchapter 4.4.

However, considering only the dimension of the matrix is limiting. In fact, as the dimension of the super-states changes, the topology of the resulting network evolves too. Figures 4.8, 4.9, 4.10 illustrates this process in terms of density. Increasing the dimension of the super-states leads to a dramatic decrease in network density.

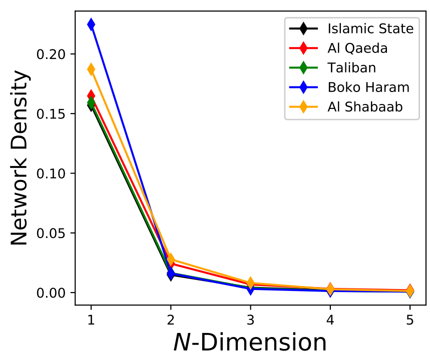

Figure 4.8: Network Density Evolution of the Stochastic Transition Matrix of Weapons Across Groups

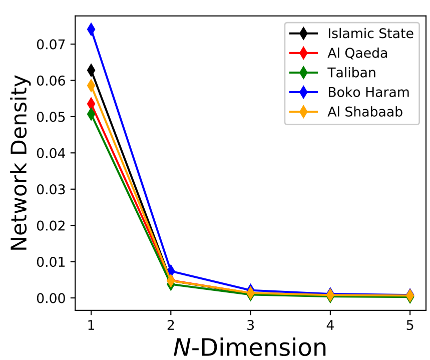

Figure 4.9: Network Density Evolution of the Stochastic Transition Matrix of Targets Across Groups

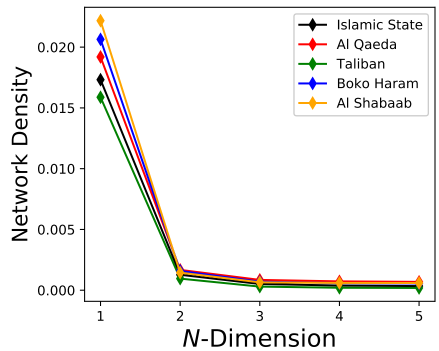

Figure 4.10: Network Density Evolution of the Stochastic Transition Matrix of Targets and Weapons Across Groups

The differences that were displayed and commented in Figures 4.5, 4.6 and 4.7 are attenuated when comparing the density trends. Network density is a global metric that captures the extent to which the nodes (in this case, the super-states) in a graph are connected. The value lies in the range 0-1, with 0 indicating a completely disconnected network and 1 a completely connected one.

Overall, the clearest pattern regards the change between 1d and 2d super-states. The density of the network decreases steeply and remains almost constant from the 2d to the 5d case, converging to values very close to zero for all the groups in the sample and in relation to all the three different dimensions: weapons, targets, and targets and weapons combined. What does this mean in practical terms? For better understanding terrorism, this finding has to be coupled with the increasing trend in matrix dimension when super-states are augmented. The number of super-states becomes considerably high when the dimension is increased, but the density of the network almost collapses right after the creation of the 2d super-states. This means that the network becomes extremely sparse and for a given cycle it will become easier to predict the next, as the number of transitions per super-state will be lower.

Figures 4.11 and 4.15 shows the 1d and 5d Boko Haram super-state transition networks from a hierarchical layout standpoint for targets.[5] The top node is the most central node of the network (in binary terms, i.e., the node with the highest number of transitions)[6]: it should be noted how the single-step transitions are fewer for the

[5] Figures A.1, A.2, A.3, A.4 and A.5 in Appendix A provide a further visual example of evolution from 1- to 5-d super-states with a circular layout.

[6] For the 1d case the top node is "Firearms", for the 5d case the top node is the super-state

5d case, meaning that starting from the top node, a very low number of other super-states are directly accessible, while the number is much higher in the 1d case. This means that increasing the dimension of the super-states provides a more complex framework in terms of the absolute number of super-states, but in the meanwhile creates a less blurry picture.

What emerges from this analysis can enrich the debate on the stability of terrorist behaviors. Notwithstanding the claims regarding the tendency of terrorists to repeat the behaviors without changing their operational characteristics over time and attacks, the framework of Markov chains and super-states provides insight on inherent patterns that aggregate statistics are not able to capture. The original data on the number of weapons employed and attacked targets described low heterogeneity. Nonetheless, processing the data in order to take into account combinations of weapons and targets and, consequently, super-states, highlighted a very different scenario. For research and policy purposes, it should be relevant to investigate how regularities occur over time and how regularities emerge, are formed and interact with more anomalous behaviors. In light of this, Markov chains and super-states can be useful in detecting rich dynamics and predict potential scenarios.

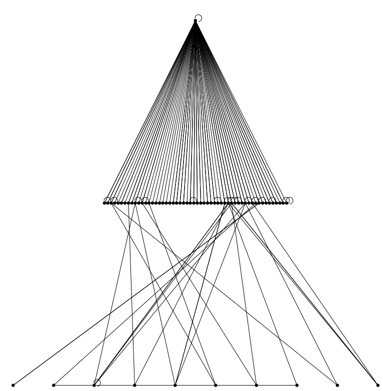

Figure 4.11: Boko Haram - Transition Network Hierarchical Layout of Targets (1-dimensional Super-States Case)

"Firearms → Firearms → Firearms → Firearms → Firearms".

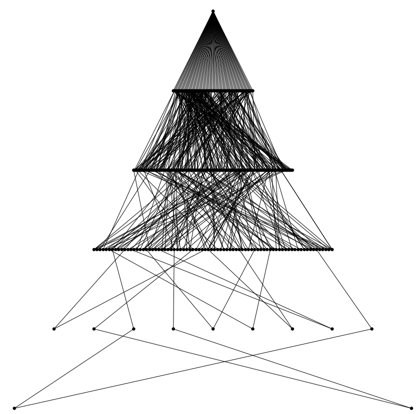

Figure 4.12: Boko Haram - Transition Network Hierarchical Layout of Targets (2-dimensional Super-States Case)

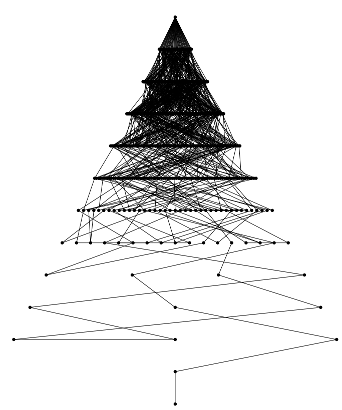

Figure 4.13: Boko Haram - Transition Network Hierarchical Layout of Targets (3-dimensional Super-States Case)

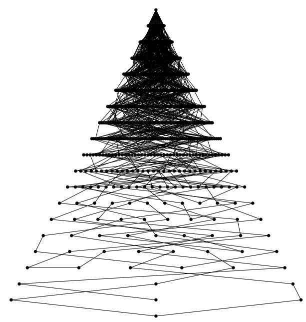

Figure 4.14: Boko Haram - Transition Network Hierarchical Layout of Targets (4-dimensional Super-States Case)

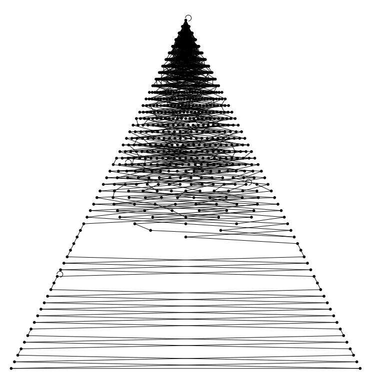

Figure 4.15: Boko Haram - Transition Network Hierarchical Layout of Targets (5-dimensional Super-States Case)

In general, while the groups in the sample tend to rely consistently on the same weapons and to constantly attack the same type of targets, their attacks demonstrate to be particularly heterogeneous and complex when combinations and subsequences are considered. Patterns exist, but they are many and, generally, relying on 1-dimensional sources of information dramatically increases the state space and decreases the possibility to understand what will come next. The challenge, for research and intelligence purposes, would be to understand the right pay-off in terms of super-states dimensionality. The 1d case is, as mentioned above, too fragmented to reliably work as it involves too many transition possibilities, while 5 (or even higher)-dimensional cases risks transforming the heterogeneity of terrorism in a trivial set of sub-sequences that are only self-connected through self-loops. Future work will better consider this trade-off. Furthermore, it will be interesting to replicate the analysis using more detailed sub-categories of weapons and targets, similarly to what Jackson and Frelinger (2008) have done, as the choice would highly likely increase the complexity of the three scenarios. While keeping this in mind, I will next introduce another alternative method to capture behavioral patterns in state transitions in terms of pairwise similarity between groups.

4.4 Normalized Transition Similarity

4.4.1 Rationale and Formalization

From the ordinary frame of Markov chains and transition networks, we can derive the concept of "network trail". Network trails are two-mode directed networks in which the behavior of source nodes is temporally ordered with respect to target nodes. Questions such as "How many times these two entities have moved in the same direction?", "How many times these two entities were in the same place together?", "Is there a mimicry dynamic between the two entities?" can be answered using trails. Trails have been used in health care, biology and scientific co-authorship networks domains (Vittori et al., 2006; Lee et al., 2010; Merrill et al., 2015) (Figure 4.16).

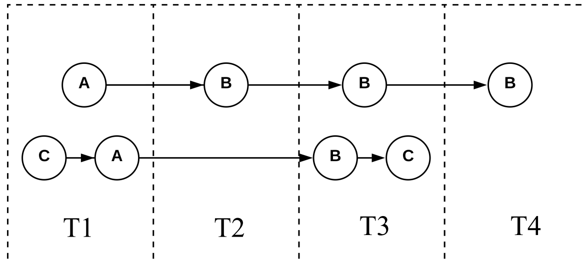

Figure 4.16: Two Sample Trails of Different Length

If we consider the formal definition of an event provided in Equation 4.1, the three types of network trails ψ I will consider in the analysis are the following:

- $\psi_{g_i}(d, w)$: a time-ordered trail of weapons for group g_i
- $\psi_{g_i}(d, t)$: a time-ordered trail of targets for group g_i
- $\psi_{g_i}(d, t, w)$: a time-ordered trail of targets and weapons for group g_i

Having set up the information framework of the work, we have developed the Normalized Transition Similarity (NTS) coefficient. To recall, a transition is a single-step change of state in the ordered sequence of attacks. For example, in the case of time-ordered sequence of targets $\psi_{g_i}(d, t)$, it is a single-step change in targets group g_i selected in two sequential attacks. To start to familiarize with the concept of transition similarity, we introduce a simple statistic that only takes into account the absolute frequency of shared transitions between two entities g_1 and g_2 (i.e., groups, in this specific experiment). This figure, for instance, is included in the dynamic network analysis software ORA (Carley, 2014), and it is calculated as:

$$\mathrm{Tr}_{common} = \sum_{i,j} \min\left[\Phi_{g_1}(s_i \to s_j), \Phi_{g_2}(s_i \to s_j)\right] \tag{4.23}$$

where s_i and s_j are two distinct generic states and Φ_{g_k} denotes the number of transitions between states s_i and s_j in the trail of group g_k. The Equation 4.23 gives us the number of common transitions between two trails expressed as the minimum sum of all single link paths (between hypothetical s_i and s_j) shared by two groups. This type of descriptive statistic can be used to evaluate the absolute frequency of shared transitions. However, it can be highly biased when analyzing trails of significantly different dimensions, as in this case (Table 4.3).

Group	Trail Length
Taliban	5,628
Islamic State	3,561
Al Qaeda	1,501
Al Shabaab	1,694
Boko Haram	1,900

Table 4.3: Trail Length per Jihadist Group

For instance, when calculating transitions for five groups it is expected that the groups with very long trails will share more common transitions in absolute terms. However, this does not mean that the highly active pair of groups share more than the other. For this specific reason, I propose a new coefficient of transition which normalizes the absolute frequencies and allows us to make pairwise comparisons to evaluate the extent to which each group is similar to another in terms of trails dynamics. This coefficient is named *Normalized Transition Similarity* and it is calculated as:

$$\mathrm{NTS}\left(g_1, g_2\right) = \frac{\sum_{i,j} \min\left[\Phi_{g_1}\left(s_i \to s_j\right), \Phi_{g_2}\left(s_i \to s_j\right)\right]}{\max\left(|A_{g_1}|, |A_{g_2}|\right) - 1} \tag{4.24}$$

NTS normalizes the number of times that each entity pair travels the same single link by the number of links of the longest trail in each pair (which is equal to the total number of states minus 1). Potential outcomes are synthesized in Equation 4.25.

$$\mathrm{NTS}\left(g_m, g_n\right): \\ \begin{cases} 0 \quad \text{if} & \sum_{i,j} \min\left[\Phi_{g_1}\left(s_i \to s_j\right), \Phi_{g_2}\left(s_i \to s_j\right)\right] = 0 \\ 1 \quad \text{if} & \sum_{i,j} \min\left[\Phi_{g_1}\left(s_i \to s_j\right), \Phi_{g_2}\left(s_i \to s_j\right)\right] = \max\left(|A_{g_1}|, |A_{g_2}|\right) - 1 \\ x & \text{otherwise} \end{cases} \tag{4.25}$$

where x can be a continuous value in the range $0 < x < 1$. Once this coefficient is calculated, to actually rescale values to take into account the relative differences between outcomes of the considered pairs, a further normalization can be performed. So, for groups g_1 and g_2 the final value would be calculated as:

$$\mathrm{NTS}\left(g_1, g_2\right)_{scaled} = \frac{\mathrm{NTS}\left(g_1, g_2\right)}{\max_{m,n \in G} \mathrm{NTS}\left(g_m, g_n\right)} \tag{4.26}$$

with G representing the set of groups in analysis (five in our case). To further explain how NTS works Figure 4.17 provides a visual representation of Equations 4.24 and

4.26.
It is worth mentioning that NTS, while allowing for intra-sample pairwise comparison, cannot be used to compare pairs belonging to different samples. In the case of two sets A and B in which an entity (e.g. a terrorist group) $g \in A \wedge B$, $NTS(g, x_A)$ cannot be compared with $NTS(g, x_B)$, where x_A and x_B are two given entities belonging to sets A and B.

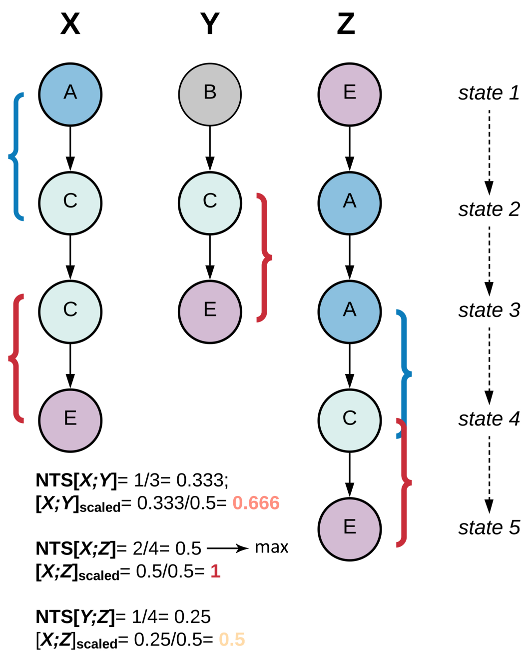

Figure 4.17: Depiction of NTS Across Three Short Sequences

4.4.2 Results

This section will showcase and explain the findings of the analysis in the following subsections: one for each trail type, with a conclusive subsection for summing up the main results.

4.4.2.1 Trails of Weapons $\psi_{g_i}(d, w)$

This first family of trails seeks to understand and investigate potential patterns in how groups change their weapons for plotting terrorist attacks. Weapons can be extremely different, and each type of weapon can denote a distinct and meaningful aspect of the consequences of an event and the power, strength, and resources of a group. Data shows that the number of unique weapon combinations is similar for all groups (ranging from a minimum of 23 to a maximum of 34 combinations). When focusing on unique transitions, the picture slightly changes. In fact, Boko Haram shows nearly double unique transitions compared to Al Shabaab (188 vs 99), demonstrating how the former group seems less predictable and stable in its operational choices. Finally, the third column further highlights evident differences between groups: the top identical subsequence of weapons for the Islamic State is significantly longer than the longest subsequences associated with all the other groups (Table 4.4).

	N **Unique Weapons Combinations**	N **Unique Transitions**	**Longest Id. Subsequence**
Taliban	34	180	21
Islamic State	33	157	110
Boko Haram	29	188	15
Al Shabaab	23	99	12
Al Qaeda	25	100	30

Table 4.4: Descriptive statistics of Transition Networks of Weapons Per Terrorist Group.

Table 4.5 presents the detailed outcomes of the NTS. Al Qaeda and Al Shabaab appear to be the most similar groups according to NTS, while their absolute number of shared transitions was not particularly relevant when looking at the mere sum. Al Shabaab and Boko Haram are the second-most similar pair, while in transition count they were the third less similar pair. Interestingly, Al Shabaab demonstrates a high degree of trail similarity with two different groups. In general, the differences between rankings highlight how NTS calculation sensibly changes the initial results. In terms of ranking (which is a measure that should be handled carefully because we do not control for relative quantitative differences), only one pair remained in the same position. Another finding is that, although they have the longest trails, therefore increasing the relative probability of sharing transitions, the Taliban and the Islamic State are only the fourth most similar pair (0.68).

4.4.2.2 Trails of Targets $\psi_{g_i}(d,t)$

The second considered trail network regards selected targets. The Taliban, also due to their longer history and sequence of events, shows the highest number of unique

Pair	Shared Trans (Count)	Count Rank	NTS	scaled NTS	NTS Rank	Rank Diff
Al Qaeda & Al Shabaab	1,150	6	0.68	1.00	1	5
Al Shabaab & Boko Haram	1,022	8	0.54	0.79	2	6
Al Qaeda & Boko Haram	891	9	0.47	0.69	3	6
Taliban & IS	2,585	1	0.46	0.68	4	-3
Al Shabaab & IS	1,398	3	0.38	0.56	5	-2
Taliban & Al Shabaab	1,736	2	0.31	0.45	6	-4
Al Qaeda & IS	1,058	7	0.29	0.43	7	0
Taliban & Boko Haram	1,383	4	0.25	0.36	8	-4
Boko Haram & IS	836	10	0.23	0.34	9	1
Taliban & Al Qaeda	1,212	5	0.22	0.32	10	-5

Table 4.5: NTS Results for Weapon Trails

targets and transitions. Specifically, in terms of the unique transition case, their total is more than three times the Al Qaeda's one, which across all groups seems to be more homogeneous, with the shortest event history overall (Table 4.6).

	N **Unique Target Combinations**	N **Unique Transitions**	**Longest Id. Subsequence**
Taliban	118	988	26
Islamic State	91	752	99
Boko Haram	65	427	29
Al Shabaab	89	638	11
Al Qaeda	92	301	10

Table 4.6: Descriptive Statistics of Transition Networks of Targets Per Terrorist Group.

Also in the target scenario, Al Qaeda and Al Shabaab prove to be the most similar groups (Table 4.7). Stability holds also for the less similar pair, namely the Taliban and Al Qaeda. Conversely, while Boko Haram and the Islamic State differed significantly in the previous analyses on weapons, here they are ranked high (fourth position). This denotes how, actually, a certain degree of similarity in a specific behavioral dimension does not imply automatically that groups are similar overall. This might suggest how, although employing and applying different methods and resources, both groups seem to have similar strategies with respect to targets. Similarly, while the Taliban and Al Shabaab were not particularly close in terms of single link transitions of weapons, they show high similarity in the choice of new targets.

Pair	**Shared Trans (Count)**	**Count Rank**	**NTS**	**scaled NTS**	**NTS Rank**	**Rank Diff**
Al Qaeda & Al Shabaab	1,011	8	0.60	1.00	1	7
Al Shabaab & Boko Haram	999	9	0.53	0.88	2	7
Taliban & IS	2,518	1	0.45	0.75	3	-2
Boko Haram & IS	1,555	4	0.43	0.71	4	0
Al Qaeda & Boko Haram	763	10	0.40	0.67	5	5
Al Shabaab & IS	1,360	6	0.37	0.62	6	0
Taliban & Al Shabaab	1,784	2	0.32	0.53	7	-5
Al Qaeda & IS	1,070	7	0.29	0.49	8	-1
Taliban & Boko Haram	1,601	3	0.28	0.48	9	-6
Taliban & Al Qaeda	1,363	5	0.24	0.41	10	-5

Table 4.7: NTS Results for Target Trails

4.4.2.3 Trails of Targets and Weapons $\psi_{g_i}(d, t, w)$

The final type of trail analysis integrates both the previously considered dimensions of terror events: weapons and targets. It relies on a much vaster quantity of possible combinations and its nature makes it potentially more informative than the previous two. In terms of basic information, while all sequences of identical combinations diminished in length in this case, the Islamic State is the only one that actually shows a very long sequence (identical to the target one). Overall, conversely, groups demonstrated their tendency to change combinations very frequently. Al Shabaab, for instance, has the longest sequence of only four consecutive identical combinations. The Taliban (followed by the Islamic State) is again the group with the largest behavioral repertoire, both in terms of unique targets and weapons states and unique transitions (Table 4.8).

In the combined setting, the Taliban and the Islamic State are found to be the most similar groups (Table 4.9). Al Qaeda and Al Shabaab, which were ranked first in the previous cases, are now ranked second (yet performing a result almost identical to the highest one). Al Shabaab appears to be very similar also to Boko Haram (third highest NTS value), while the Nigerian group seems to be significantly dissimilar not only to the Taliban but also to the Islamic State. It is interesting to note that the two pairs that yielded the second and third highest results in the NTS computation had a very low shared transition count. In terms of extreme dissimilarity, the Taliban

	N Unique Trgt and Wpn Combinations	**N Unique Transitions**	**Longest Id. Subsequence**
Taliban	363	2,102	20
Islamic State	280	1,376	99
Boko Haram	220	1,034	14
Al Shabaab	218	1,048	4
Al Qaeda	214	896	10

Table 4.8: Descriptive Statistics of Transition Networks of Weapons and Targets Per Terrorist Group.

and Al Qaeda are detected as the most dissimilar pair also when weapons and targets are considered together.

Pair	**Shared Trans (Count)**	**Count Rank**	**NTS**	**scaled NTS**	**NTS Rank**	**Rank Diff**
Taliban & IS	2,037	1	0.36	1.00	1	0
Al Qaeda & Al Shabaab	604	9	0.36	0.99	2	7
Al Shabaab & Boko Haram	624	8	0.33	0.91	3	5
Taliban & Al Shabaab	1,385	2	0.25	0.68	4	-2
Al Shabaab & IS	813	5	0.22	0.62	5	0
Al Qaeda & Boko Haram	405	10	0.21	0.59	6	4
Taliban & Boko Haram	1,066	3	0.19	0.52	7	-4
Boko Haram & IS	688	6	0.19	0.52	8	-2
Al Qaeda & IS	672	7	0.18	0.51	9	-2
Taliban & Al Qaeda	999	4	0.18	0.49	10	-6

Table 4.9: NTS Results for Target—Weapons Trails

4.4.2.4 Summary of Results

NTS was developed to correct potential biases in the simple absolute count of common transitions between two groups, controlling for the maximum probability of having a perfect identical pattern given two state sequences. Figures **??**, 4.19 and 4.19 display - for each trail - the scatter plots of the simple count of shared transitions and the unscaled NTS. Correlation values indicate how simple count is biased and NTS provides different outcomes. In the target+weapon scenario, only the pair with the highest count is also the pair with the highest NTS.

The comparative analysis indicates that the results across trails are generally

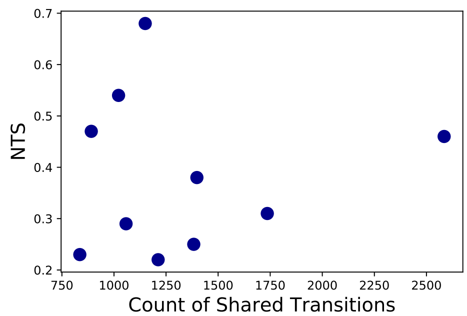

Figure 4.18: Scatter Plot: Simple Transition Count vs NTS (unscaled) for Weapon Trails (Pearson's correlation=0.056. Coefficient Statistically Significant at 99% Level.)

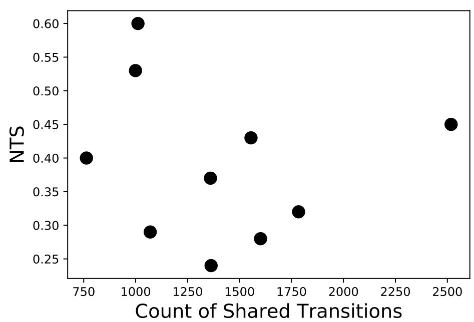

Figure 4.19: Scatter Plot: Simple Transition Count vs NTS (unscaled) for Target Trails (Pearson's correlation=-0.154. Coefficient Statistically Significant at 99% Level.)

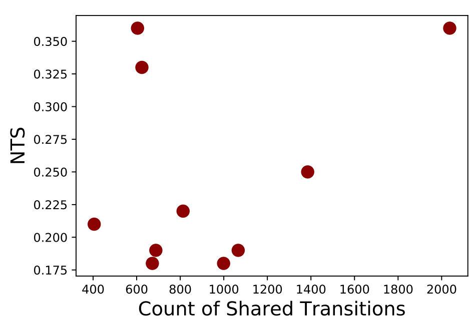

Figure 4.20: Scatter Plot: Simple Transition Count vs NTS (unscaled) for Target - Weapon Trails (Pearson's correlation=0.327. Coefficient Statistically Significant at 99% Level.)

stable. Indeed, three pairs out of ten perform standard deviation values of ranking lower than one position, as shown by Table 4.10 and Figure 4.21. Particularly, Al Qaeda and Al Shabaab are found to be the most similar groups overall, with a mean rank of 1.33: it is interesting to detect this stable similarity, considered that Al Shabaab officially became part of the Al Qaeda global network in 2012. Al Shabaab and Boko Haram are the second most similar pair. Regarding most dissimilar groups, a certain degree of stability is also shown, especially in the case of the Taliban and Al Qaeda: the pair is always ranked tenth. Little variance is exhibited by the Taliban and Boko Haram and Al Qaeda and the Islamic State. The latter pair deals with two groups that have been referenced by many as the old and the new paradigm of Islamic terrorism in the world. They do not show any particular evidence of similarity. This may propose that, besides other evident differences that span from the structural organization to the geographic scope of the operations, they also follow distinct behavioral trajectories.

	Weapon		**Target**		**Target+ Weapon**		**Mean R**	**St. Dev.**
Pair	***scaled NTS***	***NTS Rank***	***scaled NTS***	***NTS Rank***	***scaled NTS***	***NTS Rank***		
Al Qaeda & Al Shabaab	1.00	1	1.00	1	0.99	2	1.33	0.58
Al Shabaab & Boko Haram	0.79	2	0.88	2	0.91	3	2.33	0.58
Al Qaeda & Boko Haram	0.69	3	0.67	5	0.59	6	4.67	1.53
Taliban & IS	0.68	4	0.75	3	1.00	1	2.67	1.53
Al Shabaab & IS	0.56	5	0.62	6	0.62	5	5.33	0.58
Taliban & Al Shabaab	0.45	6	0.53	7	0.68	4	5.67	1.53
Al Qaeda & IS	0.43	7	0.49	8	0.51	9	8.00	1.00
Taliban & Boko Haram	0.36	8	0.48	9	0.52	7	8.00	1.00
Boko Haram & IS	0.34	9	0.71	4	0.52	8	7.00	2.65
Taliban & Al Qaeda	0.32	10	0.41	10	0.49	10	10.00	0.00
St. Dev.	0.22		0.19		0.20			

Table 4.10: Summary of NTS Results (R indicates Ranking Position)

Another relevant case regards Boko Haram and the Islamic State: the Nigerian organization is affiliated to the group led by Abu Bakr al-Baghdadi, but their similarity scores are particularly low. In fact, in the case of weapons and weapons and targets combined these groups rank among the last positions. However, in the target-only case, these differences vanish. This case is further proof of the fact that similar strategies of target selection can be coupled with distinct dynamic choices in terms of weapons.

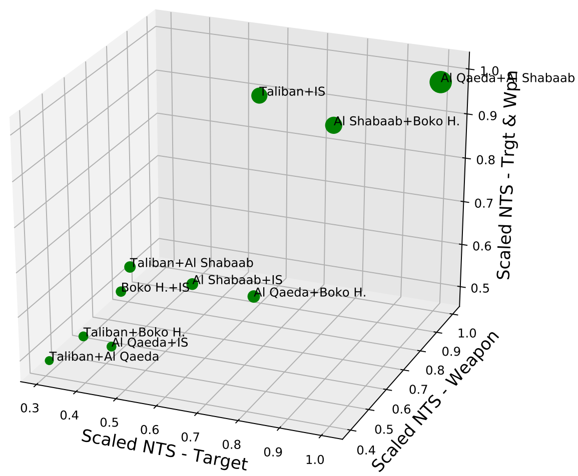

Figure 4.21: 3D Scatterplot of Scaled NTS for Group Pairs (size is scaled by the inverse of the mean R - Bigger points mean better mean ranking across trails)

Finally, to ensure that the coefficient is not biased by the skewness of the original length of the time series and the consequent temporal distribution of events, I have conducted a sensitivity analysis. This sensitivity analysis has been conducted creating two shorter time series: one taking into account events that happened from January 2007 to December 2016, and the other one considering only events that happened from January 2012 on. Pearson and Spearman correlation coefficients have been calculated to evaluate the extent to which limiting the time span would affect both the NTS coefficients and the related rankings. Table 4.11 shows that the results remain stable for all trails in both scenarios, thus suggesting that the initial choice to include all the events present in the dataset has not conducted to misleading outcomes.

Trail	**2007 censoring N=13,794**		**2012 censoring N=11,743**	
	Pearson'r R	**Spearman's Rho**	**Pearson'r R**	**Spearman's Rho**
Weapon	0.98*	0.98*	0.81*	0.86*
Target	0.89*	0.85*	0.84*	0.85*
Target+Weapon	0.99*	0.99*	0.92*	0.92*

Table 4.11: Sensitivity Test - NTS Values and Rankings Comparison Across 2007- and 2012- Censored Sequences. (*) Indicates that the Coefficient is Significant at 99.9% Level.

4.5 Conclusions and Future Work

In this study, I have proposed a two-fold structure for the quantitative study of terrorist attack sequences. First, relying on the well-established framework of Markov chains, I have introduced the concept of "super-states" for analyzing patterns in the sequence of jihadist attacks, focusing on three scenarios: sequences of weapons, sequences of targets and sequences of targets and weapons combined. The aim was to contribute to the debate regarding the nature and complexity of terrorist operational choices. The literature is divided into two main areas: most scholars agree upon the view of terrorists as "repetitive", stable and consistent in their operational processes, while others highlight their originality and diversity caused by technology advancements and shifts in objectives over time.

Starting from simple 1-dimensional states, I have created stochastic transition matrices up to the 5-dimensional case, to assess the general behavior of the networks and the complexity of combinations present in the data. The results indicate that, while drawing from relatively small sets of weapons and targets, the five groups show a highly complex repertoire of combinations in all the three considered scenarios. Nonetheless, as complexity in absolute terms (i.e., the number of existing super-states) increases, the number of accessible nodes per each super-state decreases, thus potentially reducing the challenges of predicting future behaviors. Hence, sequential heterogeneity can be studied to gain data-driven knowledge on how terrorists behave, going beyond the binary representation of terrorists as either repetitive or innovative. A final result concerns the fact that data show almost identical patterns for all the groups, potentially opening the way for generalized inferences on jihadism.

Second, I have proposed a novel coefficient, Normalized Transition Similarity (NTS), and compared the results of the analyses across groups and trails. NTS evaluates the behavioral pairwise similarity of single-link transitions between different states (again in the context of attacked targets, employed weapons and the combination of the two in each attack). It specifically uses the simple count of common transitions controlled by the potential maximum probability of perfect similarity given two random sequences associated with two considered jihadist organizations. The results showed that, across the three different networks, some stable similarities hold. Particularly, Al Qaeda & Al Shabaab (which are formally affiliated, since the former has become part of the global network in 2012) and Al Shabaab & Boko Haram are respectively ranked as the most similar pairs in two contexts out of three. At the same time, some other pairs confirm to be consistently dissimilar regardless of the transition networks that are considered. This is especially the case of Al Qaeda & Taliban. A final interesting result emerges looking at Boko Haram & Islamic State, which are very different when weapons and targets & weapons are considered but appear to be quite similar when only targets are included in the computation. This may suggest that, regardless of the proposed target, two groups may try to attack it using distinct dynamic strategies (in this case intended as weapons), thus providing potential interpretation on the scale of resources of the considered jihadist organizations.

The relevance of this exploratory analysis lies in the attempt to extract synthetic informative indications from the complex and heterogeneous behavioral dimensions of the most active jihadist groups in the global scenario. While detecting and assessing contextual differences between terror actors is valuable, it is also relevant to investigate how, if and to what extent they are similar to each other, especially when considering "state changes". Indeed, "state changes" may be fundamental sources of information for researchers and intelligence analysts, because in the frequent and apparently chaotic evolution of these behaviors lie the extreme difficulty of predicting, forecasting and countering terrorism as a violent act.

Although this is an exploratory work, it poses several policy implications. From a practical point of view, N-dimension super-states networks and similarity measures that take into account dynamic behaviors can be used by analysts to improve profiling of terrorist groups (especially if applying this methodology to larger samples involving higher number of groups) going beyond more static information, regarding for instance ideology, area of action, organizational structure. Furthermore, this general analytic approach can help inform countering strategies based on recurring sub-sequences or common state-changes. Terrorist events can be extremely harmful to societies, but every attack can be very different in scale compared with a previous or future one. For this reason, it is in the interest of institutions to understand how terrorists change their strategies and tactics. Combining additional information on attack magnitude or effects, transition networks and trails would be helpful in informing analysts and policymakers on the drivers of terrorist tactical patterns, facilitating alert tools and investigating the nature of successful (or unsuccessful) violent campaigns. With this regard, this two-fold framework is flexible in highlighting relevant evolution in groups' behavior, both for single groups and in a comparative fashion. This flexibility can be exploited in different manners, focusing on specific time windows for reducing the noisy effects of events that are distant in time or concentrating on precise dimensions of terrorist attacks.

Besides the outcomes presented in this article, the study has certainly several limitations, which call for further work that can improve research and policy applications. First, the analyses do not consider the temporal delta that occurs between two events. Given that terrorism is temporally clustered (Porter and White, 2012; White et al., 2013), and that the considered sequences are long and unequally distributed across time, not taking into account the delta that separates two events may lead to biased results that overestimate transition similarity. It would be useful to break up sequences that are more aligned to the temporal elements, considering that two events that are consecutively ordered in the original sequence may be far apart in the temporal scale, and it would be thus very risky to infer any kind of rational relation between the two.

A second layer of limitations comes from the fact that the NTS only considers single-link transitions when one-dimensional transition networks are analyzed. However, to investigate more complex patterns it would be valuable to apply the same mathematical construction of N-dimensional super-states to transition networks to understand if results are stable and hold when more complex information are employed.

A third layer of limitations is given by the fact that NTS can only assess pairwise similarity, without instead providing a global coefficient that can be applied without normalization to the whole sample, thus making it harder to interpret the results.

A fourth and final layer of limitations comes instead from the restricted sample of groups. Although working on a limited number of entities can provide more detailed insights, increasing the number of sequences to work with can control for false-positive patterns that may seem similar only due to the restricted number of pairs.

All these limitations are potentially solvable in the future, and this first exploratory study aims at opening the path towards the use of transition networks for terrorism research, showing the potential of this method that conceptually goes beyond the ordinary use cases derived from classic social network analysis.

5 Hawkes Processes of Jihadism

5.1 Introduction

One of the aims of this dissertation is to exploit the memory-like processes of terrorism to set up algorithmic architectures that can learn from these dynamics and provide reasonable and satisfactory predictions on future event characteristics. This chapter has, therefore, the primary goal to prove the existence of memory like processes in the considered sample of jihadist groups via the application of Hawkes processes.

Hawkes processes (Hawkes, 1971) are a specialized class of stochastic point processes that have gained wide success in many disciplines in the last decades. Their main feature is their self-excitability. Self-excitability posits that the occurrence of an event has a positive impact on the probability of occurrence of future ones. Technical details will be given in the following sections. Hawkes processes are thus a well-established class of models to verify whether certain events naturally cluster in time (and also space, when spatio-temporal modeling is applied) and to test the presence of memory-like processes, especially if comparing model diagnostics with homogeneous Poisson processes (Daley and Vere-Jones, 2006). In fact, Poisson processes are point processes for which the distribution of future inter-arrival times depends only on relevant information about the current time, but not on information from further in the past, while the self-excitability of Hawkes processes makes them intrinsically non-Markovian, creating dependencies between the past, the present, and the future.

Given the success demonstrated by this modeling technique for both criminal and terrorist events, I will analyze two Hawkes models for each group, taking into account the streams of events that occurred in the two most attacked countries per each organization, filtering by the most popular target attacked in each country. Testing the presence of memory is fundamental, considering that this is the dominating assumption connected to the methodological setup of the third and last analytic chapter on the use of deep learning architectures to predict jihadist future targets. Estimation diagnostics will thus be compared with baseline Homogeneous Poisson models, and detailed comments on the outcome will be provided. The rest of the chapter is outlined as follows: the "Background" section will provide a review of the main literature on the application of Hawkes process, with a focus on crime-, security- and terrorism-related problems. The "Mathematical Framework" section will describe in detail the mathematics behind Hawkes processes and the specific methodology applied in the present work. The "Experiments" section will then thoroughly present the results of the models, with a specific focus on each group. Finally, general indications arising from these analyses, potential implications, future research paths will be presented in the "Discussion and Future Work" section.

5.2 Related Work

The application of Hawkes processes spans across several domains. Their modeling flexibility has captured the attention of almost all areas in which event and sequence data are relevant sources of information for researchers and scientists. Besides research from the foundational and theoretical standpoints (Daley and Vere-Jones, 2006; Bacry et al., 2012; Eichler et al., 2017) , some of the fields in which Hawkes processes are applied are finance (Chavez-Demoulin and McGill, 2012; Hawkes, 2018), geophysics (Ogata, 1988; Türkyilmaz et al., 2013), computational social science (Kobayashi and Lambiotte, 2016) and neuroscience (Reynaud-Bouret et al., 2013; Gerhard et al., 2017) (for a more comprehensive review, see Reinhart (2018)). However, the progressive and increasing use of quantitative measurements on a variety of social phenomena in research has started to make Hawkes processes (and point processes in general) more popular also in criminology (Mohler et al., 2011; Mohler, 2013) (Mohler 2011, Hegemann 2012, Mohler 2013) and terrorism research (Porter and White, 2012; Lewis et al., 2012; Tench et al., 2016) . Besides the availability of data and the shift towards statistical approaches in the social sciences, Hawkes processes have been successful due to the well-known clustering mechanisms of certain types of crimes and violent phenomena in space and time: the temporal concentration of crime as a scientific finding has indeed long preceded the diffusion and application of Hawkes models in criminology.(Midlarsky, 1978; Midlarsky et al., 1980; Holden, 1986; Freeman et al., 1996; Braithwaite and Li, 2007; Weisburd et al., 2009; Weisburd, 2015).

For what concerns urban crimes Mohler et al. (2011) tested the clustering dynamics of crime using self-exciting point processes on residential burglary data of the city of Los Angeles. They have considered offenses that occurred in the San Fernando

Valley in the years 2004 and 2005 and fitted an unmarked model (i.e., a model in which each event is considered as equal, without any qualitative information regarding impact, damage, etc.) with a non-parametric estimation. Their study has been among the first ones to introduce and present the potential of Hawkes processes for the study and prediction of crime, emphasizing the similarities between earthquakes and repeated offenses.

In a further attempt to advance the methodology and expand the types of criminal phenomena to be analyzed, Mohler (2013) studied property and violent crimes in Chicago and terrorist attacks and casualties in Northern Ireland, Israel, and Iraq. He tested a particular type of Hawkes process with a background rate driven by a log Gaussian Cox process, proposing a Metropolis adjusted Langevin algorithm for learning the model parameters. The work clearly shows the number of events to be associated with the background rate and the component connected to the Hawkes specification, calling for additional research that may embed also spatial components. The same author continued his research on Chicago expanding the methodological framework via the application of Marked Hawkes processes to yield accurate hot-spot maps to be used to tackle gun violence. The shift towards marked Hawkes processes was performed to take into account potential triggering and precursory offenses. The model is developed using an Expectation-Maximization algorithm (Veen and Schoenberg, 2008) and shows better performance compared to other types of hot-spot prediction techniques.

For what pertains terrorism, Northern Ireland and Iraq have been two countries of particular interest for research on Hawkes processes Indeed, Tench et al. (2016) have conducted a study using data on Improvised Explosive Devices (IED) attacks carried out by the Provisional Irish Republican Army (PIRA) during “The Troubles” in Northern Ireland, also integrating information on counter-attacks plotted by the British Security Forces. The integration of these two sources of events led to the development of a multivariate Hawkes process in the attempt to understand whether besides past-dependencies between attacks also inter-dependencies among the two event sequences exist. The authors showed indeed how counter-terrorism operations lead to subsequent spikes of terrorist violence. Lewis et al. (2012) have instead focused on reported deaths of civilians in Iraq, proposing and comparing three adjustments for non-stationarity of the background rate. The study proved the best performance of models including self-exciting components with respect to both stationary and non-stationary homogeneous Poisson processes, thus demonstrating the presence of memory dynamics and the violation of the Markovian assumption.

Furthermore, White et al. (2013) concentrated their research on three Southeast Asian countries, namely Indonesia, the Philippines, and Thailand. Their work aimed at developing interpretable metrics for risk, resilience, and volatility of terrorist activity. Through the use of self-exciting point process models, they have computed measures of risk of daily expected terrorist attacks, additional attacks caused by every single attack and number of days with low risk.

Maintaining the focus on Indonesia, retrieving data on the daily number of attacks in the period 1994-2007, Porter and White (2012)formalized a dynamic model using a shot noise process for explaining the self-excitability of terrorism. Using a power-law distribution and a shot noise derived parameters, they achieved the best performance in modeling the daily number of attacks. As for most of the other works described in this section, relevant evidence is given to the promises of Hawkes and self-exciting point processes for gaining practical and useful knowledge on terrorism, going beyond the borders of academic research.

5.3 Mathematical Framework

5.3.1 Introducing Homogeneous Poisson and Hawkes Point Processes

Before describing in detail the nature of Hawkes processes, it is necessary to introduce some preliminary mathematical concepts. Given a point process $(t_i)_{i \in \mathbb{N}^*}$, then its associated counting process is defined as:

$$N(t) = \sum_{i \in \mathbb{N}^*} \mathbf{1}_{t_i \leq t} \tag{5.1}$$

with t_i being the times in which the phenomenon under analysis occurs and 1_{t_i} being the indicator function that is equal to 1 if $t_i \leq t$ and 0 otherwise. Having introduced the concept of counting process, it is worth also to provide definitions for duration and history. The duration process associated with $(t_i)_{i \in \mathbb{N}^*}$ is defined by:

$$\forall i \in \mathbb{N}^*, \delta t_i = t_i - t_{i-1} \tag{5.2}$$

while the history of events up to a given time t is given by:

$$\mathcal{H}(t) =: t_i | t_i < t \tag{5.3}$$

At this point, the conditional intensity function $\lambda(t)$ associated to the process, dependent on $\mathcal{H}(t)$ is formalized as:

$$\lambda\left(t|\mathcal{H}(t)\right) = \lim_{h \to 0} \frac{\mathbb{E}(N(t+h) - N(t)|\mathcal{H}(t))}{h} \tag{5.4}$$

Following this equation, λ is the expected number of events that should occur at each time unit t. This quantity will always depend upon $\mathcal{H}(t)$. The selected history will then act as a sort of filter of each model: given a complete sequence of events and an artificially-reduced one, the intensity will vary accordingly.

This initial background allows making the first important distinction in the realm of point processes, namely the distinction between Homogeneous Poisson and Hawkes processes. Understanding the difference between these two mathematical entities is crucial to understand the implications of this work. A point process is said to be a Homogeneous Poisson process if the intensity is positive, fixed and constant:

$$\lambda = \mu \tag{5.5}$$

More formally, a Poisson process with constant rate λ is a point process formalized as:

$$P(N(t+h) - N(t) = 1|\mathcal{H}) = \lambda h + o(h) \tag{5.6}$$

$$P(N(t+h) - N(t) > 1|\mathcal{H}) = \lambda o(h) \tag{5.7}$$

Equations 5.6 and 5.7 indicate that the intensity does not depend on the history of the process itself, with the probability of an event happening in $(t, t+h]$ being indeed independent of the filtering given by $\mathcal{H}$ and duration δt_i independent and identically distributed (i.i.d.) according to an exponential distribution parametrized by λ. In other words, Poisson processes are memoryless: their nature is intrinsically Markovian. However, in the real world, many phenomena do not respect the assumption of constant probability in fixed time windows and independence from the past. With this regard, a Hawkes process is an alternative class of model with different properties: it revolves around the idea of "self-excitability", which means that the occurrence of an event has an impact (generally, assumed positive) on the occurrence of another event in the future.

In light of this, given the background rate μ, which is the average rate of event occurrence per time unit; k_0 defined as the increase rate of events following a past one (the higher the value, the more reactive the process is to future events), and ω, the decay parameter that maps the extent to which the probability of an events decreases after a spike, then for a given set of unique times t_i, Hawkes (1971) defines the intensity function of a self-exciting process as:

$$\lambda(t) = \mu + k_0 \int_{-\infty}^{t} g(t - t_i) dZ(u) = \mu + k_0 \sum_{t > t_i} g\left(t - t_i\right) \tag{5.8}$$

where Z is the normal counting measure and the response function g is an exponential kernel in the form:

$$g(t) = \omega e^{-\omega t} \tag{5.9}$$

The equation of the intensity function would then read as:

$$\lambda(t) = \mu + k_0 \sum_{t > t_i} g\left(\omega e^{-\omega t}\right) \tag{5.10}$$

Given that ω appears in the exponent, then the higher its value, the lower the temporal effect an event has on future ones. Additionally, Lewis et al. (2012) notes that w^{-1} gives the average time length over which a spike in the rate of events occur. As pointed out by Lewis et al. (2012), a point process $N(t)$ is said to be self-excitating iff:

$$\operatorname{Cov}\left[N(t_1, t_2), N(t_2, t_3)\right] > 0 \quad \forall\, t_1 < t_2 < t_3 \tag{5.11}$$

which means that if an event occurs, another one is more likely to happen locally in time (but also space). In the case of homogeneous Poisson processes, in fact, this is not true, as

$$\operatorname{Cov}\left[N(t_1, t_2), N(t_2, t_3)\right] = 0 \quad \forall\, t_1 < t_2 < t_3 \tag{5.12}$$

5.3.2 Estimation of the Parameters and Model Comparison

Maximum Likelihood Estimation (MLE) has been performed to learn the parameters μ, k_0 and ω of each Hawkes process model. MLE aims at finding the parameters that maximize the log-likelihood function. As noted by Tench (2018), for a set of event times $\{t_i\}_{i=1}^{N}$, log-likelihood is calculated as:

$$\ln\mathcal{L}\left(\{t_i\}; \mu; k_0; \omega\right) = \sum_{i=1}^{N} \log(\lambda(t_i)) - \int_0^T \lambda(t) dt \tag{5.13}$$

Following Tench (2018), the integral of the second term can be simplified as:

$$\begin{aligned}
\int_0^T \lambda(t)dt &= \int_0^T \mu + k_0 \sum_{t_i<t} \omega e^{-\omega(t-t_i)} dt \\
&= \mu T + k_0 \sum_i \int_0^T \omega e^{-\omega(t-t_i)} \mathbf{1}_{t>t_i} dt \\
&= \mu T + k_0 \sum_i \int_{t_i}^T \omega e^{-\omega(t-t_i)} dt \\
&= \mu T + k_0 \sum_i \int_{t_i}^T \omega e^{-\omega(t-t_i)} dt \\
&= \mu T + k_0 \sum_i \left[-e^{-\omega(t-t_i)}\right]_{t_i}^T \\
&= \mu T + k_0 \sum_i \left[1 - e^{-\omega(T-t_i)}\right]
\end{aligned} \tag{5.14}$$

Substituting this final equation in the original one, provides the complete form of the log-likelihood computed as:

$$\ln\mathcal{L} = \sum_{i=1}^{N} \left[\log \left(\mu + k_0 \sum_{t_i>t_j} \omega e^{-\omega(t_i-t_j)} \right) + k_0 \left(e^{-\omega(T-t_i)} - 1 \right) \right] - \mu T \tag{5.15}$$

To obtain the maximization of the log-likelihood function, I relied on the Nelder-Mead (Nelder and Mead, 1965) method available in the R package ptproc. The algorithm in the package works to minimize the function, thus it searches for the minimization of $-\ln\mathcal{L}$. The Nelder-Mead approach is a heuristic optimization technique that uses the geometric concept of a simplex, a special case of polytop with $n+1$ vertices in n dimensions. The algorithm starts generating a random simplex, and at every iteration it proceeds to reshape and move it, iteratively one vertex at a time, trying to settle in an optimal region of the search space. Specifically, given a n-dimensional space, a simplex consists of $n+1$ points $x_1, x_2, ..., x_{n+1}$ and the algorithm tries to minimize a function $f(x)$ via several steps during each iteration. During the "Ordering" phase, all points are sorted in order to set the value of f for the first point as the lowest, and the one of the last as the highest. The indices of the worst, second-worst and best points be h, s, l. In the "Centroid Computation" phase, the algorithm considers all points but the worst one (x_h), and calculate their centroid as:

$$c = \frac{1}{n} \sum_{i \neq h} x_i \tag{5.16}$$

After this step, comes the "Transformation" phase, which consists of *Reflection*, *Expansion*, *Contraction* and *Shrink Contraction*. During the *Reflection* the algorithm computes the reflected point as:

$$x_r = c + \alpha(c - x_h) \tag{5.17}$$

where α is the reflection parameter, and x_r is a point on the line that connects c and x_h, but sufficiently away from it. This step aims at move the simplex in a direction away from the sub-optimal region around x_h. If after this step, $f(x_s) < f(x_r) \leq f(x_l)$, the algorithm substitutes x_h with x_r and proceeds to the *Expansion* step. If then the reflected point x_r is better than the current best ($f(x_r) > f(x_l$, the algorithm moves in the direction of x_r from c. The expanded point is then defined as:

$$x_e = c + \gamma(x_r - c) \tag{5.18}$$

where γ is here an expansion parameter (usually set at a value of 2). At this point, the algorithm replace x_h with the best point between x_e and x_r. In case the reflection point was worse than x_s (the second worst point), the algorithm contracts the simplex. The contraction point is then defined as:

$$x_c = c + \beta(x_h - c) \tag{5.19}$$

with β being the contraction parameter. If $f(x_c) > f(x_h)$, it means that the contracted point is actually better than the current worst, and the algorithm then replaces x_h with x_c in the simplex. In case this the relation above is not satisfied, the algorithm moves to the *Shrink Contraction* step. During this step, the algorithm only keeps the best point (x_l and re-define the other with respect to it, so that the new point is defined as:

$$x_j = x_l + \delta(x_j - x_l) \tag{5.20}$$

where δ is the shrinkage parameter. The algorithm finally terminates if (1) a pre-set number of iterations is reached, or (2) the simplex reaches a limit of minimum size, or (3) the current best solution reaches some acceptable limit. In the present work, a default limit of 500 iterations for each model has been selected.

Once the parameters have been estimated, it is necessary to check whether the model actually fits the real data. Following the approaches used in other articles, residual analysis and a consequent Kolmorogv-Smirnov test (Massey, 1951) have been performed. Given a set of event times $\{t_i\}$ associated to an Hawkes point process with intensity λ, the residuals for each i are computed through:

$$\tau_i = \int_0^{t_i} \lambda(t)dt \tag{5.21}$$

Following these residuals should be distributed as a stationary process with unit rate, therefore it can be proved that they are also exponentially distributed via:

$$\begin{aligned}
\mathrm{Y}_i = \tau_i - \tau_{i-1} &= \int_0^{t_i} \lambda(t)dt - \int_0^{t_{i-1}} \lambda(t)dt \\
&= \int_{t_{i-1}}^{t_i} \lambda(t)dt \\
&= \int_{t_{i-1}}^{t_i} \mu + k_0 \sum_{t_j<t} \omega e^{-\omega(t-t_j)} dt \\
&= \mu\left(t_i - t_{i-1}\right) + k_0 \int_{t_{i-1}}^{t_i} \sum_{t_j<t} \omega e^{-w(t-t_j)} dt \\
&= \mu\left(t_i - t_{i-1}\right) + k_0 \sum_{j=1}^{i-1} \int_{t_{i-1}}^{t_i} \omega e^{-w(t-t_j)} dt \\
&= \mu\left(t_i - t_{i-1}\right) + k_0 \sum_{j=1}^{i-1} \left[-e^{-\omega(t-t_j)}\right]_{t_{i-1}}^{t_i} \\
&= \mu\left(t_i - t_{i-1}\right) + k_0 \sum_{j=1}^{i-1} \left[e^{-\omega(t_{i-1}-t_j)} - e^{-\omega(t_i-t_j)}\right]
\end{aligned} \tag{5.22}$$

This would thus mean that U_i are uniform random variables:

$$\begin{aligned}
\mathrm{U}_i &= 1 - \exp^{-\mathrm{Y}_i} \\
&= 1 - \exp\left[-\left(\mu(t_i - t_{i-1}) + k_0 \sum_{j=1}^{i-1} \left[e^{-\omega(t_{i-1}-t_j)} - e^{-\omega(t_i-t_j)}\right]\right)\right]
\end{aligned} \tag{5.23}$$

It follows that to check whether the Hawkes process fits the data, it can be verified if U_i actually belong to a uniform distribution. Applying the same approach of

Tench et al. (2016), I have then performed Kolmogorov-Smirnov (KS) test. This very common test compares the value of a test statistic to a given critical value D_σ. The KS test statistic is computed as:

$$D_n = \max_k \left(\left| U_k - \frac{k-1}{N}, \left| \frac{k}{N} - U_k \right| \right| \right) \tag{5.24}$$

and the model is found to be fitting to the data if $D_n < D_\sigma$. Finally, to provide further evidence that the data under analyses are modeled correctly through a Hawkes process, I have compared the Hawkes models with baseline Homogeneous Poisson processes. To do so, I have compared the Akaike Information Criterion (AIC) (Akaike, 1974) value of Hawkes and homogeneous Poisson in each model. Between the two, the best model to be chosen is the one with the lowest AIC, calculated as:

$$\mathrm{AIC} = 2k - 2\ln\mathcal{L} \tag{5.25}$$

with k being the number of parameters of the model and $\ln\mathcal{L}$ being the MLE.

5.3.3 The Present Study

The present study aims at deepening the knowledge on jihadist dynamics applying Hawkes processes to evaluate self-excitability given certain attacks characteristics, and particularly targets. This methodological choice is motivated also by the need for providing empirical results on memory processes that can justify the architecture of the last analytic chapter on neural networks for predictive purposes. Showing the presence of past-dependency in the attack sequences with specific characteristics is a valuable way to suggest that further patterns can be learned when focusing on historical data. In this chapter, I have decided to perform a double-filter. First, for each group, the attacks that occurred in the two most targeted countries will be kept. Secondly, only the attacks with the two most popular targets will be selected (target frequency per each group and country is displayed in Tables 5.1, 5.2, 5.3, 5.4 and 5.5).[1] Given the necessity to provide unique time stamps to correctly run the models, when multiple attacks occurred in the same day, a number x sampled from a random uniform distribution ($x \sim U(0, 1]$) has been added in order to keep all the

[1] In this chapter, I have only taken into account the first *target type*, without considering also the potential other two. This decision reduces the level of detail of the information, but still guarantees a solid proof of concept for the aims of the work. As done in the other chapters, I have used the most general level of information out of the four different available when considering targets.

attacks and maintain relevant information that can better explain the intensity of a terrorist wave of attacks. The analyses have been performed relying on R package ptproc (Peng, 2003), and on a private Pythonanywhere website kindly provided by Stephen Tench and originating from his doctoral dissertation work (Tench, 2018). As anticipated above, the models work with exponential kernels and the Nelder-Mead algorithm (Nelder and Mead, 1965) will be used for MLE optimization to learn the parameters of each model.

Iraq			**Syria**		
Target	**N**	**%**	**Target**	**N**	**%**
Private Citizens & Prop.	1476	49.02%	Private Citizens & Prop.	202	51.53%
Police	487	16.17%	Terrorists/N. St. Militia	41	10.46%
Military	223	7.41%	Military	33	8.42%
Business	182	6.04%	Business	27	6.89%
Terrorists/N.S.Militia	144	4.78%	Journalists & Media	19	4.85%
Government (General)	126	4.18%	Police	16	4.08%
Unknown	104	3.45%	Religious Fig./Inst.	15	3.83%
Others	269	8.93%	Others	39	9.95%
Total	3011	100%	Total	392	100%

Table 5.1: Target Type Frequency (Highest in Red) for the Islamic State in Iraq and Syria

Afghanistan			**Pakistan**		
Target	**N**	**%**	**Target**	**N**	**%**
Police	2185	39.09%	Private Citizens & Prop.	23	46.94%
Private Citizens & Prop.	1238	22.15%	Military	5	10.20%
Government (General)	850	15.21%	Police	5	10.20%
Military	287	5.14%	Educational Institution	4	8.16%
Business	201	3.60%	Government (General)	3	6.12%
Unknown	184	3.29%	Business	2	4.08%
Educational Institution	135	2.42%	Terrorists/N. St. Militia	2	4.08%
Others	509	9.11%	Unknown	2	4.08%
Total	5589	100.00%	Others	3	6.20%
			Total	49	100.00%

Table 5.2: Target Type Frequency (Highest in Red) for the Taliban in Afghanistan and Pakistan

Yemen			Iraq		
Target	**N**	**%**	**Target**	**N**	**%**
Government (General)	145	24.62%	Private Citizens & Prop.	233	40.10%
Police	119	20.20%	Police	123	21.17%
Private Citizens & Prop.	81	13.75%	Government (General)	62	10.67%
Military	68	11.54%	Business	46	7.92%
Terrorists/N. St. Militia	50	8.49%	Terrorists/N. St. Militia	29	4.99%
Utilities	37	6.28%	Military	18	3.10%
Government (Diplomatic)	20	3.40%	Religious Fig./Inst.	17	2.93%
Business	19	3.23%	Others	53	9.12%
Others	50	8.48%	Total	581	100.00%
Total	589	100.00%			

Table 5.3: Target Type Frequency (Highest in Red) for Al Qaeda in Yemen and Iraq

Nigeria			Cameroon		
Target	**N**	**%**	**Target**	**N**	**%**
Private Citizens & Prop.	946	49.68%	Private Citizens & Prop.	131	65.50%
Police	228	11.97%	Religious Fig./Inst.	16	8.00%
Religious Fig./Inst.	152	7.98%	Unknown	12	6.00%
Government (General)	138	7.25%	Police	11	5.50%
Business	99	5.20%	Business	8	4.00%
Educational Institution	86	4.52%	Military	7	3.50%
Military	68	3.57%	Others	15	7.50%
Others	187	9.82%	Total	200	100%
Total	1904	100%			

Table 5.4: Target Type Frequency (Highest in Red) for Boko Haram in Nigeria and Cameroon

Somalia			Kenya		
Target	**N**	**%**	**Target**	**N**	**%**
Private Citizens & Prop.	427	30.48%	Police	99	34.26%
Government (General)	342	24.41%	Private Citizens & Prop.	55	19.03%
Military	165	11.78%	Business	47	16.26%
Police	107	7.64%	Religious Fig./Inst.	16	5.54%
Business	95	6.78%	Government (General)	15	5.19%
Journalists & Media	44	3.14%	Transportation	15	5.19%
Government (Diplomatic)	34	2.43%	Military	10	3.46%
Unknown	31	2.21%	NGO	8	2.77%
Airports & Aircraft	27	1.93%	Others	24	8.30%
Others	129	9.20%	Total	289	100.00%
Total	1401	100.00%			

Table 5.5: Target Type Frequency (Highest in Red) for Al Shabaab in Somalia and Kenya

5.4 Experiments

5.4.1 The Islamic State

The two most targeted countries from attacks carried out by the Islamic State are Iraq and Syria. However, the comparison is disproportionate given that Iraq accounted for more than 3,000 incidents, while Syria only accounts for less than 400. The first attack ever recorded in the GTD and assigned to the Islamic State was dated April 18, 2013. This testifies the extremely intense activity of the IS in the temporal frame under analysis. Besides pure event counting, the IS has carried out attacks in Iraq on 933 unique days, while in Syria days were 228. When focusing on the most popular targets, the Islamic State preferred to hit Private Citizens and Property in both countries (49.02% of all attacks in Iraq and 51.53% in Syria). The results of the models for both countries are reported in Table 5.6.

Parameters	**Iraq**	**Syria**
μ	0.562	0.090
k	0.495	0.424
ω	3.908	2.497
KS Test Stat	0.035*	0.032*
KS 95% Sig. Level	0.095	0.095
KS 99% Sig. Level	0.114	0.0114
Hawkes AIC	1283.852†	934.330†
H. Poisson AIC	2634.031	1147.258

Table 5.6: Univariate Hawkes Estimates for Islamic State Models (Iraq and Syria). † Indicates which Model Between Hawkes and H. Poisson better Explains the Process. * Indicates 95% significance of the KS Statistic, ** Indicates 99%.

Firstly, the results of the AIC tests (values are lower for the Hawkes process compared to the Homogeneous Poisson) suggest that a certain degree of self-excitation is present in both the countries and that, therefore, there exist some memory-like dynamics in the way in which the Islamic State behaves when considering Private Citizens and Property as targets. A second order of results regards the interpretation of the parameters. The parameter μ describes the average number of events at each time step: Iraq has a much higher average number of attacks per day compared to Syria. The jump factors k instead suggest that in both countries the reactivity of the process at each event is almost similar.[2] Thirdly, the inverse of the parameter ω

[2] k is bounded in the range $(0, 1]$, with higher values indicating higher reactivity.

provides information on the average length of periods in which higher rates of events occur. In the case of Iraq, the length is shorter (0.25 days) than the length in Syria (0.4 days), testifying the very high frequency of attacks in both geographical contexts, which represents a distinguishing characteristic of the Islamic State. This is visually shown also by Figures 5.5, 5.6, 5.7, 5.8, where for both groups the event streams and the estimated conditional intensities λ are displayed.

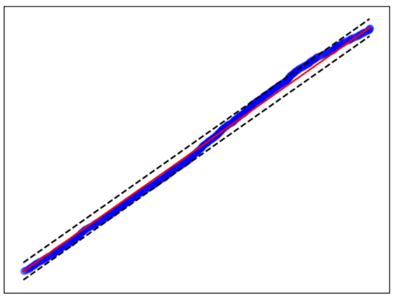

Figure 5.1: IS KS Plot - Iraq

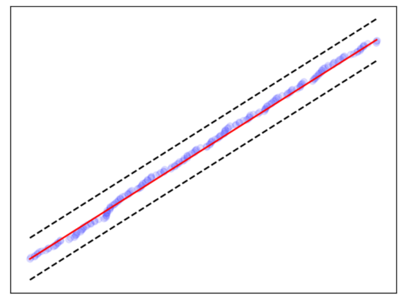

Figure 5.2: IS KS Plot - Syria

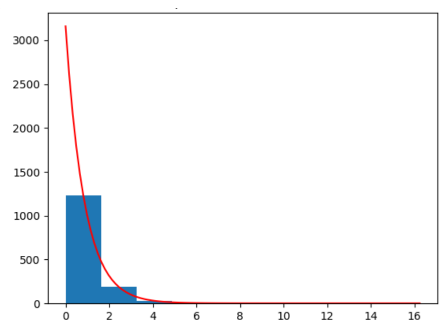

Figure 5.3: IS Inter-arrival Times - Iraq

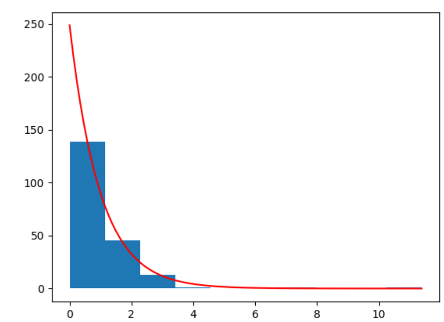

Figure 5.4: IS Inter-arrival Times - Syria

The KS test statistic gives qualitative information regarding the goodness of fit of the models at the level of confidence of 95%. Figures 5.1 and 5.2 show that most of the points are falling on the red solid line (Dashed lines represent 95% confidence boundaries). Inter-arrival times are displayed in Figure 5.3 and 5.4. The first-order differences follow an exponential decay: the model thus represents the data well. For both countries, most attacks occur within two days, further demonstrating the extreme frequency of violence carried out by the Islamic State.

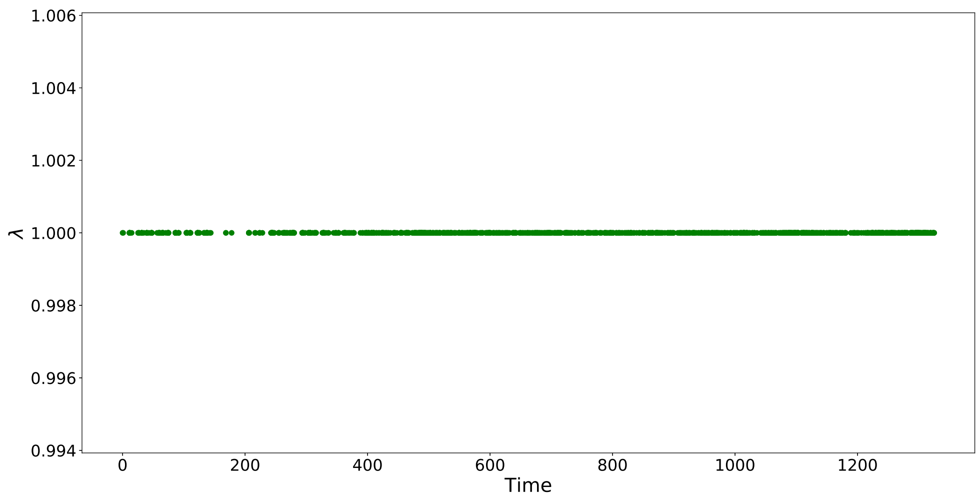

Figure 5.5: Event Stream of Islamic State in Iraq

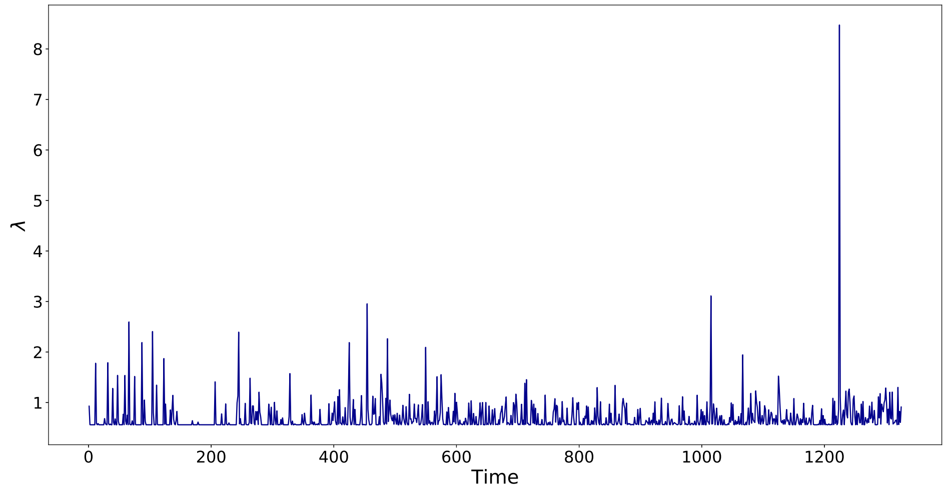

Figure 5.6: Conditional Intensity λ of Islamic State Attacks in Iraq

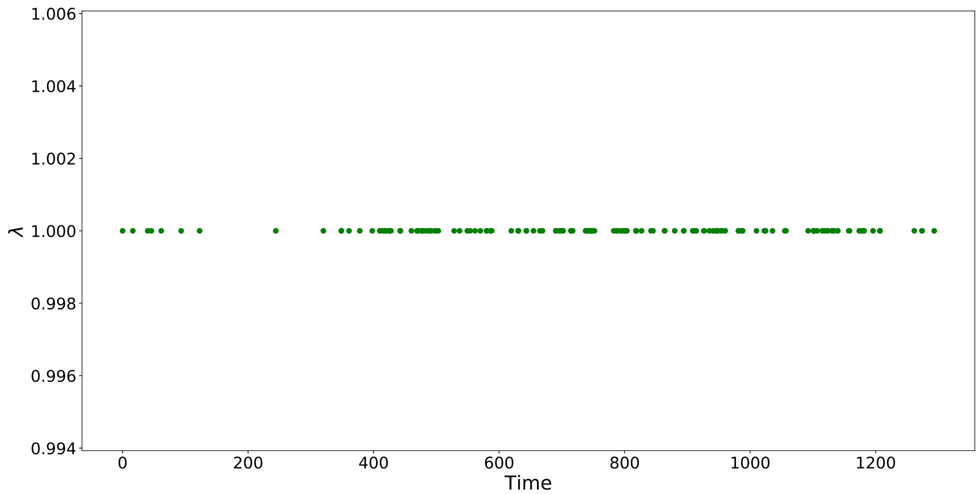

Figure 5.7: Event Stream of Islamic State in Syria

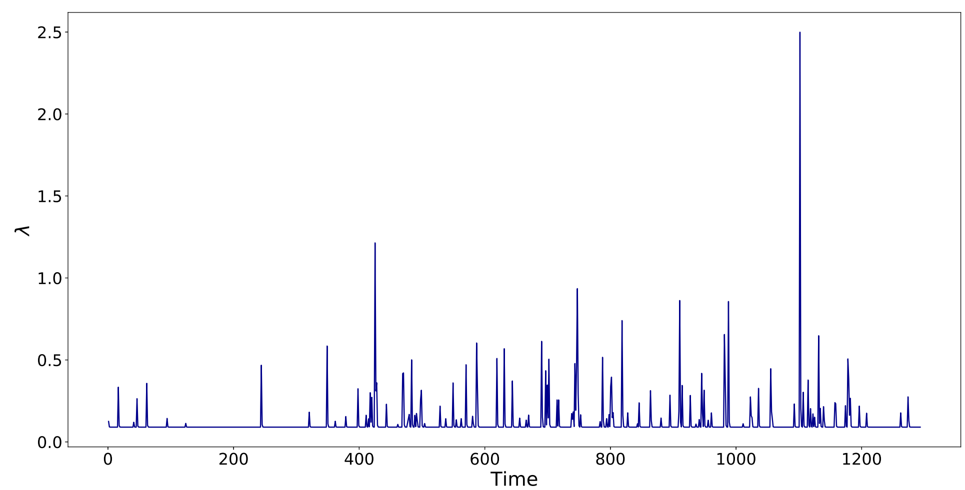

Figure 5.8: Conditional Intensity λ of Islamic State Attacks in Syria

5.4.2 The Taliban

The Taliban, during their very long history of terrorism, has only attacked two countries: Afghanistan and, marginally, Pakistan. In fact, out of a total of 5,638 attacks, only 49 have targeted Pakistan. The first recorded event in the dataset occurred in Afghanistan in April 1995. In terms of days, the Taliban have carried out terrorist attacks in Afghanistan on 2,522 unique days and in Pakistan on 47. For what concerns targets, in Afghanistan the Taliban has targeted Police in 39.09% of the attacks, testifying their strategy against the State and institutions. In Pakistan, instead, Private Citizens and Property have been attacked in 46.94% of the cases. Model results are shown in Table 5.7.

Parameters	**Afghanistan**	**Pakistan**
μ	0.023	0.002
k	0.942	0.615
ω	0.083	0.011
KS Test Stat	0.048	0.121*
KS 95% Sig. Level	0.029	0.283
KS 99% Sig. Level	0.034	0.339
Hawkes AIC	5325.685†	269.912
H. Poisson AIC	7915.538	265†

Table 5.7: Univariate Hawkes Estimates for Taliban Models (Afghanistan and Pakistan). † Indicates which Model Between Hawkes and H. Poisson better Explains the Process. * Indicates 95% significance of the KS Statistic, ** Indicates 99%.

The results for the Taliban are not as good as the ones obtained with the Islamic State. While the AIC statistic indicates that, for Afghanistan, the Hawkes model better captures the dynamics of the data, for Pakistan the Homogeneous Poisson process seems to have a better fit (although AIC values are very similar). This can be probably explained by the very low number of attacks considered for Pakistan (i.e., 23) and by their distribution in the considered time frame. The KS test provides again contrasting results: while the Pakistan model reaches a 95% significance level, the Afghanistan one fails the test. The outcomes of the KS tests are graphically provided in Figures 5.9 and 5.10. In the case of Afghanistan, the points remain within the confidence boundaries except for a deviation in the bottom-left of the plot.

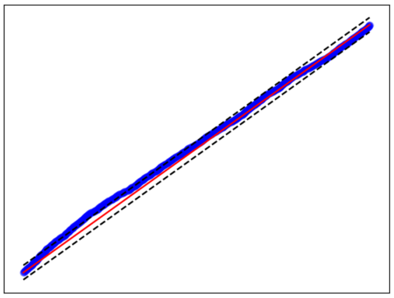

Figure 5.9: Taliban KS Plot - Afghanistan

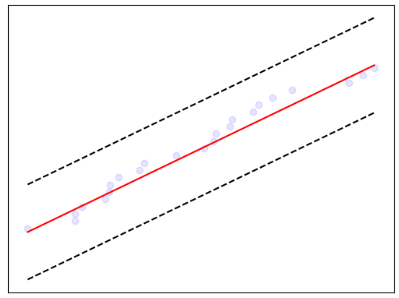

Figure 5.10: Taliban KS Plot - Pakistan

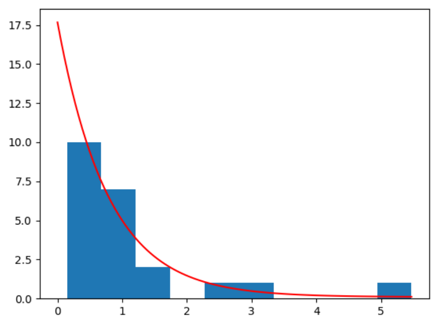

Figure 5.11: Taliban Inter-arrival Times - Afghanistan

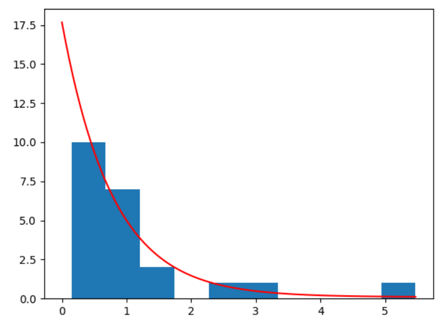

Figure 5.12: Taliban Inter-arrival Times - Pakistan

In spite of the limitations addressed in the previous lines, it is however worth to provide some context via parameter interpretation. Parameter μ shows that, on average, Afghanistan tended to experience a higher number of attacks per single day (0.023 vs 0.002). The jump factor k is very close to the limit of 1 in Afghanistan (0.942), testifying the very high level of excitability and eventually escalation. For the Pakistan case, the value is still high (0.615). Finally, ω^{-1} gives information on the average number of days over which self-excited events persist: in the case of Afghanistan, this time window lasts 12 days, while the period is much longer for Pakistan (90 days): this is discrepancy in the results between the two countries is obviously motivated by the sizeable difference in absolute numbers of attacks and consequent distribution. These aspects can be seen in Figures 5.13, 5.14, 5.15, 5.16.

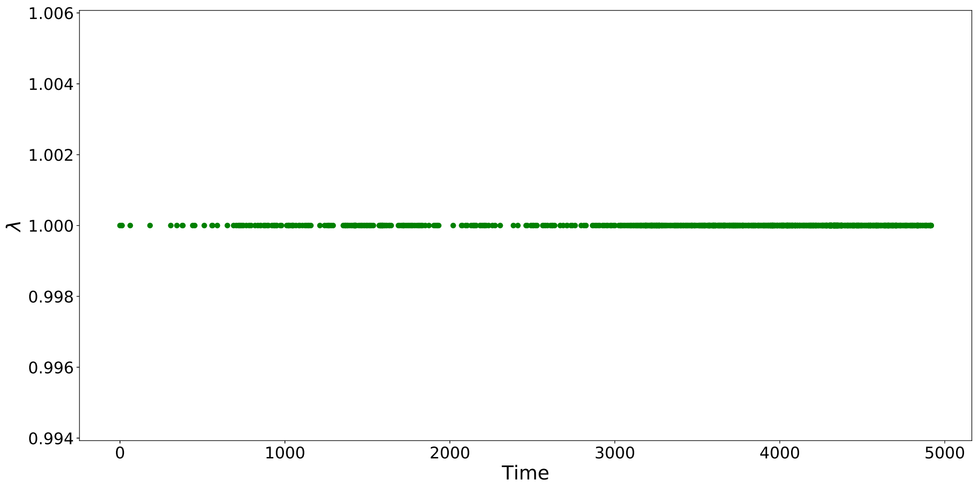

Figure 5.13: Event Stream of Taliban in Afghanistan

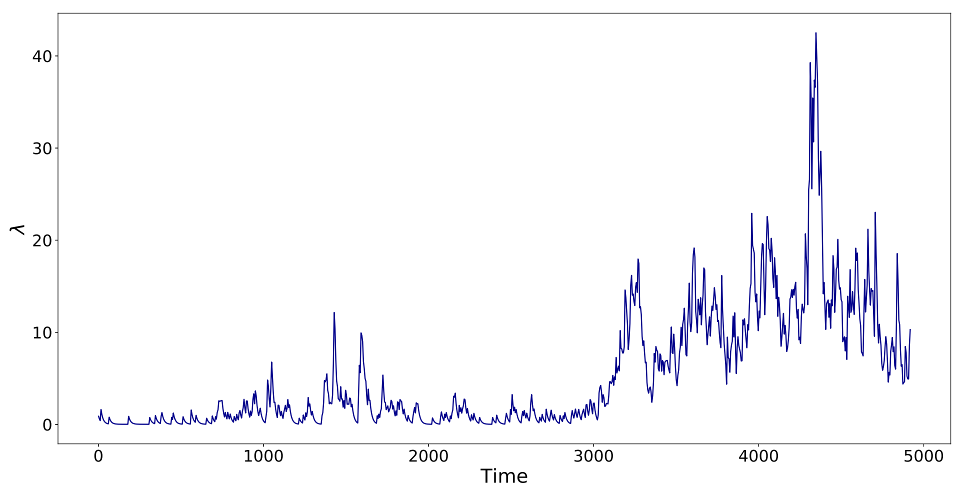

Figure 5.14: Conditional Intensity λ of Taliban Attacks in Afghanistan

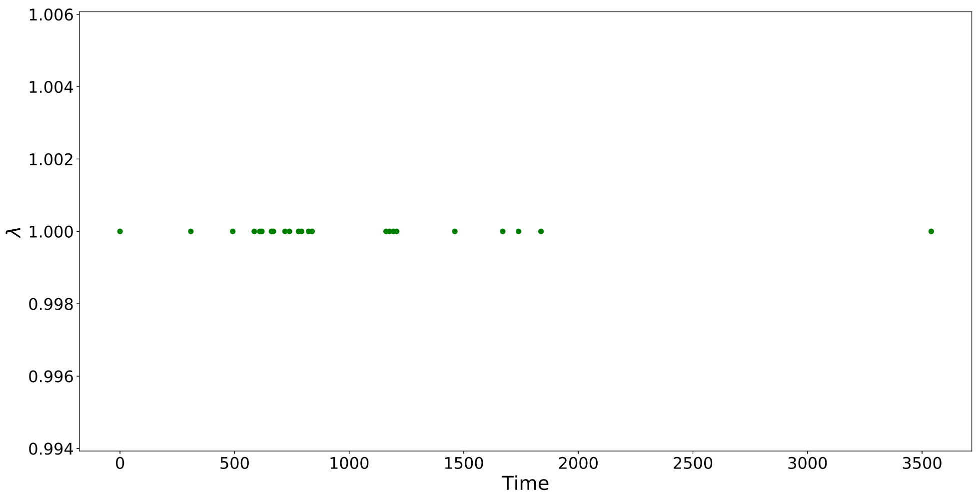

Figure 5.15: Event Stream of Taliban in Pakistan

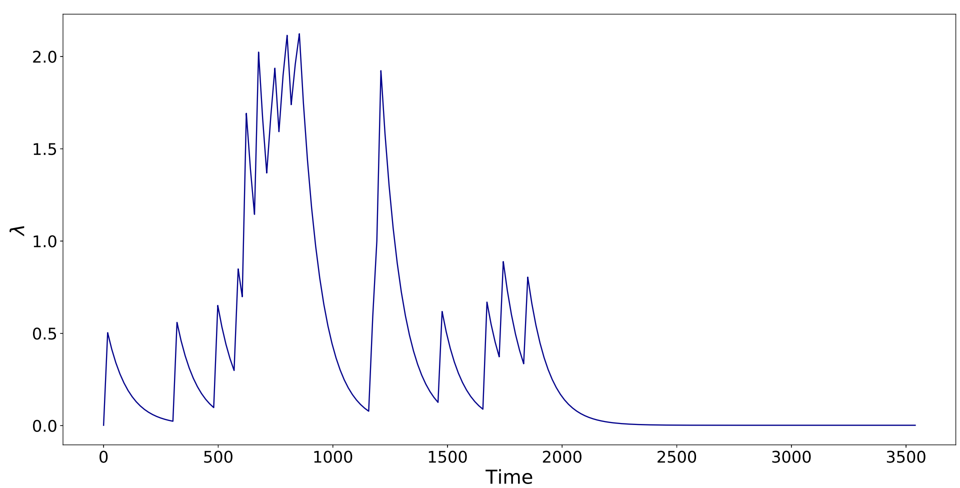

Figure 5.16: Conditional Intensity λ of Taliban Attacks in Pakistan

5.4.3 Al Qaeda

Relying on the concept of "Al Qaeda Network", as done in the other chapters of the work, also in this section a single category for the Al Qaeda group has been created in the first phase, merging together all the smaller groups and fraction that constitute the network (except for Al Shabaab which constitutes a group *per se*). The most targeted countries are Yemen (589) and Iraq (581). The first attack in Yemen has been carried out in 2005, while Al Qaeda appeared in Iraq in 2004. In Yemen, *Al Qaeda in Yemen* and *Al Qaeda in the Arabian Peninsula* are the two organizations responsible for all the attacks, in collaboration with other terrorist groups, as the *Adan-Abyan Province of the Islamic State*. In Iraq, Al Qaeda has been present through the attacks of *Al Qaeda in Iraq* and *Al Qaeda Kurdish Battalions*. In Yemen, from 2005 to 2016, attacks occurred in 461 unique days, while in Iraq 202. Shifting the focus on the most popular targets, Al Qaeda in Yemen mostly hit Government (General) buildings and/or personalities (24.62% of the events), while in Iraq the organization mainly attacked Private Citizens and Property (40.1%). Besides descriptive statistics, the results of the Hawkes models are reported in Table 5.8.

Parameters	**Yemen**	**Iraq**
μ	0.032	0.024
k	0.477	0.687
ω	0.039	3.957
KS Test Stat	0.075*	0.067*
KS 95% Sig. Level	0.112	0.089
KS 99% Sig. Level	0.135	0.107
Hawkes AIC	1076.844†	552.252†
H. Poisson AIC	1091.252	1626.547

Table 5.8: Univariate Hawkes Estimates for Al Qaeda Models (Yemen and Iraq). † Indicates which Model Between Hawkes and H. Poisson better Explains the Process. * Indicates 95% significance of the KS Statistic, ** Indicates 99%.

For both countries, the models indicate that a self-exciting component is present in the data under analysis. Indeed, AIC values for Yemen and Iraq are lower for Hawkes models compared to Homogeneous Poisson. Furthermore, also the KS tests are both significant at 95%, thus suggesting that the Hawkes specification is capable of capturing the mechanics of terrorist attacks against Government in Yemen and Private Citizens and Property in Iraq. The goodness derived from the KS test is visually represented in Figure 5.17 and Figure 5.18. Additionally, Figures 5.19 and

5.20 show that in both countries the inter-arrival times are dispersed following an exponential distribution, further demonstrating the goodness of the Hawkes models. The exponential decay is steeper in the Iraq case, with the majority of inter-arrival times being between 0 and 1 day, while in Yemen the exponent is smoother.

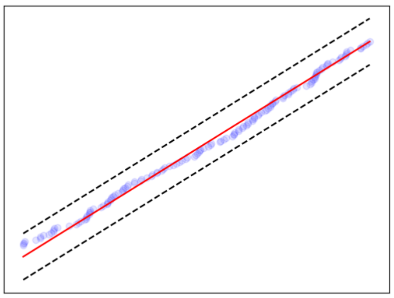

Figure 5.17: Al Qaeda KS Plot - Yemen

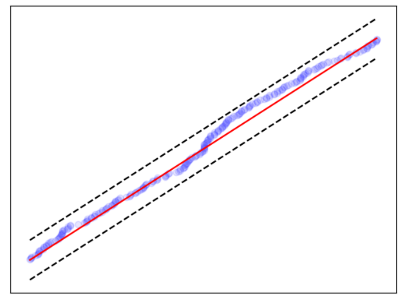

Figure 5.18: Al Qaeda KS Plot - Iraq

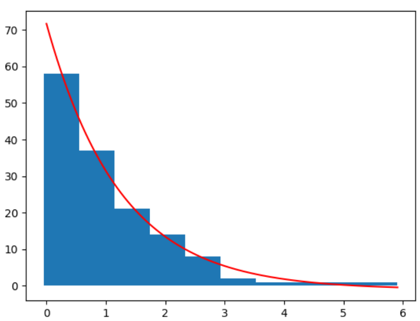

Figure 5.19: Al Qaeda Inter-arrival Times - Yemen

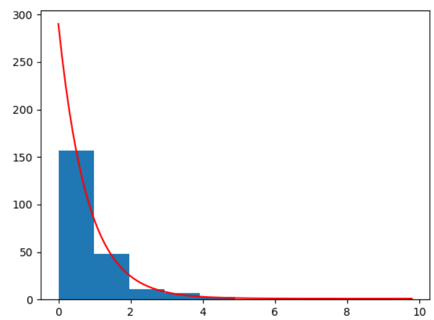

Figure 5.20: Al Qaeda Inter-arrival Times - Iraq

Learned parameters provide additional information on the dynamics of terrorist events plotted by Al Qaeda in Yemen and Syria against Government and Private Citizens respectively. The average expected number of attacks per day (μ) is higher in Yemen (0.032 against 0.024), while k suggests that the process is more reactive in Iraq. Finally, the inverse of the estimated ω parameters are extremely different. In Yemen, the average number of days in which self-excitability persists is 25, while in Iraq it lasts for less than a day. This, also considering the graphical depiction in Figure 5.22 and Figure 5.24, indicates that self-excitability is higher in Yemen.

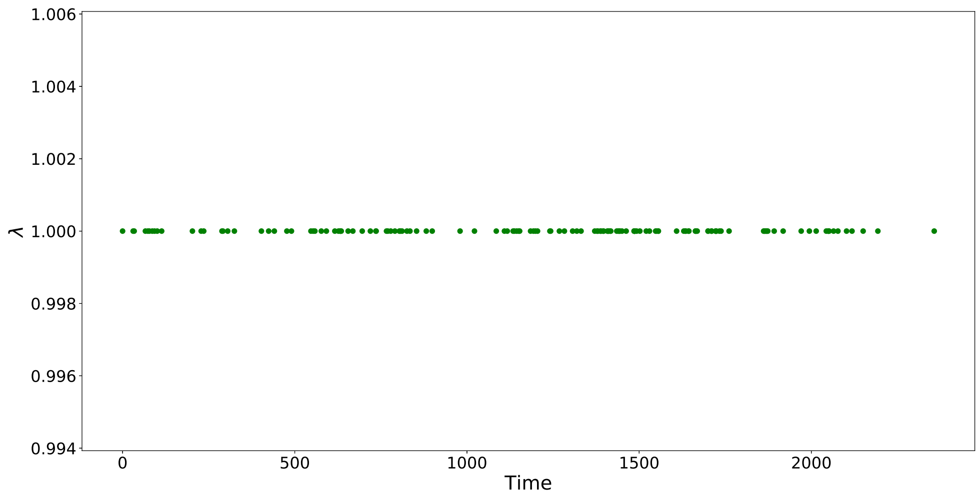

Figure 5.21: Event Stream of Al Qaeda in Yemen

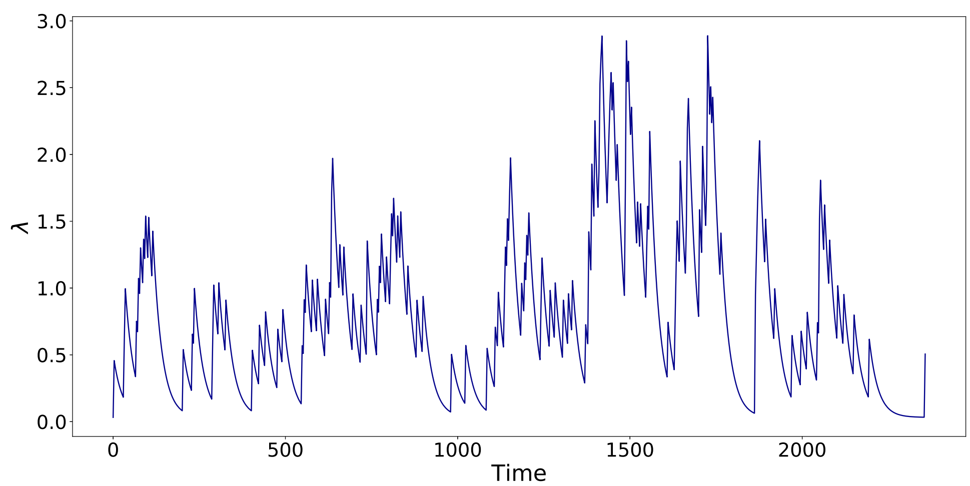

Figure 5.22: Conditional Intensity λ of Al Qaeda Attacks in Yemen

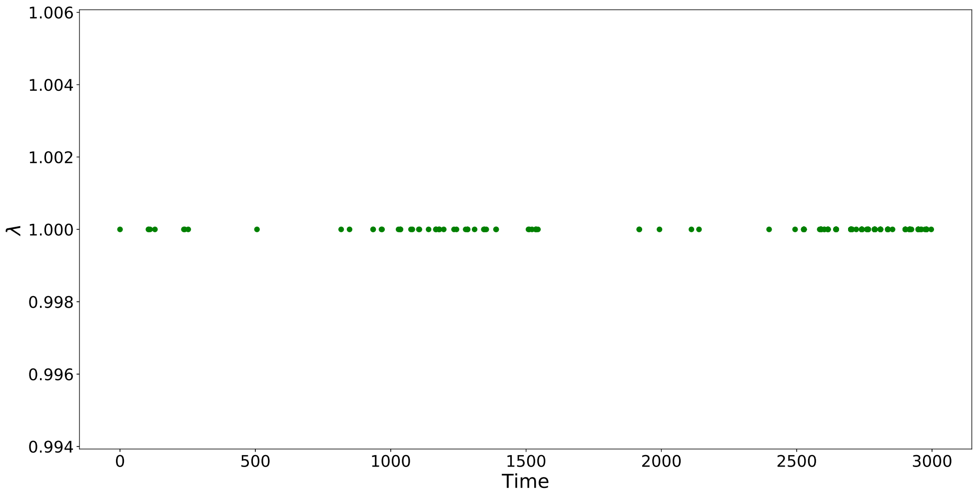

Figure 5.23: Event Stream of Al Qaeda in Iraq

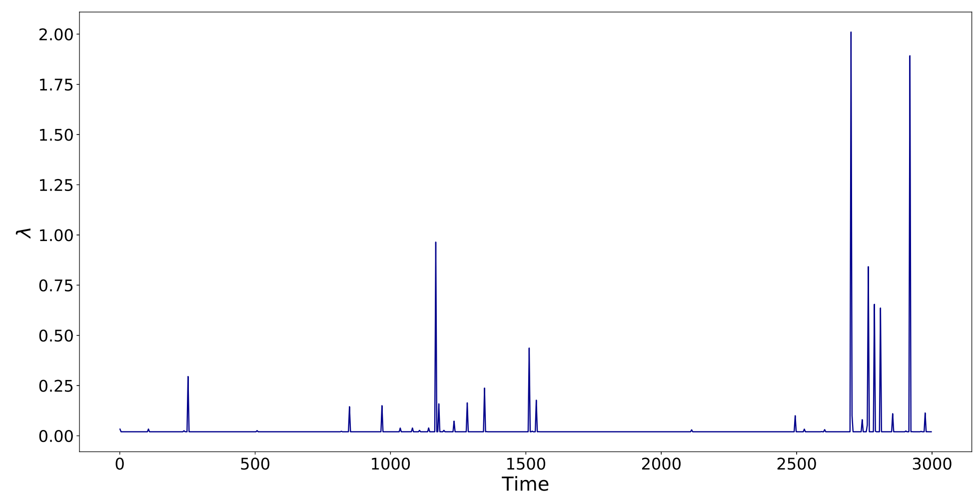

Figure 5.24: Conditional Intensity λ of Al Qaeda Attacks in Iraq

5.4.4 Boko Haram

The jihadist group Boko Haram's first attack included in the GTD dates back to 2009. Since then, six countries have been targeted: Nigeria and Cameroon, with respectively 1,904 and 200 terror events, are the most hit countries. Nigeria is the country where the group was born and accounts for the vast majority of attacks perpetrated by Boko Haram, with 982 days with at least one event between 2009 and 2016. The adjoining country Cameroon has been certainly affected by a smaller number of terrorist attacks, though having experienced them for 148 unique days from 2013 to 2016. As for most of the other groups, also Boko Haram showed a preference towards Private Citizens and Property as targets. In Nigeria and Cameroon, 49.68% and 65.5% of attacks respectively were carried out against this category. The results of the models are reported in Table 5.9.

Parameters	**Nigeria**	**Cameroon**
μ	0.037	0.075
k	0.883	0.252
ω	0.136	2.343
KS Test Stat	0.093	0.077*
KS 95% Sig. Level	0.044	0.118
KS 99% Sig. Level	0.052	0.142
Hawkes AIC	3327.452†	798.694†
H. Poisson AIC	3857.419	848.176

Table 5.9: Univariate Hawkes Estimates for Boko Haram Models (Nigeria and Cameroon). † Indicates which Model Between Hawkes and H. Poisson better Explains the Process. * Indicates 95% significance of the KS Statistic, ** Indicates 99%.

While for both models the AIC statistic suggests that the Hawkes model provides a better fit compared against a homogeneous Poisson model, the Nigeria model fails the KS test (the Cameroon model, instead, is accepted with 95% confidence). Inspecting the KS plot for Nigeria (Figure 5.25, the points deviates from the 95% confidence boundaries for a large part of the graph (it is worth to recall that a perfect fit would imply a perfect overlap on the solid red line that has an inclination of 45). The KS test is known to be very demanding (Lallouache and Challet, 2014), but there might be different explanations regarding the negative result for the Nigerian case. The most probable is that the models should be fitted using a different type of decay kernel (e.g. Power-law, Rayleigh). Another concurrent explanation is the presence of non-stationarity in the data.

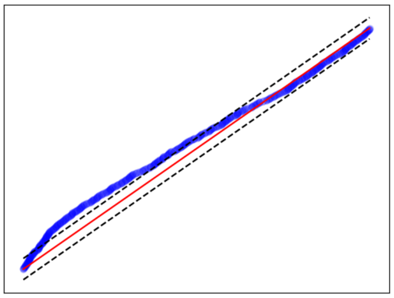

Figure 5.25: Boko Haram KS Plot - Nigeria

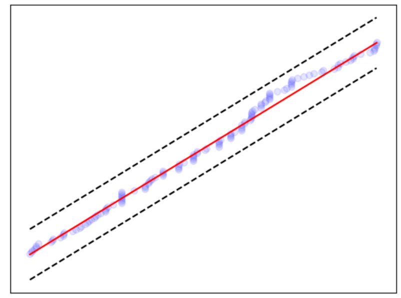

Figure 5.26: Boko Haram KS Plot - Cameroon

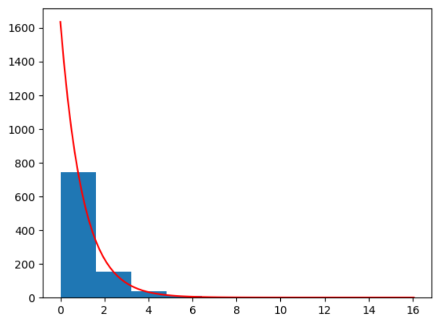

Figure 5.27: Boko Haram Inter-arrival Times - Nigeria

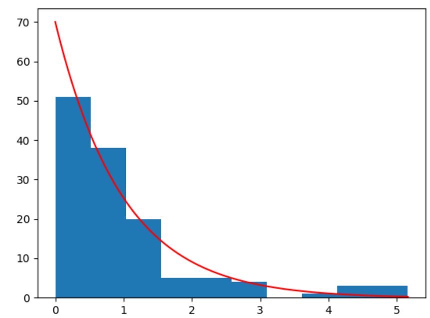

Figure 5.28: Boko Haram Inter-arrival Times - Cameroon

In spite of the limitations shown by the models, in both cases at least a certain degree of evidence is provided for the presence of memory and self-excitability (considering the AIC tests). In the considered time spans (which are different), μ parameter is higher for Cameroon than for Nigeria, while the jump factor k highlights the much higher reactivity of the self-excitation for the Nigerian model. Finally, ω^{-1} values extremely differ in the two distinct scenarios: the length of self-excitation windows lasts around 7 days in Nigeria, while in Cameroon the duration is less than a day (0.41), showing the very different nature of the two processes.

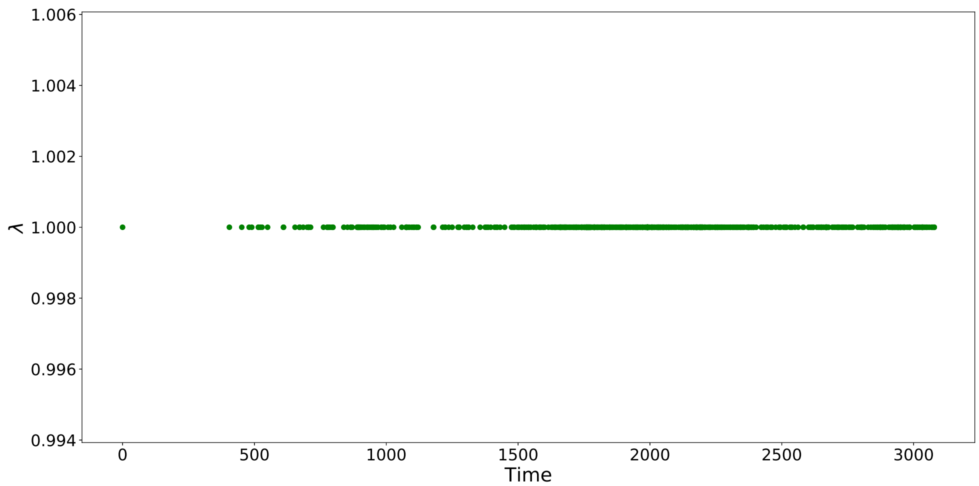

Figure 5.29: Event Stream of Boko Haram in Nigeria

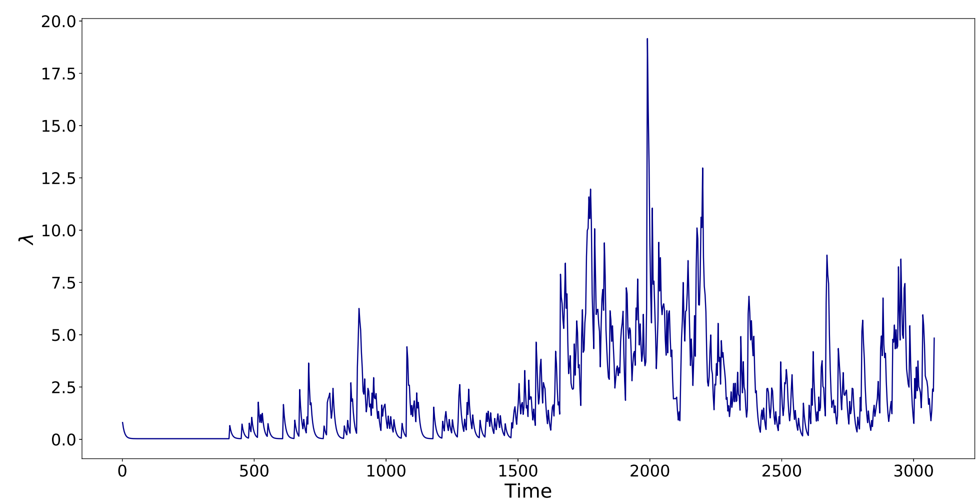

Figure 5.30: Conditional Intensity λ of Boko Haram Attacks in Nigeria

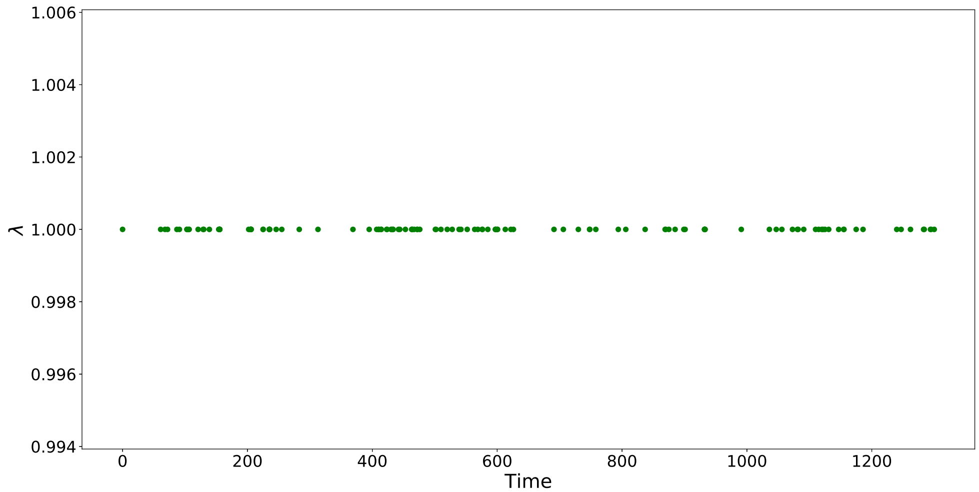

Figure 5.31: Event Stream of Boko Haram in Cameroon

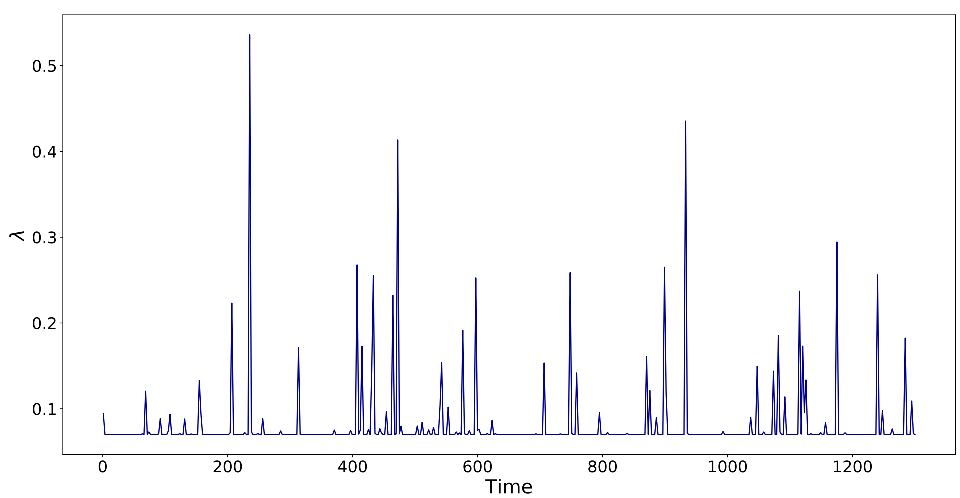

Figure 5.32: Conditional Intensity λ of Boko Haram Attacks in Cameroon

5.4.5 Al Shabaab

Back in 2007, the first attack claimed by Al Shabaab hit Somalia, which is indeed the country of origin of the group and the most targeted one. After Somalia comes to Kenya (that experienced the first attack only a few months after the first Al Shabaab attack overall). Until 2016, Somalia has been hit by 1,401 attacks (in 963 unique days), while Kenya accounted for a total of 289 (in 228 unique days). Concerning targets, Al Shabaab mostly hit Private Citizens and Property in Somalia (30.48%) and, similarly to the Taliban in Afghanistan, Police in Kenya (34.26% out of the total number of events). The results of the two separate Hawkes models are showcased in Table 5.10.

Parameters	**Somalia**	**Kenya**
μ	0.023	0.004
k	0.861	0.982
ω	0.041	0.005
KS Test Stat	0.479	0.134*
KS 95% Sig. Level	0.061	0.136
KS 99% Sig. Level	0.074	0.163
Hawkes AIC	2501.592†	806.725†
H. Poisson AIC	2781.596	820.594

Table 5.10: Univariate Hawkes Estimates for Al Shabaab Models (Somalia and Kenya). † Indicates which Model Between Hawkes and H. Poisson better Explains the Process. * Indicates 95% significance of the KS Statistic, ** Indicates 99%.

The models of Al Shabaab provide distinct outcomes. In the Somalian case, in spite of the AIC statistic being preferable for the Hawkes case compared to the Homogeneous Poisson, the KS test is largely failed. Visually, this is testified by the KS plot in Figure 5.33. The points largely deviate from the significance boundaries. Again, this can be explained by a wrong choice in the type of decay kernel being non-exponential in the natural representation of the data. These results partially confirm some of the results found in Tench (2018), where a considerable number of models, although with different data and methodology, did not pass the KS test, yet performing better than the Poisson baseline case.

Nonetheless, for what concerns Kenya, the Hawkes better captures the dynamics found in the data: the KS statistic is significant at 95% confidence level, and the distribution of inter-arrival times fits well an exponential distribution (while this was not the case for the Somalian case, as shown in Figures 5.35 and 5.36).

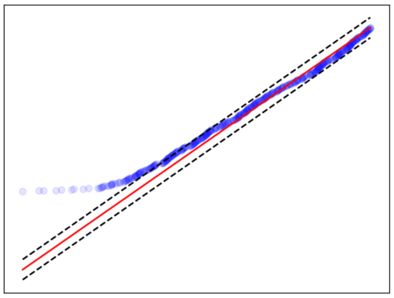

Figure 5.33: Al Shabaab KS Plot - Somalia

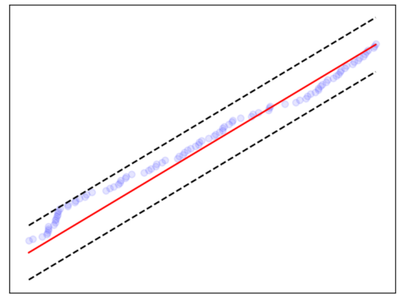

Figure 5.34: Al Shabaab KS Plot - Kenya

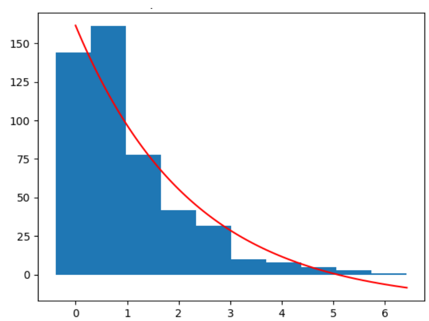

Figure 5.35: Al Shabaab Inter-arrival Times - Somalia

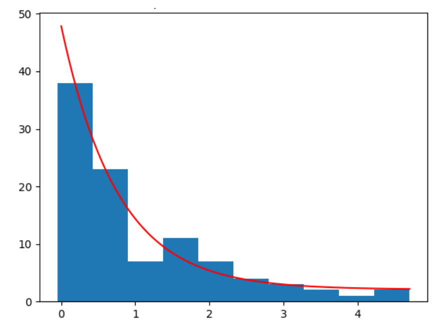

Figure 5.36: Al Shabaab Inter-arrival Times - Kenya

Inspecting the parameters, the average expected number μ is way higher for Somalia, testifying higher frequency in the number of attacks (which was indeed foreseeable given the large difference in the absolute number of attacks and the almost similar duration of the period under analysis). However, the k parameter shows that the process is much more reactive in the Kenyan case. Yet, both values are very high and quite close to the bounded limit of 1 (set to avoid the process being explosive). Finally, the average number of days over which a series of self-exciting attacks last is also quite different for the two scenarios: 24.39 days in Somalia, while 200 days for the Kenyan case. This last result might be influenced by the relatively low frequency of attacks in the first three years included in the analysis.

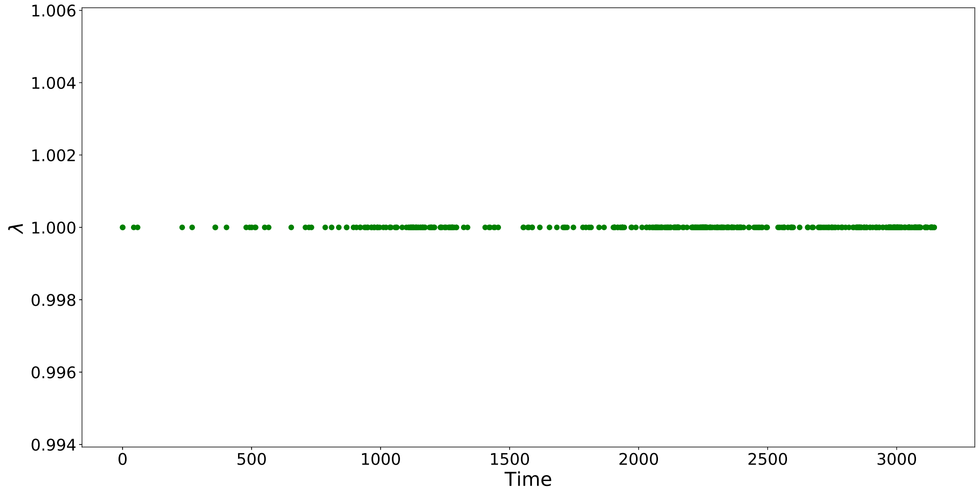

Figure 5.37: Event Stream of Al Shabaab in Somalia

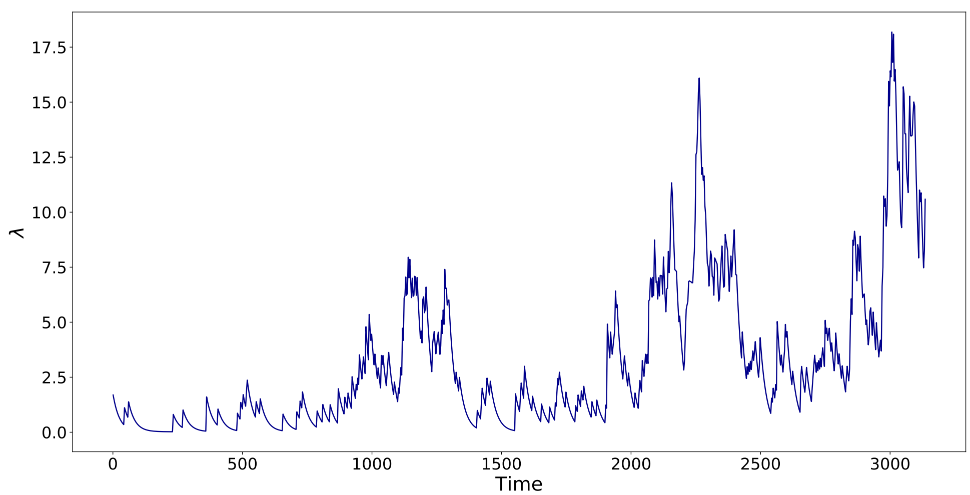

Figure 5.38: Conditional Intensity λ of Al Shabaab Attacks in Somalia

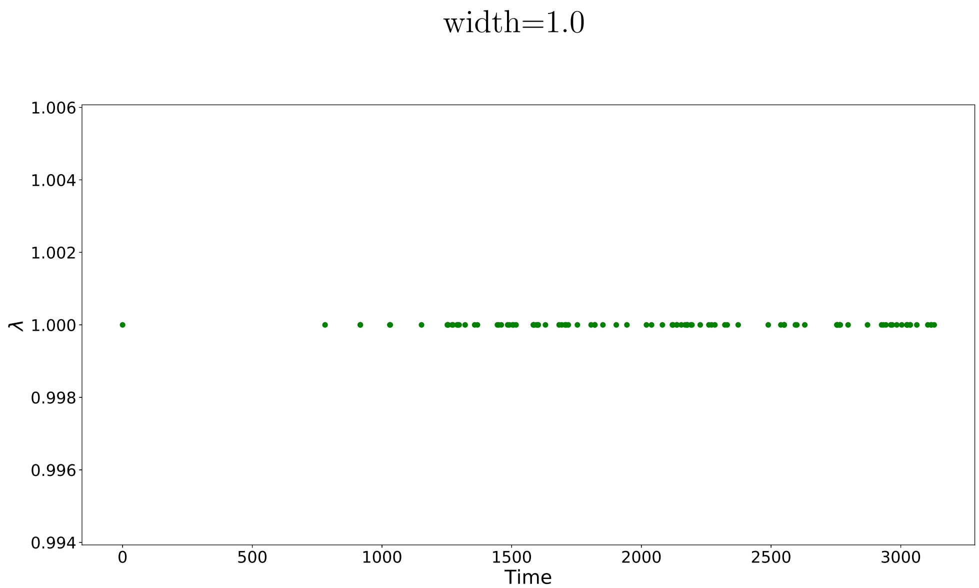

Figure 5.39: Event Stream of Al Shabaab in Kenya

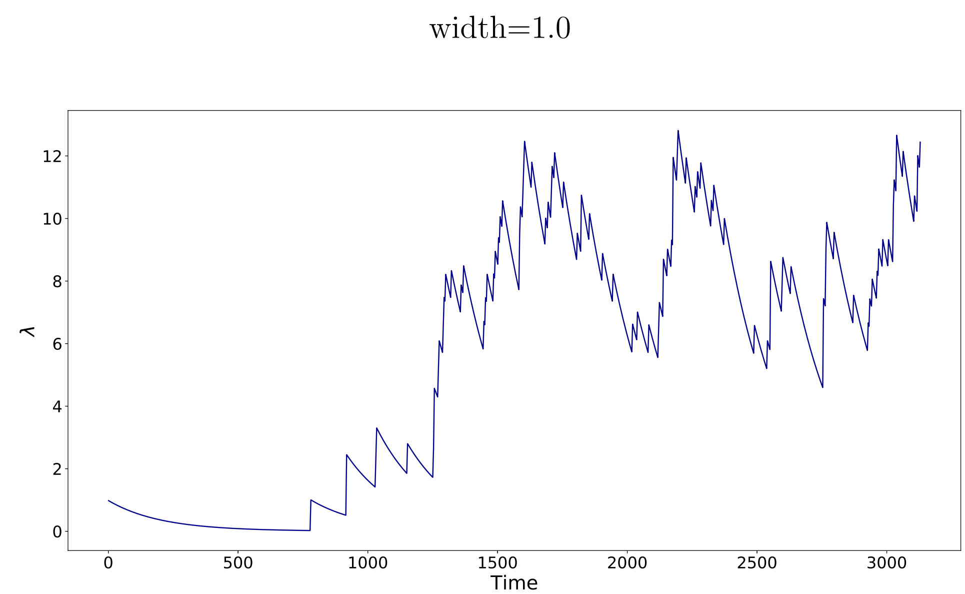

Figure 5.40: Conditional Intensity λ of Al Shabaab Attacks in Kenya

5.4.6 Summary of the Results

To synthesize the outcomes of the ten different models (two per each terrorist group), Table 5.11 provides a summary of the results beyond parameter estimation and interpretation, to assess whether this proposed modeling technique has been suitable for the data at my disposal.

Group	**Country/Target 1**		**Country/Target 2**	
	AIC selection	**KS Test**	**AIC selection**	**KS Test**
Islamic State	Hawkes	Yes	Hawkes	Yes
Taliban	Hawkes	No	Poisson	Yes
Al Qaeda	Hawkes	Yes	Hawkes	Yes
Boko Haram	Hawkes	No	Hawkes	Yes
Al Shabaab	Hawkes	No	Hawkes	Yes

Table 5.11: Summary of Results for Model Selection and Goodness of Fit

The results indicate that, in nine cases out of ten, Hawkes point processes seem to better capture the inherent nature of the data. The only case in which the Homogeneous Poisson process (considered as the null baseline alternative) has a better AIC statistic is the model on Taliban attacks in Pakistan against Private Citizens and Property. As explained in the dedicated subsection, this can be due to the very low number of attacks considered (23, making it the model with the lowest number of points under analysis) and its distribution over time, which is particularly sparse over a long period.

When focusing on KS tests, three models have been problematic, namely the Taliban in Afghanistan, Boko Haram in Nigeria and Al Shabaab in Somalia, while the other seven passed the test with a 95% level of significance. It is worth to note that the models which failed the test are all referring to sequences with a relatively high number of events (respectively 2184 for Taliban/Afghanistan, 946 for Boko Haram/Nigeria and 485 for Somalia). As anticipated in the chapter, this can be driven by the decay function not being exponential avoiding, therefore, an acceptable fit for the models. Different solutions can be tested to better investigate the causes and motivations behind these results, and some of them will be proposed in the last section. However, the summary of results provides encouraging evidence that, in the vast majority of the cases, the processes exhibit memory-dynamics, suggesting once again how jihadist attacks do not happen at random but, instead, are clustered in time.

5.5 Discussion and Future Work

Hawkes processes are a particular type of stochastic processes that have gained wide success in many different research areas. Named after Alan G. Hawkes, these types of processes have been applied in the analysis of earthquakes, financial markets, brain activity and, lately, crime and political violence. Indeed, in the last years, several scholars have modeled the spatio-temporal dynamics of either gang violence or terrorism exploiting the properties of this particular point process. The success of Hawkes processes resides in the capability of capturing the behaviors of many real-world phenomena. Earthquakes, stock selling and buying, bank defaults, crimes, and terrorist attacks naturally cluster in time and space and, moreover, events are self-exciting, meaning that the occurrence of an earthquake is likely to have a positive impact on the probability of the occurrence of an aftershock in the near future: Hawkes processes are exactly created to capture these dynamics.

Building upon the recent relevant scientific production at the intersection between criminology and statistical modeling, in this chapter I have applied Hawkes processes to test the presence of memory-dynamics and self-excitability in the data. Given that Hawkes processes are a non-Markovian extension of Poisson processes (Laub et al., 2015), they exhibit memory, which is a fundamental concept for the entire architecture of my dissertation. Going beyond the existing literature, instead of simply analyzing terrorist events, I have focused on a subset of events per each group, considering only the two most attacked countries and the most popular targets for each group.

This choice is motivated by three factors. First, the desire to investigate dynamics at a higher level of detail, discriminating events based on their characteristics, since, as testified by the results of the previous chapter, jihadism can be extremely heterogeneous in its behavior. Indeed, while a process of pure events can exhibit memory dynamics, when focusing on event characteristics via data disaggregation, these dynamics may vanish. Second, testing Hawkes processes on sequences of events related to specific targets is not only relevant for the structure of this research work but can also provide further indications in terms of counter-terrorism policies. Discriminating between attacks of different nature (or magnitude) can be of help in determining different strategies of risk mitigation, providing guidance on resource allocation for counter-terrorism campaigns. Third, the necessity to conceptually connect this chapter with the previous one on trails and transition networks and, foremost, with the last one on deep neural networks, relying on the concept of “memory”. While these analyses have given insights on the behavior of jihadist groups, still many improvements can be added. There are several pathways for future work, starting from the outcomes of this chapter.

First, to solve the issues found in the problematic models, it would be interesting to explore Hawkes models fitted with alternative kernels. Theoretical and applied research has been done (especially in finance, see for instance Hardiman et al. (2013) and Zhang (2016)) as long as programming frameworks and packages have been developed (Xu, 2018) to take into account different potential modeling specifications of kernel functions and it will thus be interesting to evaluate if these developments can contribute to a better explanation of these data.

Second, integrating temporal information on events with further information could enrich the models and results. To do so, Marked Hawkes processes should be deployed. This particular class of Hawkes processes allows associating to each event a particular feature (for a detailed theoretical and mathematical explanation, see for example Daley and Vere-Jones (2006)). They have been used in several areas, including information diffusion (Chen and Tan, 2018), finance (Lee and Seo, 2017), energy conservation research (Li and Zha, 2015) and also crime (Mohler, 2014). Their statistical characteristics would enable us to discriminate between different types of events with the same characteristics (or belonging to the same filtered subset, e.g., the subset of attacks directed against a particular target) but with different magnitude, consequences and impact. Associating a so-called "mark" to each event to capture the damages that the attack has provoked (in terms of human losses, for instance) could help in addressing further research questions, trying to investigate whether high-impact attacks can lead to higher self-excitability, or not. Third, investigating the relation (possibly also in terms of causality) between multiple point processes would be crucial in depicting the nature of more complex dynamics. Tench et al. (2016) have already tested a multidimensional model taking into account terrorist and counter-terrorist activities. For the data at my disposal, different options could be viable. One option could be investigating the dynamics between attacks in adjoining countries or, even better, provinces. Another option, instead, would be the analysis of multidimensional processes of attacks directed against different types of targets. Multivariate Hawkes processes can better explain and capture the natural dynamics of the data: however, they can also reveal causality. Many works have developed methods to infer causality (in different terms) from multidimensional Hawkes processes (Xu et al., 2016; Etesami et al., 2016; Achab et al., 2017): being able of discovering causal structures in terrorism domain would represent a crucial advancement both in terms of research and policy.

This page intentionally left blank

6 Deep Learning and Terrorism: Long Short-Term Memory Networks for Jihadist Target Forecasting

6.1 Introduction

Network science in terrorism research has been mostly applied to map the relations between individuals belonging to the same terrorist group, to highlight their organizational structure and eventually provide suggestions on how to disrupt them (Krebs, 2002; Farley, 2003; Adler, 2007; Moon and Carley, 2007; Keller et al., 2010; Gerdes, 2015; Malm et al., 2016). This work, conversely, does not deal with networks of individuals: ironically, individuals are not even included among the several entities that will be considered. Indeed, the concept of meta-networks of terrorism which has been fundamental also in Chapter 4, will constitute the cornerstone of this chapter too. In this case, meta-networks will consider different dimensions of terrorist attacks and map relations among these dimensions and consider the evolution of these relations over-time. Specifically, given the five groups under analysis, each jihadist organization will be associated with a meta-network, mapping the history of its attacks during its existence in this meta-relational way. The dimensions of the meta-networks will be locations (intended as countries in which the group operates, i.e., where it plots attacks), employed weapons, tactics and attacked targets. My intuition is that, by using the information retrieved from all the events perpetrated by the jihadist groups, it will be possible to extract deep knowledge on how these terrorists behave and even predict and forecast their future actions. Indeed, while research on terrorist events, both at explanatory and predictive level, is of no certain novelty (meaning that research on terrorist events is not innovative *per se*), the innovative element here is represented by the way in which attacks are conceptualized and treated.

Classic research on terrorist events has treated attacks as independent of one another (with some exceptions if attacks were physically connected, as in the case of 9/11). However, terrorist events may be connected in different ways. Besides pure physical relations, namely attacks that were plotted together as part of a wider strategy, as again in the case of 9/11, events can be related by hidden connections that can be conceptualized in a network-based fashion. To exemplify, we can think about the case of three hypothetical terrorist attacks: A, B, C. Attack A and C share the same target type, attack B and C share the same weapon type, attack A and B share the same location. These are all abstract relations that may be gathered from the original information set: having extensive data of this type for long sequences of attacks allow to re-create these hybrid meta-networks (hybrid because they map relations that are not only of different type but also of different intrinsic nature: physical versus abstract), and these hybrid meta-networks can monitor and possibly detect patterns that would remain hidden if information were treated in the classic way.

The concept of mutual dependency and the multiple inter-relations that this approach allows to consider are essential to grasp a deeper understanding of terrorism as a complex social phenomenon. It is a conceptual mistake to think that terror events are not connected even at this "deeper" level. In a dynamic meta-network environment, recurring regularities are easily detectable, for instance. Additionally, networks allow controlling for anomalies, helping to answer questions such as "how close an organization is to change its behavior?" or "how diverse is a group's terror strategy?". These are all information that can be crucial for research and policy purposes. Furthermore, this approach helps in extracting new knowledge from existing data that have been already analyzed and unfolded in many ways by scholars.

The fundamental seed of this part of the work starts from a single question. This question is: *"Does terrorism have memory?"*. As it is posed the question may cause confusion in the reader. The previous chapter already demonstrates that memory, intended as the existing time-dependent structure between events, is found in jihadist dynamics. This finding is in line with previous literature on the topic.

The proven fact that terrorism does not happen randomly is already an important step towards a better understanding of the phenomenon. In fact, knowing that attacks are clustered over time is relevant to the intelligence community to design counter-plans or response activities after major incidents. However, still a lot is missing. This is why the question has to be posed again. Besides the fact that events are clustered in time and are not independent one of another, we should investigate if some specific dynamics show patterns and schemes which can be related to a multidimensional concept of "memory". The idea is that, besides the analysis of mere events, we shall focus on specific dimensions of attacks (i.e. countries, tactics, weapons, and targets, in this case) to understand if the analysis of how these features are distributed over time follow specific interrelated pathways and properties.

Memory is hence defined as the situation in which, given a considerably long sequence of events, events hold a specific interconnected temporal structure that can be learned by a model and can be useful in predicting the future. If terrorism has a memory, and, besides events, features hold specific interconnected behaviors, then we can employ models that have some "memory" dimensions to make predictions about the future. The question is relevant, and if it does not appear to be so, it is my fault because it means I was not clear enough in explaining the whole reasoning. For the sake of simplicity, the reader may just think about the concept of memory as a deeper extension of the concept of spatio-temporal concentration of terrorist events. "Deeper" means that memory will consider multiple dimensions of terrorism assuming non-randomness across them, thus distinguishing events based on their characteristics instead of treating them as all equal.

The whole investigation poses further interesting theoretical questions. If memory existed, understanding why it exists would be of indisputable value. Does it exist because terrorists are rational agents? Does it exist because each terror group has some specific expertise and tends to demonstrate it through the repetition of certain actions over time? I will not directly answer these questions, but the reader shall keep them in mind as it advances in the chapter, because the dissertation will indirectly look at them as potential future work.

The investigation on memory and interdependence will be directed towards the integration of network science and deep learning for forecasting future terrorist targets.

The chapter develops as follows: the next section will be dedicated to the background of the analyses, presenting related work and explaining why it is relevant to address the problem of forecasting terrorist targets.

The third section will thoroughly describe the methodological framework, starting from the concept of "dynamic meta-network" up to the description of Long Short-Term Memory network, a class of deep learning algorithms designed for handling sequence and time-series data.

The fourth section will investigate the properties of the graph-derived time series of jihadist groups, specifically dealing with stationarity, randomness and temporal dynamics of targets hit by the jihadist groups.

The fifth section will present the results of the model in detail, also providing a summary of the performance of the deep learning models.[1]

The sixth section will deal with the problems of weak and rare signals in terrorism research, trying to propose a potential solution to avoid the risk of missing crucial events when applying computational methods for prediction and forecasting.

Finally, the seventh section will discuss the main results and implications of the study, also outlining potential research directions for the future.

6.2 Background

6.2.1 Related Work

Humans have been fascinated with the idea of predicting and forecasting the future for centuries. In Ancient Rome, haruspices[2] could supposedly give instructions about future events, gathering information from the entrails of sacrificed animals. A long time has passed since then: during the centuries, divination has been replaced by more empirical and scientific methods, but the attention and efforts towards the prediction of the future have even increased.

Nowadays, the science of prediction covers almost every academic and scientific discipline. What particularly strikes scientists is the idea of forecasting human behavior, as a way to better understand how individuals think, act and make decisions in a wide set of distinct realms (Pentland and Liu, 1999; Armstrong, 2001; Subrahmanian and Kumar, 2017). Hundreds of research groups around the world are working every day to disentangle and illuminate the mysterious nature of humans, with an eye directed to the future, focusing on specific aspects and contexts. Scholars have tried to predict human behavior in terms of political voting (Lewis-Beck and Rice, 1984; Kou and Sobel, 2004; Fowler and Dawes, 2008), consumer choices (Goel et al., 2010), social media activity (Ruths and Pfeffer, 2014) and health conditions (De Choudhury et al., 2013; Choudhury et al., 2013).

[1] Further details on model results can be found in Annex B.

[2] An haruspex was a person trained to practice the divination activity of “haruspicy”. The concept derives from the Etruscan religion. Forms of divination have been found even earlier in history, as in the case of Babylonians.

In this frame, as already noted in this dissertation, the increasing access to large datasets and the progress made in mathematical and statistical modeling have played a central role in the growing interest of the scientific community towards the investigation of future human dynamics. Specifically, two methodological areas (that are getting closer every day) have gained popularity and demonstrated their potential in the effort to better predict what humans (or communities made by humans) will likely do in the future: network science (Börner et al., 2007; Barabási, 2011) and artificial intelligence (Russell and Norvig, 2010; Nilsson, 2014).

However, mankind not only chooses between Republicans and Democrats, Socialists and Conservatives, not only purchase clothes and book hotels online, and not only post holiday pictures, romantic songs and newspaper articles on their social media accounts. In fact, humans also commit crimes. Under the word "crime" resides a tremendously heterogeneous world of actions that span from very low levels of severity (e.g., traffic misdemeanors) to atrocious forms of violence (e.g., genocides). It is not within the scope of this work to dive into conceptual and theoretical discussions regarding the limited generalizability of the definition of certain crimes, and into the dependencies between political and social contexts and the subsequent inclusion or exclusion of certain acts into the set of criminal activities. It is instead worth to reason about the implications that this extreme heterogeneity of actions have for empirical and quantitative research: the higher the complexity, the higher the difficulty to extract patterns of common behaviors, the higher the need for research to investigate the criminal behavior of members of the humankind. Also in this case, network science and artificial intelligence have been employed - with various degrees of intensity - as precious methodological and technical frames. The beyond-research implications of crime prediction are intrinsically related to the pragmatic importance of providing policy-makers with tools or instructions to eventually prevent and reduce crime. Studying how crime occurs naturally leads to the attempt to anticipate it as much as possible to design and deploy efficient strategies for reducing the real and perceived insecurity of human societies.

For what concerns the study of crime, network science has on one side helped in better understanding how criminals interact with each other, merging together the flexibility of networks as mathematical representations of reality and well-established criminological theoretical frameworks (Papachristos, 2014). In organized crime studies, for instance, social network analysis has proved to be able to highlight how criminal groups work and are structured (Mainas, 2012; Calderoni et al., 2014; Smith and Papachristos, 2016; Calderoni et al., 2017) and as a tool for assisting intelligence analysts

in tasks such as link prediction when information is noisy or incomplete (Berlusconi et al., 2016).
On the other side, scholars in the last decades have tested the potential benefits of statistical modeling and machine learning for predicting - among the others - future crime locations, time, characteristics, recidivism risk (Nagin and Tremblay, 2005; Weisburd et al., 2009; Neuilly et al., 2011; Favarin, 2018). Beyond classical quantitative methods, statistical and machine learning, two interrelated dimensions within the broader field of Artificial Intelligence, have achieved a growing popularity also due to the public debate that has spread from the United States regarding the use - and misuse - of mathematical and computer models for predictive policing purposes and criminal justice risk assessment models (Shapiro, 2017; Berk, 2019). The practical application of such models and the flaws detected within them (as low fairness, bias, feedback loops) have called the scientific community not only to invest in the development and deployment of sophisticated methods for predicting and forecasting criminal activities, but even to extensively reason about the future perspectives posed by data and algorithms for criminology (Berk, 2008; Brennan and Oliver, 2013; Ozkan, 2019) and the ethical, legal, and political consequences of corrupted predictive systems for human society itself (Saunders et al., 2016; Yeung, 2018; Hannah-Moffat, 2018; Berk et al., 2018; PAI, 2019). While, on one hand, researchers have focused on the implications (both in terms of potential and threats) posed by the novel applications of these methods within the realm of criminology, on the other hand scientists (especially coming from fields as statistics, computer science, physics and mathematics) have either developed or tested new algorithms for the study of crime or experimented the use of new types of data gathered from the digital footprints that every human leaves every day on the internet.
With regard to the former aspect, machine and deep learning have experienced a rise in their popularity and applicability on several problems as recidivism prediction and spatio-temporal modeling (Kang and Kang, 2017; Zeng et al., 2017; Wang et al., 2017; Aglietti et al., 2018; Stec and Klabjan, 2018; Marchant et al., 2018; Stalidis et al., 2018; Huang et al., 2018; Balocchi and Jensen, 2019). For what concerns the latter, instead, scientists have started to use massive information gathered from unconventional sources as mobile tools, social media activity and even satellite images for predicting or forecasting future crimes (Wang et al., 2012; Bogomolov et al., 2014; Chen et al., 2015; Najjar et al., 2018).
As a specific type of crime, terrorism is no exception in reference to the attempts of the scientific community to predict or forecast it. All the works that are being pub-

lished today adjacent to this topic can be somehow traced back to the seminal works of Lewis Fry Richardson, an English mathematician, psychologist, and physicist that made invaluable contributions to the study of meteorology and conflicts (Richardson et al., 1960; Richardson, 1960; Hess, 1995). He is indeed considered the father of the mathematical study of conflicts and wars. The statistical attempts to inspect the nature of adversarial actions between countries (that are, trivially, made by humans) still inspires scientists nowadays (Cederman, 2003; Schrodt, 2006; Clauset et al., 2007; Clauset and Woodard, 2013; White, 2013). Conflicts and wars are different from terrorism in the strict sense (in wars there can be terrorism, and terrorism is certainly a conflictual type of behavior, but the phenomena cannot be considered as equivalent) and the academic community has different views with regard to the strengths and the future perspectives of mathematical modeling of conflicts and wars (Ward et al., 2013; Cederman and Weidmann, 2017): nonetheless, the legacy of Richardson's vision has spread over, also inspiring and influencing the quantitative study of terrorism events. The intersection between new data and new methods has then allowed the researcher to concentrate not only on predicting terrorist attacks but also on studying human relations and predicting their nature regarding terrorist violence. Social network analysis in the first phase (Krebs, 2002; Fellman, 2008; Mainas, 2012; Fellman and Wright, 2014; Malm et al., 2016) and consequently more sophisticated approaches from the fields of complex networks and multi-agent systems (Latora and Marchiori, 2004; Moon and Carley, 2007; Keller et al., 2010; Desmarais and Cranmer, 2013; Fellman and Wright, 2014; Campedelli et al., 2019b; Skillicorn et al., 2019) have showed promising directions and highlighted patterns that could only be discovered through the mathematical study of physical and abstract connections between individuals, groups, entities, countries as fundamental components for explaining terrorism.

The contributions that quantitative methods have made for studying and predicting terrorism come from a variety of sub-fields related to computational sciences (Subrahmanian, 2012), and have forced scholars to start to think about the new challenges that counter-terrorism can take, exploiting this revolution (Thuraisingham, 2003; Ganor, 2019). Particularly, in the last years, fragmented attempts have tested the performance of machine and deep learning algorithms on issues related to terrorism. However, less scientific production can be found in comparison with studies that address other types of crimes. Among the methods applied for these purposes are Hidden Markov models, Random Forests, Artificial Neural Networks, Support Vector Machines (Sun et al., 2003; Raghavan et al., 2013; Ding et al., 2017; Kang and Kang, 2017).

In spite of these signs of progress, to the best of my knowledge, still no studies investigate the potential of the integration between network science and artificial intelligence for the study of terrorism. While foundational research has already addressed the problem of learning graph representations through machine or deep learning algorithms (Tian et al., 2014; Henaff et al., 2015; Jain et al., 2015; Cao et al., 2016; Monti et al., 2017; Huang and Carley, 2019), terrorism has not yet been investigated via the exploitation of these two scientific realms for prediction or forecasting purposes. Notably, only a recent study by Liu et al. (2016) addressed the problem of predicting the next location of an attack plotted by the whole set of groups present in the Global Terrorism Database using deep learning architectures that take into account spatio-temporal dimensions. The authors proposed the definition of a Spatial Temporal Recurrent Neural Network (STRNN) as an alternative to classic Recurrent Neural Networks (RNN): STRNN would be able to incorporate time interval information via time-specific transition matrices and geographic transition matrices for mapping distances between locations of attacks. According to the authors, this alternative architecture was necessary to overcome the limitation of RNN in modeling continuous time intervals. Besides the elegant and sophisticated mathematical architecture of the model, and its ability to perform well on this problem, the authors (as typically happens when scholars from other fields dive into the terrorism/crime realm) do not point out how they treated and cleaned the data at their disposal. In fact, they claim to consider all the groups included in the GTD: however, the number is on a scale of thousands, with a vast majority of actors that have plotted either one or very few attacks. This leaves unanswered questions about how this problem has been solved in practice, before running the experiments.

This work, as anticipated in earlier chapters of this dissertations, aims at placing itself in the focal point of integration between network science and deep learning, exploiting my criminological background to set up a reasonable and grounded information space that relies on the assumed existence of memory and interdependence between terrorist events which, in turn, are concepts framed within the theories on the spatio-temporal concentration of crime and violence and the rational decision-making processes of terrorist groups.

6.2.2 Why Terrorist Targets?

Terrorist target selection has been a long-standing feature of interest for scientific research (Sandler and Lapan, 1988; Wilkinson, 1990; Hoffman, 1993; Drake, 1998a; Eyerman, 1998; Clarke and Newman, 2006; Krueger and Laitin, 2008; Asal et al., 2009; Pizam, 2010; Toft et al., 2010; Brandt and Sandler, 2010; Santifort et al., 2013; Hastings and Chan, 2013; Asal and Hastings, 2015; Morris, 2015; Abrahms et al., 2018). Indeed, its importance is related to the fact that shining a light on them can help in designing prevention policies and allocating resources to protect sensible and potential future targets (Clarke and Newman, 2006; Bier et al., 2007). Targets have been studied from different perspectives in the literature. Among the many approaches, Sandler and Lapan (1988) have first relied on a game-theoretic formal model assuming rational behavior of agents to demonstrate that when intelligence sharing is not linked to deterrence coordination between different countries, policies for protecting likely targets become of little help or, even worst, completely useless.

Shifting from formal modeling to data-driven analyses, other scholars have applied Bayesian models to detect dynamics and key-points of terrorist target selection processes (Brandt and Sandler, 2010; Santifort et al., 2013). Brandt and Sandler (2010) employed Bayesian models to detect dynamics and key-points of terrorist target selection processes. They have specifically employed Bayesian Poisson changepoint regression models to investigate how transnational terrorists adjust their selection of targets in response to target hardening. Their study, conducted for attacks that occurred from 1968 to 2007, identified four separate periods and three underlying covariates that can explain this clustering over-time, namely the dominant terrorist influence, countermeasures, and terrorist state-sponsorship considerations.

Santifort et al. (2013), instead, compared diversity in target choice among domestic and transnational terrorism during a 40 years time range. Relying on the Global Terrorism Database, they have used a Bayesian Reversible Jump Markov Chain Monte Carlo model obtaining arrival rate changes in both types of terrorism and evaluating the extent to which their target selection is diverse, also from a temporal standpoint. Diversity was calculated using Herfindahl indexes for different specifications of attacks. One of the main findings of the study was that bombings of private parties have become the preferred target-attack at transnational and domestic levels during the period taken into account, positing an improvement of homeland security resources to counter these dynamics.

Focusing specifically on soft targets (like individuals) in the period 1998-2005, Asal et al. (2009) have detected how ideology, and specifically religion, is the decisive factor driving the choice to turn towards target civilians. Furthermore, they have highlighted how there is no relation between democratic regimes and undefended civilians, contesting a widespread assumption of political science and international relations which posits that regime type is a relevant dimension in explaining terrorism.

Formalizing countries as targets and starting from the assumption of regime type as a determinant of terrorism, Ivanova and Sandler (2006) tested instead whether, based on the regime, the likelihood of chemical, biological, radiological and nuclear incidents increases. In their study, they have found that there seems to be a positive relation between democracies and the risk for these attacks.

This non-exhaustive brief review of the state of the art in terrorism research regarding targets and related dynamics demonstrates that there exists a tiny part of the scientific community which concentrates on this dimension, and most of the contributors point out how research in this regard can be extremely important for real-world applications in counter-terrorism and defense programs. Furthermore, an interesting point is that, besides differences in samples and actual research questions, several studies highlight the existence of certain patterns over time (either cycles or keypoints), this thus suggests that there is a degree of temporal structure in the way in which terrorism occurs.

6.3 Methodological Framework

6.3.1 Dynamic Meta-Networks of Terrorism

This work is founded on a conceptual intuition that I believe has relevance if a researcher's desire is to rethink how terrorism happens and evolves over time. Specifically, this conceptual intuition revolves around the idea of "meta-network". A dynamic meta-network, as defined in this dissertation, is a complex network that is characterized by three main aspects: 1. Multi-modality; 2. Multiplexity; 3. Dynamicity.

Multi-modality (also known as multi-entity) means that the networks are comprised of different types of nodes, representing substantially different entities. Multiplexity means that there are different levels of links, as we can think of a multiplex network as the union of separated simple one-link-type networks. Finally, dynamicity is the element that introduces a temporal and evolutionary dimension, meaning that the meta-network is mapped across different (discrete, in most cases) timestamps that allow the researcher to assess and analyze changes over time within the meta-network itself. In a general way, given a temporal vector of discrete time-stamps $T = \{t_1, t_2, ..., t_i\}$ where to each temporal element are associated $|k + m|$ networks that take the form of mathematical graphs G and each graph can be either one-mode (monopartite) in the form $G = \langle N, E \rangle$ or two-mode (bipartite) in the form $G =$

$\langle U, N, E\rangle$ where $N, U \subset \mathbb{N}$ are two different sets nodes and $E = \{(i, j) : i, j \in N, U\}$, I define a meta-network a meta-network for the time unit t_i as:

$$\mathrm{M}_{ti} =: \bigcup \left[\bigcup_{k=1}^{l} G_l \langle N_l, E_l \rangle , \bigcup_{m=1}^{n} G_l \langle U_l, N_l, E_l \rangle \right] \tag{6.1}$$

therefore, a unified dynamic meta-network for a whole given temporal vector can be defined as:

$$\mathfrak{M}_T =: \bigcup_{t=1}^{T} \mathrm{M}_t \tag{6.2}$$

This brief introduction already highlights how meta-networks are extremely different if compared with common networks used in most SNA research. Usually, networks in SNA studies (also in the fields of criminology and terrorism research) are one-mode (one type of nodes) and not multiplex, with cases in which networks combine two (but not more) types of nodes, mapping a dual entity system, thus generally defined only as $G = \langle N, E\rangle$. Scholars in criminology have not yet exploited the whole potential of dynamic or multi-mode techniques because they have mainly relied on networks solely comprising individuals, and often discretizing these networks through artificial time-stamps is not possible or too subjective. Additionally, networks of individuals in criminology many times rely on judicial files or open-sources, therefore posing problems of certainty of network boundaries and missing information (Berlusconi, 2013; Campana, 2016).

6.3.2 Graph-derived Multivariate Time Series

As reminded many times throughout the work, the intent is to use dynamic meta-networks to feed multiple neural network architectures to predict future likely targets attacked by jihadist groups. It is thus necessary to point out what are the technical steps that will allow performing this task. Moving from the conceptualization of a meta-network, for each terrorist group, we first define:

$$\mathfrak{G}^N := \left\langle (V_1, V_2, ..., V_m), (E_{1,2}, ..., E_{m,n}), \left(W_{E_{1,2}}, ..., W_{E_{m,n}}\right)\right\rangle \tag{6.3}$$

As a multipartite graph (also called manifold) that contains N partitions describing relations between different sets of nodes V_m and V_n. These relations are formalized as edges $E_{m,n}$ that are weighted by $W \in \mathbb{R}_{>0}$. Within this context, a single mode

in the multipartite graph is represented as $G_{m,n} = \langle V_m, V_n, E_{m,n}, W_{E_{m,n}} \rangle$. For each group, the original graph has terrorist events identified as source nodes, while the different partitions are locations of attack (intended as targeted countries), employed weapons, tactics and targets. As already mentioned, each attack may have multiple features within the single partition (specifically, up to three targets and tactics and up to four weapons). Weights, within single modes, are at this point the number of times a single feature is targeted or employed. Additionally, terrorist events $A_1, A_2, ..., A_k$ hold a temporal attribute which represents the day of the attack. Specifically, for each A_i the relation $\forall\ A_i\ \exists\ t \in T = (t_1, ..., t_n)$ is generally verified (where T represents a vector of time units in daily format, t_1 coincides with the first attack (the oldest) attack plotted by the group, while t_n coincides with the day of the last (newest) attack. When the relation is not verified, i.e. the day is missing, I have imputed the modal day for that month.
For the purposes of the analyses, however, I need to restructure the multipartite graphs to aggregate time units not to end up with extremely sparse matrices, at the same time preserving the necessary richness of the dataset without incurring in the "curse of dimensionality". Considering that the terrorist groups in the sample have very different histories, compressing the temporal element in too wide time units would have led to extremely small datasets in some cases (i.e.: the Islamic State, which appeared only in 2013). Moreover, for intelligence and policy purposes it is more useful to create predictive models on restricted time units. This is part of the innovation of this work, which tries to build a model that is not only useful for scientific purposes but might be interesting also from practical uses. In particular, developing three-day-based time series forecasting models seemed a good compromise between detail, computational feasibility, and applied usefulness. Therefore, for each group, I have aggregated the data using three-day time stamps. Doing this I have obtained nested multipartite graphs for each group where source nodes are still events, but nesting is performed using timestamps. This allowed me to create a dynamic multipartite graph for each group. Each feature within each mode associated with each new timestamp is now simply the frequency of occurrence of that feature in the events that in the original graph happened in days belonging to that specific three-day time range. To obtain this, I have operated matrix algebra.
Given that $\mathbf{G}^{a \times n}$ is the weighted adjacency matrix of dimension $M \times N$ that formalizes the two-mode sub-graph $G_{a,n} \in \mathfrak{G}^N$, with a indexing the source node, namely events, I can obtain a one-mode square symmetric matrix $\mathbf{M}^{n \times n}$ where the unique existing mode is given by n, namely the feature of interest among the four considered

via

$$\mathbf{G}^{\mathrm{T}}\mathbf{G} = \mathbf{M} \tag{6.4}$$

where $\mathbf{G}^{\mathrm{T}}$ is the transpose matrix. Through this operation, four one-mode matrices associated with each nested multipartite graph are created. These matrices show the recurrence of each feature within each mode in terms of frequency. Frequencies, however, highly varied across different time units (since terrorist attacks are not equally distributed over time) and especially across groups. I, therefore, had to extract comparable knowledge both for internal and external validity. To achieve this goal, I have computed normalized weighted total degree centrality for each mode. Total degree centrality is the most common node-level metric used in network analysis. For a focal node (feature) i in weighted one-mode matrices at each time unit t, the metric is calculated as:

$$C_D^W(i)_t = \sum_j^N \mathbf{wM}_{i,j} \tag{6.5}$$

where $\mathbf{wM}_{i,j}$ a weighted adjacency matrix, where entries are greater than 0 if feature i is connected to j, with N being the total number of features. Further, for each t, the value has been normalized such that:

$$\mathrm{norm}\, C_D^W(i)_t = \frac{C_D^W(i)_t}{\max C_D^W(N)_t} \tag{6.6}$$

The value of norm $C_D^W(i)_t$ can only lie in the range $[0, 1]$: this operation allowed to obtain scaled normalized metrics for each group, each feature and each time stamp at the same time, in order to make data comparable over time and across features, controlling for the variation originating from the high variance in terrorist attacks per time unit.

The last step to create the final version of the multipartite graph was then to collapse the nested architecture to obtain a classic time-series data structure. We achieved this fixing time stamps as the source mode and maintaining the original four modes as target nodes. In this final step, the entry $i_{t,n}$ represents the centrality of that specific feature within that specific time-frame. After all these computations, this metric has to be interpreted as the degree to which that specific feature was targeted or employed by a given terrorist group. High values of degree centrality signal that a given feature was very popular during that time frame, while 0 means that that the feature was not attacked or employed by the considered group. A short visual explanation of this process is shown in Figure 6.1.

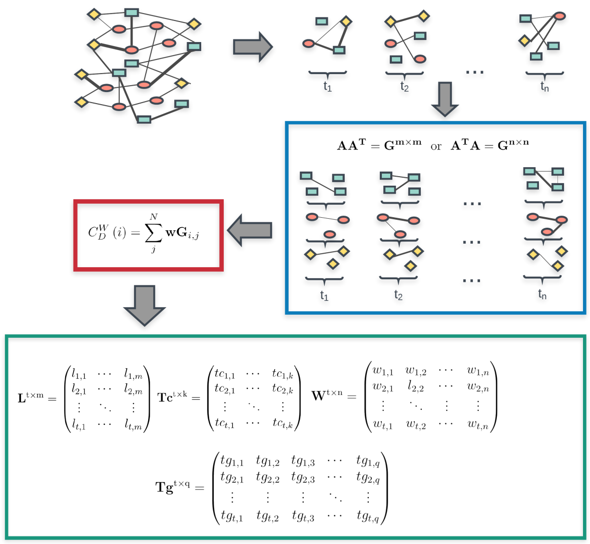

Figure 6.1: Simplified Visual Depiction of Graph-Derived Time Series Extraction

This type of process and data transformation provide a framework in which actual values are abstract representations of reality. For instance, comparing across two quantities A and B that are also smaller than 1 in two distinct time units does not mean anything. Even the comparison of the same two values in the same time unit does not give any information in the absolute sense. However, the relative comparison is important, and that is why in the models, the prediction will be evaluated in terms of ranking. Ranking will be useful to provide risk-based outputs regardless of the absolute centrality values which are, again, an abstraction. It will help to capture the preferences and recurring patterns of jihadist attacks over-time, providing an easy-to-understand and interpretable measure of risk. After having explained the steps behind the processing and manipulation of the data, the next subsection will start to introduce the proper algorithms that will be used for modeling purposes.

6.3.3 A Brief Introduction to Neural Networks

Machine Learning (in the work also referred to as ML) can be defined as the science of building computer models that can learn and improve their learning capabilities over time through the use of information and data. The expression “Machine Learning” is supposed to be coined by Harmon (1959). ML evolved from the area of pattern recognition in Artificial Intelligence (AI) and has seen an explosion of theoretical innovations and applications in dozens of fields in the last years, spanning from pure research to industry, intelligence and government.

The main distinction within the realm of ML is between supervised and unsupervised learning. In the case of supervised learning, the computer program is provided with inputs and outputs, and the main goal is to find a function that accurately maps inputs to outputs. Conversely, in the case of unsupervised learning, the programmer does not provide the program with labels, therefore there is no distinction between inputs and outputs. The aim of an unsupervised learning system is generally to discover hidden patterns in the data. Besides these two distinct tasks that ML algorithms should solve, there is a third one (which sometimes is included in the former category of supervised learning) that is worth to mention. This third category is called "reinforcement learning", and it deals with the problem of finding actions to take in a given situation to maximize a certain reward (Bishop, 2006) Besides the different tasks, ML algorithms can be distinguished also by their applications. Within supervised learning, indeed, we can divide two main applications: classification and regression problems. Classification is related to the learning problems in which the output (also called target or response variable) is categorical. When, conversely, the output is real-valued or continuous, the application is called regression (Murphy, 2012). Within unsupervised learning, two main applications are clustering, density estimation, and visualization. Clustering aims at discovering similar groups within the data (Xu and Wunsch, 2005), density estimation at determining the distribution of data across inputs and dimensionality reduction at projecting data to a two or three-dimensional space, starting from a high-dimensional one (Saul and Roweis, 2013).

The list of ML algorithms is extremely vast and this work does not aim at surveying them. Instead, it specifically focuses on a particular family of algorithms: Neural Networks (also called Artificial Neural Networks). The expression "Neural Network" originates from the pioneering work of McCulloch and Pitts (1943). The American scientists tried to reproduce the information processes of biological systems via mathematical modeling and proposed a model to simulate how neurons behave. After them, many others tried to refine and improve this mathematical formalization (Rosenblatt, 1958; Harmon, 1959; Widrow and Hoff, 1960; Block et al., 1962; Rumelhart et al., 1986). Over the course of decades, scientists have found that, besides some exaggerate claims regarding the accuracy of artificial neurons in representing the way real ones act, these mathematical and statistical models were efficient in pattern recognition. This finding led to the development of a myriad of different models – neural networks

– that aimed at learning to simulate how the biological brain does.
From the definitional point of view, scholars have tried in many ways to describe what a NN is. No way has been found to universally define what NNs consist of. A good review of the definition has been compiled by Guresen and Kayakutlu (2011). For the sake of simplicity, I will here adopt the definition provided by Haykin (1994). Haykin defines NN as massively parallel combinations of simple processing units that can acquire knowledge from the environment (data) through a learning process and store this knowledge in its connections. This definition is extremely concise, though it is simple enough to explain clearly what a NN fundamentally is. NNs are organized in different layers that are made up of many interconnected nodes (i.e. neurons). These nodes contain an activation function and existing patterns are presented to the NN through an input layer that is connected and communicates with at least one hidden layer. A hidden layer is where mathematical processing is done via weighted connections. This hidden layer is also connected to the output layer, where the actual outputs of the model are shown (Figure 6.2).

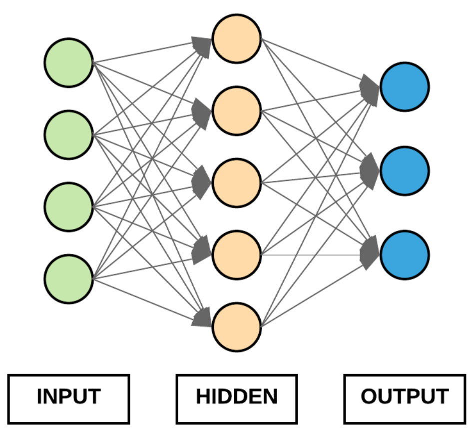

Figure 6.2: A Simple Neural Network Structure with a Single Hidden Layer

Basic research has seen the development of dozens of different neural network architectures. The most straightforward ones are feed-forward neural networks and perceptrons (Rosenblatt, 1958), while other very popular typologies are Boltzmann Machines (Ackley et al., 1985; Hinton, 2012), Convolutional Neural Networks (Lawrence et al., 1997; Lecun et al., 1998), Generative Adversarial Networks (Goodfellow et al., 2014) and Support Vector Machines (Cortes and Vapnik, 1995). The developments in this direction and the creation of neural networks with multiple (generally more than two) hidden layers gave birth to the so-called area of deep learning. Deep learning is indeed a sub-area of machine learning that relies on multiple levels of representation (i.e., hidden layers) (Goodfellow et al., 2016) to find patterns and representations in the data. In this work, I will specifically deal with the use of a specific deep class of neural networks capable of handling sequence-shaped data, namely Long Short-Term Memory network.

6.3.4 Long Short-Term Memory Networks: an Overview

Although several algorithms capable of handling temporally ordered or sequences data exist (and many more are currently being developed), the most common architecture in the literature is represented by Long Short-Term Memory (LSTM) networks. LSTM has been created to improve performance and solve the issues of Simple Recurrent Neural Networks (SRNN).
Simple Recurrent Neural Networks (SRNNs) were first developed by Elman (1990). His work started with the question of how to represent time in connectionist models, as NNs are. Elman worked in the realm of cognitive science and computational linguistics, and its architecture was first suited for lexical and textual data, specifically dealing with the problems of words and sentence emergence. SRNNs, as anticipated by the label, is made up of a simple architecture that basically consists of a three-layer feed-forward backpropagation network. These networks specifically hold a time twist. This means that they have connections through time: therefore, the order in which a network is fed really matters. Besides their simplicity, however, they present the so-called "vanishing (or exploding) gradient problem" (Bengio et al., 1994; Pascanu et al., 2012). This problem is related to RNNs iterative nature and practically means that information gets rapidly lost over time. Thus, SRNNs are not capable of handling very long memory processes. As pointed out by Chung et al. (2014), to formally define a SRNN we can consider a sequence $\mathbf{x} = (x_1, x_2, ..., x_T)$: the SRNN updates its recurrent hidden state h_t as follows:

$$h_t = \begin{cases} 0 & ; \quad t = 0 \\ \phi\left(h_{t-1}, x_t\right) & ; \quad otherwise \end{cases} \tag{6.7}$$

where ϕ represents a non-linear function (e.g. a logistic sigmoid) with an affine transformation. In the same work, Chung et al. say that the network can optionally have an output $\mathbf{y} = (y_1, y_2, ..., y_T)$ of variable length. Usually, the update showed in the previous equation is formalized as:

$$h_t = g\left(W_{X_T} + U_{h_{t-1}}\right) \tag{6.8}$$

where g represents a smooth, bounded function (again, as a logistic sigmoid or a hyperbolic tangent), W is a weight matrix and U is the hidden state-to-hidden state matrix, which has the form of a transition matrix, similar to the one that we can find in Markov chains. It is worth to mention that h_t represents the process of carrying "memory" in a mathematical fashion. Generative SRNNs outputs the probability distribution over the following elements of a sequence, given its current state h_t. Indeed, the sequence probability can be decomposed applying the general product rule of probability as follows:

$$\begin{gathered} p(x_1, ..., x_T) = p(x_1)p\left(x_2 \mid x_1\right)\left(x_3 \mid x_1, x_2\right) \cdots p\left(x_T \mid x_1, ..., x_{T-1}\right) = \\ p\left(\bigcap_{k=1}^{T-1} x_k\right) = \prod_{k=1}^{T-1} p\left(x_k \mid \bigcap_{j=1}^{T-2} x_j\right) \end{gathered} \tag{6.9}$$

where the last element represents a special output symbol in the form of a end-of-sequence value. Finally, we model each conditional probability distribution as follows:

$$p\left(x_T \mid x_1, ..., x_{T-1}\right) = g\left(h_t\right) \tag{6.10}$$

Regarding applications, this architecture has been used to solve problems in which sequence matters but not in the pure "temporal" nature, for example in computational linguistics, where the order of words is fundamental but there is not an actual timeline and, additionally, a researcher is not interested in the delta between two inputs or items. Besides computational linguistics, SRNNs have been employed, among the others, also in classic time series problems (Connor et al., 1994; Ho et al., 2002; Han et al., 2004) and speech recognition (Graves et al., 2013).

To overcome the issues associated with SRNNs, Hochreiter and Schmidhuber (1997) have developed Long Short-Term Memory Networks. This particular form of RNNs was able to control the vanishing gradient problem: this improvement explains their success. The way in which these networks combat the mentioned issue is by introducing gates and explicit memory cells. Each node (neuron) has three gates: input, output and forget. The input gate decides the degree to which the information coming from the previous layer gets stored in the cell. The output later, conversely, decides how much information regarding the present cell will be passed to the next layer. Finally, the forget gate helps in discarding information which is not useful and that should not be stored in the network. The flexibility of these networks and the introduction of explicit memory cells have proved to be very powerful in learning complex patterns and sequences. LSTM can handle and learn long-term dependencies and this is the main reason why they can learn complex patterns better than what can be achieved by SRNNs. This improvement marked the wide use of LSTM in a wide range of disciplines.

Although different versions of LSTM have been proposed by researchers, a commonly used architecture is the one proposed by Graves (2013). In his paper, Graves provides an LSTM where H, the hidden layer function, is implemented by a composite function. This composite function comprises five elements: input gate, forget gate, output gate, cell, cell input activation vectors, hidden vector. All these elements are of the same length as the latter element. The input target i_t is given by:

$$i_t = \sigma\left(W_{x_i} x_t + U_{h_i} h_{t-1} + V_{c_i} c_{t-1} + b_i\right) \tag{6.11}$$

where σ represent a logistic sigmoid function that maps nonlinear relations, W_{x_i} is the weight matrix of the input, U_{h_i} is the hidden-input gate matrix, V_{c_i} are the cell-gate diagonal weight matrices all at the previous state, b_i is the bias term of the input gate. The forget gate f_t is given by:

$$f_t = \sigma\left(W_{x_f} x_t + U_{h_f} h_{t-1} + V_{c_f} c_{t-1} + b_i\right) \tag{6.12}$$

withW_{x_f} representing the input-forget weight matrix, U_{h_f} is the hidden-forget weight matrix, V_{c_f} is the cell-forget diagonal weight matrix all at the previous state, and b_f is the bias term of the forget gate. The cell gate c_t is represented as:

$$c_t = f_t c_{t-1} + i_t \tanh\left(W_{x_c} x_t + U_{h_c} h_{t-1} + b_c\right) \tag{6.13}$$

where c_{t-1} is the old state, W_{x_c} is the input-cell weight matrix and U_{h_c} is the hidden-cell weight matrix all at the previous state, and b_c is the bias term of the cell. Finally, the output (activation) function of the LSTM unit h_t is defined as:

$$h_t = o_t \tanh\left(c_t\right) \tag{6.14}$$

where o_t is the output gate that modulates the amount of memory that is computed as:

$$o_t = \sigma\left(W_{x_o} x_t + U_{h_o} h_{t-1} + V_{c_o} c_{t-1} + b_o\right) \tag{6.15}$$

where W_{x_o} is the input-output weight matrix, U_{h_o} is the hidden-output weight matrix, V_{c_o} is the cell-output weight diagonal matrix all at previous state, and again b_o is the bias term of the output. It has to be noted that for the modelling of the research problem, a slightly different configuration has been used in this dissertation, as detailed in Subsection 6.3.5.
Besides time series (Gers et al., 2002), LSTM have been applied to face recognition problems (Levada et al., 2008), emotions modelling in audiovisual settings (Wöllmer et al., 2013), language modelling (Soutner and Müller, 2013), and medical diagnoses (Lipton et al., 2015).

6.3.5 Deep LSTM Configuration

The multi-partite graph processing phase led to the creation of dynamic networks for each group, with data shaped on a three-day unit basis. Since the groups cover extremely different time-spans, it is highly expectable that neural networks will operate and perform accordingly. Indeed, while some groups have a long-standing presence in the global scenario (e.g. Taliban), others are more recent (and far more active in terms of attacks), thus a first hypothesis is that the algorithms hyperparameters will have to be set depending on the dimension of the data at my disposal. Table 6.1 displays the number of time units per group, including the number of units with no attacks.

Group	**N Time Units**	**N of Units with No Attacks**
Islamic State	453	34 (7.50%)
Taliban	1949	665 (34.12%)
Al Qaeda	1946	1284 (65.98%)
Al Shabaab	1096	411 (37.46%)
Boko Haram	905	323 (35.65%)

Table 6.1: Time Units per Group

As introduced above, the length of time-series is variable across groups. Furthermore, it is not just the length of time series which differentiates them. Looking at the number of time units with no attack, we detect that classes are highly unbalanced. Given its high frequency, for instance, the Islamic State has the shortest time-line, but the lower percentage of inactive units. Conversely, the Taliban and Al Qaeda show longer time-lines, but the actual number of active units is extremely low. These figures involve potential high impact differences in the way the algorithms will be set and working. The proper modeling process involves the comparison of the performance of different configurations of LSTM networks.[3] Each configuration has been tuned setting different numbers of batch sizes and look back. These concepts are presented below. Additionally, a description of other relevant elements of the models is also provided. Elements include the number of layers, number of neurons, regularization, activation function, train size, and optimizer.

6.3.5.1 Architecture of the LSTM

Layers The number of layers in a neural network generally reports the actual number of hidden layers that are comprised in the model. A vivid debate in the field of AI focused on whether neural network should have more than a single hidden layer. Until 2006, most of the field believed that one hidden layer was sufficient due to the Universal Approximation Theorem proposed by Cybenko (1989) and then expanded by Hornik (1991). However, in 2006, due to the considerable shift towards more and more complex datasets (including ones designed for time-series), Hinton et al. (2006) posited that multiple (and eventually dense) hidden layers can improve the performance of the algorithms. In my models, each deep neural network will comprise three hidden layers, as preliminary testing showed little learning capacity for networks with just one or two hidden layers. Additionally, networks of higher complexity did not reach a statistically better performance than the three-layered architecture.

Neurons The number of neurons involves a crucial decision in the topology of the network. Indeed, the number of neurons can highly influence the performance of the algorithm. The decision between too few or too many neurons is tightly related to the problems of under-fitting (for which the algorithm is not capable of detecting patterns and signals) and over-fitting (for which the network capacity by far exceeds the complexity of the data structure). Furthermore, neurons have an impact on the time required to train the network: as the number increases, time does increases too and can lead to the impossibility of proper training, therefore nullifying the utility

[3] The entire modelling part has been deployed via Keras (Chollet, 2015), using TensorFlow backend (Abadi et al., 2015).

of the network. After extensive preliminary experiments, each network will have an input number of neurons equal to the number of features of the whole input space, the three hidden layers will respectively have 256, 128 and 64 neurons, and the output layer will be again equal to the input space.

Regularization: Dropout and Early Stopping Dropout is a form of regularization applied to neural networks that was introduced by Srivastava et al. (2014). It is a technique that aims at preventing a neural network from overfitting. The idea behind this technique is to drop units and related connections in random order while the network is training, preventing its excessive co-adaptation. The values that can be set for dropout fall in the range $[0, 1]$, where 0 means no dropout at all. The experiments have been carried out applying dropout regularization of 0.5 for each of the three hidden layers. Extensive preliminary experiments demonstrated that lower dropout was not able to avoid overfitting in the networks, while higher values prevented the networks to learn efficiently. In addition, an "early stop" option is also included. This option automatically interrupts the learning process after several epochs (i.e., 20) with no detected change (either increase or decrease) in the chosen loss function during testing.

Activation The activation function of a node in neural network models the output originating from an input for that given node. There are many different activation functions, and the most popular are the binary step, logistic sigmoid, hyperbolic tangent (tanh), rectified linear unit (ReLu), Leaky rectified linear unit (Leaky ReLU) and softmax. For each of the models, two different functions have been applied: the logistic sigmoid for the input and output layers, while in the hidden ones ReLu has been used. The logistic sigmoid is a monotonic non-linear activation function given by the equation:

$$\sigma = \frac{1}{1 + e^{-x}} \tag{6.16}$$

This function guarantees a smooth gradient (preventing spikes in the outputs), bounded values between 0 and 1 (which constitute a perfect fit for this specific problem, given that all the data are within this range). Unfortunately, this function also shows disadvantages, as the vanishing gradient that can dramatically decrease the learning procedure or severely impact the speed of the computations. For this reason, to avoid these risks, the hidden layers are activated via ReLu. ReLu is also a non-linear monotonic function modeled by the equation:

$$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases} \tag{6.17}$$

or, much more simplistically, $f(x) = max(x, 0)$. Compared to the logistic sigmoid, ReLU is much more computationally efficient (it converges fast) and it is sparsely activated, a characteristic that fits well with the extreme sparsity of the data used for the analysis of this chapter. It is now very common in the deep learning literature and the models will then exploit its strengths in the hidden layers, where the majority of neurons is located.

Loss Function Neural networks involve an optimization process that seeks to minimize a given loss function. In other words, loss functions seek to minimize the prediction error, comparing real and model-generated data. The choice of the specific loss function is dependent upon the specific problem setting but can be *a priori* divided into two main families: loss functions for classification and loss functions for regression problems. Given the original regression nature of the experiments (that will be then transformed into a ranking problem through the introduction of two accuracy measures), two loss functions have been monitored to evaluate the fitness of the models and detect potential overfitting (or underfitting) issues. The first loss function that has been monitored is Mean Squared Error (MSE). The standard form of MSE is given by:

$$\mathcal{L} = \frac{1}{n}\sum_{i=1}^{n}(y^{(i)} - \hat{y}^{(i)})^2 \tag{6.18}$$

where $(y^{(i)} - \hat{y}^{(i)})$ is called residual between the actual and the predicted value of y and the objective of the function is to minimize the residual sum of squares. Additionally, also Mean Absolute Error (MAE) has been used to assess the performance of the LSTM networks. MAE is calculated as:

$$\mathcal{L} = \frac{1}{n}\sum_{i=1}^{n}|y^{(i)} - \hat{y}^{(i)}| \tag{6.19}$$

where $|\bullet|$ denotes the absolute difference between the actual and the predicted value of y. MAE is less efficient than MSE in terms of computation, but it is more robust to outliers since, compared to MSE, it does not use the square. These two functions then map the loss in different ways with respect to the magnitude of the error, and this is the reason why their combined use as a choice to control either aspect (i.e. smaller and bigger errors) of the prediction processes.

Look Back The look back is a component that has been included in the models to set the length of the vector of inputs from which the network has to learn and forecast future values. Indeed, it can be set fixing a certain number N of input vectors (namely, time units). The name itself suggests that this hyperparameter “force” the network to only used previous N time steps to detect patterns. This represents an intriguing way to take memory into account. Of course, for larger N, the training will be slower. Considered that each time unit includes data for three days, the look back thresholds will be set accordingly to test meaningful hypotheses regarding “seasonal pattern”. In the experiments, look back sizes of 1, 2, 3, 10, 20, 30, 50 (equivalent to 2, 6, 9, 30, 90 and 150 days of data prior the present attack) have been considered.[4]

Train Set The train size is the amount of data that is fed into the network to detect patterns before the proper test phase. Generally, the larger is the train size, the more the network can learn from its inputs. For neural networks with no sequence or time-dependencies, the split between train and test size is made upon percentage thresholds off-the-shelf. However, in the case of time series, the distinction is a bit more delicate. Indeed, fixing a % for models related to groups that have very different lengths and dimensions has to be evaluated carefully. As a baseline approach for this dissertation, I have used a 90/10 split. This unbalanced choice is motivated by two main reasons. First, almost all groups exhibit very few attacks in the first part of the temporal windows. A more balanced (e.g., 65/35) split would have potentially posed the risk of training the network on a set of data that does not represent the current situation highlighted in the right-end of the time-frames. Second, the time series are not particularly long (especially in the cases of Boko Haram and the Islamic State), thus providing more data to the networks could practically reduce the risk of under-fitting. Table 6.2 shows the relative temporal scale of the 90/10 split for each group.

[4] For the Islamic State, the maximum lag will be set to 40, as 50 would have exceeded the actual length of the test set in terms of data points.

Group	**Total N of Data Points**	**Training Time Units**	**Testing Time Units**
Islamic State	453	3.35 years	4.53 months
Taliban	1949	14.41 years	1.6 years
Al Qaeda	1946	14.39 years	1.59 years
Boko Haram	905	6.69 years	9.05 months
Al Shabaab	1096	8.10 years	10.96 months

Table 6.2: Relative Temporal Scale of the 90/10 Split - per Group

Optimizer An optimization algorithm seeks to minimize (or, alternatively, maximize) an objective function (i.e. loss function) to improve the training process of a Neural Network. There are generally two types of optimization algorithms: first-order (that use first-order derivatives) and second-order (which use second-order derivatives). The former is less expensive, while the latter is slower and costly. Generally, the most common technique for optimizing a neural network is the so-called Gradient Descent. Gradient Descent is based on a convex function and updates its parameters iteratively to minimize a given function (again, loss as an example) to ist local minima. There is a wide range of different gradient descent optimization algorithms. The most common are: Adam, Adagrad, Adadelta and Nesterov Accelerated Gradient. Although further experiments will test multiple types of optimization algorithms, my models have been run using Adam (Kingma and Ba, 2014). Adam is a first-order optimization algorithm that combines AdaGrad (Duchi et al., 2011) and RMSProp (Tieleman and Hinton, 2012) and the reason behind its use as optimization algorithm is that it works well with sparse, noisy and even non-stationary data. Technically, given the sequence of the gradients at each timestep, $g_1, g_2, \cdots, g_T$ it computes the exponential average of the gradient as :

$$v_t = \beta_1 \cdot v_{(t-1)} - (1 - \beta_1) \cdot g_t \tag{6.20}$$

Furthermore, it also calculates the squares of the gradient for each of the parameters w as:

$$s_t = \beta_2 \cdot s_{(t-1)} - (1 - \beta_2) \cdot g_t^2 \tag{6.21}$$

where g_t^2 is the elementwise square $g_t \odot g_t$. After the first steps, it chooses the learning step through:

$$\Delta\omega_t = -\eta \frac{v_t}{\sqrt{s_t + \epsilon}} \cdot g_t \tag{6.22}$$

where ω gives the model weights, η is the learning rate. The update ω_t is then equal to $(\omega_t)+\Delta\omega_t$. It is worth to mention that the authors recommend to keep β_1 equal to 0.9, β_2 equal to 0.999 and ϵ equal to 1e-10. The learning rate is fixed at the common value of 0.001. The learning rate of a neural network is the hyper-parameter which sets the threshold of how much the algorithm should adjust weights with respect to loss gradient. On one side, if the learning rate is fixed at a low level, the training phase will be longer due to the high amount of time needed for convergence. However, training will be more reliable. On the other side, if the learning rate is high the training will be much faster but the network will face the risk of divergence. In the case of Adam algorithm, for convention, in many implementations the learning rate has to be fixed at 0.001 and kept at that level. The experiments rely on this default level.

Batch Size The batch size indicates the number of examples taken from the training set utilized in a single iteration. There are three main distinctions in terms of batch size: 1. *batch mode*, which means that the batch size is equal to the dataset; 2. *mini-batch mode*, which means that the batch size is greater than one training example but smaller than the dataset; 3. *stochastic mode*: where batch size is equal to one. Research showed how mini-batch is usually preferred because of less computational cost and better performance in gradient descent. More precisely, in a recent paper, Masters and Luschi (2018) proved how mini-batch sizes better perform when it is kept in the range between size 2 and 32, contrasting the idea that using very large (also in the sizes of thousands) mini batch sizes provides better results in terms of generalization performance. The performance is better also in terms of computation expensiveness. The models will test performance on batch sizes respectively equal to 2, 32, 64 and 100.

Epochs An epoch indicates when the whole dataset is passed forward and backward through the network once. In the mini-batch size, an epoch is comprised of multiple batches. The number of epochs used to train a neural network is usually large. It is straightforward that the higher the number, the slower the training phase. There is no *a priori* rule regarding the number of epochs to set in a neural network. Given the relatively small dimension of the dataset, and prior experiments on the data, models with 500 epochs have been run for each group.
To sum up what written in the last paragraphs, Table 6.3 synthetizes the hyperparameter combinations that were used to run the models per each group. Overall, multiplying the number of tested batch sizes and the number of epochs, a total of 24

models per group have been run.

Configuration	Tested
N of Hidden Layers	3
N of Neurons	n of features, 256, 128
Dropout	0.5
Activation Function	Sigmoid and ReLU
Batch Size	2, 32, 64, 100
Look Back	1, 2, 3, 10, 20, 30, 50
Train Size	90%
Optimizer	Adam
Learning Rate	0.001
N of Epochs	500

Table 6.3: Tested Configurations for Neural Networks

6.3.5.2 An "Everything to Everything" Learning System

In these particular experiments, the LSTM will use all the variables to predict all the variables. In a simple and general mathematical notation, this can be formalized as $X_{real_{t-k}} \mapsto Y_{pred}$. In a more precise way, we can think of this type of architecture as a set of parallel predictions that are run on the same neural network such that if $X = (x_1, .., x_n)$ is an input vector and $Y = (y_1, .., y_n)$ is an output vector and both have the same length n, the neural network will work using the whole X to predict the whole Y at $(t + 1)$, and this Y will then be incorporated in X for predicting the next step, and so on.

It can be defined, somehow, as a sort of distributed teacher-forcing (Bengio et al., 2015; Lamb et al., 2016) system to train the network in which each output for a given feature uses all the other features plus the same feature ground-truth value at the previous time step (again, previous time steps can be of order higher than just one, depending on the look back). This choice was privileged due to its flexibility for other applications in the future (e.g., predicting tactics instead of targets via a little modification of the code), but other alternatives that could be experimented are (in order of ascending expected utility): 1. Targets to Targets (forecasting future targets using only previous data points on targets); 2. Input Space to Targets (which uses the information on hit countries, weapons, and tactics to predict future targets) and, finally, 3. Input Space and Targets to Targets (the output prediction is done using information on the input space and targets at previous time steps).

6.3.6 Performance Evaluation

MSE and MAE are useful for evaluating the learning process in terms of minimization of the error and in the attempt to diagnose potential over- and under-fitting, considering that the behavior of the training and test curves of the loss functions is precious in detecting flaws in the deep learning architecture, as diverging curves indicate over-fitting, while distant parallel curves suggest that the model is actually under-fitting. Nonetheless, these measures provide little information on the central task of the models: predicting the correct continuous central values is not the primary scope of the statistical learning procedure. The main goal is instead to correctly predict the most central (and therefore popular, and therefore at risk) targets at a given time unit. To evaluate the performance of the models two metrics have been developed ad hoc for this work, namely element-wise and set-wise accuracy.

Element-wise Accuracy Element-wise accuracy Φ is the most simple metric among the two. Given the sequence of time units $T_{n\neq 0} = t_1, t_2, \cdots, t_k$, where at least one terrorist attack has occurred and t_k represent the last (more recent) unit with attacks in the sequence, and the sets S and $\hat{S}$ that map the actual set of most central targets (up to three[5] in each t) and the predicted set of most central targets (again, up to three elements), I define the element-wise accuracy for t_1, ϕ_{t_i} as:

$$\phi_{t_i} =: \begin{cases} 1 & \text{if } \hat{S} \cap S \neq \emptyset \\ 0 & \text{if } \hat{S} \cap S = \emptyset \end{cases} \tag{6.23}$$

The Equation means that if the sets have at least one element in common, then ϕ_{t_i} is equal to one, while if the two sets are disjoint the value will be equal to zero. For the entire history of considered attacks $T_{n\neq 0}$, then, the overall EA accuracy Φ_T is computed as:

$$\Phi_T = \frac{1}{T_{n\neq 0}} \sum_{i=1}^{T} \phi_{t_i} \tag{6.24}$$

with Φ_T being the ratio between the sum of single unit binary accuracies ϕ_{t_i} and the total number of time units T with at least an attack.

[5] When less than three targets have been hit in a given t, the set comprises either two (if present) or one entity. When two or more targets are *ex aequo* ranked within the three most central targets in the set, a random procedure select only one.

Set-wise Accuracy Set-wise accuracy Γ is more complicated and further tests the ability of the deep neural networks to identify and predict the correct set S of most central targets. Going a bit more in detail regarding S, the cardinality of the set is bounded in the range $0 < |S| \leq 3$. Thus, for a given time unit t_i, single γ is defined as:

$$\gamma_{t_i} = \begin{cases} 1 & \text{if } \hat{S} = S \\ x & \text{if } \hat{S} \cap S \neq \emptyset \\ 0 & \text{if } \hat{S} \cap S = \emptyset \end{cases} \tag{6.25}$$

The singular γ_{t_i} is then equal to 1 if the two set are perfectly identical (as in any set, it is worth to note, the order does not matter), is 0 when the two sets are disjoint and can take a real value x when there is an intersection between S and . This value x is calculated via:

$$x = \frac{1}{|S|} \sum_{i=1}^{|S|} \ddot{s}_{\in \hat{S} \wedge S} \tag{6.26}$$

where $\ddot{s}$ maps an element which is part of both S and $\hat{S}$. In the case in which $\ddot{s}$ is exactly equal to the cardinality $|S|$, the value of x becomes 1 as it would mean that $\hat{S} = S$. Finally, the overall metric Γ for the sequence $T_{n \neq 0}$ is given by:

$$\Gamma_{T_{n \neq 0}} = \frac{1}{\sum_{t=1}^{T} |S|} \sum_{t=1}^{T} \ddot{s}_t \tag{6.27}$$

$\Gamma_{T_{n \neq 0}}$ is computed through the product of the inverse of the sum of the elements in each $|S|$ present in the sequence T and the sum of all elements $\ddot{s}$ that, in each time unit of T, are both part of sets S and $\hat{S}$. The metric is, as anticipated, more challenging for the algorithm and aims at providing more comprehensive information to researchers and - potentially - policymakers on the riskiest targets in a future time unit. The threshold of 3 has been set arbitrarily upon previous analysis of the average number of targets hit in each unit with at least one attack. However, it can be modified in the future based on a particular necessity or in relation to different research problems analyzed with different types of data (i.e., less sparse).

6.4 On the Properties of Time Series of Jihadist Groups

In this section, two key properties of time series are investigated (i.e., stationarity and randomness) and additional information specifically on targets hit by terrorist groups is provided to shed some light on how they behave over time, seeking to highlight some patterns that might be relevant also in the actual modeling part.

6.4.1 Investigating Stationarity

Understanding the structure of the data over time is fundamental when dealing with time series. Besides merely descriptive statistics, one of the statistical steps which are required for assessing patterns in data and verifying assumptions in classical statistical methods is to test for stationarity. A stationary process is defined as a stochastic process that has unconditional joint probability distribution that does not change when time-shifting is considered, and therefore parameters as mean and variance remain stable over time. We could thus say that a time series is stationary if its properties do not depend upon the time at which the series is observed. There are two main typologies of stationarity: strict and weak (also known as covariance stationarity or second-order stationarity, commonly used in signal processing). In mathematical terms, a process is said to be strictly stationary if all its finite order distributions are time-invariant, meaning the joint cumulative distribution functions of:

$$X(t_1), X(t_2), ..., X(t_k) \text{ and } X(t_1+\tau), X(t_2+\tau), ..., X(t_k+\tau) \tag{6.28}$$

are the same for all k, $t_1, t_2, ..., t_k$ and τ. Relaxing the assumptions used for defining a strictly stationary process, the conditions for weak stationarity are three:

1. The second moment of x_t is finite for all t, which means that:

$$E\left|x_t\right| < \infty \;\; \forall\, t \tag{6.29}$$

2. The first moment of x_t is independent of t, which means that:

$$E\left(x_t\right) = \mu \;\; \forall\, t \tag{6.30}$$

3. The cross moment $E\left(x_{t_1}, x_{t_2}\right)$ depends only on $t_1 - t_2$ that is:

$$cov\left(x_{t_1}, x_{t_2}\right) = cov\left(x_{t_{1+h}}, x_{t_{2+h}}\right) \;\; \forall\, t_1, t_2, h \tag{6.31}$$

The concept of stationarity is extremely important in statistics and econometrics because most time series models require data to be stationary to be meaningfully performed. Although in the case of neural networks stationarity is not strictly required, because the algorithm can handle non-linear relations and it is more flexible in processing the data, I have checked for stationarity in the network-derived time series for shedding some light on the shape on the potential patterns that may arise from the analyses. To do so, I have applied one of the most common statistical tests for stationarity checking in time series, which is the Augmented Dickey Fuller (ADF) test (Dickey and Fuller, 1979). Given an autoregressive process of first order, AR(1), written in the form:

$$y_t = \rho y_{t-1} + u_t \tag{6.32}$$

where y_t represents the dependent variable, t is the time index, ρ is a coefficient and u_t is the related error term, we say that a unit root is present, i.e. the process is non-stationary, iff $\rho = 1$. Starting from the model above, the regression model can be then written as:

$$\Delta y_t = (\rho - 1)\, y_{t-1} + u_t \tag{6.33}$$

where δ indicated the first difference operator. In this case, the unit root can be tested fixing $\delta = 0$ (where $\delta = \rho - 1$). Considering the fact that the test is computed over the residual term, the statistic t is characterized by a distribution known as Dickey-Fuller table. There are three main versions of the original Dickey-Fuller test:

1. Test for a unit root with no drift and deterministic time trend:

$$\Delta y_t = \delta y_{t-1} + u_t \tag{6.34}$$

2. Test for a unit root with drift and no deterministic time trend:

$$\Delta y_t = a_0 + \delta y_{t-1} + u_t \tag{6.35}$$

3. Test for a unit root with drift and deterministic time trend:

$$\Delta y_t = a_0 + a_1 t + \delta y_{t-1} + u_t \tag{6.36}$$

The ADF test simply adds lagged dependent variables to the test equation, and it is therefore applied to the model:

$$\Delta y_t = a_0 + \beta_t + \gamma y_{t-1} + \delta \Delta y_{t-1} + ... + \delta_{k-1} \Delta y_{t-k+1} + u_t = \\ a_0 + \beta_t - \sum_{k=1}^{m} \delta \Delta y_{t-k} \tag{6.37}$$

with a_0 representing a constant, β as the coefficient on a time trend and k is the lag order that may allow for higher-order autoregressive processes. The unit root is tested as the null hypothesis $\gamma = 0$ against the alternative hypothesis of $\gamma < 0$ which would detect stationarity. The test statistic has the form:

$$ADF = \frac{\hat{\gamma}}{SE\left(\hat{\gamma}\right)} \tag{6.38}$$

and if it is less than the larger negative critical value, then the null hypothesis can be rejected, and the unit root can be excluded.

To test for stationarity of the data related to the five groups, I have therefore performed the ADF test controlling for different lags. Since my data are divided into three-days units, I have run the test for $1 \leq k \leq 50$. Results are provided in Figure 6.3. The test, iterated through lags of different dimensions, yielded clear results. Indeed, what is directly noticeable from the figure is that two processes remain stationary when all lag orders are tested. This is verified in the case of the Taliban and Al Qaeda (this latter only highlights a risible fluctuation of statistical significance). Al Shabaab, similarly, starts to highlight non-stationary time series only after $k > 30$. In the case of Islamic State and Boko Haram, lags of very high order fail to reject the null hypotheses: k=10 for the Islamic State already detects unit root processes, the number further increases up to around 80% of the time series when k is set to 50. For what concerns Boko Haram, the tests starts to fail after k=10, and reaches a maximum of non-stationary features around k=50, with a ratio of 0.5 over the total number of time series.

One additional indication has to be drawn from the analysis: the lookback in the neural networks will have to consider the results of this test. Indeed, fixing a lookback that is too large may prevent the algorithm from correct learning. As stationarity is found for lags that do not exceed a certain threshold the interpretation of the results will take into account this analysis.

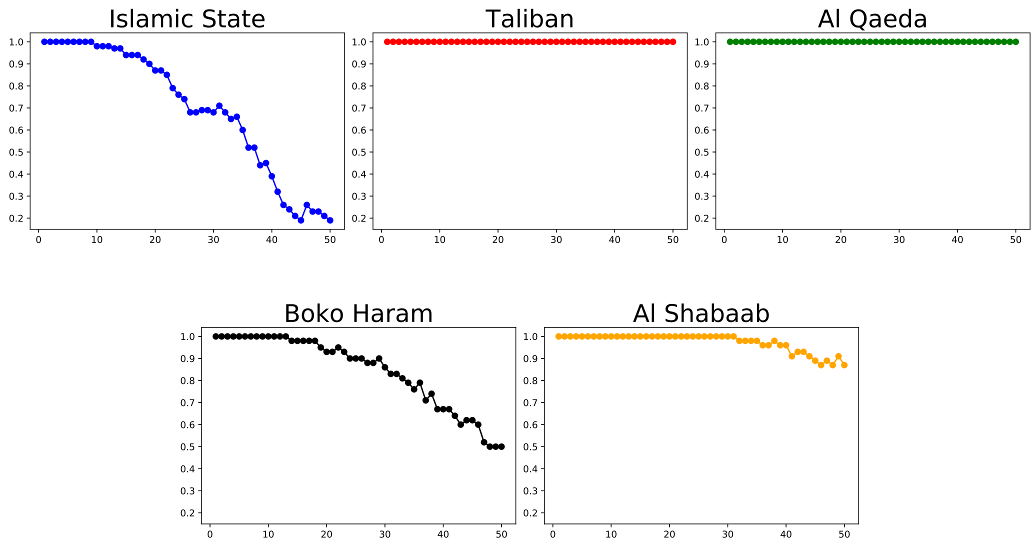

Figure 6.3: Ratio of Stationary Time Series per Each Group with Lag from 1 to 50

6.4.2 Investigating Randomness

After having verified the presence of stationarity in the data, it is useful to check for another property to understand the structure of the data: this property is randomness. In statistics, a "truly random" process is intended as a process that can produce independent and identically (i.i.d.) distributed samples. This means that if a given value in a sequence is influenced by its position, or by the prior data points, the process is not truly random. There are many possible solutions for testing randomness in time-series data (Moore and Wallis, 1943; Cox and Stuart, 1955; Mateus and Caeiro, 2013), however I will here apply the Bartel's Rank Test (Bartels, 1982), which is a modified rank-based version of von Neumann's Ratio Test for randomness (von Neumann, 1941).

Differently from other tests, instead of comparing the magnitude of the observations with their prior data points, Bartels' Rank Test ranks all the observations from the smallest to the largest. If the null hypothesis of randomness if verified, rank arrangements from the whole set of $n!$ possible combinations have the same probability. Considered the rank $R(X_i)$ as the sequential number of X_i, the probability for the test statistic RVN is given by:

$$RVN = \frac{\sum_{i=1}^{n-1} \left(R_i - R_{i-1}\right)^2}{\sum_{i=1}^{n-1} \left[R_i - \left(n+1\right)/2\right]^2} \tag{6.39}$$

It follows that $\frac{RVN-2}{\sigma}$ is asitomptically standard normal, where σ is the variance and it is equal to $\sigma = \frac{4(n-2)\left(5n^2-2n-9\right)}{5n(n+1)(n-1)^2}$. Among the many different possibilities, this test that uses ranks is particularly suitable for our data considered two factors. The first factor that it is worth taking into account is related to the very nature of the data which are – as reminded several times – abstractions of real-world dynamics. We cannot tell what a particular value for a particular feature at a given time is specifically saying. Also, since values across timestamps are not independent, but depend on the number of features that are also existing in a given timestamp it is difficult to compare absolute value through time, even in the extreme case where the same feature is equal to 1 in two consecutive points. In fact, 1 in a case may indicate that the given feature was the only one that was employed or targeted, while in the other case, 1 may indicate that all attacks the time unit only implied that feature. The second factor is tightly related to the development of the model: neural networks will process absolute values, but predictions will be evaluated on a ranking base. Therefore, it is far more useful to evaluate potential randomness from the point of view of ranked sequences rather than focus on their values which are less important, considering that if there are three features that are respectively equal to 1, 0.05 and 0 we are not concerned about the differences between these features, but we are seeking to find a perfect overlap in terms of rankings for forecasting risks.
There are three potential alternative hypotheses that one can use to run the test: right-sided, left-sided and two-sided. When using a right-sided test, the null hypothesis is confronted against the alternative of a systematic oscillation; when using the left-sided, the alternative hypothesis regards the presence of a trend; finally, when using the two-sided specification, the alternative hypothesis is proper non-randomness. The latter is the specification that I have used. The results of the tests are provided in Table 6.4.

The results are of the Bartels Rank test for randomness are interesting and demonstrates that the neural networks will deal with potentially very complex prediction problems. Indeed, the random component of processes is widely present in each group. The Islamic State is the jihadist organization that presents the wider random component, with 69.35% of the features that fail to reject the null hypothesis of the test. Conversely, Al Qaeda shows the smallest presence of randomness in its features (13.8%).

Group	Random Features (ρ>0.1)	Non-Random (ρ<0.1)	Non-Random (ρ<0.05)	Non-Random (ρ<0.01)
Islamic State	43 (69.35%)	3 (4.41%)	4 (5.88%)	12 (17.64%)
Taliban	13 (30.9%)	0 (0.00%)	1 (2.38%)	28 (66.66%)
Al Qaeda	10 (13.8%)	0 (0.00%)	0 (0.00%)	62 (86.2%)
Al Shabaab	25 (55.5%)	2 (4.44%)	5 (11.11%)	13 (37.77%)
Boko Haram	19 (45.2%)	2 (4.76%)	4 (9.52%)	17 (40.47%)

Table 6.4: Results of Bartels Rank Test for Randomness - per Group

The presence of randomness in data can be considered an issue: indeed, how is it possible to make good predictions out of training data that do not seem to have any oscillatory or seasonal patterns? This is one of the most challenging points of terrorist events forecast, especially when such a time series is processed using very detailed time units, as in this case.

6.4.3 Temporal Dynamics of Targets

The previous paragraphs aimed at verifying stationarity and potential randomness and served as an empirical analysis to grasp as much information as possible from the data before properly applying algorithms and analyzing the results. What we now know, from the macro perspective at least (i.e.: without looking at specific features) are general properties that should be considered before and after the modeling step. That said, it is therefore useful to focus on the micro-level. Micro-level, in this case, is the analysis (though at a first descriptive glance) of temporal dynamics of targets, which represents the most important and relevant portion of data at my disposal, since models will be evaluated based on their ability to correctly forecast them. Table 6.5 provides average values for all targets that are present in the data, i.e. all targets that have at least a mention in at least one group. Figures, in this fashion, are a valuable instrument to understand which are the most frequent (or popular) targets overall, without taking into account – or visualizing – trends or temporal dynamics. Comparing across groups in terms of absolute values does not make sense, but – again – thinking in terms of ranking is a helpful workaround to make more sense of the data.

Target	**Islamic State**	**Taliban**	**Al Qaeda**	**Al Shabaab**	**Boko Haram**
Business	0.043	0.056	0.021	0.047	0.033
Government (Diplomatic)	0.033	0.019	0.024	0.012	0.003
Private Citizens	0.273	0.247	0.083	0.217	0.300
Refugee Camp	0.005	0.001	NA	0.002	0.007
Maritime	0.000	NA	0.001	0.001	NA
Private Security Company	0.003	NA	NA	0.001	0.002
Police	0.113	0.243	0.006	0.068	0.085
Journalists & Media	0.011	0.008	0.003	0.022	0.001
Religious Figures/Institutions	0.017	0.017	0.007	0.016	0.050
Military	0.069	0.063	0.038	0.079	0.033
Terrorists/Non State Militia	0.078	0.012	0.021	0.008	0.009
Unknown	0.015	0.035	NA	0.008	0.016
Ambulance	0.001	0.001	0.001	0.001	NA
NGO	0.006	0.024	0.002	0.011	0.001
Transportation	0.009	0.020	0.007	0.010	0.008
Utilities	0.006	0.003	0.001	0.002	NA
Educational Institution	0.003	0.039	0.005	0.008	0.018
Violent Political Party	0.001	0.001	0.002	NA	0.001
Airports & Aircraft	0.001	0.009	0.004	0.011	0.001
Government (General)	0.033	0.167	0.069	0.130	0.045
Other	0.001	NA	0.000	NA	NA
Food or Water Supply	0.001	0.002	0.001	0.003	0.001
Demilitarized Zone	0.000	0.001	0.001	NA	NA
Tourists	0.000	0.001	0.005	0.001	0.001
Fire Fighters	NA	NA	0.001	NA	0.001
Telecommunication	NA	0.006	0.001	0.002	0.002

Table 6.5: Average Centrality Values for all Targets Present in the Data - per Group

The Islamic State proves to attack more Private Citizens, Police, and Terrorists or Non-State Militias. Taliban highlights similar preferences, but instead of targeting terrorists or Non-State Militias, they tend to attack Government (General). Al Qaeda has a tendency against Government (General) too, coupled with Military (and Private Citizens), showing how their efforts have been put into actions against institutions. Al Shabaab has a very similar profile to Al Qaeda's one, as also detected in Chapter 4. Finally, a first look at Boko Haram's data points out that the group plots attack against Religious Institutions as a probable consequence of the hybrid religious setting of the geographic area in which the organization operates. "Stock" data give a first summary of groups' preferences, however much more information may be provided visualizing and commenting proper temporal dynamics.

Islamic State Observing Figure 6.4, several things can be detected. First of all, it is evident why for so many features (also comprising locations, tactics, and weapons) randomness was detected. In fact, there is a consisted part of targets that seem not to follow logical or predictable trends. This is the case of targets such as Tourists, Airports and Aircraft, NGO, Transportation, and Utilities. In the case of Tourists, relevance looks risible, considering the very low centrality values and frequency of attacks against this category. Airports and Aircraft, NGOs and Transportation highlight sudden peaks that are generally not followed by close (in time) consequent attacks, therefore an algorithm (though powerful) might find it difficult to learn structure considered the anomaly of such events.
On the other side, other targets appear to be more patterned in time. This is especially the case for Private Citizens and Properties, Police, Terrorists and Non-State Militias and Military. There are quite long sequences of attacks that are characterized by continuous targeting of such categories. In some cases, we can identify very high recurrent peaks followed by decreasing trends that tend to converge over low levels of "popularity". Private Citizens and Properties were particularly attacked in the first phase of the existence of the Islamic State, probably coinciding with the expansion strategy over physical territories. After this first phase, the Islamic State seems to differentiate across targets (whether this is an endogenous or exogenous process remains an open question) and starts to attack other Terrorists or Non-State Militias and Military forces. One hypothesis is that these two target categories represent the organized effort of structured or semi-structured bodies to counter the jihadist group. It is interesting to note how Government (General) has never been a highly recurring target in the Islamic State's short history. This, again, might be the reflection of the fact that the jihadist organization has been able to infiltrate, expand and spread across territories were the institutional stability and power were weak and not sufficiently present.
Finally, looking at the overall picture, it can be noticed how the most diversity and complexity of targets (high number of high-level trends in parallel) is clustered from 2013 to half 2014, meaning that in the first years of its existence, the group was probably able to carry out larger and heterogeneous attacks.

The Taliban Figure 6.5 shows the trends of the different targets attacked by the Taliban from the first recorded attack in the GTD to the last. As it will be for Al Qaeda, during the first years (until the early 2000s) recorded attacks are very few. Besides 1993, it would be really relevant to understand whether the groups were not properly active or if this constitutes an additional problem of missing data in the database compiling process.

That said, the figure seems to distinguish between extremely frequent attacked targets and other minor categories that look more like random (from a temporal point of view). On one hand, Business, Government (General), Private Citizens and Properties, Police and lately also Military are the most recurrent targets that are persistently present (with high centrality values) across the temporal spectrum, from 2005 circa on (with the exception of Military, which are shown to be highly targeted from 2012 on).

On the other hand, while Telecommunication, Tourists, Food and Water Supply seem almost randomly distributed in time, for other categories patterns are similar to microcycles or medium-term persistent shocks (namely recurrent attacks against a given category that persists for some weeks or months). This is the case of Transportation, with microcycles between 2010 and 2012 and persistent shocks between 2008 and 2009, NGOs (medium-term persistent shocks at the end of 2008 and 2009) and Educational Institutions (the most evident medium-term persistent shock is at the end of 2005).

At a first look, it seems that from 2011 on the Taliban has continuously plotted very diverse attacks, with multiple targets, proving a capacity to handle logistical complexity.

Al Qaeda Figure 6.6 looks in part similar to Figure 6.5 because of the absence (or very low frequency) of attacks in the first part of the plots. It is worth to highlight that the parallel peaks in Business, Government (General), Private Citizens and Properties and Airports and Aircraft at the end of 2001 represent the 9/11 attacks.

The frequency of events increases especially after 2004: interestingly, we can observe medium- and long-term persistent shocks in attacks against Police and Military, while Private Citizens and Properties and Government (General) are quite stably present with high levels of popularity across the last ten years spectrum.

The fact that Al Qaeda is represented here is a set of different actors (belonging to the Al Qaeda network) that may explain the popularity that the Government (General) category has. In fact, the ideology and final goal of the Al Qaeda Network can be

linked to the necessity to target institutional symbols of declared enemies to obtain enough echo and provoke potential sympathetic reactions by possible recruits.
It is furthermore worth to focus on some microcycles and medium-term recurring shocks in attacks against utilities, which may indicate particular strategies (maybe to illegally obtain resources) in specific geographic areas (trends should be disaggregated by a single group to actually understand micro-dynamics, this will thus be a good path for potential future work).
Finally, as it was found for Islamic State and Taliban, some targets do not seem to follow any type of trend or pattern. In the case of Tourists, Food and Water Supply, Demilitarized Zones and Fire Fighters, either events are too few or too distant in time. This will represent a considerable challenge for the algorithms, and it has to be taken in mind.

Boko Haram Figure 6.7 shows the target dynamics referred to attacks plotted by the Nigerian group Boko Haram. As it was already shown for other types of analysis (trails, for instance), Boko Haram holds several characteristics that are different from all the other groups. This consideration can be corroborated here by two target categories which are prominently present (though not with universal stability overtime) in Boko Haram temporal spectrum and that are not equally important for other groups: Religious Institutions and Educational Institutions.
The motivation behind the popularity of Religious Institutions in the Boko Haram attack strategies lies in the fact that the region in which the group operates has a consistent percentage of Christian population. Furthermore, it has to be noted that Islamic communities have denounced and refused the ideology of Boko Haram as written by Aghedo and Osumah (2012), thus Religious Institutions are enemies of the jihadist groups on a regardless of the specific religion.
A line can be drawn which connects Religious Institutions to Educational Institutions, a second relevant target category for Boko Haram. Indeed, Boko Haram attacks schools, colleges and universities (Catholic and not) that are in line with the Western idea of education completely refused and repudiated by Boko Haram. Educational Institutions, in general, are also used by Boko Haram as effective places for kidnapping girls and women that are used as hostages or are smuggled across adjoining African countries, like Chad and Cameroon (Peters, 2014).
Besides Religious and Educational Institutions, Boko Haram tends to mostly attack Private Citizens and seemed to stop targeting Police, since from late 2013 centrality has dramatically reduced (except for certain short cycles).

Al Shabaab Finally, figure 6.8 points out several things regarding the dynamics of targeting conducted by Al Shabaab. In the first three subplots, it is immediately noticeable how Business, Government (General) and Private Citizens and Properties represent by far the most preferred (also because of permanent recurrence overtime) targets by the Somalian group.

Looking at less persistent mechanics, Police and Military, two categories that are generally related if not mathematically at least conceptually, present medium- and long-term recurring shocks that are identifiable in the plots (for the former, between 2012 and 2013 and in early 2015, while for the latter a quite long term is exhibited during 2014). These two targets hold interesting dynamics that can indicate and highlight time windows where either countering actions forced the group to react massively or, conversely, direct strategies of the group to reach a particular goal.

Finally, Government (Diplomatic), Transportation, Telecommunication and Terrorists or Non-State Militias are certainly not amongst the most frequent targets, and the distribution may seem random if we concentrate on the whole spectrum, nonetheless, there are cases in which high spikes are actually followed by decreasing centrality values, indicating that, somehow, the given category has been the subject of attacks for a short, but detectable, period of time. This type of behavior certainly represents a challenge for the deep learning models, but it is more solvable than pure singular peaks that in statistical terms are defined as "outliers".

In these three cases we have instead short cycles that might remember of the behaviour that can be captured by Hawkes Processes, where - as seen in Chapter 5 - a first shock is followed by other shocks of minor intensity for a certain period of time, until the frequency and intensity converges over 0 in a certain amount of time (Ogata, 1988, 1998).

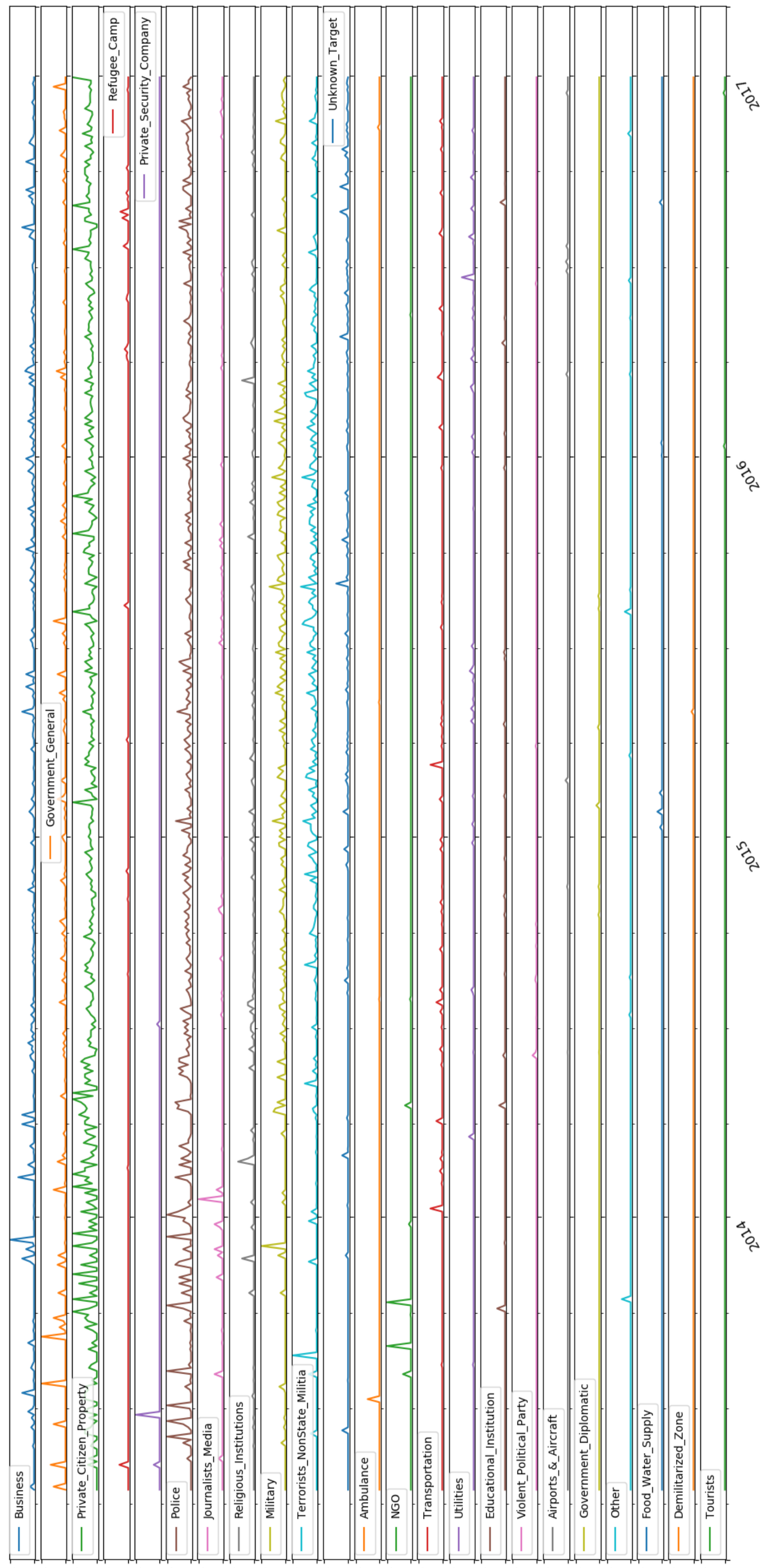

Figure 6.4: Centrality of Targets Over Time - Islamic State

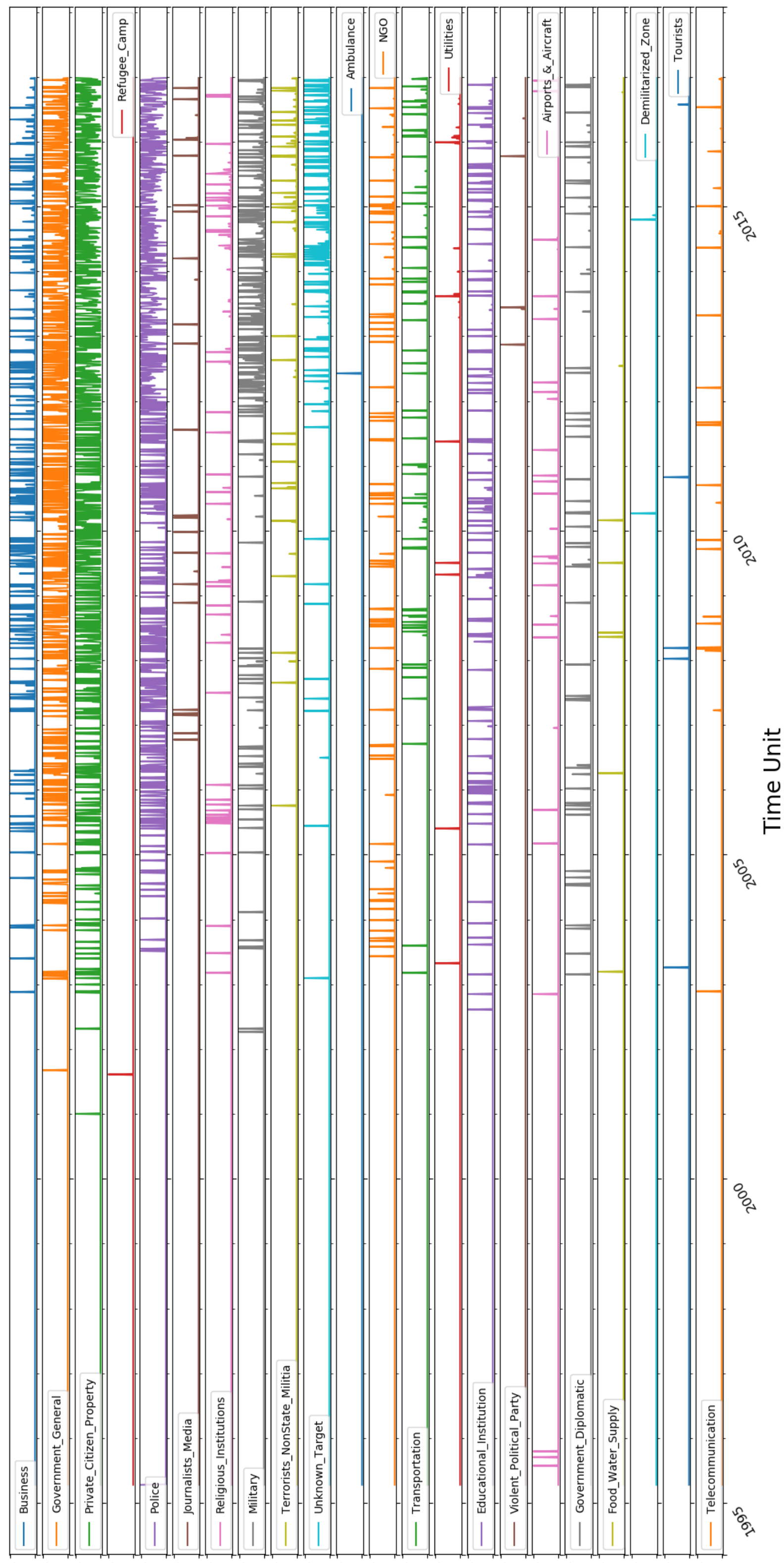

Figure 6.5: Centrality of Targets Over Time - Taliban

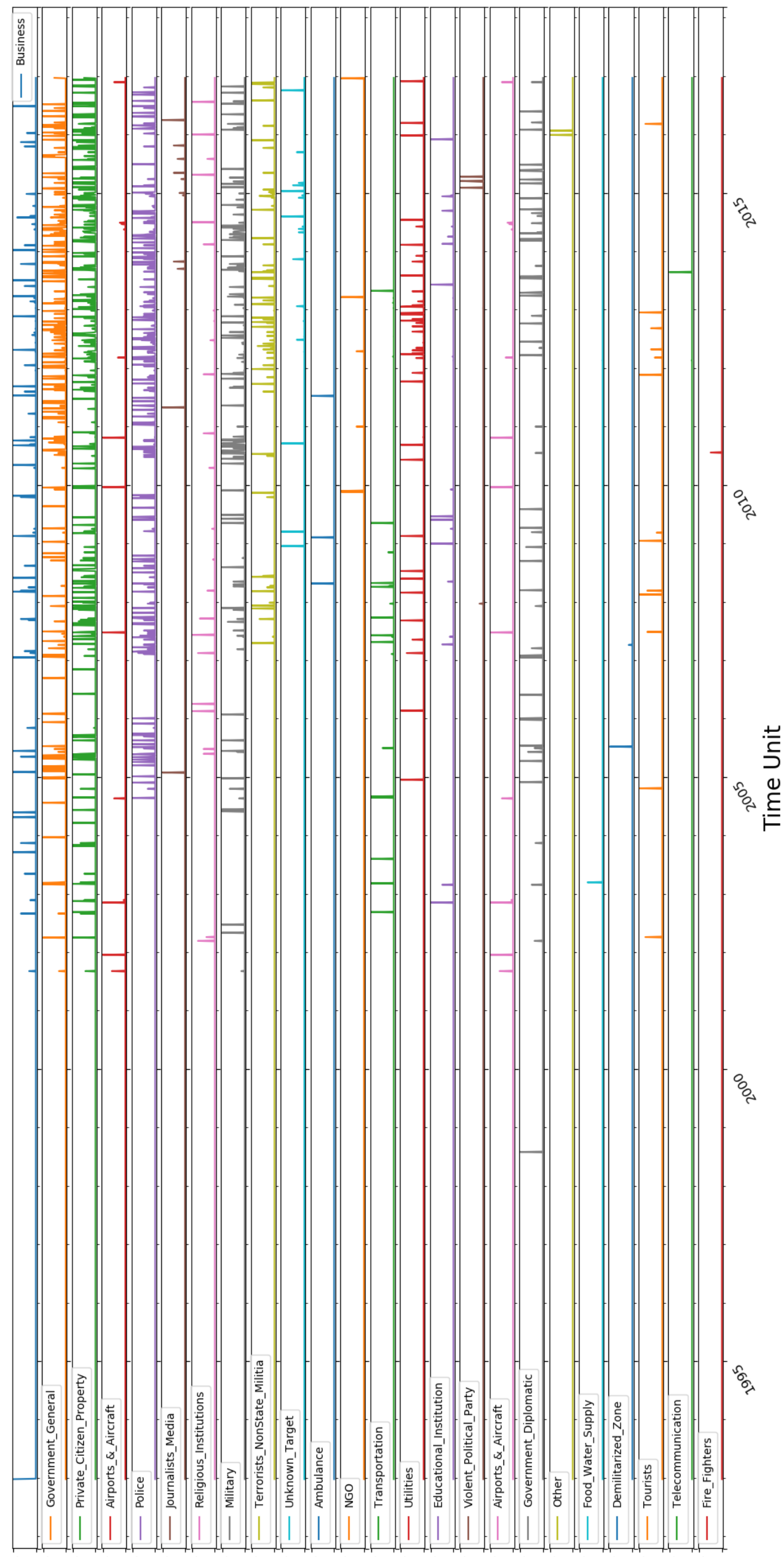

Figure 6.6: Centrality of Targets Over Time - Al Qaeda

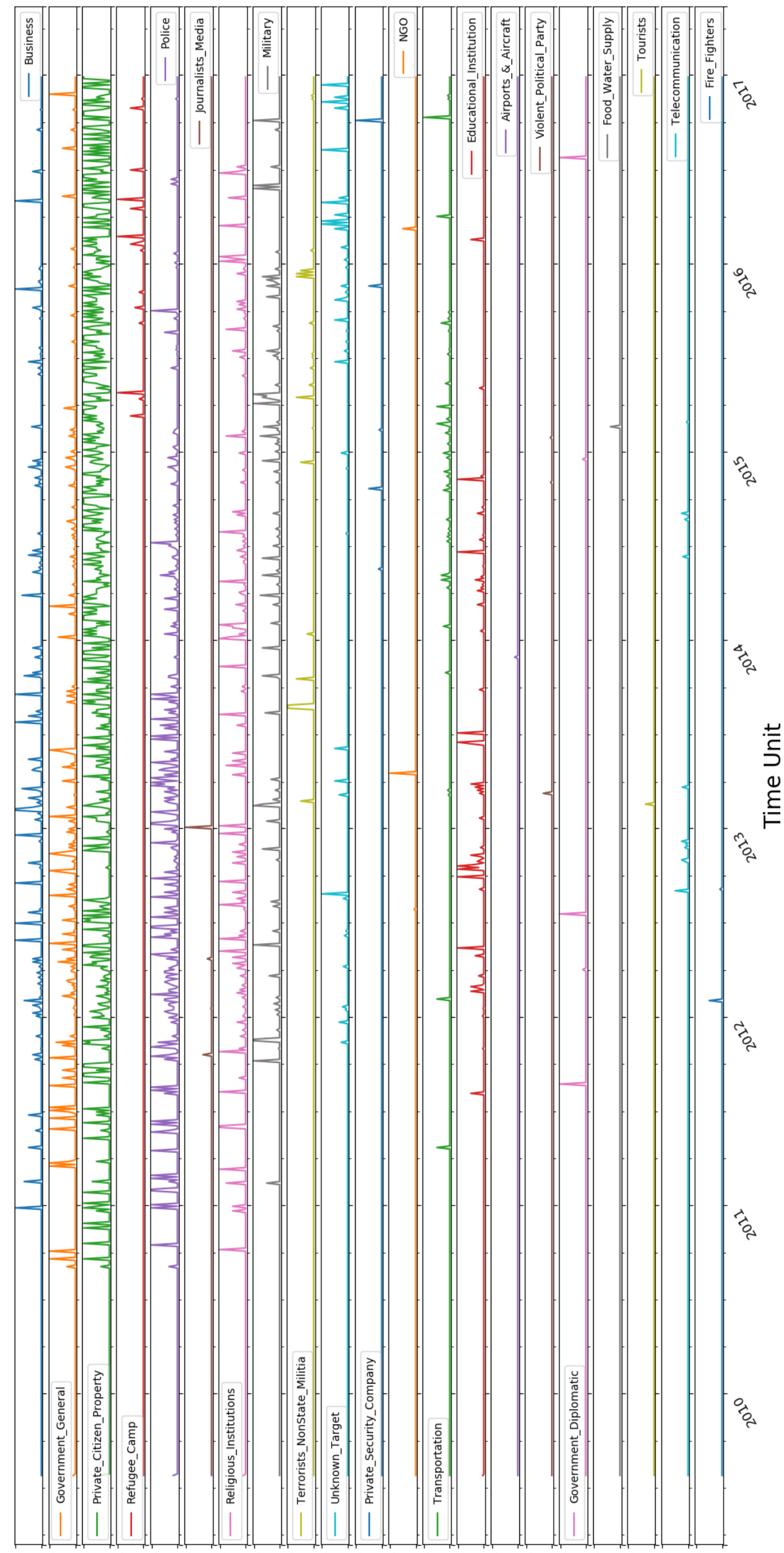

Figure 6.7: Centrality of Targets Over Time - Boko Haram

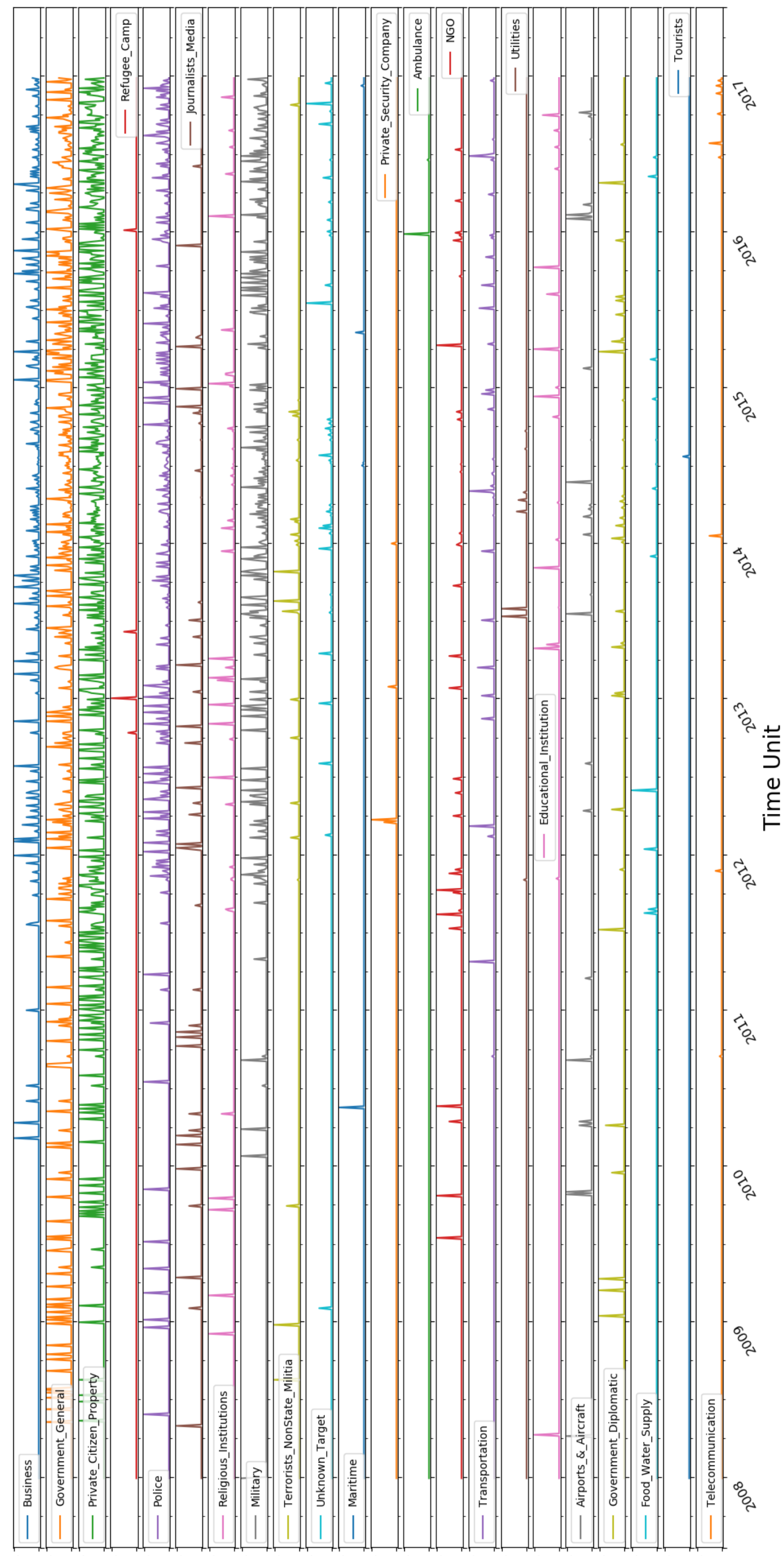

Figure 6.8: Centrality of Targets Over Time - Al Shabaab

6.5 Results of the Models

6.5.1 The Islamic State

The Islamic State with the shortest time sequence overall: however, when taking into account their activity, the group geographically originating in Iraq and Syria shows the highest frequency of attacks. The feature space comprising targeted countries, employed weapons, adopted tactics and hit targets is the less temporally sparse overall. Figure 6.9 provides a graphic visualization of the correlation between each column vectors (i.e., Syria, Firearms, Government (General) and so on) to preliminarily inspect the existing relation between the variables. The plot shows that few features actually show very high (either positive or negative) correlation values in the 62 × 62 matrix. This suggests that the prediction problem might be particularly challenging.

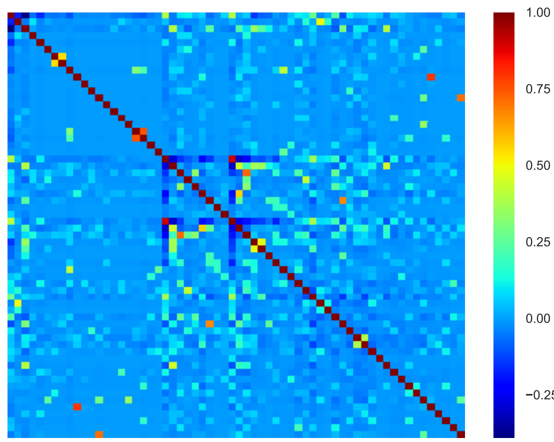

Figure 6.9: Correlation Matrix of Centrality Values (All Features) - Islamic State

To further inspect the correlations to provide more specific knowledge on the issue, Table 6.6 lists the ten highest correlation values between the variables. Trivially, the tactic "Bombing/Explosion" and weapon "Explosives/Bombs/Dynamite" almost reach a perfect correlation of 1. Following, it is interesting to note how the territories of the West Bank/Gaza Strip are strongly related to the targeting of Diplomatic

Figures. Additionally, Indonesia and the Philippines also are particularly correlated, indicating that often when one country is targeted in a given t, the other one also experiences an attack, and vice-versa. With regards to targets, attacks in Jordan are more likely to be directed against Tourists, Journalists/Media are targeted via Facility/Infrastructure attacks and terrorist events in Turkey are generally related to the presence of Refugee Camps, considering that Turkey has shared borders with both Syria and Iraq, the two most attacked countries by the Islamic State.

Feature 1	**Feature 2**	r
Bombing/Explosion	Explosives/Bombs/Dynamite	0.911
West Bank/Gaza Strip	Government (Diplomatic)	0.795
Indonesia	Philippines	0.755
Armed Assault	Firearms	0.726
Jordan	Tourists	0.710
Facility/Infrastructure	Journalists/Media	0.668
Unknown	Unknown	0.571
Egypt	Lybia	0.547
Turkey	Refugee Camp	0.501
Melee	Other	0.480

Table 6.6: Ten Highest Correlation Coefficients Between Features - Islamic State

Before commenting on the results of the models, the reader may appreciate the information provided by Figure 6.10. The heatmap aims to visualize the patterns of each column vector (to be read as "feature") over time in terms of centrality. In the case of the Islamic State, this heatmap clearly shows several things. First, as anticipated above, the Islamic State is responsible for a dramatic frequency of attacks over the considered period. Second, the matrix values are clustered within a small number of vectors. Third, and related to the first point, Iraq (belonging to the Countries dimension of the manifold) Bombing/Explosions (belonging to tactics), Explosives/Bombs/Dynamite (weapons), and Private Citizens and Property (targets) are the four leading vectors in terms of variance within the manifold. This can anticipate that the prediction problem might not be extremely challenging, in the end, at least in terms of $\Phi(T)$ accuracy.

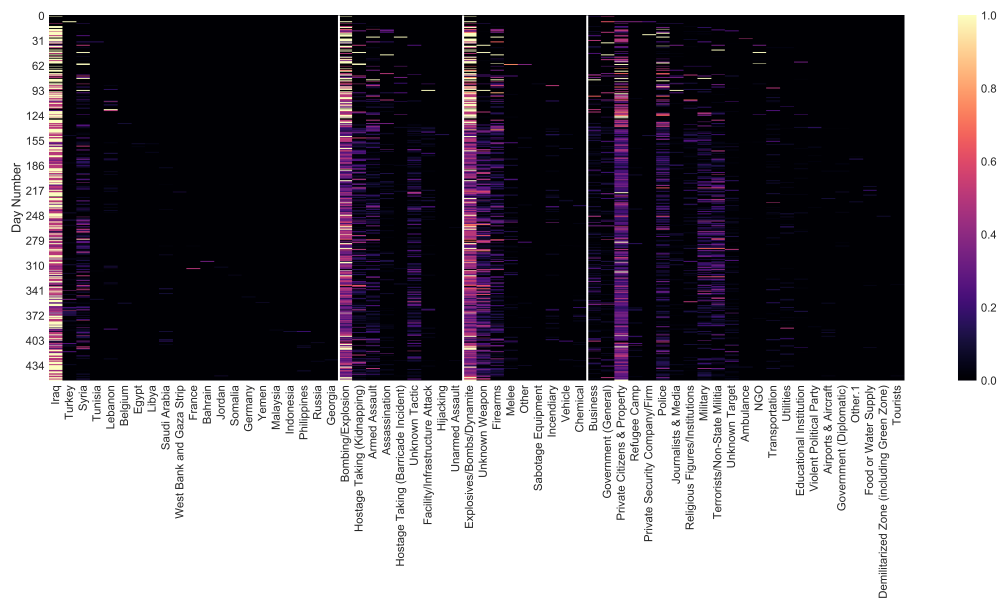

Figure 6.10: Temporal Heatmap - Centrality Over Time (Islamic State). Vertical White Lines Separate Modes

The results obtained from the models are particularly promising. Indeed, all models (with no exception) can identify at least one of the three most central targets in each time unit ($\Phi(T)$). This might be due to the extreme regularity of attacks against Private Citizens and Property that the Islamic State exhibits in the period under observation.

Nonetheless, the results of setwise accuracy $\Gamma(T)$. As noted in the presentation of the metric, it is much more challenging compared to $\Gamma(T)$, and this is showed by the existing disparity between the two, for this particular jihadist group. However, the range in which the prediction fall ($\sim$ 0.4 - 0.75) demonstrates that the deep neural network has considerable power in detecting and revealing the most probable future targets, beyond the easily predictable regularity of Private Citizens and Property.

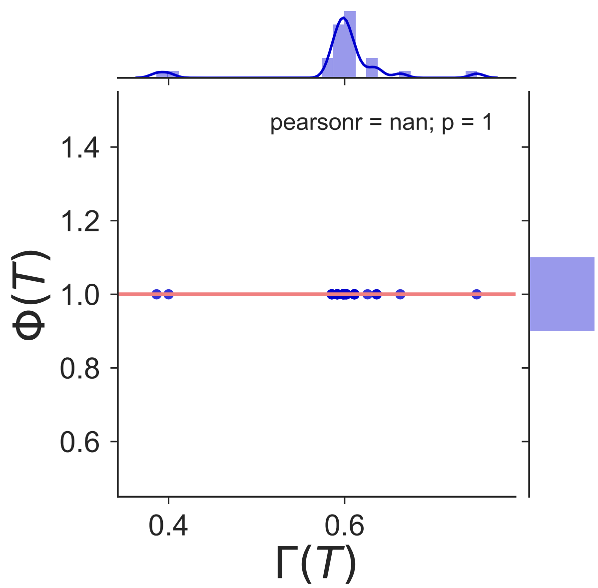

Figure 6.11: Bivariate Relation Between Model Performance in $\Gamma(T)$ and $\Phi(T)$ - Islamic State

Specifically focusing on the best model[6], the parameters and characteristics show that a batch size of 32 performs better than the two other bigger alternatives (64 and 100) and that, in spite of the non-stationarity of considerable number of features around k=30 ($\sim$ 35%), the model performs better when it is fed with data coming from the last 30 points ($\sim$ 3 months).

Parameter	**Value**
Batch Size	32
Look back	30
$\Phi(T)$	1
$\Gamma(T)$	0.75
MSE	0.005
MAE	0.026
Execution Time	$\sim$ 11 min.

Table 6.7: Best Model Performance and Results - Islamic State

[6]For each of the jihadist groups, the best model will be selected based on the best performance in terms of $\Gamma(T)$, as the most demanding metric.

Furthermore, Figure 6.12 and Figure 6.13 show how the loss functions converged quite fast and the models stopped to learn after 140 epochs due to the imposed constraint of the "early stop" option to prevent over-fitting.

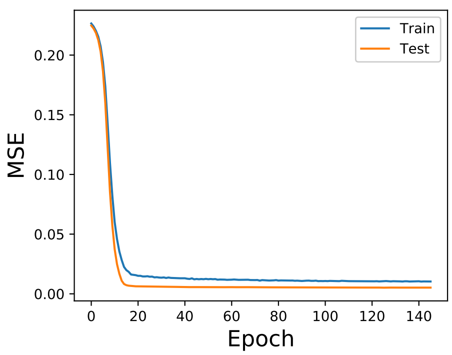

Figure 6.12: MSE - Islamic State

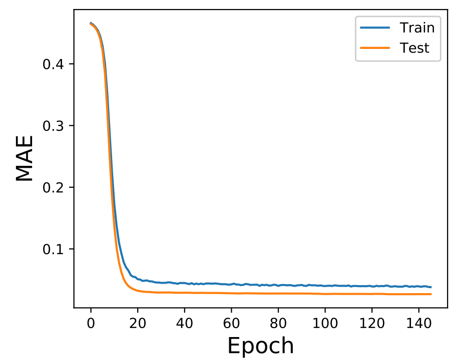

Figure 6.13: MAE - Islamic State

6.5.2 The Taliban

The visual inspection of Figure 6.14 reveals how the Taliban presents some strong relations across the feature vectors of the 39 × 39 matrix. Overall, the Taliban are associated with the lowest number of features to learn from and forecast, this is partially due to the extreme concentration of attacks in Afghanistan (and, marginally, Pakistan).

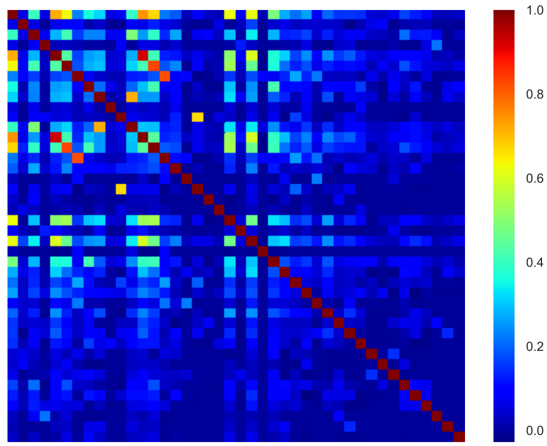

Figure 6.14: Correlation Matrix of Centrality Values (All Features) - Taliban

In detail, the strong role of Afghanistan within the matrix is testified by the information included in Table 6.8: out of the ten highest correlation coefficients, half are related to it. This strong clustering dynamic might lead to two divergent results. It could either reduce the range of information to learn from, bounding variance and therefore posing challenging problems to the forecasting or, contrarily, mark well-determined patterns facilitating the task of the LSTM.

Feature 1	**Feature 2**	r
Bombing/Explosion	Explosives/Bombs/Dynamite	0.927
Armed Assault	Firearms	0.834
Facility/Infrastructure	Incendiary	0.822
Afghanistan	Explosives/Bombs/Dynamite	0.743
Afghanistan	Bombing/Explosion	0.718
Unknown	Unknown.1	0.716
Afghanistan	Firearms	0.671
Unarmed Assault	Chemical	0.669
Afghanistan	Armed Assault	0.633
Afghanistan	Police	0.628

Table 6.8: Ten Highest Correlation Coefficients Between Features - Taliban

Figure 6.15 further allows noting the prominent presence of Afghanistan over time.

The color map also shows how the country has been the most central nearly in each unit of the sequence. Less clear patterns can be detected in the other modes, and especially in the target dimension which accounts for half of the column vectors. Indeed, besides the clear persistent centrality of Police and Private Citizens and Property, other types of targets are much more sparse, volatile or clustered in limited time frames, with high spiking centrality that vanishes after relatively short periods. This is the case of Business-related targets or Religious Figures/Institutions. With regards to tactics, bombing and explosions and armed assaults are stably central over-time, in association with related weapons (i.e., Explosives, Bombs, Dynamite and Firearms).

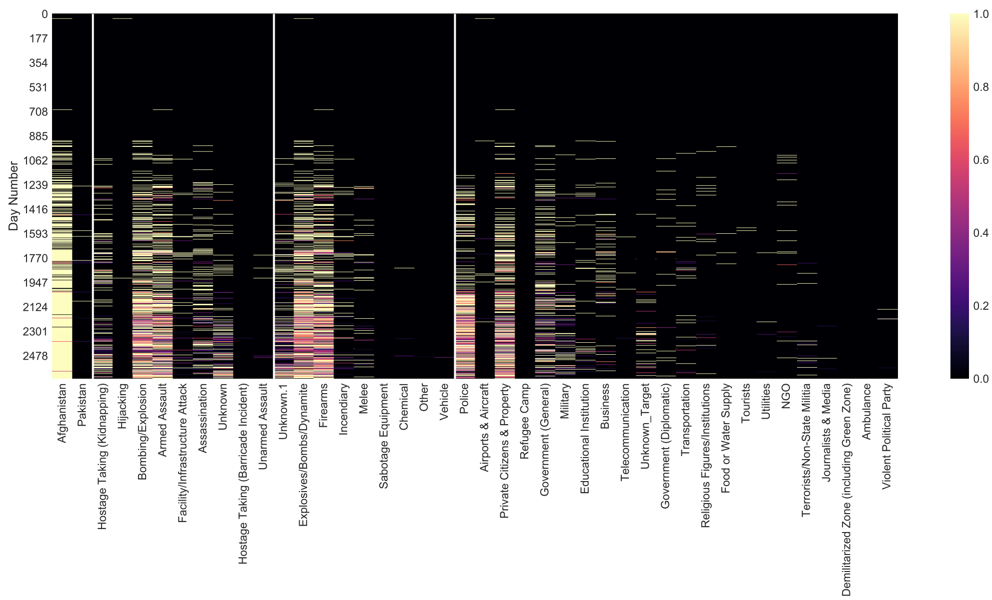

Figure 6.15: Temporal Heatmap - Centrality Over Time (Taliban). Vertical White Lines Separate Different Modes.

Also in the case of the Taliban, the results are encouraging. Figure 6.16 shows that the performance of each model is concentrated between ~ 0.95 and ~ 0.98 for $\Phi(T)$ and ~ 0.75 and ~ 0.76 for $\Gamma(T)$. The relation between the two metrics is positive linear: as one metric increases in its power, so does the other.

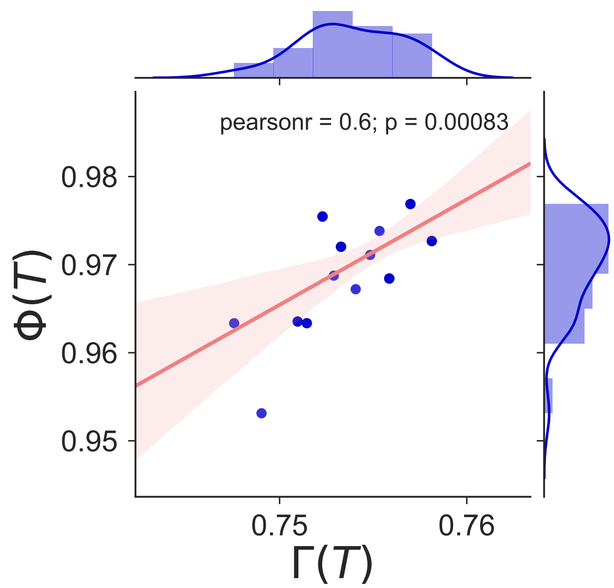

Figure 6.16: Bivariate Relation Between Model Performance in $\Gamma(T)$ and $\Phi(T)$ - Taliban

The best model and its characteristics are illustrated in Table 6.9: the deep LSTM which achieved the highest set-wise accuracy was trained using a look back of 10: thirty days of history are sufficient for learning patterns and forecast them in the most accurate way across all the twenty-eight networks. The implication of this result is that Taliban hold medium-range dependencies between the present and the past: the Taliban seem to be consistent and homogeneous in their behaviour for some weeks.

Parameter	**Value**
Batch Size	32
Look back	10
$\Phi(T)$	0.758
$\Gamma(T)$	0.973
MSE	0.053
MAE	0.126
Execution Time	$\sim$ 15 min.

Table 6.9: Best Model Performance and Results - Taliban

The behavior of the loss functions shows no particular sign of over-fitting. Contrarily, the performance of MAE might even suggest that there is still room for improvement since the curves are parallel and do not converge over the epochs, as early stop parameters stopped the process after only 80 epochs.

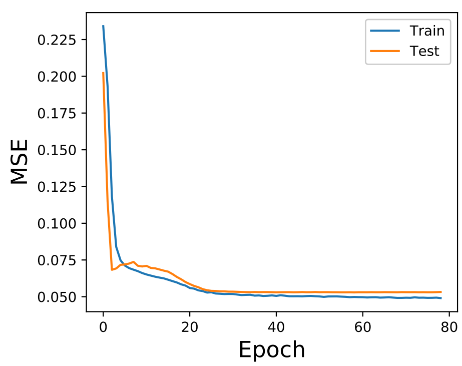

Figure 6.17: MSE - Taliban

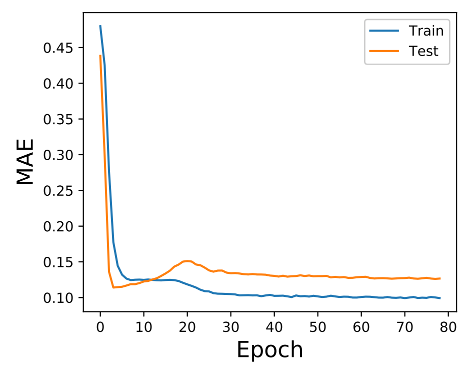

Figure 6.18: MAE - Taliban

6.5.3 Al Qaeda

Al Qaeda has demonstrated to be the most challenging group overall in terms of forecasts. The 72 × 72 matrix is the most sparse and the one with most features among all the considered groups: the dimensionality of the data certainly played a role in the performance of the models. Figure 6.19 shows how most of the vectors do not show any relation. Notably, very few pairwise relations show values higher than 0.5. The upper left region of the matrix is almost perfectly uncorrelated. This introductory description already anticipates some of the challenges that the models had to face in reaching their goals.

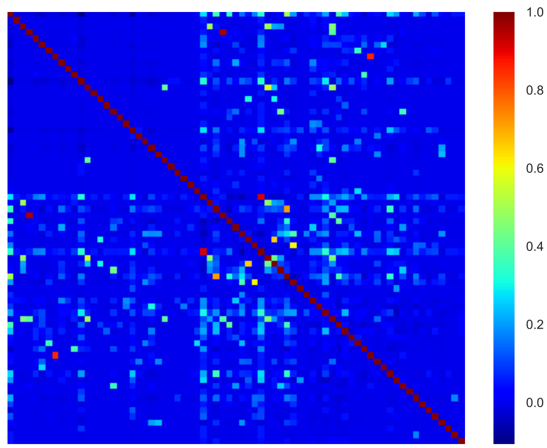

Figure 6.19: Correlation Matrix of Centrality Values (All Features) - Al Qaeda

Focusing on the highest correlations (Table 6.10, and going beyond the most trivial and expectable relations, among the top ten ones only one actually refers to a specific target. Indeed, Airports and Aircraft are highly associated with the United Kingdom (r=0.495). It is also interesting to note that when information on tactics is not available for an attack, generally the same applies to information on weapons. Data also show a prominent tendency of Al Qaeda to plot facility/infrastructure attack by means of incendiary tactics.

Feature 1	**Feature 2**	r
France	Unarmed Assault	0.957
Bombing/Explosion	Explosives/Bombs/Dynamite	0.909
International	Maritime	0.857
Armed Assault	Firearms	0.718
Facility/Infrastructure	Incendiary	0.659
Unknown Tactic	Unknown Weapon	0.608
United Kingdom	Vehicle	0.567
Yemen	Firearms	0.509
United Kingdom	Airports and Aircraft	0.495
United States	Vehicle	0.491

Table 6.10: Ten Highest Correlation Coefficients Between Features - Al Qaeda

Figure 6.20 shows how that, in spite of the dimension of the matrix, the majority of information is clustered around a relatively small number of feature vectors. In fact, the first mode of the manifold, which includes countries, is scarcely populated (with the most attacks plotted in Iraq and Yemen). More heterogeneity can be found in the mode containing tactics and weapons. However, the information might not be sufficiently rich to meaningfully infer patterns within the target mode, which is instead more diverse (at least, from a visual inspection of the heatmap).

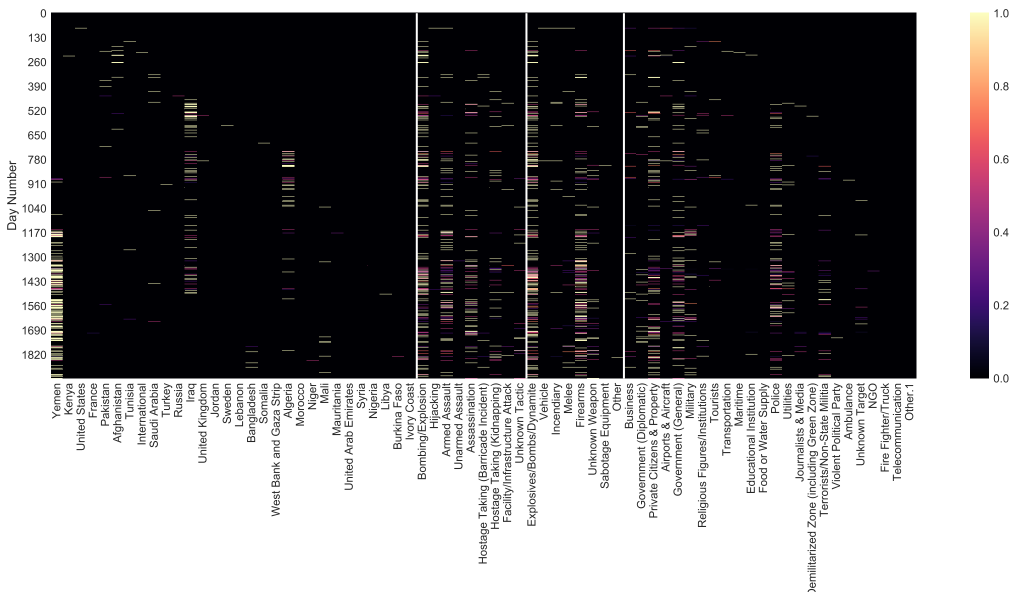

Figure 6.20: Temporal Heatmap - Centrality Over Time (Al Qaeda). Vertical White Lines Separate Different Modes.

Figure 6.21 shows the bivariate relation between the two different types of accuracy metrics used to assess the model predictive ability. It is worth to note how there is a small number of models that perform poorly in both measures. However, much of the models get results that are clustered in the upper region of the plot. Nonetheless, the forecasting models for Al Qaeda are certainly the most problematic ones. Both $\Gamma(T)$ and $\Phi(T)$, also when only considering the best model, have the lowest values across all groups. A connection might be drawn from this findings to the conclusion of Moghadam (2013), who claims that the more decentralized a group is, the more it innovates in terms of actions and attacks: Al Qaeda is certainly the most decentralized group in this analysis (as it contains several sub-groups), and innovation might be here seen as the degree of sparsity and low-correlation between feature vectors. Furthermore, as noted by Pham (2011), the Al Qaeda in the Islamic Maghreb faction has become over time more pragmatic also in terms of attack planning (and not only in relation to its geographical scope of action), and this aspect also can be reflected in the data.

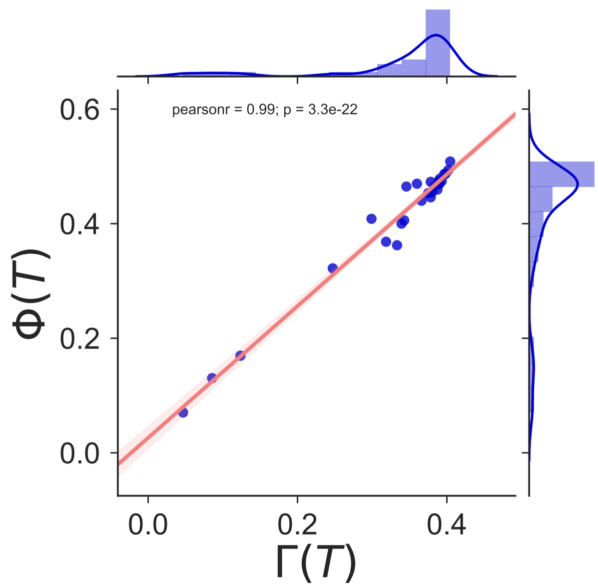

Figure 6.21: Bivariate Relation Between Model Performance in $\Gamma(T)$ and $\Phi(T)$ - Al Qaeda

The best model for Al Qaeda is obtained via a minibatch size of 32, as seen with the Islamic State and the Taliban (Table 6.11). The deep LSTM achieves the highest result in $\Gamma(T)$ when using a wide time range to infer future behaviors: the look back is equal to 50 (150 past days), meaning that the network is more efficient when can access information on quite long past sequences. This result may suggest how Al Qaeda tends to frequently change its behavior (mapped by the 72 multivariate times series) over time, preventing the network to learn efficiently on short windows. Figures 6.22 and 6.23 interestingly show a fast convergence towards a minima after very few epochs. In the MAE plot, it can be noted of the curves started to diverge after 45 epochs, however, the early stop parameter stopped the learning process after 80 epochs, avoiding further risks of over-fitting.

Parameter	Value
Batch Size	32
Look Back	50
$\Phi(T)$	0.508
$\Gamma(T)$	0.404
MSE	0.020
MAE	0.044
Execution Time	$\sim$ 11 min.

Table 6.11: Best Model Performance and Results - Al Qaeda

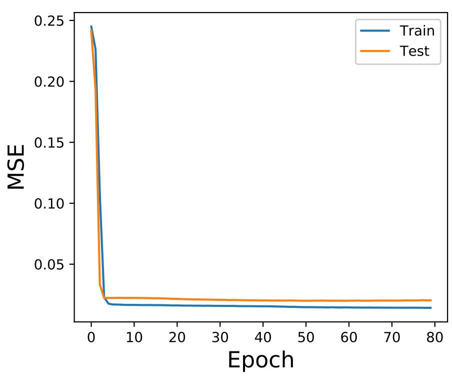

Figure 6.22: MSE - Al Qaeda

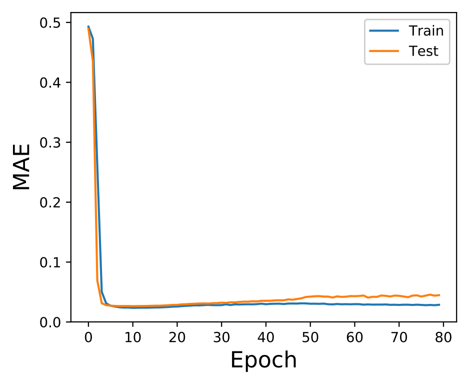

Figure 6.23: MAE - Al Qaeda

6.5.4 Boko Haram

The introductory graphic visualization of the correlation between the feature vector comprised in the 42 $\times$ 42 matrix of Boko Haram illustrates how, in spite of the vast majority of vectors not showing any correlation, there exist some regions in which a relatively strong relation holds for certain variables (see, for instance, the bottom left or the top right of the matrix). On the contrary, the bottom right of the lower triangle, where targets are located, is almost completely uncorrelated: this may suggest how, in any given time unit, Boko tend not to combine different types of targets for their terrorist attacks.

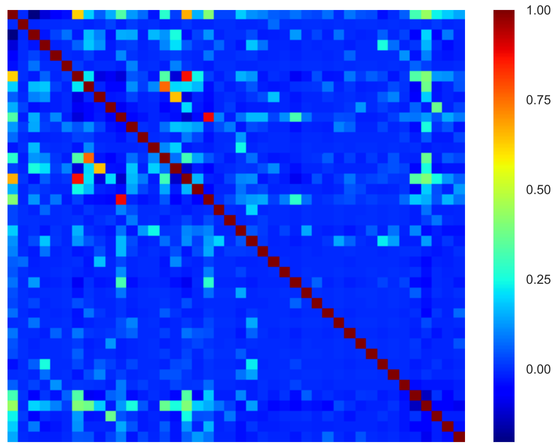

Figure 6.24: Correlation Matrix of Centrality Values (All Features) - Boko Haram

Nigeria - the country that was most affected by the attacks of Boko Haram - is strongly related to events against Private Citizens and Property. This target is also connected to the use of firearms through armed assault tactics. (Table 6.12).

Feature 1	**Feature 2**	r
Bombing/Explosion	Explosives/Bombs/Dynamite	0.957
Armed Assault	Firearms	0.909
Facility/Infrastructure	Incendiary	0.857
Nigeria	Firearms	0.718
Unknown Tactic	Unknown Weapon	0.659
Nigeria	Armed Assault	0.608
Nigeria	Private Citizens and Property	0.567
Nigeria	Explosives/Bombs/Dynamite	0.509
Armed Assault	Private Citizens and Property	0.495
Firearms	Private Citizens and Property	0.491

Table 6.12: Ten Highest Correlation Coefficients Between Features - Boko Haram

The temporal heatmap in Figure 6.25 indeed shows the consistent high centrality of Nigeria: other countries started to appear later (more than one year after the first attack in Nigeria). The attacks of Boko Haram are carried out using a variety of tactics, as visible in the second mode of the manifold: besides armed assault and bombings or explosions, which are common to most groups, many time units experience significant centrality of tactics such as Hostage Taking, Incendiary attacks and also Facility/Infrastructure. The heterogeneity of tactics does not reflect into the heterogeneity of weapons: Boko Haram consistently oscillates between Firearms, Explosives, and Melee. Private Citizens and Property dominates in the mode of targets. The mode also shows sparse and rare signals on several targets, with apparently no evident time clustering.

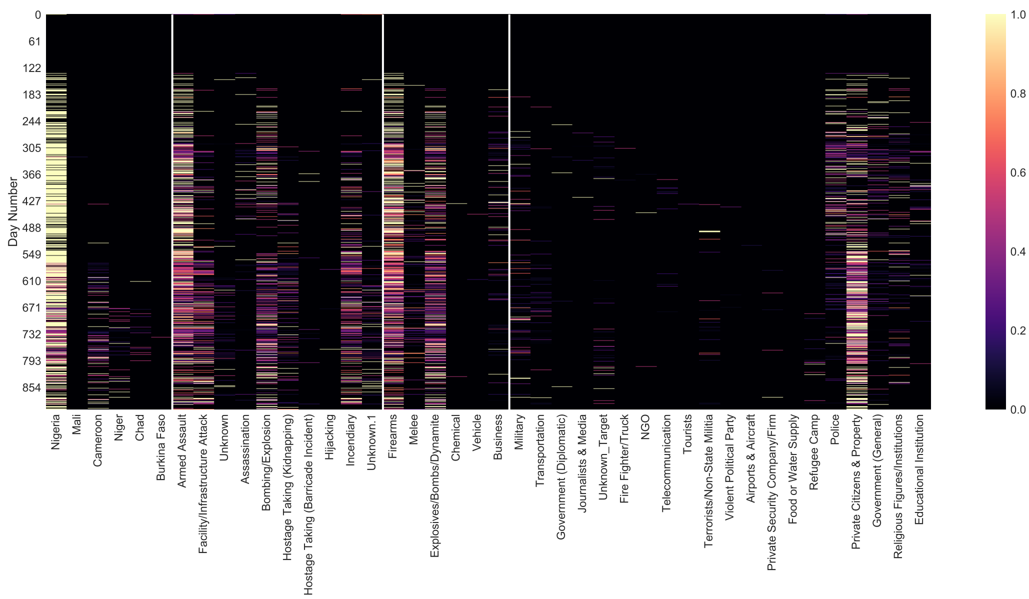

Figure 6.25: Temporal Heatmap - Centrality Over Time (Boko Haram). Vertical White Lines Separate Modes

Concentrating on the models, their behavior seems promising in spite of the problems with stationarity failing fastly when wider lags are introduced. Nevertheless, the relation between $\Phi(T)$ and $\Gamma(T)$ is anomalous: in the two-dimensional space, the points representing the models almost form a circle: the relation is not certainly linear. As a matter of fact, the detected Pearson coefficient is not statistically significant. Those models that perform better in terms of $\Phi(T)$ are not automatically among the subset of those that also perform better in $\Gamma(T)$.

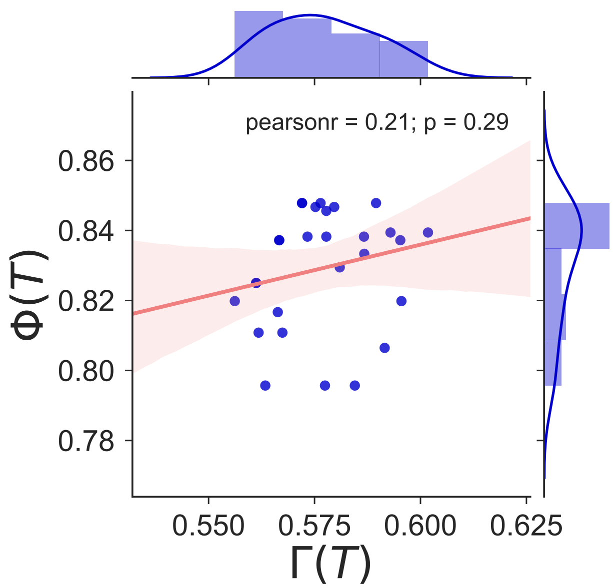

Figure 6.26: Bivariate Relation Between Model Performance in $\Gamma(T)$ and $\Phi(T)$ - Boko Haram

The model with the best performance in $\Gamma(T)$ has different characteristics from the other best models commented for the other groups. First, its batch size is equal to 64, going against the results of the other models and the suggestions contained in the paper by Masters and Luschi (2018).

In addition, it relied on a short look back: this might suggest that Boko Haram changes frequently its behavior that might have a microcycle-shaped nature. For this reason, the LSTM only needs data from the past 6 days to infer the future targets, as then more distant information in time could not be relevant or even noisy and disruptive of the learning process.

Parameter	Value
Batch Size	64
Look back	2
$\Phi(T)$	0.839
$\Gamma(T)$	0.601
MSE	0.046
MAE	0.103
Execution Time	$\sim$ 5 min.

Table 6.13: Best Model Performance and Results - Boko Haram

The MAE loss function decay (Figure 6.28 again testifies the important role of the "early stop" option to avoid overfitting: 100 epochs are sufficient to the LSTM to achieve the results. In the case of MSE (Figure 6.27, the learning curves are overlapping up to epoch 20, and then separate and remain in parallel, with the same trend.

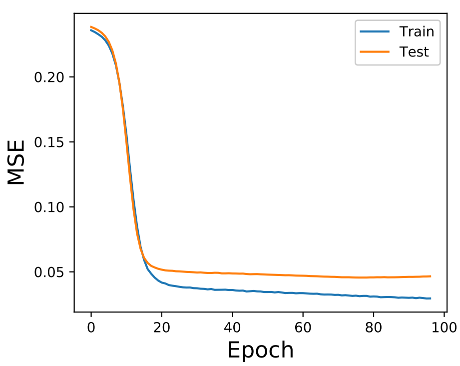

Figure 6.27: MSE - Boko Haram

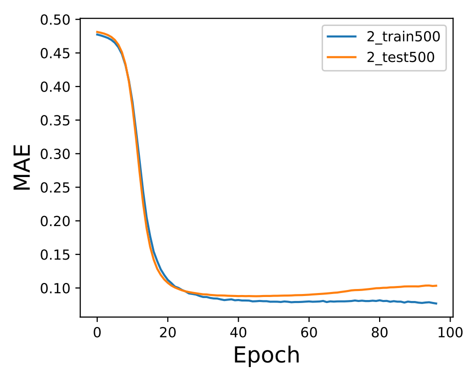

Figure 6.28: MAE - Boko Haram

6.5.5 Al Shabaab

Al Shabaab is the last group of the sample. The correlation plot between the feature vectors that make up the 45 $\times$ 45 matrix shows more clear correlated sub-regions than the previous ones: though the extremely high values seem not so frequent, there are several clustered areas where the coefficient floats around 0.25-0.5 (Figure 6.29).

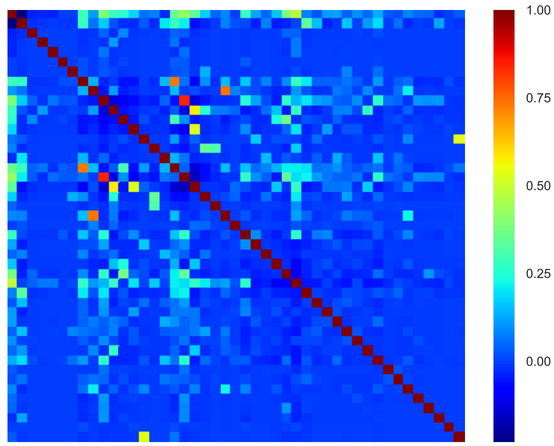

Figure 6.29: Correlation Matrix of Centrality Values (All Features) - Al Shabaab

The list of the highest correlation coefficients within the matrix shows how, in terms of targets, Somalia (which is the country that accounts for the most attacks plotted by Al Shabaab) is strongly related to events against Private Citizens and governmental entities (Government (General)). Indeed, the first column vector in the lower triangle of Figure 6.29 maps the relation of Somalia with other variables and it clearly shows how the country accounts for a large number of positive strong associations. Another strong pairwise connection is the one between the Hijacking tactic and Ambulance targets (Table 6.14).

Feature 1	**Feature 2**	r
Bombing/Explosives	Explosives	0.844
Incendiary	Facility/Infrastructure	0.740
Armed Assault	Firearms	0.732
Hostage Taking	Unknown Weapon	0.578
Hijacking	Ambulance	0.539
Unknown Tactic	Unknown Weapon	0.533
Private Citizens and Property	Somalia	0.481
Explosives/Bombs/Dynamite	Somalia	0.426
Bombing/Explosives	Somalia	0.389
Government (General)	Somalia	0.382

Table 6.14: Ten Highest Correlation Coefficients Between Features - Al Shabaab

The temporal heatmap below (Figure 6.30) has similarities with the Boko Haram one: indeed, almost half of the manifold is occupied by targets, as a consequence of the actions of a group that shows no particular variety of adopted tactics and weapons and that act on a regional scale. Somalia again demonstrates its consistent stability in the time series, with a marginal role of Kenya, especially in the second half of the considered period. Two tactics are prominently present: armed assaults and bombing/explosions, but it is worth to note how Al Shabaab seems to plan attacks or short campaigns that combine more than one tactic, as testified by the color of the temporal units of tactics: many are darker colors, indicating lower centrality values and, therefore, a non unitary behavior. The same, similarly, can be said for Firearms and Explosives. The mode occupied by targets, instead, is less patterned. In spite of the recurrent centrality of Military, Government (General) and Private Citizens, Al Shabaab also directed its attack to other types of targets that, although sparsely, were noticeably central in certain time windows.

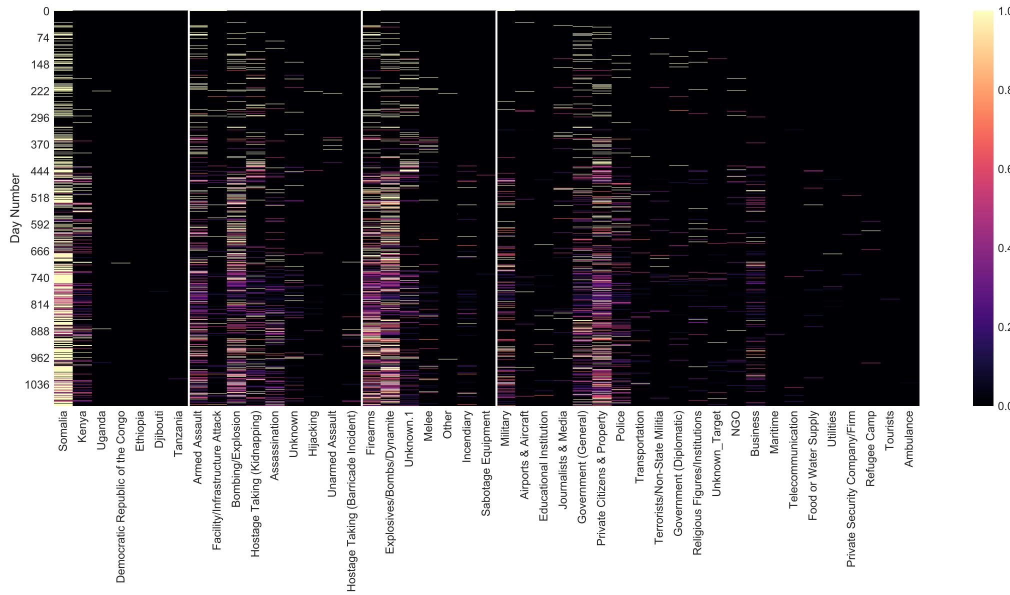

Figure 6.30: Temporal Heatmap - Centrality Over Time (Al Shabaab). Vertical White Lines Separate Modes

The behavior of the models in terms of $\Phi(T)$ and $\Gamma(T)$ is showcased in the regression scatter plot in Figure 6.31. Besides an outlier model that performed significantly worse than all the others in both metrics, most of the models are tightly clustered in the upper right region. The values of $\Phi(T)$ fall within the 0.8-0.9 range, demonstrating a very high capacity of forecasting at least one of the most central targets in the next time unit. In terms of $\Gamma(T)$, the LSTM show even small variance: most points are centered around ~ 0.6, which is also a promising outcome considering the challenging construction of the metric.

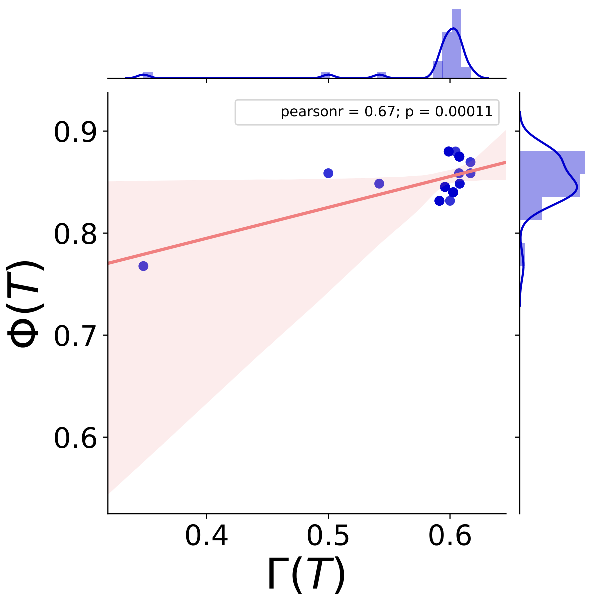

Figure 6.31: Bivariate Relation Between Model Performance in $\Gamma(T)$ and $\Phi(T)$ - Al Shabaab

In spite of the similarities detected comparing Al Shabaab's and Boko Haram's temporal heatmaps, the best model for the Somali jihadist organization has different characteristics (though the performance is similar between the two groups). Al Shabaab is the group that achieved the best result through the smallest batch size overall (2): this result justifies the longest execution time ($\sim$ 49 minutes). Furthermore, the best look back is the same as the Taliban: the last month of data is the right trade-off for the LSTM to obtain the most accurate predictions.

Finally, Figures 6.32 and 6.33 reports the learning process behaviour through the

Parameter	**Value**
Batch Size	2
Look back	10
$\Phi(T)$	0.859
$\Gamma(T)$	0.617
MSE	0.026
MAE	0.081
Execution Time	$\sim$ 49 min.

Table 6.15: Best Model Performance and Results - Al Shabaab

trend of MSE and MAE: the deep neural networks converge on a very low local minima after one epoch for train set, while it takes about 10 epochs in the MSE case for the test set: the training curve then continues to decrease, while the learning process of train set curves almost stops or deteriorates over time. The test curve in the Figure 6.33 displays an anomalous trend, since it starts very low, increasing then the error around epoch 10, and subsequently improving its performance up to the final epoch. The "early stop" option prevents over-fitting by stopping the process around epoch 50.

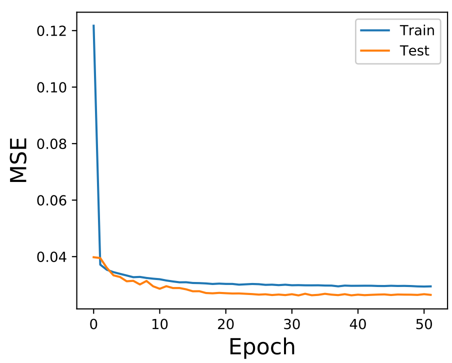

Figure 6.32: MSE - Al Shabaab

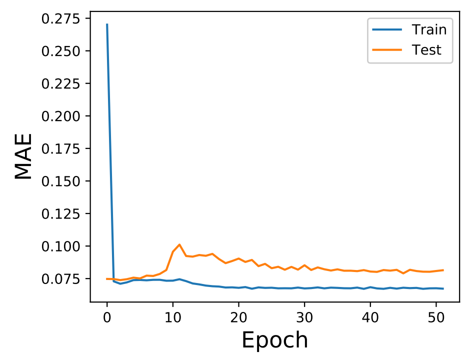

Figure 6.33: MAE - Al Shabaab

6.5.6 Performance Summary

This last subsection aims at summarizing the results across the models, to more comprehensively show the performance of the models (with a special focus on Look Back) and to possibly anticipate future directions.

Figure 6.34 provides a 2d Kernel Density Estimation plot to show how, regardless of the group, the models performed overall to understand if there exists a general underlying pattern in terms of $\Phi(T)$ and $\Gamma(T)$. The plot displays the presence of a positive linear relationship: while some models performed poorly (bottom left), most data point concentrates in the upper-right region, falling in the $\sim$ 0.5-0.6 range for $\Gamma(T)$ and 0.8-1 for $\Phi(T)$. While this result has been partially introduced in the previous subsections specifically dedicated to each group, it is worth to note that, at a more aggregated level, this finding supports the optimistic nature of the LSTM networks for target forecasting.

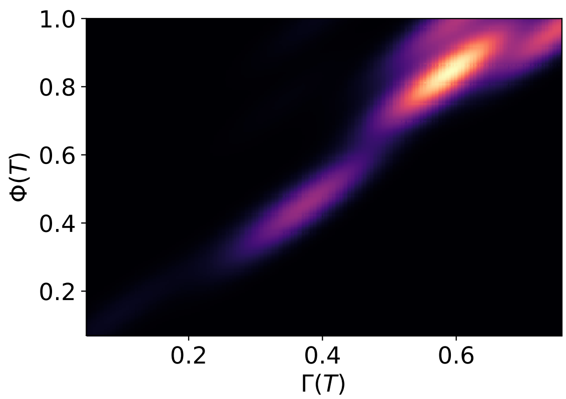

Figure 6.34: Two-dimensional Kernel Density Estimation of $\Phi(T)$ and $\Gamma(T)$ Across All Models

Finally, Figures 6.35 and 6.36 report the distribution of each look back size (regardless of the jihadist group) across models. Interestingly, the distributions for $\Phi(T)$ look very similar from a look back equal to 1 up to 30. The 40 case is specifically dedicated to the Islamic State since due to the small size of the test set I could not rely on the ordinary maximum look back of 50. This latter look back indeed shows the lowest median value overall, while 30 slightly outperforms all the others. This finding might suggest that, when only accounting for one out of three most central targets, there is no necessary need to consider long past sequences. Instead, the pay-off of choosing a small portion of the group's behavioral history is sufficient to achieve very good results.
The situation for what concerns $\Gamma(T)$ is, instead, less clear, probably due to the fact that the challenging nature of the metric does not lead to a "one size fits all" result, making it highly dependent upon each group's history. Overall, what emerges is that look back of 1 and 2 have significantly more compressed distributions, and that the median value is almost identical across all configurations. What changes is the range between lowest and highest performances: 3, 30 and 50, in this case, reach the highest peaks of $\Gamma(T)$, but at the same time their worst results are the worst overall. This may indicate that the bigger the look back, the higher the risk of noisy variability that may affect the learning process.

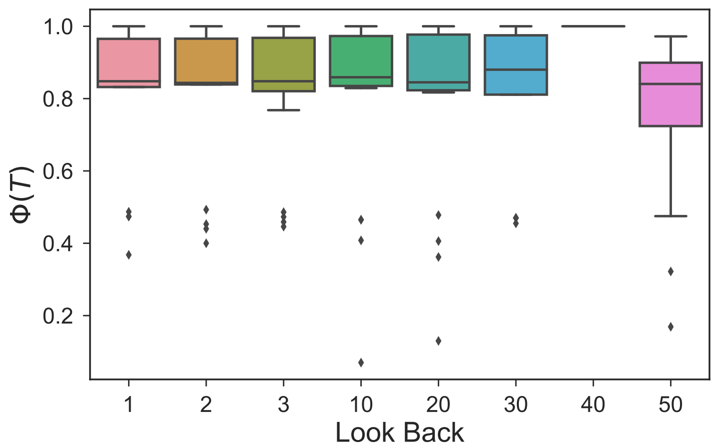

Figure 6.35: Box Plot of Look Back Performance in Relation to $\Phi(T)$ Across all Models and Groups

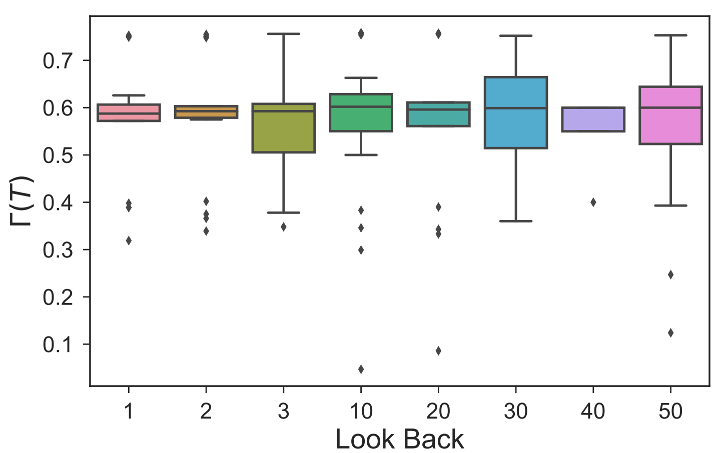

Figure 6.36: Box Plot of Look Back Performance in Relation to $\Gamma(T)$ Across all Models and Groups

6.6 Overcoming Issues on Weak and Rare Signals

There are opposing views about the ability of predictive analytics to prevent terrorism. When discussing about predictive analytics we can include all those mathematical or computer models that try to predict or forecast terrorist events, perpetrators, dynamics. Scholars are divided into different factions: some are more optimistic (with various degree of cautiousness) (Chen et al., 2008; Haberman and Ratcliffe, 2012; Siegel, 2016), while others have less positive expectations on the power of these computational instruments to actually help the fight against terrorism (and, more generally, crime) (Weber and Bowling, 2014; Bouchard, 2017; Hälterlein and Ostermeier, 2018). It is not just a matter of the debate fostered around the unethical behavior of certain algorithms and the damages that these instruments can pose to certain parts of the societies (Berk et al., 2018; Hannah-Moffat, 2018; PAI, 2019). It is, instead, a debate that revolves around some reasoning on the statistical power of these models. The negative opinions are generally centered around three critical points: (1) the lack of meaningful and solid data to be used by the scientific community (Sageman, 2014), (2) the insufficient training or computational skills of intelligence analysts training or skills of intelligence analysts (Sageman, 2014; Britten, 2018) and, finally, (3) the intrinsic and non-solvable problems of forecasting things that are simply unpredictable. Specifically, concerning the latter, Munk (2017) claimed that the quality of the data and other statistical problems make it very difficult to develop and deploy reliable models for predicting, for instance, potential terrorist actors within a population. The author estimates a potential cost of 100,000 false positives per every real terrorist found by an algorithm. While this estimate should be carefully verified, and predicting terrorism is not just about "predicting" the next lone-wolf or the next radicalized individual, it is surely true that the scientific community should invest much more on researching ways to reduce the cost of false positives (and even false negatives) and, in general, should focus on finding ways to detect those events, people or entities that, in the strict statistical sense, we call "outliers".

In my experiments, the risk of losing outliers is high. I will focus on the shortcomings of my analyses in the next section, but this consideration led me to reason about how to frame, evaluate and potentially solve this issue in the future. The data that I use only map the centrality of entities over time, starting from the single event as the original bit of information. As a recap, an event can be abstractly and compactly defined as a combination of (1) a certain group responsible for the attack, (2) a certain country that was hit by the event, (3) an employed tactic (up to three) to carry

out the attack, (4) a weapon (up to four) used in the tactical plan and (5) a target (up to three), against which the event is directed. This is a flexible and meaningful way to define an attack: it potentially incorporates rich features that - as seen by the model results - can be learned by an algorithm to make inferences about the future. Nonetheless, criminologists and terrorist researchers might reasonably raise a crucial point, and tightly connected to the issue of outliers. I am referring to the absence of any information regarding the impact of the attack. In fact, in the time series, every attack is only distinguishable based on the abovementioned features, but no distinction is made based upon the number of fatalities, casualties or the economic damages that follows a violent event.
This is extremely relevant: leaving that kind of information out of the picture poses the risk of losing large-scale attacks in the stream of data that are fed into the models. Figure 6.37 helps illustrate the problem.

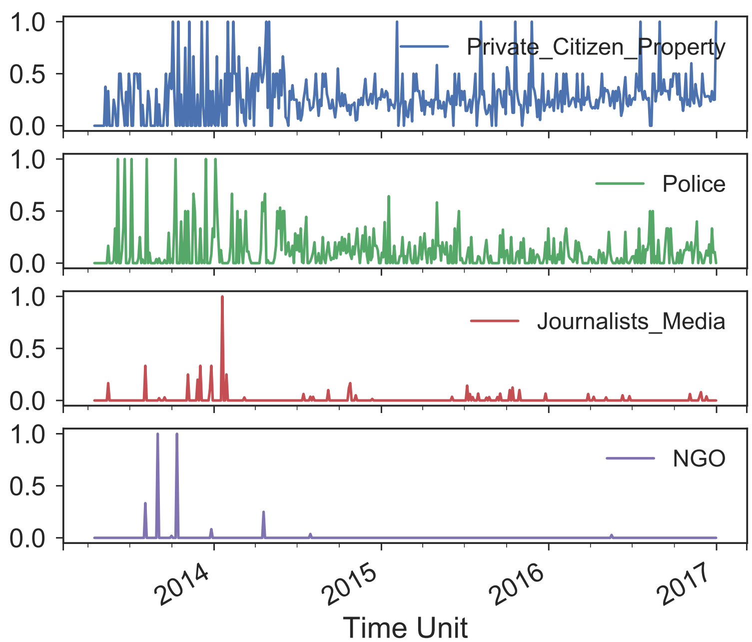

Figure 6.37: Islamic State Sample Comparison of Two High-Frequency and Two Weak Signals

The two top subplots (Private Citizens and Property and Police) are examples of strong signals that are persistently present across the whole time frame. The other two subplots, instead, describe a very different situation. The case of Journalists and Media can be defined as an example of a "weak" signal: the signal, in fact, is not as persistent as the previous two and, additionally, its centrality tends to be very low in magnitude also in those periods where the specific target category raises in popularity (e.g., second semester of 2015). Finally, NGO is a perfect example of a "rare" signal. Over the years, very few time units experience attacks against non-governmental organizations, and the distribution of the centrality values is too sparse to reasonably expect meaningful inferences.

As anticipated, being a weak or rare signal can hide numerous and serious pitfalls. What if the constant and persistent signals are associated with small-scale and low-impact attacks, while events belonging to rare signals are correlated with massive and high-impact plots? (Martens et al., 2014). The algorithms, fed only with the information I have described earlier in the chapter, cannot capture the underlying dynamics that govern such hidden bits of information.
As a matter of fact, 9/11 belongs to this latter family of signals: here resides the importance of putting efforts and attention to address this problem. Interesting and very well-grounded research exist in terms of estimating the probability of rare events (Glasserman et al., 1999; Haan and Sinha, 1999; King and Zeng, 2001; Reijsbergen et al., 2013), also in terrorism (with a specific attention also to sever ones) (Clauset et al., 2007, 2010; Johnson et al., 2011; Clauset and Woodard, 2013), as well as extensive research on outlier detection and modelling (Rousseeuw and Leroy, 1987; Hodge and Austin, 2004; Ben-Gal, 2005; Campos et al., 2016). Nonetheless, I am here proposing a heuristic method that might serve as a baseline for future work on weak and rare signals in temporal dynamics within the realm of terrorism research.
This method is inspired by how humans memorize and learn how to deal with rare but impactful events. Though in the context of artificial intelligence the expression "memorizing is not learning" resonates, memorizing can be a crucial first step for selecting meaningful information that should then be learned subsequently by a model. Going back to humans, we are generally particularly good at memorizing the effect that impactful rare events may have on our lives, as to modify our future behaviors. This may happen with rare diseases: once an individual gets curated, it will always remember how that disease changed his or her existence. Lotteries are another example: the winner of a monetary prize (a reasonably high prize) will always carry memory about the happiness and excitation felt after the announcement.
These are extremely rare events that are likely not to get repeated over one's life. However, there are other types of events - such as winning sports competitions, get-

ting job promotions, falling in love - that can occur rarely but more than once in our lives. These might belong to the class of "weak" signals, and nonetheless we, as humans, are able to memorize the event, its consequences, our feelings so that, eventually, we may adapt our behaviour in the future according to that experience (e.g., a soccer player that wins a Champions League with a non-top tier team will not only memorize the glory and joy of the moment but also all those behaviors and facts that, according to his/her vision, led to the final result). This process can be associated with the so-called "attentional boost effect" found in psychology and cognition experiments (Swallow and Jiang, 2012).

In light of this, a terrorist attack i should be evaluated based on two components: (1) its rarity and (2) its impact. Not all rare events are impactful (many terrorist attacks that follow no historical patterns end up with no physical damages to the selected human targets, for instance), and not all impactful events are rare (as there exist, in certain regions of the world, very long sequences of attacks associated to very high number of deaths), this is the reason behind the necessity to incorporate both information in the evaluation. This conceptualization can lead to the formalization of a measure that provides a numerical quantification of the rarity and impact of an event.

Given x which captures the number of classes to which events i belongs (in the case of our manifold, a single target (e.g., NGO) belongs to a set of x classes, where (x-1) are the other types of targets in the mode), N which is the total number of sampled events (do we want to evaluate rarity over the entire history of a single group? Or do we want to evaluate rarity over the entire history of attacks plotted against a specific location?), n which is the total number of events in which the considered x has happened, and, finally, $(\alpha)_i$, which is the impact of the given event, that can be formalized as an arbitrary multivariate function of k variables, as:

$$\alpha_i = f(d, c, e, \cdots, k) \tag{6.40}$$

with d, c and e mapping, to exemplify, the number of deaths, the number of injured people and the economic loss or damage, then the rarity/impact indicator Ω_i is computed as:

$$\Omega_i = \ln \left[x_i^{-1} \left(\frac{N}{n_i} \right) \right]^{\ln(\alpha_i)} \tag{6.41}$$

The equation multiplies the reciprocal of the number of classes to which the event belongs (so that the wider the set of classes, the smaller the quantity) by the ratio

between the total number of sampled events and the actual number of attacks belonging to that class (e.g., attacks against NGO). This quantity is then raised to the power of the natural logarithm of α, the impact of the event.
Three simulations on mock data have been run to picture the behavior of Ω. I have let vary the three key components of the equation, namely n (Figure 6.38), x (Figure 6.39) and α (Figure 6.40).

The three different plots indicate that Ω reports three well-patterned behaviors. Firstly, and trivially, the equation let rare events to be weighted more: this is visible from Figure 6.38. In a sequence with a total of 3000 sampled events, when n is set to a very small quantity Ω increases fastly, and vice-versa.
Secondly, rare events raise Ω when the number of classes is lower. This means that if we are analyzing the history of a certain group and our data space divides targets into three classes then a rare event falling in one of those three classes will be weighted more than a rare event belonging to a class out of potentially hundreds. This is to capture a form of relative rarity in terms of the likelihood of an event: given that we know the set of all potential outcomes (e.g., classes) *a priori*, the rarity in a binary-space is more important than the rarity in multidimensional one.
Thirdly, rare events with high impact α are weighted exponentially, as visible in Figure 6.40. This allows distinguishing between rare events that have not had a considerable impact and all those attacks that have had large-scale consequences (it is worth to note that the composition of α function may depend upon the specific problem and research or policy setting).

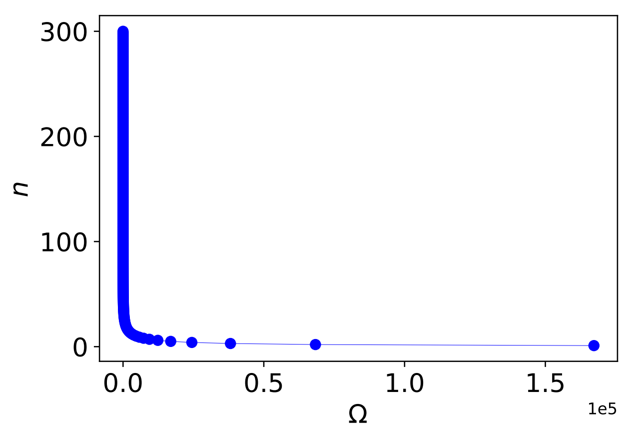

Figure 6.38: Ω Trend With Different n (N=3000)

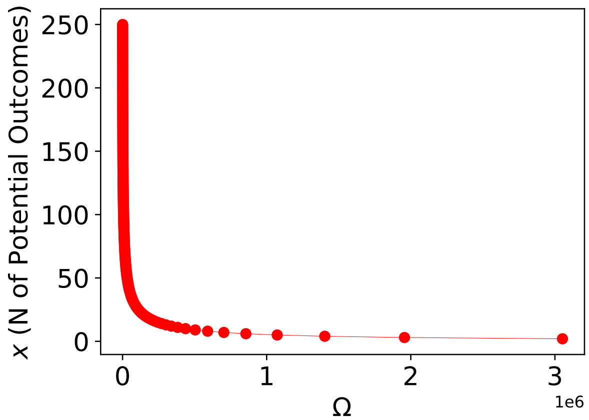

Figure 6.39: Ω Trend With Different Number of Potential Classes (Max=250)

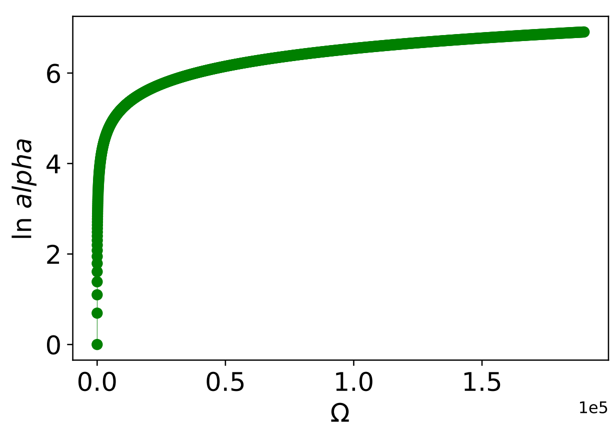

Figure 6.40: Ω Trend With Impact Levels (ln α)

Once the metric has been calculated for a single event, it has to be incorporated into the models and distributed over-time. From the theoretical point of view, Ω should act as a sort of memory-keeper such that its permanence into the data space $T(M)$ is proportional to its own value:

$$T(M)_i \propto \Omega_i \tag{6.42}$$

The proper kernel to be used for the decay has to be empirically tested and may be subjected to the particular problem to be investigated, however, this construction

allows to introduce a mathematical quantity potentially able to reduce the risk of missing relevant rare patterns that cannot be captured solely looking at strict and static event characteristics.
Further work will certainly test the performance of Ω on real-datasets such as the GTD: in the meanwhile, this section aimed at theoretically introducing the whole concept drawing a line that connects human behavior, outlier detection, and data science. It may sound or seem a naive approach to some readers, but, in my humble opinion, this section addresses a crucial problem that still prevents researchers and policymakers to obtain meaningful predictions in a large number of sub-domains falling within intelligence analysis. In a field such as terrorism, where weak and rare signals may hide thousands of human losses, outliers are never solely statistical constructs.

6.7 What is This All About? Notes to Potential Critiques

After having read this chapter, a person who is not familiar with network science, machine learning and mathematics might be asking what is this all about and what would be - in the end - the added value of the whole machinery for the field of criminology, and terrorism research more precisely. The concerns are, in principle, more than reasonable, and this section will try to clarify different aspects of this work. Indeed, I will try to anticipate and answer two possible critical questions posed by readers and potential reviewers.

What is the criminological contribution of this work? The strong methodological accent of this work has been mentioned many times throughout the previous sections. However, I am convinced that this work has not only relevance because it tries to merge network science and artificial intelligence, proposing a possible modeling framework derived from these two fields, but also because it tries to advance the knowledge on terror dynamics, specifically. My sensation is that it can contribute to the field investigating two interrelated concepts that link all the parts of this work: *interdependence* and *memory*.

Interdependence represents the foundation of some of the intuitions that allowed me to think about the benefits of using the science of complex networks as the first basis for the present dissertation. Indeed, the physical and abstract interdependence between terrorism features in a complex setting that involves multiple distinct dimensions guided the creation of dynamic meta-networks of jihadist groups and the fostered the idea that hidden knowledge can be extracted from data, stepping out of the comfort zone of purely physical relations, that is the core of social network analysis. In criminology, researchers tend to limit their analyses to certain common types of interdependencies (e.g. spatial ones and autocorrelations), without considering also other types of relations that might not be always self-evident and naturally visible. This work thus seeks to provide new insights into the benefits of deeply looking at interdependencies between distinct entities that are part of the same evolving phenomenon to highlight hidden schemes that may better explain strategic and rational dynamics.

Although the concept of "memory" cannot be explicitly found in criminological studies on terrorism, a variation of this concept is intrinsically connected to the studies that investigate the distribution of terror events over time and the strategic behavior of terrorist groups. However, in the present thesis, memory has to be considered not only as a proof of the spatio-temporal concentration of attacks but also as a more complex concept. Indeed, when discussing about memory I also intend to deal with the non-random recurrence of multi-dimensional patterns over time (e.g., a group attack a target y in a country c, with a tactic x and a weapon z for a certain length of time t, and then changes its behavior modifying the characteristics of the events according to new campaigns and aims). This leads to the hypothesis that sequence-handling algorithms can be trained to infer and detect patterns to forecast feature behaviors, assuming that not only events themselves do not behave randomly, but also locations, tactics, weapons, and targets follow some specific schemes that are interdependent and recur in time. Memory, in this fashion, goes in parallel with the concept of interdependence, and a further hypothesis is that only by looking at the whole realm of interconnections between event characteristics we can search for meaningful memory like-processes. Thought it is not certainly a proof of these ideas, data in Section 4.2 demonstrate how, out of the infinite possible combinations in which terrorism could happen, only a minimal part actually takes place. This is somehow a suggestion that some regularities exist, and that we should focus on them to build informative models.

The criminological relevance lies, indeed, in the combination of the strategic perspective and the spatio-temporal patterns of concentration of terrorist violence which, combined in such a novel methodological framework, can help in enhancing the knowledge on the violent histories of these groups.

Why using network science and machine learning instead of classic time series analysis? A question that might be posed regards the motivation behind the use of this specific methodological framework instead of the application of more traditional techniques. The analysis of time series is one of the most developed and longstanding fields in statistical sciences. Applications of time series analysis vary across a wide number of disciplines and sciences. Within this frame, there exist some very common techniques: autoregressive (AR) models, moving average (MA) models, integrated (I) models, autoregressive moving average (ARMA) models, autoregressive integrated moving average (ARIMA) models, autoregressive fractionally integrated moving average (ARFIMA) models. Besides these, others exist for dealing with multivariate analysis, as vector autoregression models (VAR) and Hidden Markov models (HMM).

All these different modeling techniques are solidly grounded in literature and have been used also in criminological or terrorism studies (Enders et al., 1992; Li and Schaub, 2004; Brandt and Sandler, 2010). However, in this work, I have not applied any of the abovementioned ones for two main reasons. The first one is related to some general advantages that machine and deep learning have over classic statistical methods. The second one is related to the specific use case of my application.

Regarding the first one, I have chosen LSTM because DL models are generally more flexible and can handle nonlinear relations between variables, while classic time-series models require more assumptions on data to be made. The ability to find non-linear relations (which theoretically increases when data increases, though requiring a relatively high number of observations) is indeed one of the main advantages of LSTM networks in general (Giles et al., 2001; Bontempi et al., 2013).

Concerning the second motivation, the specific application required a flexible algorithm able to perform sequence modeling on many multivariate time-series. This can be technically defined as a multi-label prediction learning problem (Zhang and Zhou, 2014), meaning that for each time unit, given a certain definition of the feature space, we want our algorithm to be trained to obtain different output prediction, terrorist targets in our case, via real-values assignment representing the forecasted centrality values. In notational terms, given a feature space $X = \mathbb{R}^d$ which is d-dimensional and an output space $Y = (y_1, ..., y_q)$ that includes q possible labels, the task that it is required to a multi-label learning algorithm is to find and learn a function $h : X \mapsto 2^Y$ from the multi-label training set $D = \{(x_i, Y_i)\,|1 \leq i \leq m\}$. This type of complex prediction problem can be more easily achieved using algorithms such as LSTM, while classic time-series methods are usually used for predicting single outputs y. However, though the forecasting task has not relied on classic time-series techniques, preliminary analyses have been run via popular stationarity and randomness tests that are valid across the whole spectrum of modeling techniques.

6.8 Discussion and Future Work

The pervasiveness of artificial intelligence has contaminated a huge number of scientific fields. This is no surprise, considering the incredible results that methods coming from machine, deep and reinforcement learning have reached on practical problems (Weiss et al., 2012; Jean et al., 2016; Xie et al., 2016; Silver et al., 2016; Krittanawong et al., 2017; Brown and Sandholm, 2018; Fang et al., 2019). Nonetheless, criminology and terrorism research still struggle in finding appropriate ways to exploit the strength of intelligent algorithms for the study of criminal human behavior (potentially, this also holds for the exploitation of traditional statistical methods, as noted by LaFree and Freilich (2012)). This is due to a variety of reasons, as the lower availability of data compared to other disciplines and the lack of sufficient quantitative and computational training of scholars belonging to the field.
With the aim to accelerate the process of bridging AI and criminology, this chapter has presented a methodological framework that combines the science of complex networks and deep learning to forecast future most likely targets hit in attacks plotted by five jihadist groups, namely the Islamic State, the Taliban, Al Qaeda, Boko Haram and Al Shabaab. Beyond the technical aspects, the work has been founded on a two-fold theoretical framework. This theoretical framework relied on (1) the theories on the spatio-temporal concentration of crime and (2) the strategic perspective of terrorism decision-making.
The spatio-temporal concentration of crime has been largely investigated with regard to several crimes and offenses. Terrorist violence is no exception. Many studies (Porter and White, 2012; Braithwaite and Johnson, 2012; Tench et al., 2016), starting from the seminal work of Midlarsky et al. (1980), have analyzed how terrorism clusters in time and in space, leading for instance to diffusion, contagion and microcycles dynamics (Midlarsky, 1978; Behlendorf et al., 2012). Terrorism is a patterned criminal phenomenon: its apparent chaotic and random nature hides, instead, schemes that can be captured by mathematical and statistical modeling, especially when terrorism is studied in terms of its temporal and geographic components.
Many theories have been developed and proposed to explain how terrorists decide to act. Studies on the decision-making processes of terrorist organizations have formal-

ized different explanations that many times are incomplete pictures of a multi-faceted phenomenon. This is also the case of the strategic theories on terrorism, which are built on some assumptions (such as rationality) that may exclude other relevant components of terrorist dynamics, as the psychological or purely organizational sides of the phenomenon. Nonetheless, for the aims of this study, the strategic explanations fit the problem to describe how jihadist groups recursively combine sets of tactics and weapons to hit some specific countries and targets.
The present chapter, indeed, extracts graph-derived time-series for each group from the GTD, mapping the over-time centrality of each feature comprised into the four modes of the manifold (i.e., Countries, Tactics, Weapons and Targets) and feed LSTM networks - algorithms specifically designed to handle sequence-dependent data - to provide forecasts of targets at highest risk of being hit in the future. The mathematical abstraction of the jihadist behavior relies in the first phase on the construction of meta-networks that aggregate terrorist attacks based on three-day time units via the creation of connections that map relations between events. This allows taking into consideration hidden relations that may not be captured using models that work on the assumption of i.i.d. data.
In the second phase, the models rely on the existence of memory-like processes and on the assumption of the strategic behavior of the jihadists to evaluate the performance of a very popular class of algorithms for the abovementioned forecasting purposes.
In spite of the promising results of the models, this work is just a very first exploratory step towards more meaningful and accurate models. The actual strength of this chapter actually resides in its incompleteness. The vibrant community of AI is continuously working on developing new ideas and methods, with an increasing interest in the area of social good (i.e., all those fields of application that address social and societal problems): this is encouraging. Besides the technical reasoning that led to the creation of graph-derived time series, the employed algorithms have nothing really innovative and, additionally, not all the potential configurations have been run due to computational expensiveness and time constraints (the models included in this thesis have run on my personal computer for sessions of 11, 9 and 7 days respectively). This indicates that there is potentially much more to explore.
Furthermore, this chapter only deals with a very restricted and particular sample of jihadist groups. It could be plausible that, applying the same framework to groups belonging to other ideologies and fighting for other motivations, the results might be different not only numerically but also conceptually (e.g., which type, if any, of memory does an Asian nationalist group have?). The choice beyond this restricted

sample was to center the focus on jihadism, in the first place and was also related to the necessity of having a sufficient amount of data to use to train the algorithms.
Data are crucial. Not only with regards to their quality (Sageman, 2014), but also in relation to their quantity, especially in AI. These algorithms work better when they can dive in and explore massive amounts of information. This is certainly a limitation of my study: I am using what are nowadays called "small data" since events are tracked by day and many groups do not have a longstanding history. What if we can apply the same framework to more detailed, precise and vast data? This question remains now open, but the hope is to have the chance to answer to it, one day.
There are potentially hundreds of pathways for future work, starting from this chapter: the integration of socioeconomic and political contextual data to control for exogenous components that may impact terrorist behavior and the disaggregation of information at a lower level (e.g., from weapons categories to weapons subcategories, from countries to provinces) are just two of the most promising directions. However, what really needs to be taken away from this work is that AI can help in the study of terrorism, and might provide really interesting insights and suggestions also in terms of practical applicability.

This page intentionally left blank

7 Concluding Remarks

There is something strikingly fascinating associated with apparently chaotic, random, unpredictable phenomena. The human history is full of scientists that have tried to capture a sort of order behind the visible layer of chaos using models and numbers to characterize phenomena that generally escape from our understanding. This addicting fascination is what brought me here, today, as a computational criminologist. After three years of research, there are many more unsolved questions on my table rather than answers. Three years ago I was naively thinking that ending a Ph.D. would have instead provided me with much stronger conclusions and certainties. I was wrong, and I have realized it as getting closer to grasp the inherent order of terrorism. There have been answers to some of my questions, but these answers fatally generated further questions, which grew exponentially over time, until today. In spite of the disorienting distance between my curiosity and what I was able to understand, an even stronger fascination now dwells in me. The hypothesis that even the most terrible actions of humans on this planet can be modeled and described through the use of numbers and mathematics is not only captivating but also, in a way, heartening. I do not believe in the perspective of a future in which we will be able to completely understand how every component of human behavior works, and why, but I am fully confident that many signs of progress will be made in order to unveil many of the obscure dynamics that daily governs our world. Since the second part of the Twentieth century, enlightened scholars have chased the order behind the chaos in the study of terrorism, and many are still doing it now.

This doctoral dissertation, entitled *On Meta-Networks, Deep Learning, Time and Jihadism* is my humble attempt to contribute to the cause and marks my efforts to highlight the potential connected to the combination of complex networks and artificial intelligence for the study of the behavior of jihadist groups.

The first chapter illustrates the challenges of the conceptualization of terrorism, proposing a focus on four specific dimensions of terrorism to overcome the issues with its unitary definition. Furthermore, it also presents the two-fold theoretical framework of the dissertation: the spatio-temporal concentration of crime and the theories of strategic terrorism as the backbones.

The second chapter outlines an overview of the aims of the work at a general level. It also proposes a reflection on the need to rethink research in criminology and terrorism in light of the spread of novel available computational methods, following the massive popularity gained by Artificial Intelligence in many academic disciplines.

The third chapter describes the origins and main characteristics of the jihadist groups taken into consideration in the thesis, namely the Islamic State, the Taliban, Al Qaeda, Boko Haram and Al Shabaab and it also provides details on the Global Terrorism Database, the dataset used in the analyses. Besides the first introductory ones, the dissertation comprises three analytic chapters.

The fourth chapter, entitled *Stochastic Matrices of Terrorism: Complexity and Heterogeneity of Jihadist Behavior*, is indeed a study on the use of stochastic transition matrices and trail networks to compare jihadist groups in terms of their strategic behavior. The chapter develops a two-fold analysis. First, it assesses the complexity in the combination of weapons, targets and targets and weapons together through a novel approach based on N-dimensional super-states: the approach allows us to consider cycles and sub-sequences of attacks as a new tool for highlighting terrorist dynamics. Second, it develops a coefficient that maps the pairwise similarity between each pair of groups considering their transitions between different types of weapons, targets and weapons and targets combined. With regard to the first part, the results of the study shows that all the terrorist groups have a very complex repertoire of combinations and configurations in the use of the same weapons and targets and that, as the dimension of the transition matrices is increased, clearer patterns emerge, as each sub-sequence (defined by super-states) is connected to few others only.

For what concerns the second part, Al Qaeda and Al Shabaab are found to be the most similar groups overall, while, on the contrary, the Taliban and Al Qaeda are the least similar. Another interesting finding is that, while groups can be similar when only targets are considered, they can show distinct strategic behavior with regard to weapons and vice-versa, suggesting how there exist different ways to reach a certain terrorist goal and that relying solely on one source of information can be misleading. The work, which has also been published in a shorter version as a research article in the Journal of Computational Social Science, is a first exploratory study in the use of N-dimensional chains and trails as meta-networks for the analysis of sequential patterns and call for further work in the direction of integrating multiple data sources to extrapolate additional knowledge.

The fifth chapter, entitled *Hawkes Processes of Jihadism*, explicitly aims at investigating the presence of memory-like processes in the temporal patterns of jihadist attacks. For each group, two models are developed using data sequences on the two most attacked countries per each group, limiting further the scope to the most popular target category hit. Building upon the blooming literature on point processes in terrorism research and criminology, Hawkes models are used to detect the presence of memory in terrorist patterns. Hawkes models are a particular class of stochastic processes that are characterized by the presence of self-excitability. Self-excitability captures the extent to which a certain event can increase (or even decrease) the probability of the future occurrence of another event. Contrarily to ordinary Poisson processes, Hawkes processes are not randomly distributed. They hold memory properties meaning that the present is generally dependent upon the past. This type of analysis is innovative in the sense that, contrarily to most literature, does not treat all events as equal but, instead, disaggregates by country and target to control for memory-like processes also at a finer-grained level. Comparing the performance of each Hawkes process against a Homogeneous Poisson model, statistical outcomes show how most of the sequences actually hold memory-like processes (nine out of ten). The finding goes against the cornerstone concept of "asymmetric warfare" as random and unpredictable violence as posited by Matusitz (2012) in its definition of terrorism.

Terrorist attacks against most popular targets are thus clustered in time and not randomly distributed, meaning that there is a time-dependent structure that can be captured and analyzed to characterize the behavior of each jihadist organization. The results are in line with previous research on terrorist events and, also in this case, call for future work. The field of point processes is vibrant and developments are continuously made also from the theoretical and foundational points of view: there are dozens of potential new directions to take, starting from multivariate modeling for studying Granger causality between multiple time-series and the use of marked-Hawkes processes to further distinguish between events of different magnitude and impact. At a general level, this chapter preciously corroborated the presence of memory, which is a fundamental concept for the third and last analytic chapter.

The sixth chapter, *Deep Learning and Terrorism: Long Short-Term Memory Networks for Target Forecasting*, is the last and the longest of the entire dissertation. It specifically proposes a methodological framework that combines the science of complex meta-networks with deep learning to unfold the temporal patterns of jihadist groups. From a dynamic manifold, I have extracted multivariate time-series mapping the centrality of attacked countries, tactics, weapons and targets for each group and then fed Long Short-Term Memory deep neural networks with the time-series. The deep learning algorithms are hence employed to learn the existing time-dependent structure in the data based on the inter-dependence among event characteristics and the presence of memory in terrorist behavior.

Following the definition of two measures of accuracy, the performance of the algorithms is evaluated based on their ability to correctly rank the most central - and thus probable - target in each future time-step. The quantitative results are encouraging and highlight different levels of predictability for each group, providing insights also on the evolution of their behavior over time. A potential solution against the problem of rare and weak signals is also proposed. This first theoretical step towards a more efficient and effective combination of complex networks and artificial intelligence automatically poses new challenges and perspectives for future research, including the test of less conventional algorithms for time-dependent data or the integration of contextual socioeconomic and political data to control for external sources of variation and influence in jihadist behavior.

This work inspired in me also general consideration regarding the current and future perspectives in terrorism research. When I started my doctorate, back in 2016, the world was just stepping out from very hard times in terms of terrorist activity. Most of the resonance of terrorism worldwide was caused by the actions of jihadist groups that, operating in different regions and by means of different tactics, diffused violence systematically. The peak of deaths reached in 2014 has been followed by a decline that still persists: data shows how this decline is associated with the battleground defeats of the Islamic State (Institute for Economics and Peace, 2018). The Islamist group which once governed over many areas of Iraq and Syria has now lost all its outposts and many of its resources and militant fighters (Tønnessen, 2019). However, as noted by Ineichen (2018) and Dawson (2018), it would be an error to simplistically consider the Islamic State threat as solved, since many issues are still in place in the global scenario, including the displacement of the group in smaller units and dispersal of the remaining Western foreign fighters.

The Islamic State is capable of driving alone most of the aggregated statistics on terrorism, especially when concentrating on the European situation. In fact, in the last two years, Europe has experienced a significant reduction in terms of fatalities, as a positive byproduct of the deteriorated strength of ISIS. In this statistic breathes and lives a dangerous trap.

Research survives thanks to grants, private and public funding and scholarships, which are many times driven by the policy-agenda of institutions and governments of the Western countries. As jihadism started to become less present in the daily chronicles due to the decrease of large-scale and fatal attacks in Western countries, suddenly the availability of resources to conduct research on this type of terrorism almost vanished, too. The last two years have seen the rise of other types of priorities for policy-makers, including the study of how social media can modify, influence or distort political opinions, fueling extremism and misinformation (Del Vicario et al., 2016; Vosoughi et al., 2018; Wu and Liu, 2018; Carley et al., 2018) and the renovated attention towards political terrorism, with a special mention to far-right and racist ideologies (Freilich et al., 2018; Ravndal, 2018; Fahey and Simi, 2019). It goes without saying that these topics are extremely relevant and that I am not claiming that there might be an intrinsic moral superiority associated with the study of jihadism. What I am suggesting is, instead, that relegating jihadism in the list of the marginal priorities of today might pose serious consequences for tomorrow.

Schuurman (2019) claims that the over-representation of jihadism in terrorism studies is not only incorrect but rather dangerous and the process underestimates the threats posed by other types of terrorism. The assumption of a sort of mutually exclusive allocation of attention is wrong and the reasoning misses a focal point: jihadist terrorism has been studied more than other types of terrorism because it has posed much bigger threats and inflicted much more damages than all the other types of political or religious violence, in the last twenty years. This, again, should not be reflected in a sort of "monopoly" in research. Nonetheless, certain comparisons should be carefully made based on historical facts.

Terrorism research has received massive funding after the 9/11 attacks and the European events of Madrid and London: more than ten years of research have, according to many, produced very little knowledge and responded to very few crucial questions regarding jihadism. Agreeing with this position or not, suspending or interrupting research in this area is certainly the best way to slow down the progress that has been made and that still could be made in the attempt to better understand jihadist terrorism and to provide policy-makers and institutions with meaningful tools, suggestions, guidelines to mitigate the problem.

Furthermore, we all risk falling into an overly Western-centered vision, where the ranking of priorities is tailored solely upon the necessities and issues experienced by developed countries. This is not political speculation (I am a researcher dedicated to science): it is, instead, a fact.

While Europe has experienced a decrease in the number of fatal attacks, there are many areas in the world in which jihadist groups still organize and plot terrible attacks that are capable of killing hundreds of people at a time. Regional groups such as Boko Haram and Al Shabaab attract less attention in the Western world because their actions are confined to specific regions of the African continent, and do not, therefore, pose any risk to the security of European, American, Australian borders. However, terrorism should not be a strict matter of physical borders. History has taught us that the globalized world is a powerful weapon also for terrorists, and treating regional groups as if they were only an issue of those African countries is not only partially immoral and hypocrite, it is also a good way to underestimate the risks associated to these groups. The Islamic State has shown how it is not necessary to be trained or officially affiliated to the group to act in its name. Why should it be different for other groups that are not acting (so far, at least) on a global scale? Another point that should be considered is that, even hypothesizing a complete defeat of the most important jihadist groups that are still operating nowadays, jihadist terrorism may strike back in the future years under different symbols, flags, and acronyms. Why, then, should we stop or cut research in terrorism?

The exact point is that we should not. This dissertation is, indeed, a modest attempt to show that there is a tremendous need for data, resources, training, and cross-disciplinary collaboration and that these elements combined can play a fundamental role in advancing our understanding of the phenomenon. This understanding is not solely confined to academia and scientific research, it can be, instead, directly applicable for policy or counter-terrorism intelligence purposes.

My thesis has dealt with a small region of the overwhelmingly vast complexity of jihadist terrorism. It has focused on a small sample of jihadist groups, and it has focused on events, not on people, individuals, organizational structures and psychological motives. It has treated every group as an entity, without considering the internal components that drive the decision-making process of each organization, it has only focused on general event characteristics, without going too much in depth in terms of geographic profiling, for instance. It has addressed some precise research questions that, in turn, leave out many other relevant ones, as "what drives jihadism?", or "why do individuals resort to jihadist terrorist violence?".

This is to say that this work does not aim to be comprehensive and universal. It has, indeed, many limitations. It is only the first incomplete step towards a long path of further research. It, however, poses the basis for the integration of two fields

- network science and artificial intelligence - for the study of jihadism, exploiting the strength of both fields. The ability of network science to map, detect, recognize, interpret the complex underlying physical and abstract relations between events and the power of artificial intelligence to flexibly handle multidimensional and non-linear data, going beyond the strong assumptions required by ordinary statistical models. This dissertation is an attempt to show that these two fields can be meaningfully combined to highlight patterns, motifs and recurrent behaviors and that the research community should not be worried about opening its doors to a new idea of contamination.

Terrorist research has, over the years, benefited from the dialogue between distinct disciplines (Ross, 2006; Richardson, 2013). While first originating from international relations and political science (Jenkins, 1974; Jongman, 2017), it has then attracted the interest of psychology (Silke, 2003; Horgan, 2005; Bongar et al., 2006), sociology (Turk, 2004; Tilly, 2004), economics (Lakdawalla and Zanjani, 2005; Llusa and Tavares, 2007; Berman, 2011) and, lately, criminology (LaFree and Freilich, 2016). This process led to a hybrid debate which certainly brought new ideas, perspectives and answers to a dimension of human behavior that, in the last decades, has affected the entire world in dramatic and horrific ways. Opening the doors towards further contamination from fields such as complex networks, artificial intelligence, and computer science could bring added values to the community itself and, in turn, to research on the phenomenon.

Schuurman (2018) found that the prevalence of “one-time” contributors is one of the most relevant issues of the terrorist research community. This is exactly one of the negative loops that can arise in the absence of a well-structured, defined and solid multidisciplinary community. On one hand, computer scientists, statisticians and scholars from the field of complexity science address the problem of terrorism as a laboratory where to experiment new algorithms and techniques, without posing any attention over the real implications of certain results or without carefully evaluate assumptions or data limitations, giving birth to sophisticated and yet sterile forms of science. On the other hand, terrorist researchers struggle to explore the landscape of state-of-the-art computational methods because they lack technical expertise and training. This process leads to sparse research, which is not followed up during the years and cannot meaningfully address the crucial points that are still without an answer, avoiding solid theoretical explanations and extensive reasoning on. Furthermore, the hype around artificial intelligence poses the virtual risk of future funding allocation exclusively to data scientists, engineers, computer scientists, ruling out social scientists from the policy-oriented research arena.

This perspective is alarming. The power of data and the revolution of artificial intelligence have started to change the world, and the benefits experienced by every one of us every day are uncountable. But this revolution also calls for new responsibilities, opening ethical and moral debates on the role that machines have in impacting our future. The exclusive allocation of power, funding and resources to algorithmic systems to predict crime or recidivist behavior has already shown its tremendous flaws and drawbacks. This is why criminologists and social scientists in general cannot be excluded by this progress, as they can act as guarantors and barriers against the misuse of data and computational methods in such critical areas. Terrorism research should proactively behave for its renovation in light of these aspects because wasting the opportunities of this new era would be not only mindless, but even dangerous.

"The time is gone, the song is over
thought I'd something more to say"
Pink Floyd, "Time"

This page intentionally left blank

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.

Abrahms, M. (2006). Why Terrorism Does Not Work. *International Security*, 31(2):42–78.

Abrahms, M. (2008). What Terrorists Really Want: Terrorist Motives and Counterterrorism Strategy. *International Security*, 32(4):78–105.

Abrahms, M., Ward, M., and Kennedy, R. (2018). Explaining Civilian Attacks: Terrorist Networks, Principal-Agent Problems and Target Selection. *Perspectives on Terrorism*, 12(1).

Abubakar, D. (2017). From Sectarianism to Terrorism in Northern Nigeria: A Closer Look at Boko Haram. In *Violent Non-State Actors in Africa*, pages 17–47. Palgrave Macmillan, Cham.

Achab, M., Bacry, E., Gaïffas, S., Mastromatteo, I., and Muzy, J.-F. (2017). Uncovering Causality from Multivariate Hawkes Integrated Cumulants. *J. Mach. Learn. Res.*, 18(1):6998–7025.

Ackerman, W. V. and Murray, A. T. (2004). Assessing spatial patterns of crime in Lima, Ohio. *Cities*, 21(5):423–437.

Ackley, D., Hinton, G., and Sejnowski, T. (1985). A learning algorithm for boltzmann machines. *Cognitive Science*, 9(1):147–169.

Adler, R. M. (2007). A Dynamic Social Network Software Platform for Counter-Terrorism Decision Support. In *2007 IEEE Intelligence and Security Informatics*, pages 47–54.

Aghedo, I. and Osumah, O. (2012). The Boko Haram Uprising: How should Nigeria respond? *Third World Quarterly*, 33(5):853–869.

Aglietti, V., Damoulas, T., and Bonilla, E. (2018). Efficient Inference in Multi-task Cox Process Models. *arXiv:1805.09781 [cs, stat]*. arXiv: 1805.09781.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723.

Akinola, O. (2015). Boko Haram Insurgency in Nigeria: Between Islamic Fundamentalism, Politics, and Poverty. *African Security*, 8(1):1–29.

Alakoc, B. P. (2017). Competing to Kill: Terrorist Organizations Versus Lone Wolf Terrorists. *Terrorism and Political Violence*, 29(3):509–532.

Aly, A., Macdonald, S., Jarvis, L., and Chen, T. (2016). *Violent Extremism Online: New Perspectives on Terrorism and the Internet*. Routledge.

Anderson, D. M. and McKnight, J. (2015). Understanding al-Shabaab: Clan, Islam and insurgency in Kenya. *Journal of Eastern African Studies*, 9(3):536–557.

Armstrong, J. S. (2001). *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Springer Science & Business Media.

Asal, V. and Hastings, J. V. (2015). When Terrorism Goes to Sea: Terrorist Organizations and the Move to Maritime Targets. *Terrorism and Political Violence*, 27(4):722–740.

Asal, V. H., Rethemeyer, R. K., Anderson, I., Stein, A., Rizzo, J., and Rozea, M. (2009). The Softest of Targets: A Study on Terrorist Target Selection. *Journal of Applied Security Research*, 4(3):258–278.

Bacry, E., Dayri, K., and Muzy, J. F. (2012). Non-parametric kernel estimation for symmetric Hawkes processes. Application to high frequency financial data. *The European Physical Journal B*, 85(5).

Balocchi, C. and Jensen, S. T. (2019). Spatial Modeling of Trends in Crime over Time in Philadelphia. *arXiv:1901.08117 [stat]*. arXiv: 1901.08117.

Barabási, A.-L. (2011). The Network Takeover. *Nature Physics*, 8:14–16.

Barfield, T. (2010). *Afghanistan: A Cultural and Political History*. Princeton University Press.

Bartels, R. (1982). The Rank Version of von Neumann's Ratio Test for Randomness. *Journal of the American Statistical Association*, 77(377):40–46.

Basra, R. and Neumann, P. R. (2016). Criminal Pasts, Terrorist Futures: European Jihadists and the New Crime-Terror Nexus. *Perspectives on Terrorism*, 10(6).

Bates, R. (2012). Dancing With Wolves: Today's Lone Wolf Terrorists. *The Journal of Public and Professional Sociology*, 4(1).

Behlendorf, B., LaFree, G., and Legault, R. (2012). Microcycles of Violence: Evidence from Terrorist Attacks by ETA and the FMLN. *Journal of Quantitative Criminology*, 28(1):49–75.

Ben-Gal, I. (2005). Outlier Detection. In Maimon, O. and Rokach, L., editors, *Data Mining and Knowledge Discovery Handbook*, pages 131–146. Springer-Verlag, New York.

Bengio, S., Vinyals, O., Jaitly, N., and Shazeer, N. (2015). Scheduled Sampling for Sequence Prediction with Recurrent Neural Networks. *arXiv:1506.03099 [cs]*.

Bengio, Y., Simard, P., and Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 5(2):157–166.

Benigni, M. and Carley, K. M. (2016). From Tweets to Intelligence: Understanding the Islamic Jihad Supporting Community on Twitter. In *Social, Cultural, and Behavioral Modeling*, Lecture Notes in Computer Science, pages 346–355. Springer, Cham.

Benmelech, E. and Klor, E. (2018). What Explains the Flow of Foreign Fighters to ISIS? *Terrorism and Political Violence*, pages 1–24.

Bennett Moses, L. and Chan, J. (2018). Algorithmic prediction in policing: assumptions, evaluation, and accountability. *Policing and Society*, 28(7):806–822.

Bergesen, A. J. and Lizardo, O. (2004). International Terrorism and the World-System. *Sociological Theory*, 22(1):38–52.

Berk, R. (2008). How you can tell if the simulations in computational criminology are any good. *Journal of Experimental Criminology*, 4(3):289.

Berk, R. (2019). *Machine Learning Risk Assessments in Criminal Justice Settings*. Springer International Publishing, Cham.

Berk, R., Heidari, H., Jabbari, S., Kearns, M., and Roth, A. (2018). Fairness in Criminal Justice Risk Assessments: The State of the Art. *Sociological Methods & Research*, page 004912411878253.

Berlusconi, G. (2013). Do all the pieces matter? Assessing the reliability of law enforcement data sources for the network analysis of wire taps. *Global Crime*, 14(1):61–81.

Berlusconi, G., Calderoni, F., Parolini, N., Verani, M., and Piccardi, C. (2016). Link Prediction in Criminal Networks: A Tool for Criminal Intelligence Analysis. *PLOS ONE*, 11(4):e0154244.

Berman, E. (2011). *Radical, Religious, and Violent: The New Economics of Terrorism*. MIT Press.

Bessner, D. and Stauch, M. (2010). Karl Heinzen and the Intellectual Origins of Modern Terror. *Terrorism and Political Violence*, 22(2):143–176.

Bickenbach, F. and Bode, E. (2003). Evaluating the Markov Property in Studies of Economic Convergence. *International Regional Science Review*, 26(3):363–392.

Bier, V., Oliveros, S., and Samuelson, L. (2007). Choosing What to Protect: Strategic Defensive Allocation against an Unknown Attacker. *Journal of Public Economic Theory*, 9(4):563–587.

Bishop, C. M. (2006). *Pattern Recognition And Machine Learning*. Springer Verlag, New York, new ed edizione edition.

Block, H. D., Knight, B. W., and Rosenblatt, F. (1962). Analysis of a Four-Layer Series-Coupled Perceptron. II. *Reviews of Modern Physics*, 34(1):135–142.

Bogomolov, A., Lepri, B., Staiano, J., Oliver, N., Pianesi, F., and Pentland, A. (2014). Once Upon a Crime: Towards Crime Prediction from Demographics and Mobile Data. In *Proceedings of the 16th International Conference on Multimodal Interaction - ICMI '14*, pages 427–434, Istanbul, Turkey. ACM Press.

Bongar, B., Brown, L. M., Beutler, L. E., Breckenridge, J. N., and Zimbardo, P. G. (2006). *Psychology of Terrorism*. Oxford University Press.

Bontempi, G., Ben Taieb, S., and Le Borgne, Y.-A. (2013). Machine Learning Strategies for Time Series Forecasting. In Aufaure, M.-A. and Zimányi, E., editors, *Business Intelligence: Second European Summer School, eBISS 2012, Brussels, Belgium, July 15-21, 2012, Tutorial Lectures*, Lecture Notes in Business Information Processing, pages 62–77. Springer Berlin Heidelberg, Berlin, Heidelberg.

Bouchard, M., editor (2017). *Social Networks, Terrorism and Counter-terrorism*. Routledge, London, 1 edition.

Bowen, W. Q. (2004). Deterrence and asymmetry: non-state actors and mass casualty terrorism. *Contemporary Security Policy*, 25(1):54–70.

Braithwaite, A. and Johnson, S. D. (2012). Spatio-temporal modeling of insurgency in Iraq. *Journal of Quantitative Criminology*, 28(1):31–48.

Braithwaite, A. and Li, Q. (2007). Transnational Terrorism Hot Spots: Identification and Impact Evaluation. *Conflict Management and Peace Science*, 24(4):281–296.

Brandt, P. T. and Sandler, T. (2010). What Do Transnational Terrorists Target? Has It Changed? Are We Safer? *Journal of Conflict Resolution*, 54(2):214–236.

Brennan, T. and Oliver, W. L. (2013). The Emergence of Machine Learning Techniques in Criminology: Implications of Complexity in our Data and in Research Questions. *Criminology & Public Policy*, 12(3):551–562.

Britten, S. (2018). Intelligence Failures Are Analytical Failures. *Counter Terrorist Trends and Analyses*, 10(7):12–18.

Brown, M. A. (1982). Modelling the Spatial Distribution of Suburban Crime. *Economic Geography*, 58(3):247–261.

Brown, N. and Sandholm, T. (2018). Superhuman AI for heads-up no-limit poker: Libratus beats top professionals. *Science*, 359(6374):418–424.

Buckley, M. E. A. and Fawn, R. (2003). *Global Responses to Terrorism: 9/11, Afghanistan, and Beyond*. Psychology Press.

Burgess, R. L. and Akers, R. L. (1966). A Differential Association-Reinforcement Theory of Criminal Behavior. *Social Problems*, 14:128.

Byman, D. (2013). Fighting Salafi-Jihadist Insurgencies: How Much Does Religion Really Matter? *Studies in Conflict & Terrorism*, 36(5):353–371.

Byman, D. (2014). Buddies or Burdens? Understanding the Al Qaeda Relationship with Its Affiliate Organizations. *Security Studies*, 23(3):431–470.

Börner, K., Sanyal, S., and Vespignani, A. (2007). Network science. *Annual Review of Information Science and Technology*, 41(1):537–607.

Calderoni, F., Brunetto, D., and Piccardi, C. (2017). Communities in criminal networks: A case study. *Social Networks*, 48:116–125.

Calderoni, F., Skillicorn, D., and Zheng, Q. (2014). Inductive Discovery of Criminal Group Structure Using Spectral Embedding. *Information & Security: An International Journal*, 31:49–66.

Campana, P. (2016). Explaining criminal networks: Strategies and potential pitfalls. *Methodological Innovations*, 9:205979911562274.

Campedelli, G. M., Bartulovic, M., and Carley, K. M. (2019a). Pairwise similarity of jihadist groups in target and weapon transitions. *Journal of Computational Social Science*.

Campedelli, G. M., Cruickshank, I., and Carley, K. M. (2019b). Detecting Latent Terrorist Communities Testing a Gower's Similarity-Based Clustering Algorithm for Multi-partite Networks. In Aiello, L. M., Cherifi, C., Cherifi, H., Lambiotte, R., Lió, P., and Rocha, L. M., editors, *Complex Networks and Their Applications VII*, volume 812, pages 292–303. Springer International Publishing, Cham.

Campos, G. O., Zimek, A., Sander, J., Campello, R. J. G. B., Micenková, B., Schubert, E., Assent, I., and Houle, M. E. (2016). On the evaluation of unsupervised outlier detection: measures, datasets, and an empirical study. *Data Mining and Knowledge Discovery*, 30(4):891–927.

Cannon, B. J. and Ruto Pkalya, D. (2017). Why al-Shabaab Attacks Kenya: Questioning the Narrative Paradigm. *Terrorism and Political Violence*, pages 1–17.

Cao, S., Lu, W., and Xu, Q. (2016). Deep Neural Networks for Learning Graph Representations. In *Thirtieth AAAI Conference on Artificial Intelligence*.

Carley, K. M. (2014). ORA: A Toolkit for Dynamic Network Analysis and Visualization. In Alhajj, R. and Rokne, J., editors, *Encyclopedia of Social Network Analysis and Mining*, pages 1219–1228. Springer New York, New York, NY.

Carley, K. M., Cervone, G., Agarwal, N., and Liu, H. (2018). Social Cyber-Security. In Thomson, R., Dancy, C., Hyder, A., and Bisgin, H., editors, *Social, Cultural, and Behavioral Modeling*, Lecture Notes in Computer Science, pages 389–394. Springer International Publishing.

Cederman, L.-E. (2003). Modeling the Size of Wars: From Billiard Balls to Sandpiles. *The American Political Science Review*, 97(1):135–150.

Cederman, L.-E. and Weidmann, N. B. (2017). Predicting armed conflict: Time to adjust our expectations? *Science*, 355(6324):474–476.

Chavez-Demoulin, V. and McGill, J. (2012). High-frequency financial data modeling using Hawkes processes. *Journal of Banking & Finance*, 36(12):3415–3426.

Chen, F. and Tan, W. H. (2018). Marked Self-Exciting Point Process Modelling of Information Diffusion on Twitter. *arXiv:1802.09304 [stat]*. arXiv: 1802.09304.

Chen, H., Reid, E., Sinai, J., Silke, A., and Ganor, B. (2008). *Terrorism Informatics: Knowledge Management and Data Mining for Homeland Security*. Springer Science & Business Media.

Chen, X., Cho, Y., and Jang, S. Y. (2015). Crime prediction using Twitter sentiment and weather. In *2015 Systems and Information Engineering Design Symposium*, pages 63–68, Charlottesville, VA, USA. IEEE.

Chollet, F. (2015). Keras.

Choudhury, M. D., Gamon, M., Counts, S., and Horvitz, E. (2013). Predicting Depression via Social Media. In *Seventh International AAAI Conference on Weblogs and Social Media*.

Chung, J., Gulcehre, C., Cho, K., and Bengio, Y. (2014). Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. *arXiv:1412.3555 [cs]*.

Clarke, R. V. G. and Newman, G. R. (2006). *Outsmarting the Terrorists*. Greenwood Publishing Group.

Clauset, A., Heger, L., Young, M., and Gleditsch, K. S. (2010). The strategic calculus of terrorism: Substitution and competition in the Israel—Palestine conflict. *Cooperation and Conflict*, 45(1):6–33.

Clauset, A. and Woodard, R. (2013). Estimating the historical and future probabilities of large terrorist events. *The Annals of Applied Statistics*, 7(4):1838–1865.

Clauset, A., Young, M., and Gleditsch, K. S. (2007). On the Frequency of Severe Terrorist Events. *Journal of Conflict Resolution*, 51(1):58–87.

Connor, J., Martin, R., and Atlas, L. (1994). Recurrent neural networks and robust time series prediction. *IEEE Transactions on Neural Networks*, 5(2):240–254.

Corsi, J. R. (1981). Terrorism as a Desperate Game: Fear, Bargaining, and Communication in the Terrorist Event. *Journal of Conflict Resolution*, 25(1):47–85.

Cortes, C. and Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, 20(3):273–297.

Cox, D. R. and Stuart, A. (1955). Some Quick Sign Tests for Trend in Location and Dispersion. *Biometrika*, 42(1/2):80.

Crenshaw, M. (1987). Theories of terrorism: Instrumental and organizational approaches. *Journal of Strategic Studies*, 10(4):13–31.

Crenshaw, M. (2010). Innovation: Decision Points in the Trajectory of Terrorism. In Rasmussen, M. and Hafez, M., editors, *Terrorist innovations in weapons of mass effect: Preconditions, causes and predictive indicators*, (Report No. ASCO 2010-019).

Cronin, A. K. (2015). ISIS Is Not a Terrorist Group: Why Counterterrorism Won't Stop the Latest Jihadist Threat. *Foreign Affairs*, 94:87.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2(4):303–314.

Daley, D. J. and Vere-Jones, D. (2006). *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods*. Springer Science & Business Media.

Dawson, L. L. (2018). The Demise of the Islamic State and the Fate of Its Western Foreign Fighters: Six Things to Consider.

De Choudhury, M., Counts, S., and Horvitz, E. (2013). Predicting postpartum changes in emotion and behavior via social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, page 3267, Paris, France. ACM Press.

de Montclos, M.-A. P., editor (2015). *Boko Haram: Islamism, Politics, Security, and the State in Nigeria*. Tsehai Publishers, Los Angeles.

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, H. E., and Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, 113(3):554–559.

Dershowitz, A. M. (2003). *Why Terrorism Works: Understanding the Threat, Responding to the Challenge*. Yale University Press.

Desmarais, B. A. and Cranmer, S. J. (2013). Forecasting the locational dynamics of transnational terrorism: a network analytic approach. *Security Informatics*, 2(1):8.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, 74(366a):427–431.

Ding, F., Ge, Q., Jiang, D., Fu, J., and Hao, M. (2017). Understanding the dynamics of terrorism events with multiple-discipline datasets and machine learning approach. *PLOS ONE*, 12(6):e0179057.

Dobson, C. and Payne, R. (1979). *The Terrorists: Their Weapons, Leaders and Tactics*. Facts on File.

Dolnik, A. (2007). *Understanding Terrorist Innovation: Technology, Tactics and Global Trends*. Routledge.

Drake, C. J. M. (1998a). *Terrorists' target selection*. St. Martin's Press, New York. OCLC: 759111922.

Drake, C. J. M. (1998b). *Terrorists' Target Selection*. Palgrave Macmillan UK, London.

Duchi, J., Hazan, E., and Singer, Y. (2011). Adaptive Subgradient Methods for Online Learning and Stochastic Optimization. *Journal of Machine Learning Research*, 12(Jul):2121–2159.

Dugard, J. (1973). Towards the Definition of International Terrorism. *The American Journal of International Law*, 67(5):94–100.

Dugard, J. (1974). International Terrorism: Problems of Definition. *International Affairs (Royal Institute of International Affairs 1944-)*, 50(1):67–81.

Dunham, R. G. and Alpert, G. P. (2015). *Critical Issues in Policing: Contemporary Readings, Seventh Edition*. Waveland Press.

Dutt, A. K. and Venugopal, G. (1983). Spatial patterns of crime among Indian cities. *Geoforum*, 14(2):223–233.

Eichler, M., Dahlhaus, R., and Dueck, J. (2017). Graphical Modeling for Multivariate Hawkes Processes with Nonparametric Link Functions. *Journal of Time Series Analysis*, 38(2):225–242.

Elman, J. (1990). Finding structure in time. *Cognitive Science*, 14(2):179–211.

Enders, W., Parise, G. F., and Sandler, T. (1992). A time-series analysis of transnational terrorism: Trends and cycles. *Defence Economics*, 3(4):305–320.

Enders, W. and Sandler, T. (2006). Distribution of Transnational Terrorism Among Countries by Income Class and Geography After 9/11. *International Studies Quarterly*, 50(2):367–393.

Enders, W., Sandler, T., and Gaibulloev, K. (2011). Domestic versus transnational terrorism: Data, decomposition, and dynamics. *Journal of Peace Research*, 48(3):319–337.

Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(1):32–64.

Etesami, J., Kiyavash, N., Zhang, K., and Singhal, K. (2016). Learning Network of Multivariate Hawkes Processes: A Time Series Approach. In *Proceedings of the Thirty-Second Conference on Uncertainty in Artificial Intelligence*, UAI'16, pages 162–171, Arlington, Virginia, United States. AUAI Press. event-place: Jersey City, New Jersey, USA.

Evans, D. J. and Herbert, D. T. (1989). *The Geography of Crime*. Routledge.

Eyerman, J. (1998). Terrorism and democratic states: Soft targets or accessible systems. *International Interactions*, 24(2):151–170.

Fahey, S. and Simi, P. (2019). Pathways to violent extremism: a qualitative comparative analysis of the US far-right. *Dynamics of Asymmetric Conflict*, 12(1):42–66.

Fang, F., Tambe, M., Dilkina, B., and Plumptre, A. J. (2019). *Artificial Intelligence and Conservation*. Cambridge University Press.

Farley, J. D. (2003). Breaking Al Qaeda Cells: A Mathematical Analysis of Counterterrorism Operations (A Guide for Risk Assessment and Decision Making). *Studies in Conflict & Terrorism*, 26(6):399–411.

Farwell, J. P. (2014). The Media Strategy of ISIS. *Survival*, 56(6):49–55.

Favarin, S. (2018). This must be the place (to commit a crime). Testing the law of crime concentration in Milan, Italy. *European Journal of Criminology*, 15(6):702–729.

Feldman, M. (2013). Comparative Lone Wolf Terrorism: Toward a Heuristic Definition. *Democracy and Security*, 9(3):270–286.

Fellman, P. V. (2008). The Complexity of Terrorist Networks. In *2008 12th International Conference Information Visualisation*, pages 338–340.

Fellman, P. V. and Wright, R. (2014). Modeling Terrorist Networks, Complex Systems at the Mid-range. *arXiv:1405.6989 [physics]*.

Felson, M. (1987). Routine Activities and Crime Prevention in the Developing Metropolis. *Criminology*, 25(4):911–932.

Fishman, B. (2016). Defining ISIS. *Survival*, 58(1):179–188.

Fleming, M. (1980). Propaganda by the deed: Terrorism and anarchist theory in late nineteenth-century Europe. *Terrorism*, 4(1-4):1–23.

Fowler, J. H. and Dawes, C. T. (2008). Two Genes Predict Voter Turnout. *The Journal of Politics*, 70(3):579–594.

Freeman, S., Grogger, J., and Sonstelie, J. (1996). The Spatial Concentration of Crime. *Journal of Urban Economics*, 40(2):216–231.

Freilich, J. D., Chermak, S. M., Gruenewald, J., Parkin, W. S., and Klein, B. R. (2018). Patterns of Fatal Extreme-Right Crime in the United States. *Perspectives on Terrorism*, 12(6):38–51.

Fromkin, D. (1975). The Strategy of Terrorism. *Foreign Affairs*, 53(4).

Fussey, P. (2011). An economy of choice? Terrorist decision-making and criminological rational choice theories reconsidered. *Security Journal*, 24(1):85–99.

Ganor, B. (2002). Defining Terrorism: Is One Man's Terrorist another Man's Freedom Fighter? *Police Practice and Research*, 3(4):287–304.

Ganor, B. (2008). Terrorist Organization Typologies and the Probability of a Boomerang Effect. *Studies in Conflict & Terrorism*, 31(4):269–283.

Ganor, B. (2019). Artificial or Human: A New Era of Counterterrorism Intelligence? *Studies in Conflict & Terrorism*, pages 1–20.

Garrison, A. H. (2004). Defining terrorism: philosophy of the bomb, propaganda by deed and change through fear and violence. *Criminal Justice Studies*, 17(3):259–279.

Gerdes, L. M. (2015). *Illuminating Dark Networks: The Study of Clandestine Groups and Organizations*. Cambridge University Press.

Gerhard, F., Deger, M., and Truccolo, W. (2017). On the stability and dynamics of stochastic spiking neuron models: Nonlinear Hawkes process and point process GLMs. *PLOS Computational Biology*, 13(2):e1005390.

Gers, F. A., Eck, D., and Schmidhuber, J. (2002). Applying LSTM to Time Series Predictable Through Time-Window Approaches. In Tagliaferri, R. and Marinaro, M., editors, *Neural Nets WIRN Vietri-01*, Perspectives in Neural Computing, pages 193–200. Springer London.

Giles, C. L., Lawrence, S., and Tsoi, A. C. (2001). Noisy Time Series Prediction using Recurrent Neural Networks and Grammatical Inference. *Machine Learning*, 44(1):161–183.

Gill, P., Horgan, J., Hunter, S. T., and Cushenbery, L. D. (2013). Malevolent Creativity in Terrorist Organizations. *The Journal of Creative Behavior*, 47(2):125–151.

Gilli, A. and Gilli, M. (2014). The Spread of Military Innovations: Adoption Capacity Theory, Tactical Incentives, and the Case of Suicide Terrorism. *Security Studies*, 23(3):513–547.

Glasserman, P., Heidelberger, P., Shahabuddin, P., and Zajic, T. (1999). Multilevel Splitting for Estimating Rare Event Probabilities. *Operations Research*, 47(4):585–600.

Goel, S., Hofman, J. M., Lahaie, S., Pennock, D. M., and Watts, D. J. (2010). Predicting consumer behavior with Web search. *Proceedings of the National Academy of Sciences*, 107(41):17486–17490.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press. http://www.deeplearningbook.org.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative Adversarial Networks. *arXiv:1406.2661 [cs, stat]*.

Goodson, L. P. (2001). *Afghanistan's Endless War: State Failure, Regional Politics, and the Rise of the Taliban*. University of Washington Press.

Gorr, W. L. and Olligschlaeger, A. M. (2010). Weighted Spatial Adaptive Filtering: Monte Carlo Studies and Application to Illicit Drug Market Modeling. *Geographical Analysis*, 26(1):67–87.

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

Graves, A. (2013). Generating Sequences With Recurrent Neural Networks. *arXiv:1308.0850 [cs]*.

Graves, A., Mohamed, A.-r., and Hinton, G. (2013). Speech recognition with deep recurrent neural networks. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 6645–6649, Vancouver, BC, Canada. IEEE.

Gunaratna, R. and Oreg, A. (2010). Al Qaeda's Organizational Structure and its Evolution. *Studies in Conflict & Terrorism*, 33(12):1043–1078.

Guresen, E. and Kayakutlu, G. (2011). Definition of artificial neural networks with comparison to other networks. *Procedia Computer Science*, 3:426–433.

Haan, L. d. and Sinha, A. K. (1999). Estimating the probability of a rare event. *The Annals of Statistics*, 27(2):732–759.

Haberman, C. P. and Ratcliffe, J. H. (2012). The Predictive Policing Challenges of Near Repeat Armed Street Robberies. *Policing*, 6(2):151–166.

Hamilton, L. C. and Hamilton, J. D. (1983). Dynamics of Terrorism. *International Studies Quarterly*, 27(1):39.

Han, M., Xi, J., Xu, S., and Yin, F.-L. (2004). Prediction of Chaotic Time Series Based on the Recurrent Predictor Neural Network. *IEEE Transactions on Signal Processing*, 52(12):3409–3416.

Hannah-Moffat, K. (2018). Algorithmic risk governance: Big data analytics, race and information activism in criminal justice debates. *Theoretical Criminology*, page 1362480618763582.

Hansen, S. J. (2013). *Al-Shabaab in Somalia: The History and Ideology of a Militant Islamist Group*. Oxford University Press.

Hardiman, S. J., Bercot, N., and Bouchaud, J.-P. (2013). Critical reflexivity in financial markets: a Hawkes process analysis. *The European Physical Journal B*, 86(10):442.

Harmon, C. C. (2001). Five Strategies of Terrorism. *Small Wars & Insurgencies*, 12(3):39–66.

Harmon, L. D. (1959). Artificial Neuron. *Science*, 129(3354):962–963.

Harries, K. (2006). Extreme spatial variations in crime density in Baltimore County, MD. *Geoforum*, 37(3):404–416.

Hashim, A. S. (2014). The Islamic State: From al-Qaeda Affiliate to Caliphate. *Middle East Policy*, 21(4):69–83.

Hastings, J. V. and Chan, R. J. (2013). Target Hardening and Terrorist Signaling: The Case of Aviation Security. *Terrorism and Political Violence*, 25(5):777–797.

Hawkes, A. G. (1971). Spectra of Some Self-Exciting and Mutually Exciting Point Processes. *Biometrika*, 58(1):83.

Hawkes, A. G. (2018). Hawkes processes and their applications to finance: a review. *Quantitative Finance*, 18(2):193–198.

Haykin, S. S. (1994). *Neural Networks: A Comprehensive Foundation*. Macmillan ; Maxwell Macmillan Canada ; Maxwell Macmillan International, New York : Toronto : New York.

Hechter, M. and Kanazawa, S. (1997). Sociological Rational Choice Theory. *Annual Review of Sociology*, 23(1):191–214.

Heckathorn, D. D. (2002). Respondent-Driven Sampling II: Deriving Valid Population Estimates from Chain-Referral Samples of Hidden Populations. *Social Problems*, 49(1):11–34.

Heger, L., Jung, D., and Wong, W. H. (2012). Organizing for Resistance: How Group Structure Impacts the Character of Violence. *Terrorism and Political Violence*, 24(5):743–768.

Hegghammer, T. (2013). Should I Stay or Should I Go? Explaining Variation in Western Jihadists' Choice between Domestic and Foreign Fighting. *American Political Science Review*, 107(1):1–15.

Heinzen, K. (1853). *Mord und Freiheit*. New York.

Henaff, M., Bruna, J., and LeCun, Y. (2015). Deep Convolutional Networks on Graph-Structured Data. *arXiv:1506.05163 [cs]*. arXiv: 1506.05163.

Henry, V. E. (2002). *The COMPSTAT Paradigm: Management Accountability in Policing, Business, and the Public Sector*. Looseleaf Law Publications.

Hess, G. D. (1995). An Introduction To Lewis Fry Richardson and His Mathematical Theory of War and Peace. *Conflict Management and Peace Science*, 14(1):77–113.

Hinton, G. E. (2012). A Practical Guide to Training Restricted Boltzmann Machines. In Montavon, G., Orr, G. B., and Müller, K.-R., editors, *Neural Networks: Tricks of the Trade*, volume 7700, pages 599–619. Springer Berlin Heidelberg, Berlin, Heidelberg.

Hinton, G. E., Osindero, S., and Teh, Y.-W. (2006). A Fast Learning Algorithm for Deep Belief Nets. *Neural Computation*, 18(7):1527–1554.

Ho, S., Xie, M., and Goh, T. (2002). A comparative study of neural network and Box-Jenkins ARIMA modeling in time series prediction. *Computers & Industrial Engineering*, 42(2-4):371–375.

Hochreiter, S. and Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780.

Hodge, V. and Austin, J. (2004). A Survey of Outlier Detection Methodologies. *Artificial Intelligence Review*, 22(2):85–126.

Hoffman, B. (1993). Terrorist targeting: Tactics, trends, and potentialities. *Terrorism and Political Violence*, 5(2):12–29.

Hoffman, B. (1997). The confluence of international and domestic trends in terrorism. *Terrorism and Political Violence*, 9(2):1–15.

Hoffman, B. (1998). *Inside Terrorism*. Columbia University Press.

Hoffman, B. (2002). Rethinking Terrorism and Counterterrorism Since 9/11. *Studies in Conflict & Terrorism*, 25(5):303–316.

Holbrook, D. (2015). Al-Qaeda and the Rise of ISIS. *Survival*, 57(2):93–104.

Holden, R. T. (1986). The Contagiousness of Aircraft Hijacking. *American Journal of Sociology*, 91(4):874–904.

Holland, T. R. and McGarvey, B. (1984). Crime specialization, seriousness progression, and Markov chains. *Journal of Consulting and Clinical Psychology*, 52(5):837–840.

Horgan, J. (2005). *The Psychology of Terrorism*, volume 20054641 of *Cass Series on Political Violence*. Routledge.

Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2):251–257.

Huang, B. and Carley, K. M. (2019). Inductive Graph Representation Learning with Recurrent Graph Neural Networks. *arXiv:1904.08035 [cs, stat]*. arXiv: 1904.08035.

Huang, C., Zhang, J., Zheng, Y., and Chawla, N. V. (2018). DeepCrime: Attentive Hierarchical Recurrent Networks for Crime Prediction. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, CIKM '18, pages 1423–1432, New York, NY, USA. ACM. event-place: Torino, Italy.

Hull, C. L. (1943). *Principles of Behavior: An Introduction to Behavior Theory*. D. Appleton-Century Company, Incorporated.

Hyman, A. (2016). *Afghanistan under Soviet Domination, 1964–91*. Springer.

Hälterlein, J. and Ostermeier, L. (2018). Special Issue: Predictive Security Technologies. *European Journal for Security Research*, 3(2):91–94.

Ineichen, B. (2018). ISIS and the Caliphate: triumph, defeat, and dispersal. *Mental Health, Religion & Culture*, 21(3):319–324.

Ingiriis, M. H. (2018). The invention of Al-Shabaab in Somalia: Emulating the anti-colonial dervishes movement. *African Affairs*, 117(467):217–237.

Institute for Economics and Peace (2016). Global Terrorism Index Report. Technical report, Institute for Economics and Peace.

Institute for Economics and Peace (2017). Global Terrorism Index Report. Technical report, Institute for Economics and Peace.

Institute for Economics and Peace (2018). Global Terrorism Index Report. Technical report, Institute for Economics and Peace.

Ivanova, K. and Sandler, T. (2006). CBRN Incidents: Political Regimes, Perpetrators, and Targets. *Terrorism and Political Violence*, 18(3):423–448.

Iyekekpolo, W. O. (2016). Boko Haram: Understanding the context. *Third World Quarterly*, 37(12):2211–2228.

Jabareen, Y. (2015). The emerging Islamic State: Terror, territoriality, and the agenda of social transformation. *Geoforum*, 58:51–55.

Jackman, S. (2004). Bayesian Analysis for Political Research. *Annual Review of Political Science*, 7(1):483–505.

Jackson, B. A. and Frelinger, D. R. (2008). Rifling Through the Terrorists' Arsenal: Exploring Groups' Weapon Choices and Technology Strategies. *Studies in Conflict & Terrorism*, 31(7):583–604.

Jain, A., Zamir, A. R., Savarese, S., and Saxena, A. (2015). Structural-RNN: Deep Learning on Spatio-Temporal Graphs. *arXiv:1511.05298 [cs]*.

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., and Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301):790–794.

Jenkins, B. M. (1974). International Terrorism: A New Kind of Warfare. Technical Report P-5261, Rand Corporation - Santa Monica, California.

Johnson, N., Carran, S., Botner, J., Fontaine, K., Laxague, N., Nuetzel, P., Turnley, J., and Tivnan, B. (2011). Pattern in Escalations in Insurgent and Terrorist Activity. *Science*, 333(6038):81–84.

Johnson, T. H. (2013). Taliban adaptations and innovations. *Small Wars & Insurgencies*, 24(1):3–27.

Jongman, A. (2017). *Political Terrorism: A New Guide to Actors, Authors, Concepts, Data Bases, Theories, and Literature*. Routledge, 1 edition.

Jongman, A. J. and Schmid, A. P. (1988). *Political Terrorism: A New Guide To Actors, Authors, Concepts, Data Bases, Theories, And Literature*. Transaction Publishers.

Joosse, P., Bucerius, S. M., and Thompson, S. K. (2015). Narratives and Counternarratives: Somali-Canadians on Recruitment as Foreign Fighters to Al-Shabaab. *British Journal of Criminology*, 55(4):811–832.

Judge, G. G. and Swanson, E. (1962). Markov Chains: Basic Concepts and Suggested Uses in Agricultural Economics. *Australian Journal of Agricultural Economics*, 6(2):49–61.

Kang, H.-W. and Kang, H.-B. (2017). Prediction of crime occurrence from multi-modal data using deep learning. *PLOS ONE*, 12(4):e0176244.

Kassel, W. (2009). Terrorism and the International Anarchist Movement of the Late Nineteenth and Early Twentieth Centuries. *Studies in Conflict & Terrorism*, 32(3):237–252.

Keller, J. P., Desouza, K. C., and Lin, Y. (2010). Dismantling terrorist networks: Evaluating strategic options using agent-based modeling. *Technological Forecasting and Social Change*, 77(7):1014–1036.

Kijima, M. (2016). *Stochastic Processes with Applications to Finance, Second Edition*. Chapman and Hall/CRC.

King, G. and Zeng, L. (2001). Explaining Rare Events in International Relations. *International Organization*, 55(3):693–715.

Kingma, D. P. and Ba, J. (2014). Adam: A Method for Stochastic Optimization. *arXiv:1412.6980 [cs]*. arXiv: 1412.6980.

Klausen, J. (2015). Tweeting the Jihad: Social Media Networks of Western Foreign Fighters in Syria and Iraq. *Studies in Conflict & Terrorism*, 38(1):1–22.

Klein, G. A., Orasanu, J., Calderwood, R., and Zsambok, C. E. (1993). *Decision making in action: Models and methods*. Decision making in action: Models and methods. Ablex Publishing, Westport, CT, US.

Kleiner, J. (2000). The Taliban and Islam. *Diplomacy & Statecraft*, 11(1):19–32.

Kobayashi, R. and Lambiotte, R. (2016). TiDeH: Time-Dependent Hawkes Process for Predicting Retweet Dynamics. In *Tenth International AAAI Conference on Web and Social Media*.

Kou, S. G. and Sobel, M. E. (2004). Forecasting the Vote: A Theoretical Comparison of Election Markets and Public Opinion Polls. *Political Analysis*, 12(3):277–295.

Krebs, V. (2002). Mapping Networks of Terrorist Cells. *Connections*, 24(3):43–52.

Krittanawong, C., Zhang, H., Wang, Z., Aydar, M., and Kitai, T. (2017). Artificial Intelligence in Precision Cardiovascular Medicine. *Journal of the American College of Cardiology*, 69(21):2657–2664.

Krueger, A. B. and Laitin, D. D. (2008). Kto Kogo?: A Cross-Country Study of the Origins and Targets of Terrorism. In Keefer, P. and Loayza, N., editors, *Terrorism, Economic Development, and Political Openness*, pages 148–173. Cambridge University Press, Cambridge.

Kurowski, L. and Sussman, D. (2011). Domestic and International Terrorism. In *Investment Project Design: A Guide to Financial and Economic Analysis with Constraints*, pages 1–1. John Wiley & Sons, Inc.

Kydd, A. H. and Walter, B. F. (2006). The Strategies of Terrorism. *International Security*, 31(1):49–80.

LaFree, G. (2010). The Global Terrorism Database: Accomplishments and Challenges. *Perspectives on Terrorism*, 4(1).

LaFree, G. and Dugan, L. (2007). Introducing the Global Terrorism Database. *Terrorism and Political Violence*, 19(2):181–204.

LaFree, G., Dugan, L., Xie, M., and Singh, P. (2012). Spatial and Temporal Patterns of Terrorist Attacks by ETA 1970 to 2007. *Journal of Quantitative Criminology*, 28(1):7–29.

LaFree, G. and Freilich, J. D. (2012). Editor's Introduction: Quantitative Approaches to the Study of Terrorism. *Journal of Quantitative Criminology*, 28(1):1–5.

LaFree, G. and Freilich, J. D. (2016). Bringing Criminology into the Study of Terrorism. In LaFree, G. and Freilich, J. D., editors, *The Handbook of the Criminology of Terrorism*, pages 1–14. John Wiley & Sons, Inc.

Lakdawalla, D. and Zanjani, G. (2005). Insurance, self-protection, and the economics of terrorism. *Journal of Public Economics*, 89(9-10):1891–1905.

Lake, D. A. (2002). Rational Extremism: Understanding Terrorism in the Twenty-first Century. *Dialogue IO*, 1(1):15–28.

Lallouache, M. and Challet, D. (2014). The limits of statistical significance of Hawkes processes fitted to financial data. *arXiv:1406.3967 [q-fin]*. arXiv: 1406.3967.

Lamb, A., Goyal, A., Zhang, Y., Zhang, S., Courville, A., and Bengio, Y. (2016). Professor Forcing: A New Algorithm for Training Recurrent Networks. *arXiv:1610.09038 [cs, stat]*.

Laqueur, W. (1987). *The Age of Terrorism*. Little, Brown.

Latora, V. and Marchiori, M. (2004). How the science of complex networks can help developing strategies against terrorism. *Chaos, Solitons & Fractals*, 20(1):69–75.

Laub, P. J., Taimre, T., and Pollett, P. K. (2015). Hawkes Processes. *arXiv:1507.02822 [math, q-fin, stat]*. arXiv: 1507.02822.

Lawrence, S., Giles, C., Ah Chung Tsoi, and Back, A. (Jan./1997). Face recognition: A convolutional neural-network approach. *IEEE Transactions on Neural Networks*, 8(1):98–113.

Le Gallo, J. (2004). Space-Time Analysis of GDP Disparities among European Regions: A Markov Chains Approach. *International Regional Science Review*, 27(2):138–163.

Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324.

Lee, D., Goh, K.-I., Kahng, B., and Kim, D. (2010). Complete trails of coauthorship network evolution. *Physical Review E*, 82(2).

Lee, K. and Seo, B. K. (2017). Marked Hawkes process modeling of price dynamics and volatility estimation. *Journal of Empirical Finance*, 40:174–200.

Levada, A. L. M., Correa, D. C., Salvadeo, D. H. P., Saito, J. H., and Mascarenhas, N. D. A. (2008). Novel approaches for face recognition: Template-matching using Dynamic Time Warping and LSTM neural network supervised classification. In *2008 15th International Conference on Systems, Signals and Image Processing*, pages 241–244.

Lewis, E., Mohler, G., Brantingham, P. J., and Bertozzi, A. L. (2012). Self-exciting point process models of civilian deaths in Iraq. *Security Journal*, 25(3):244–264.

Lewis-Beck, M. S. and Rice, T. W. (1984). Forecasting presidential elections: A comparison of naive models. *Political Behavior*, 6(1):9–21.

Li, L. and Zha, H. (2015). Energy Usage Behavior Modeling in Energy Disaggregation via Marked Hawkes Process. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*.

Li, Q. and Schaub, D. (2004). Economic Globalization and Transnational Terrorism: A Pooled Time-Series Analysis. *Journal of Conflict Resolution*, 48(2):230–258.

Lind, J., Mutahi, P., and Oosterom, M. (2017). 'Killing a mosquito with a hammer': Al-Shabaab violence and state security responses in Kenya. *Peacebuilding*, 5(2):118–135.

Linse, U. (1982). 'Propaganda by Deed' and 'Direct Action': Two Concepts of Anarchist Violence. In Mommsen, W. J. and Hirschfeld, G., editors, *Social Protest, Violence and Terror in Nineteenth- and Twentieth-century Europe*, pages 201–229. Palgrave Macmillan UK, London.

Lipton, Z. C., Kale, D. C., Elkan, C., and Wetzel, R. (2015). Learning to Diagnose with LSTM Recurrent Neural Networks. *arXiv:1511.03677 [cs]*.

Liu, Q., Wu, S., Wang, L., and Tan, T. (2016). Predicting the Next Location: A Recurrent Model with Spatial and Temporal Contexts. In *Thirtieth AAAI Conference on Artificial Intelligence*.

Llusa, F. and Tavares, J. (2007). The economics of terrorism: A synopsis. *The Economics of Peace and Security Journal*, 2(1).

Loimeier, R. (2012). Boko Haram: The Development of a Militant Religious Movement in Nigeria. *Africa Spectrum*, 47(2/3):137–155.

Mahood, S. and Rane, H. (2017). Islamist narratives in ISIS recruitment propaganda. *The Journal of International Communication*, 23(1):15–35.

Mainas, E. D. (2012). The Analysis of Criminal and Terrorist Organisations as Social Network Structures: A Quasi-Experimental Study. *International Journal of Police Science & Management*, 14(3):264–282.

Malet, D. (2010). Why Foreign Fighters? Historical perspectives and solutions. *Obis Journal of Foreign Affairs*, 54(1):97–114.

Malet, D. (2013). *Foreign Fighters: Transnational Identity in Civil Conflicts*. Oxford University Press.

Malm, A., Nash, R., and Moghadam, R. (2016). Social Network Analysis and Terrorism. In LaFree, G. and Freilich, J. D., editors, *The Handbook of the Criminology of Terrorism*, pages 221–231. John Wiley & Sons, Inc.

Marchant, R., Haan, S., Clancey, G., and Cripps, S. (2018). Applying machine learning to criminology: semi-parametric spatial-demographic Bayesian regression. *Security Informatics*, 7(1).

Martens, A., Sainudiin, R., Sibley, C. G., Schimel, J., and Webber, D. (2014). Terrorist Attacks Escalate in Frequency and Fatalities Preceding Highly Lethal Attacks. *PLoS ONE*, 9(4):e93732.

Martin, A. D. and Quinn, K. M. (2002). Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953–1999. *Political Analysis*, 10(2):134–153.

Martin, S. and Perliger, A. (2012). Turning to and from Terror: Deciphering the Conditions under which Political Groups Choose Violent and Nonviolent Tactics. *Perspectives on Terrorism*, 6(4/5):21–45.

Massey, F. J. (1951). The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association*, 46(253):68.

Masters, D. and Luschi, C. (2018). Revisiting Small Batch Training for Deep Neural Networks. *arXiv:1804.07612 [cs, stat]*. arXiv: 1804.07612.

Mateus, A. and Caeiro, F. (2013). Comparing several tests of randomness based on the difference of observations. pages 809–812, Rhodes, Greece.

Matusitz, J. (2012). *Terrorism & Communication: A Critical Introduction*. Sage Pubns, Thousand Oaks.

May, W. F. (1974). Terrorism as Strategy and Ecstasy. *Social Research*, 41(2):277–298.

McCauley, C., Moskalenko, S., and Van Son, B. (2013). Characteristics of Lone-Wolf Violent Offenders: A Comparison of Assassins and School Attackers. *Perspectives on Terrorism*, 7(1).

McCormick, G. H. (2003). Terrorist Decision Making. *Annual Review of Political Science*, 6(1):473–507.

McCulloch, W. S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4):115–133.

Medina, R. and Hepner, G. (2008). Geospatial Analysis of Dynamic Terrorist Networks. In *Values and Violence*, Studies in Global Justice, pages 151–167. Springer, Dordrecht.

Merari, A. (1978). A classification of terrorist groups. *Terrorism*, 1(3-4):331–346.

Merari, A. (1999). Terrorism as a strategy of struggle: Past and future. *Terrorism and Political Violence*, 11(4):52–65.

Merrill, J. A., Sheehan, B. M., Carley, K. M., and Stetson, P. D. (2015). Transition Networks in a Cohort of Patients with Congestive Heart Failure: A novel application of informatics methods to inform care coordination. *Applied Clinical Informatics*, 06(03):548–564.

Midlarsky, M. I. (1978). Analyzing Diffusion and Contagion Effects: The Urban Disorders of the 1960s. *American Political Science Review*, 72(03):996–1008.

Midlarsky, M. I., Crenshaw, M., and Yoshida, F. (1980). Why Violence Spreads: The Contagion of International Terrorism. *International Studies Quarterly*, 24(2):262–298.

Migaux, P. (2007). Al Qaeda. In Chaliand, G. and , A., editors, *The History of Terrorism: From Antiquity to Al Qaeda*. University of California Press.

Moffitt, T. E. (2003). Life-course-persistent and adolescence-limited antisocial behavior: A 10-year research review and a research agenda. In *Causes of conduct disorder and juvenile delinquency*, pages 49–75. The Guilford Press, New York, NY, US.

Moghadam, A. (2013). How Al Qaeda Innovates. *Security Studies*, 22(3):466–497.

Mohler, G. (2013). Modeling and estimation of multi-source clustering in crime and security data. *The Annals of Applied Statistics*, 7(3):1525–1539.

Mohler, G. (2014). Marked point process hotspot maps for homicide and gun crime prediction in Chicago. *International Journal of Forecasting*, 30(3):491–497.

Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P., and Tita, G. E. (2011). Self-Exciting Point Process Modeling of Crime. *Journal of the American Statistical Association*, 106(493):100–108.

Monroe, K. R. and Maher, K. H. (1995). Psychology and Rational Actor Theory. *Political Psychology*, 16(1):1.

Monti, F., Boscaini, D., Masci, J., Rodola, E., Svoboda, J., and Bronstein, M. M. (2017). Geometric Deep Learning on Graphs and Manifolds Using Mixture Model CNNs. pages 5115–5124.

Moon, I. C. and Carley, K. M. (2007). Modeling and Simulating Terrorist Networks in Social and Geospatial Dimensions. *IEEE Intelligent Systems*, 22(5):40–49.

Moore, G. H. and Wallis, W. A. (1943). Time Series Significance Tests Based on Signs of Differences. *Journal of the American Statistical Association*, 38(222):153.

Morris, N. A. (2015). Target Suitability and Terrorism Events at Places: Terrorism Target Suitability. *Criminology & Public Policy*, 14(2):417–426.

Mueller, J. C. (2018). The Evolution of Political Violence: The Case of Somalia's Al-Shabaab. *Terrorism and Political Violence*, 30(1):116–141.

Munk, T. B. (2017). 100,000 false positives for every real terrorist: Why anti-terror algorithms don't work. *First Monday*, 22(9).

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. The MIT Press, Cambridge, MA, 1 edition edition.

Murray, A. T., McGuffog, I., Western, J. S., and Mullins, P. (2001). Exploratory Spatial Data Analysis Techniques for Examining Urban CrimeImplications for Evaluating Treatment. *The British Journal of Criminology*, 41(2):309–329.

Myers, D. J. (2000). The Diffusion of Collective Violence: Infectiousness, Susceptibility, and Mass Media Networks. *American Journal of Sociology*, 106(1):173–208.

Nagin, D. S. and Tremblay, R. E. (2005). What Has Been Learned from Group-Based Trajectory Modeling? Examples from Physical Aggression and Other Problem Behaviors. *The ANNALS of the American Academy of Political and Social Science*, 602(1):82–117.

Najjar, A., Kaneko, S., and Miyanaga, Y. (2018). Crime Mapping from Satellite Imagery via Deep Learning. *arXiv:1812.06764 [cs]*. arXiv: 1812.06764.

Nath, S. V. (2006). Crime Pattern Detection Using Data Mining. In *2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology Workshops*, pages 41–44, Hong Kong. IEEE.

Naylor, J. C., Pritchard, R. D., and Ilgen, D. R. (2013). *A Theory of Behavior in Organizations*. Academic Press.

Nelder, J. A. and Mead, R. (1965). A Simplex Method for Function Minimization. *The Computer Journal*, 7(4):308–313.

Nettler, R. L. (1996). Guidelines for the Islamic community: Sayyid Qutb's political interpretation of the Qur'an. *Journal of Political Ideologies*, 1(2):183–196.

Neuilly, M.-A., Zgoba, K. M., Tita, G. E., and Lee, S. S. (2011). Predicting Recidivism in Homicide Offenders Using Classification Tree Analysis. *Homicide Studies*, 15(2):154–176.

Neumann, D. P. R. and Smith, D. M. L. R. (2005). Strategic terrorism: The framework and its fallacies. *Journal of Strategic Studies*, 28(4):571–595.

Nilsson, N. J. (2014). *Principles of Artificial Intelligence*. Morgan Kaufmann.

Norris, J. R. (1998). *Markov Chains*. Cambridge University Press.

Ogata, Y. (1988). Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes. *Journal of the American Statistical Association*, 83(401):9–27.

Ogata, Y. (1998). Space-Time Point-Process Models for Earthquake Occurrences. *Annals of the Institute of Statistical Mathematics*, 50(2):379–402.

Onuoha, F. C. (2010). The Islamist challenge: Nigeria's Boko Haram crisis explained. *African Security Review*, 19(2):54–67.

Ozkan, T. (2019). Criminology in the age of data explosion: New directions. *The Social Science Journal*, 56(2):208–219.

PAI (2019). Report on Algorithmic Risk Assessment Tools in the U.S. Criminal Justice System. Technical report.

Papachristos, A. V. (2014). The Network Structure of Crime: Networks & Crime. *Sociology Compass*, 8(4):347–357.

Pape, R. (2005). *Dying to Win: The Strategic Logic of Suicide Terrorism*. Random House Publishing Group.

Pascanu, R., Mikolov, T., and Bengio, Y. (2012). On the difficulty of training Recurrent Neural Networks. *arXiv:1211.5063 [cs]*.

Peng, R. (2003). Multi-dimensional Point Process Models in R. *Journal of Statistical Software*, 8(1):1–27.

Pentland, A. and Liu, A. (1999). Modeling and Prediction of Human Behavior. *Neural Computation*, 11(1):229–242.

Perry, M. and Negrin, H. E. (2008). *The Theory and Practice of Islamic Terrorism: An Anthology*. Palgrave Macmillan.

Perry, W. L. (2013). *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*. Rand Corporation.

Peters, M. A. (2014). 'Western Education is Sinful': Boko Haram and the Abduction of Chibok Schoolgirls. *Policy Futures in Education*, 12(2):186–190.

Pettiway, L. E., Dolinsky, S., and Grigoryan, A. (1994). The drug and criminal activity patterns of urban offenders: A Markov chain analysis. *Journal of Quantitative Criminology*, 10(1):79–107.

Pham, J. P. (2011). The Dangerous "Pragmatism" of Al-Qaeda in the Islamic Maghreb. *The Journal of the Middle East and Africa*, 2(1):15–29.

Piazza, J. A. (2009). Is Islamist Terrorism More Dangerous?: An Empirical Study of Group Ideology, Organization, and Goal Structure. *Terrorism and Political Violence*, 21(1):62–88.

Piazza, J. A. (2012). The Opium Trade and Patterns of Terrorism in the Provinces of Afghanistan: An Empirical Analysis. *Terrorism and Political Violence*, 24(2):213–234.

Picco, G. (2004). Tactical and Strategic Terrorism. *Journal of Aggression, Maltreatment & Trauma*, 9(1-2):71–78.

Pitcher, B. L., Hamblin, R. L., and Miller, J. L. L. (1978). The Diffusion of Collective Violence. *American Sociological Review*, 43(1):23.

Pizam, A. (2010). Hotels as tempting targets for terrorism attacks. *International Journal of Hospitality Management*, 29(1):1.

Polo, S. M. and Gleditsch, K. S. (2016). Twisting arms and sending messages: Terrorist tactics in civil war. *Journal of Peace Research*, 53(6):815–829.

Porter, M. D. and White, G. (2012). Self-exciting hurdle models for terrorist activity. *The Annals of Applied Statistics*, 6(1):106–124.

Price, H. E. (1977). The Strategy and Tactics of Revolutionary Terrorism. *Comparative Studies in Society and History*, 19(1):52–66.

Quetelet, A. (1831). *Research on the Propensity for Crime at Different Ages.* Anderson, Cincinnati, Ohio.

Raghavan, V., Galstyan, A., and Tartakovsky, A. G. (2013). Hidden Markov models for the activity profile of terrorist groups. *The Annals of Applied Statistics*, 7(4):2402–2430. arXiv: 1207.1497.

Rashid, A. (2002). *Taliban: Islam, Oil and the New Great Game in Central Asia.* I.B.Tauris.

Ravndal, J. A. (2018). Explaining right-wing terrorism and violence in Western Europe: Grievances, opportunities and polarisation. *European Journal of Political Research*, 57(4):845–866.

Regens, J. L., Mould, N., Vernon, E., and Montgomery, A. (2016). Operational Dynamics of Boko Haram's Terrorist Campaign Following Leadership Succession. *Social Science Quarterly*, 97(1):44–52.

Reid, E. F. and Chen, H. (2007). Mapping the contemporary terrorism research domain. *International Journal of Human-Computer Studies*, 65(1):42–56.

Reijsbergen, D., de Boer, P.-T., Scheinhardt, W., and Haverkort, B. (2013). Automated Rare Event Simulation for Stochastic Petri Nets. In Joshi, K., Siegle, M., Stoelinga, M., and D'Argenio, P. R., editors, *Quantitative Evaluation of Systems*, Lecture Notes in Computer Science, pages 372–388. Springer Berlin Heidelberg.

Reinhart, A. (2018). A Review of Self-Exciting Spatio-Temporal Point Processes and Their Applications. *Statistical Science*, 33(3):299–318.

Revuz, D. (2008). *Markov Chains.* Elsevier.

Reynaud-Bouret, P., Rivoirard, V., and Tuleau-Malot, C. (2013). Inference of functional connectivity in Neurosciences via Hawkes processes. In *2013 IEEE Global Conference on Signal and Information Processing*, pages 317–320.

Richardson, L. (2013). *The Roots of Terrorism*. Routledge.

Richardson, L. F. (1960). *Arms and insecurity: a mathematical study of the causes and origins of war.* Boxwood Press, Pittsburgh. OCLC: 567792.

Richardson, L. F., Wright, Q., and Lienau, C. C. (1960). *Statistics of Deadly Quarrels*. Boxwood Pr, Pacific Grove.

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386–408.

Ross, J. I. (2006). *Political Terrorism: An Interdisciplinary Approach*. Peter Lang.

Rousseeuw, P. J. and Leroy, A. M. (1987). *Robust Regression and Outlier Detection*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, USA.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088):533–536.

Russell, S. J. and Norvig, P. (2010). *Artificial Intelligence: A Modern Approach*. Pearson College Div, Upper Saddle River, 3rd edition.

Ruths, D. and Pfeffer, J. (2014). Social media for large studies of behavior. *Science*, 346(6213):1063–1064.

Sageman, M. (2014). The Stagnation in Terrorism Research. *Terrorism and Political Violence*, 26(4):565–580.

Sandler, T. (2003). Collective Action and Transnational Terrorism. *The World Economy*, 26(6):779–802.

Sandler, T. and Lapan, H. E. (1988). The calculus of dissent: An analysis of terrorists' choice of targets. *Synthese*, 76(2):245–261.

Santifort, C., Sandler, T., and Brandt, P. T. (2013). Terrorist attack and target diversity: Changepoints and their drivers. *Journal of Peace Research*, 50(1):75–90.

Saul, L. K. and Roweis, S. T. (2013). Think globally, fit locally: Unsupervised learning of low dimensional manifolds. *The Journal of Machine Learning Research*, 4.

Saunders, J., Hunt, P., and Hollywood, J. S. (2016). Predictions put into practice: a quasi-experimental evaluation of Chicago's predictive policing pilot. *Journal of Experimental Criminology*, 12(3):347–371.

Schelling, T. C. (1980). *The Strategy of Conflict*. Harvard University Press.

Schelling, T. C. (1991). What purposes can "international terrorism" serve? In Frey, R. G. and Morris, C. W., editors, *Violence, Terrorism and Justice*. Cambridge University Press.

Schmid, A. P. (2014). Comments on Marc Sageman's Polemic "The Stagnation in Terrorism Research". *Terrorism and Political Violence*, 26(4):587–595.

Schrodt, P. A. (2006). Forecasting Conflict in the Balkans using Hidden Markov Models. In Trappl, R., editor, *Programming for Peace*, pages 161–184. Springer Netherlands, Dordrecht.

Schuurman, B. (2018). Research on Terrorism, 2007–2016: A Review of Data, Methods, and Authorship. *Terrorism and Political Violence*, pages 1–16.

Schuurman, B. (2019). Topics in terrorism research: reviewing trends and gaps, 2007-2016. *Critical Studies on Terrorism*, 0(0):1–18.

Schwartz, S. J., Dunkel, C. S., and Waterman, A. S. (2009). Terrorism: An Identity Theory Perspective. *Studies in Conflict & Terrorism*, 32(6):537–559.

Shapiro, A. (2017). Reform predictive policing. *Nature News*, 541(7638):458.

Shaw, C. R., Zorbaugh, F. M., McKay, H. D., and Cottrell, L. S. (1929). *Delinquency Areas: A Study of the Geographic Distribution of School Truants, Juvenile Delinquents, and Adult Offenders in Chicago*. University of Chicago Press.

Sherman, L. W., Gartin, P. R., and Buerger, M. E. (1989). Hot Spots of Predatory Crime: Routine Activities and the Criminology of Place*. *Criminology*, 27(1):27–56.

Siebeneck, L. K., Medina, R. M., Yamada, I., and Hepner, G. F. (2009). Spatial and Temporal Analyses of Terrorist Incidents in Iraq, 2004–2006. *Studies in Conflict & Terrorism*, 32(7):591–610.

Siegel, E. (2016). *Predictive analytics: the power to predict who will click, buy, lie, or die*. Wiley, Hoboken, New Jersey, revised and updated edition edition.

Silke, A. (2003). *Terrorists, Victims and Society: Psychological Perspectives on Terrorism and its Consequences*. John Wiley & Sons.

Silke, A. (2008). Research on Terrorism. In Chen, H., Reid, E., Sinai, J., Silke, A., and Ganor, B., editors, *Terrorism Informatics: Knowledge Management and Data Mining for Homeland Security*, Integrated Series In Information Systems, pages 27–50. Springer US, Boston, MA.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., and Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587):484–489.

Simon, H. A. (1987). Rationality in psychology and economics. In *Rational choice: The contrast between economics and psychology*, pages 25–40. University of Chicago Press, Chicago, IL, US.

Simons, A. and Tucker, D. (2007). The misleading problem of failed states: A 'socio-geography' of terrorism in the post-9/11 era. *Third World Quarterly*, 28(2):387–401.

Skillicorn, D. B., Walther, O., Leuprecht, C., and Zheng, Q. (2019). The Diffusion and Permeability of Political Violence in North and West Africa. *Terrorism and Political Violence*, pages 1–23.

Smith, C. M. and Papachristos, A. V. (2016). Trust Thy Crooked Neighbor: Multiplexity in Chicago Organized Crime Networks. *American Sociological Review*, 81(4):644–667.

Sorensen, A. B. (1978). Mathematical Models in Sociology. *Annual Review of Sociology*, 4(1):345–371.

Soufan Group (2015). Foreign Fighters: An Updated Assessment of the Flow of Foreign Fighters into Syria and Iraq. Technical report.

Soutner, D. and Müller, L. (2013). Application of LSTM Neural Networks in Language Modelling. In Habernal, I. and Matoušek, V., editors, *Text, Speech, and Dialogue*, Lecture Notes in Computer Science, pages 105–112. Springer Berlin Heidelberg.

Spaaij, R. (2010). The Enigma of Lone Wolf Terrorism: An Assessment. *Conflict & Terrorism*, 33(9).

Spaaij, R. (2011). *Understanding Lone Wolf Terrorism: Global Patterns, Motivations and Prevention*. Springer Science & Business Media.

Springer, D. R., Regens, J. L., and Edger, D. N. (2009). *Islamic Radicalism and Global Jihad*. Georgetown University Press.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15:1929–1958.

Stalidis, P., Semertzidis, T., and Daras, P. (2018). Examining Deep Learning Architectures for Crime Classification and Prediction. *arXiv:1812.00602 [cs, stat]*. arXiv: 1812.00602.

Stander, J., Farrington, D. P., Hill, G., and Altham, P. M. E. (1989). Markov Chain Analysis and Specialization in Criminal Careers. *The British Journal of Criminology*, 29(4):317–335.

START (2017a). Global Terrorism Database (Data file).

START (2017b). GTD Codebook: Inclusion Criteria and Variables. Technical report, University of Maryland.

Stec, A. and Klabjan, D. (2018). Forecasting Crime with Deep Learning. *arXiv:1806.01486 [cs, stat]*. arXiv: 1806.01486.

Stergiou, D. (2016). ISIS political economy: Financing a terror state. *Journal of Money Laundering Control; London*, 19(2):189–207.

Subrahmanian, V. S. (2012). *Handbook of Computational Approaches to Counterterrorism*. Springer Science & Business Media.

Subrahmanian, V. S. and Kumar, S. (2017). Predicting human behavior: The next frontiers. *Science*, 355(6324):489–489.

Sun, A., Naing, M.-M., Lim, E.-P., and Lam, W. (2003). Using Support Vector Machines for Terrorism Information Extraction. In Chen, H., Miranda, R., Zeng, D. D., Demchak, C., Schroeder, J., and Madhusudan, T., editors, *Intelligence and*

Security Informatics, Lecture Notes in Computer Science, pages 1–12. Springer Berlin Heidelberg.

Sutherland, E. H., Cressey, D. R., and Luckenbill, D. F. (1992). *Principles of Criminology*. Rowman & Littlefield.

Swallow, K. M. and Jiang, Y. V. (2012). Goal-relevant events need not be rare to boost memory for concurrent images. *Attention, Perception, & Psychophysics*, 74(1):70–82.

Tar, U. A. and Mustapha, M. (2017). Al-Shabaab: State Collapse, Warlords and Islamist Insurgency in Somalia. In Varin, C. and Abubakar, D., editors, *Violent Non-State Actors in Africa*, pages 277–299. Springer International Publishing, Cham.

Taylor, M. (2014). If I Were You, I Wouldn't Start From Here: Response to Marc Sageman's "The Stagnation in Terrorism Research". *Terrorism and Political Violence*, 26(4):581–586.

Tench, S. (2018). *Space-Time Modelling of Terrorism and Counter-Terrorism*. Doctoral Dissertation, University College London - Department of Security and Crime Sciences; Department of Advanced Spatial Analysis.

Tench, S., Fry, H., and Gill, P. (2016). Spatio-temporal patterns of IED usage by the Provisional Irish Republican Army. *European Journal of Applied Mathematics*, 27(3):377–402.

Thruelsen, P. D. (2010). The Taliban in southern Afghanistan: A localised insurgency with a local objective. *Small Wars & Insurgencies*, 21(2):259–276.

Thuraisingham, B. (2003). *Web Data Mining and Applications in Business Intelligence and Counter-Terrorism*. CRC Press.

Tian, F., Gao, B., Cui, Q., Chen, E., and Liu, T.-Y. (2014). Learning Deep Representations for Graph Clustering. In *Twenty-Eighth AAAI Conference on Artificial Intelligence*.

Tieleman, T. and Hinton, G. (2012). Lecture 6.5 - RMSProp. Technical report.

Tilly, C. (2004). Terror, Terrorism, Terrorists. *Sociological Theory*, 22(1):5–13.

Toft, P., Duero, A., and Bieliauskas, A. (2010). Terrorist targeting and energy security. *Energy Policy*, 38(8):4411–4421.

Turk, A. T. (2004). Sociology of Terrorism. *Annual Review of Sociology*, 30(1):271–286.

Tønnessen, T. H. (2019). The Islamic State after the Caliphate. *Perspectives on Terrorism*, 13(1):2–11.

Türkyilmaz, K., van Lieshout, M. N. M., and Stein, A. (2013). Comparing the Hawkes and Trigger Process Models for Aftershock Sequences Following the 2005 Kashmir Earthquake. *Mathematical Geosciences*, 45(2):149–164.

Veen, A. and Schoenberg, F. P. (2008). Estimation of Space–Time Branching Process Models in Seismology Using an EM–Type Algorithm. *Journal of the American Statistical Association*, 103(482):614–624.

Vidino, L. (2014). European foreign fighters in Syria: Dynamics and responses. *European View*, 13(2):217–224.

Vittori, K., Talbot, G., Gautrais, J., Fourcassié, V., Araújo, A. F., and Theraulaz, G. (2006). Path efficiency of ant foraging trails in an artificial network. *Journal of Theoretical Biology*, 239(4):507–515.

von Neumann, J. (1941). Distribution of the Ratio of the Mean Square Successive Difference to the Variance. *The Annals of Mathematical Statistics*, 12(4):367–395.

Vosoughi, S., Roy, D., and Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380):1146–1151.

Wang, B., Yin, P., Bertozzi, A. L., Brantingham, P. J., Osher, S. J., and Xin, J. (2017). Deep Learning for Real-Time Crime Forecasting and its Ternarization. *arXiv:1711.08833 [cs, math, stat]*. arXiv: 1711.08833.

Wang, X., Gerber, M. S., and Brown, D. E. (2012). Automatic Crime Prediction Using Events Extracted from Twitter Posts. In Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J. M., Mattern, F., Mitchell, J. C., Naor, M., Nierstrasz, O., Pandu Rangan, C., Steffen, B., Sudan, M., Terzopoulos, D., Tygar, D., Vardi, M. Y., Weikum, G., Yang, S. J., Greenberg, A. M., and Endsley, M., editors, *Social Computing, Behavioral - Cultural Modeling and Prediction*, volume 7227, pages 231–238. Springer Berlin Heidelberg, Berlin, Heidelberg.

Ward, M. D., Metternich, N. W., Dorff, C. L., Gallop, M., Hollenbach, F. M., Schultz, A., and Weschle, S. (2013). Learning from the Past and Stepping into the Future: Toward a New Generation of Conflict Prediction. *International Studies Review*, 15(4):473–490.

Weber, L. and Bowling, B. (2014). *Stop and Search: Police Power in Global Context*. Routledge.

Weeraratne, S. (2017). Theorizing the Expansion of the Boko Haram Insurgency in Nigeria. *Terrorism and Political Violence*, 29(4):610–634.

Weiberg, L., Pedahzur, A., and Hirsch-Hoefler, S. (2004). The Challenges of Conceptualizing Terrorism. *Terrorism and Political Violence*, 16(4):777–794.

Weisburd, D. (2015). The Law of Crime Concentration and the Criminology of Place*: The Law of Crime Concentration. *Criminology*, 53(2):133–157.

Weisburd, D. and Amram, S. (2014). The law of concentrations of crime at place: the case of Tel Aviv-Jaffa. *Police Practice and Research*, 15(2):101–114.

Weisburd, D., Bernasco, W., and Bruinsma, G. J., editors (2009). *Putting Crime in its Place*. Springer New York, New York, NY.

Weisburd, D., Mastrofski, S. D., Mcnally, A. M., Greenspan, R., and Willis, J. J. (2003). Reforming To Preserve: Compstat and Strategic Problem Solving in American Policing*. *Criminology & Public Policy*, 2(3):421–456.

Weiss, J. C., Natarajan, S., Peissig, P. L., McCarty, C. A., and Page, D. (2012). Machine Learning for Personalized Medicine: Predicting Primary Myocardial Infarction from Electronic Health Records. *AI Magazine*, 33(4):33.

Wheeler, A. P., Worden, R. E., and McLean, S. J. (2016). Replicating Group-Based Trajectory Models of Crime at Micro-Places in Albany, NY. *Journal of Quantitative Criminology*, 32(4):589–612.

White, G. (2013). Discussion Of “Estimating The Historical And Future Probabilities Of Large Terrorist Events” By Aaron Clauset And Ryan Woodard. *The Annals of Applied Statistics*, 7(4):1876–1880.

White, G., Porter, M. D., and Mazerolle, L. (2013). Terrorism Risk, Resilience and Volatility: A Comparison of Terrorism Patterns in Three Southeast Asian Countries. *Journal of Quantitative Criminology*, 29(2):295–320.

Widrow, B. and Hoff, M. E. (1960). Adaptive switching circuits. Technical Report TR-1553-1, Stanford Univ Ca Stanford Electronics Labs.

Wilkinson, P. (1990). *Terrorist targets and tactics: new risks to world order*. Research Institute for the Study of Conflict and Terrorism.

Wöllmer, M., Kaiser, M., Eyben, F., Schuller, B., and Rigoll, G. (2013). LSTM-Modeling of continuous emotions in an audiovisual affect recognition framework. *Image and Vision Computing*, 31(2):153–163.

Wu, L. and Liu, H. (2018). Tracing Fake-News Footprints: Characterizing Social Media Messages by How They Propagate. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, WSDM '18, pages 637–645, New York, NY, USA. ACM. event-place: Marina Del Rey, CA, USA.

Xie, M., Jean, N., Burke, M., Lobell, D., and Ermon, S. (2016). Transfer Learning from Deep Features for Remote Sensing and Poverty Mapping. In *Thirtieth AAAI Conference on Artificial Intelligence*.

Xu, H. (2018). PoPPy: A Point Process Toolbox Based on PyTorch. *arXiv:1810.10122 [cs, stat]*. arXiv: 1810.10122.

Xu, H., Farajtabar, M., and Zha, H. (2016). Learning Granger Causality for Hawkes Processes. *arXiv:1602.04511 [cs, stat]*. arXiv: 1602.04511.

Xu, R. and Wunsch, D. (2005). Survey of Clustering Algorithms. *IEEE Transactions on Neural Networks*, 16(3):645–678.

Yeung, K. (2018). Algorithmic regulation: A critical interrogation. *Regulation & Governance*, 12(4):505–523.

Zehr, N. A. (2017). *The War Against Al-Qaeda: Religion, Policy, and Counter-Narratives*. Georgetown University Press.

Zeng, J., Ustun, B., and Rudin, C. (2017). Interpretable classification models for recidivism prediction. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 180(3):689–722.

Zhang, C. (2016). Modeling High Frequency Data Using Hawkes Processes with Power-law Kernels1. *Procedia Computer Science*, 80:762–771.

Zhang, M.-L. and Zhou, Z.-H. (2014). A Review on Multi-Label Learning Algorithms. *IEEE Transactions on Knowledge and Data Engineering*, 26(8):1819–1837.

This page intentionally left blank

Appendices

A Transition Networks of N-Dimension Super-States

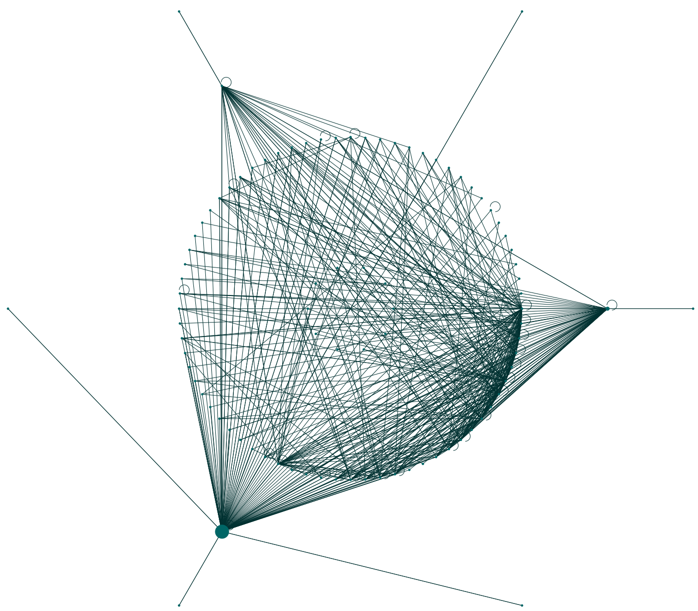

Figure A.1: Example Transition Graph - 1-Dimensional Super-States for Islamic State Targets Transitions (Nodes Sized by In-Degree Centrality)

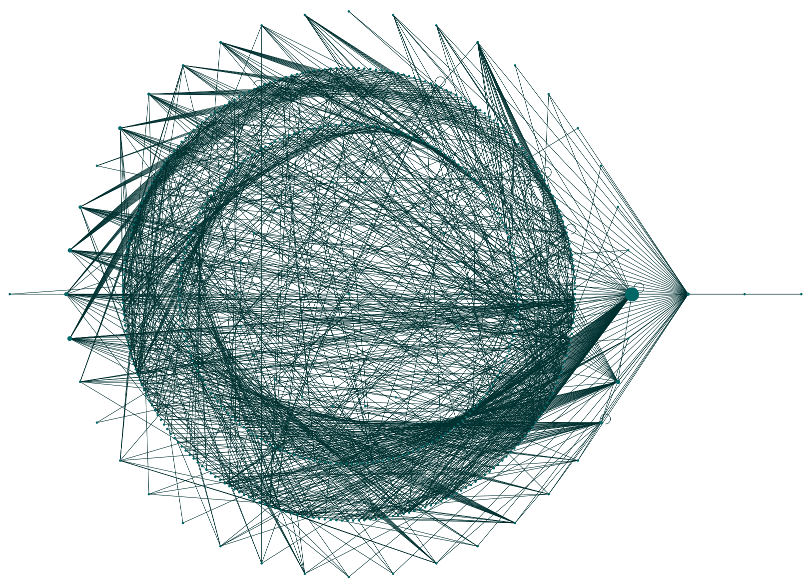

Figure A.2: Example Transition Graph - 2-Dimensional Super-States for Islamic State Targets Transitions (Nodes Sized by In-Degree Centrality)

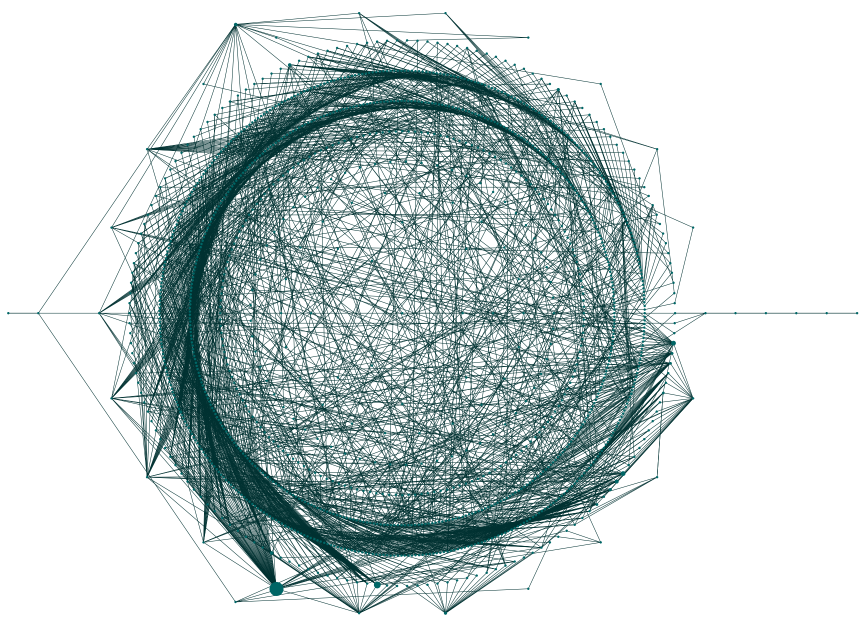

Figure A.3: Example Transition Graph - 3-Dimensional Super-States for Islamic State Targets Transitions (Nodes Sized by In-Degree Centrality)

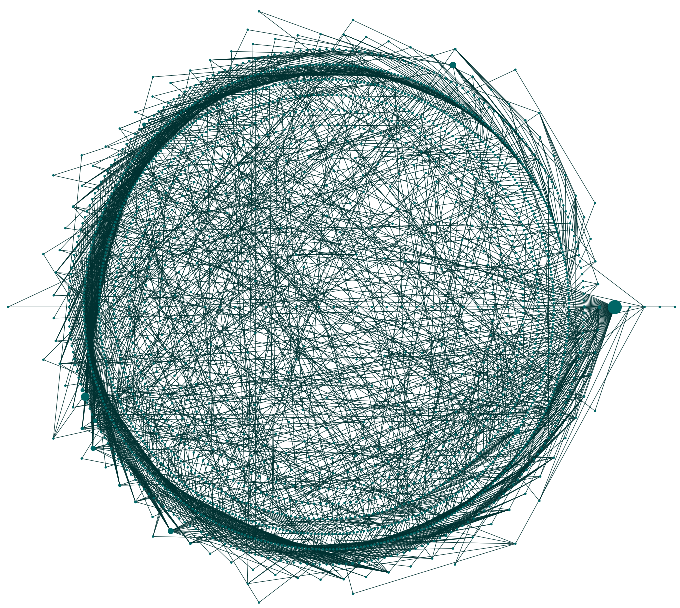

Figure A.4: Example Transition Graph - 4-Dimensional Super-States for Islamic State Targets Transitions (Nodes Sized by In-Degree Centrality)

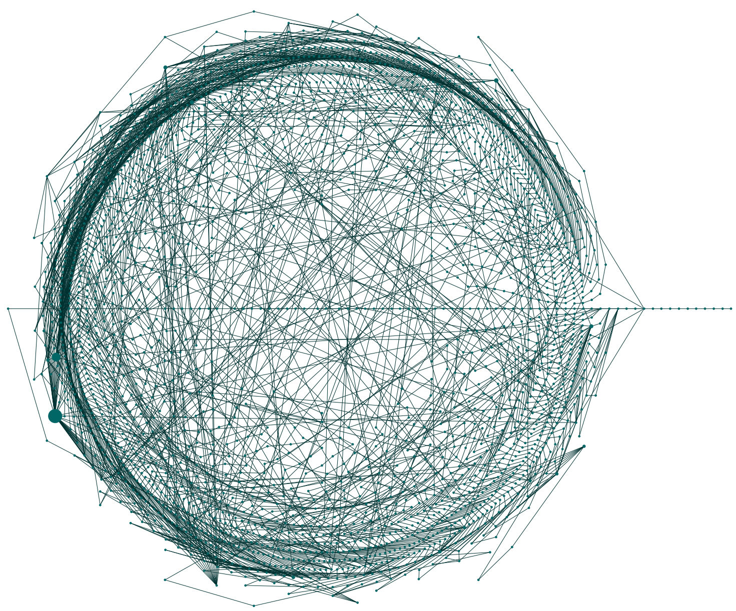

Figure A.5: Example Transition Graph - 5-Dimensional Super-States for Islamic State Targets Transitions (Nodes Sized by In-Degree Centrality)

B Additional Results of LSTM Models

Layer (Type)	Output Shape	N of Parameters
LSTM	(None, 62)	31,000
Dropout 1	(None, 62)	0
Dense 1	(None, 256)	16,128
Dropout 2	(None, 256)	0
Dense 2	(None, 128)	32,896
Dropout 2	(None, 128)	0
Dense 3	(None, 64)	8,256
Dense 4	(None, 62)	4,030
Total Parameters: 92,310		
Trainable Parameters: 92,310		
Non-trainable Parameters: 0		

Table B.1: Best Model Summary - Layers and Parameters (Islamic State)

Layer (Type)	Output Shape	N of Parameters
LSTM	(None, 62)	12,324
Dropout 1	(None, 62)	0
Dense 1	(None, 256)	10,240
Dropout 2	(None, 256)	0
Dense 2	(None, 128)	32,896
Dropout 2	(None, 128)	0
Dense 3	(None, 64)	8,256
Dense 4	(None, 62)	2,535
Total Parameters: 66,251		
Trainable Parameters: 66,251		
Non-trainable Parameters: 0		

Table B.2: Best Model Summary - Layers and Parameters (Taliban)

Layer (Type)	Output Shape	N of Parameters
LSTM	(None, 62)	42,924
Dropout 1	(None, 62)	0
Dense 1	(None, 256)	18,944
Dropout 2	(None, 256)	0
Dense 2	(None, 128)	32,896
Dropout 2	(None, 128)	0
Dense 3	(None, 64)	8,256
Dense 4	(None, 62)	4,745
Total Parameters: 107,765		
Trainable Parameters: 107,765		
Non-trainable Parameters: 0		

Table B.3: Best Model Summary - Layers and Parameters (Al Qaeda)

Layer (Type)	Output Shape	N of Parameters
LSTM	(None, 62)	14,280
Dropout 1	(None, 62)	0
Dense 1	(None, 256)	11,008
Dropout 2	(None, 256)	0
Dense 2	(None, 128)	32,896
Dropout 2	(None, 128)	0
Dense 3	(None, 64)	8,256
Dense 4	(None, 62)	2,730
Total Parameters: 69,170		
Trainable Parameters: 69,170		
Non-trainable Parameters: 0		

Table B.4: Best Model Summary - Layers and Parameters (Boko Haram)

Layer (Type)	Output Shape	N of Parameters
LSTM	(None, 45)	16,380
Dropout 1	(None, 45)	0
Dense 1	(None, 256)	11,776
Dropout 2	(None, 256)	0
Dense 2	(None, 128)	32,896
Dropout 2	(None, 128)	0
Dense 3	(None, 64)	8,256
Dense 4	(None, 45)	2,925
Total Parameters: 72,233		
Trainable Parameters: 72,233		
Non-trainable Parameters: 0		

Table B.5: Best Model Summary - Layers and Parameters (Al Shabaab)

Batch	**Look Back**	$\Gamma(T)$	$\Phi(T)$
32	30	0.750	1
32	10	0.663	1
2	30	0.636	1
64	30	0.636	1
100	1	0.626	1
2	20	0.611	1
32	20	0.611	1
64	20	0.611	1
100	20	0.611	1
2	10	0.602	1
64	10	0.602	1
100	10	0.602	1
2	40	0.600	1
32	40	0.600	1
100	40	0.600	1
2	3	0.598	1
32	3	0.598	1
64	3	0.598	1
100	3	0.598	1
2	2	0.592	1
32	2	0.592	1
64	2	0.592	1
100	2	0.592	1
2	1	0.585	1
32	1	0.585	1
64	1	0.585	1
64	40	0.400	1
100	30	0.386	1

Table B.6: Deep Neural Network Models - Results Ordered by Descending $\Gamma(T)$ (Islamic State)

Batch	**Look Back**	$\Gamma(T)$	$\Phi(T)$
32	10	0.758	0.973
64	10	0.758	0.973
100	10	0.758	0.973
2	20	0.757	0.977
64	20	0.757	0.977
100	20	0.757	0.977
2	3	0.756	0.968
32	3	0.756	0.968
64	3	0.756	0.968
100	3	0.756	0.968
100	2	0.755	0.974
32	20	0.755	0.971
2	10	0.754	0.967
2	50	0.753	0.972
32	50	0.753	0.972
64	50	0.753	0.972
100	50	0.753	0.972
2	1	0.753	0.969
2	30	0.752	0.975
32	30	0.752	0.975
64	30	0.752	0.975
100	30	0.752	0.975
32	2	0.751	0.963
64	2	0.751	0.963
32	1	0.751	0.964
64	1	0.751	0.964
100	1	0.749	0.953
2	2	0.748	0.963

Table B.7: Deep Neural Network Models - Results Ordered by Descending $\Gamma(T)$ (Taliban)

Batch	**Look Back**	$\Gamma(T)$	$\Phi(T)$
32	50	0.404	0.508
32	2	0.402	0.493
32	1	0.398	0.487
2	3	0.396	0.486
2	50	0.393	0.475
64	20	0.390	0.478
2	30	0.390	0.470
100	30	0.390	0.470
64	1	0.389	0.474
100	1	0.389	0.474
32	3	0.387	0.459
2	10	0.383	0.465
32	30	0.380	0.455
64	3	0.378	0.446
100	3	0.378	0.473
100	2	0.375	0.453
64	2	0.366	0.440
64	30	0.360	0.470
32	10	0.346	0.465
2	20	0.343	0.406
2	2	0.339	0.400
32	20	0.333	0.362
2	1	0.319	0.368
100	10	0.299	0.408
64	50	0.247	0.322
100	50	0.124	0.169
100	20	0.086	0.130
64	10	0.047	0.070

Table B.8: Deep Neural Network Models - Results Ordered by Descending $\Gamma(T)$ (Al Qaeda)

Batch	**Look Back**	$\Gamma(T)$	$\Phi(T)$
64	2	0.602	0.839
2	30	0.596	0.820
2	10	0.595	0.837
2	2	0.593	0.839
2	50	0.592	0.806
2	1	0.590	0.848
2	20	0.587	0.833
2	3	0.587	0.838
16	50	0.585	0.796
16	10	0.581	0.829
16	2	0.580	0.847
16	3	0.578	0.846
64	3	0.578	0.838
32	50	0.577	0.796
64	1	0.576	0.848
32	2	0.575	0.847
32	3	0.573	0.838
16	1	0.572	0.848
32	1	0.572	0.848
64	30	0.567	0.811
32	10	0.567	0.837
64	10	0.567	0.837
64	20	0.566	0.817
64	50	0.563	0.796
16	30	0.562	0.811
16	20	0.561	0.825
32	20	0.561	0.825
32	30	0.556	0.820

Table B.9: Deep Neural Network Models - Results Ordered by Descending $\Gamma(T)$ (Boko Haram)

Batch	**Look Back**	$\Gamma(T)$	$\Phi(T)$
2	10	0.617	0.859
64	10	0.617	0.870
2	3	0.608	0.848
32	3	0.608	0.848
2	50	0.608	0.875
32	50	0.608	0.875
64	50	0.608	0.875
100	50	0.608	0.875
32	10	0.607	0.859
100	30	0.605	0.880
2	2	0.603	0.840
32	2	0.603	0.840
64	2	0.603	0.840
100	2	0.603	0.840
100	1	0.600	0.832
2	30	0.599	0.880
32	30	0.599	0.880
64	30	0.599	0.880
2	20	0.596	0.845
32	20	0.596	0.845
64	20	0.596	0.845
100	20	0.596	0.845
2	1	0.591	0.832
32	1	0.591	0.832
64	1	0.591	0.832
100	3	0.542	0.848
100	10	0.500	0.859
64	3	0.348	0.768

Table B.10: Deep Neural Network Models - Results Ordered by Descending $\Gamma(T)$ (Al Shabaab)