Modeling and Synthetic Data Simulation on Desegregated Terrorist Profiles

Andreea Sistrunk,[1] Melody Walker,[2] Subhodip Biswas,[1] Vanessa Cedeno[1]
[1] Department of Computer Science, Virginia Tech
[2] Department of Mathematics, Virginia Tech
{sistrunk, melody, subhodip, vcedeno}@vt.edu

Abstract

Research of terrorist activities is dynamic, yet remains mainly qualitative since their rare incidence produces insufficient data to model patterns and predict their occurrence. Traditional synthetic data generation on population follows a normal distribution. However, rare events reside in the long tails of recorded human interaction, therefore generating synthetic data on a Gaussian distribution doesn't apply to this problem. A combined methodology of induction and deduction theory, known as Agent Based modeling (ABM) is known to produce good results for unknown distributions. We are seeing a methodology to produce data that can be used as input for traditional modeling, anomaly detection, and rare events extraction. Therefore goal is to preserve distribution characteristics of synthetic data in sync with recorded events. This solution could be applied to eider missing or sensitive data, nonexistent or unavailable for release, especially in the classified government setting. Our work presents an iterative process of a holistic societal approach to synthetic data generation. First we examine collected records of known terrorist activities and characterise mathematical ways to describe meaningful metrics in data. We define interactions between roles. The environment parameters are set on statistically recorded world data for education, religion, marital status, legislative/police power, population size of recorded past events. We create a set of simple rules to model interactions. The resulting synthetic data set shows similar mathematical measures to recorded data set, and can be used to augment collected events and in turn as input to subsequent quantitative research.

Introduction

In most countries, terrorism accounts on average of 0.01 % of the total deaths Figure 1. (Ritchie et al. 2013). Yet every attack causes serious negative psychological impact on society and constitutes significant disruption. For example, 9/11 cost the US over 100 billion dollars and terrorist attacks between 2004-2016 cost the European Union over 180 billion Euros(Corporation 2018). Its limited understanding comes partially from being a *rare events problem*. Yet, despite being a rare occurrence, it can produce sizable agonizing loss. Therefore recorded data are insufficient for quantitative studies, essential in the modeling, establishing reliable patterns, and generating analysis in the way to prevention and early conflict mitigation. As with other instances where we do not have sufficient collected data, scientists are used to augmenting with synthetic data, following main parameters and features.

To produce synthetic data in the absence of an established distribution, scientists most usually use the Gaussian distribution. However, the Gaussian distribution would not produce long tails. Rare events reside in the long tails of the distribution, within three σ standard deviations, with approximately 0.15% or less(Galarnyk 2019). The only known distribution is unable to capture the inflections of rare events, *traditional methodology for synthetic data generation does not apply modelling this problem.*

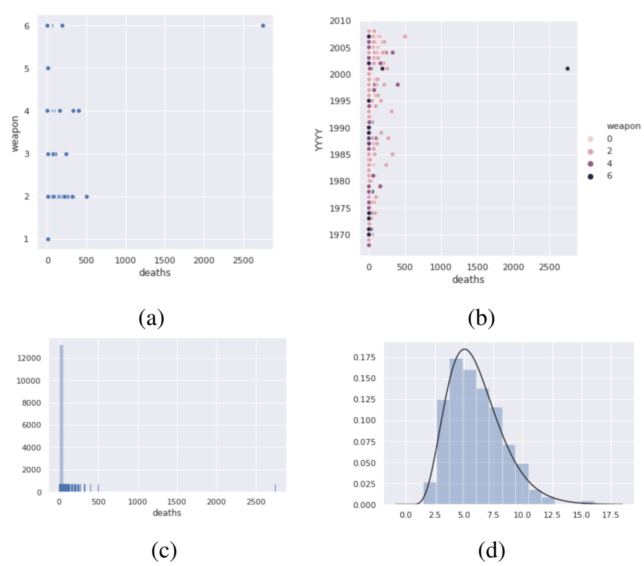

Figure 1: Visualization of the RAND-MIPT data (a) distribution of deaths (total 13274 records) in relation to weapon-type involved in the attack, (b) distribution of deaths between 1968-2007; on x-axis are the number of deaths per attack, on the y-axis is the years it took place. The legend shows the a weapon type used. Only one event in 2011 had more than 2500 deaths. (c) x-axis recorded deaths, y-axis recorded incidents. Most of the incidents recorded have no (zero) casualties (deaths), one recorded event has over 2000 deaths. (d) kernel density - gamma distribution fit for death tolls.

Producing synthetic data that exhibits similar threats as real data would be useful to augment collected events. Subsequently it could provide an **additional avenue to**

study sensitive information, otherwise unavailable due to scarcity or classification. Being able to produce synthetic data that follows the characteristics of recorded data would allow the ability to be further analyze with modern machine learning techniques, unavailable to this point.

Background and Motivation.

Many studies have been performed for the prevention of crime. However recorded information on crime is highly variable depending on type of crime and location. In addition to this constraint, grave or organized crime is a sporadic occurrence, with sparse geographical focus. One initial question we asked in this research was if there is the possibility of transfer knowledge between other crimes and organized crime for the scope of modeling and prediction. *Crime vs Organized Crime.* At surface all crime, quantified based on fatalities and property loss, comes across with similar consequences. There is research (Mullins 2009) showing many similarities as well as individuals convicted of other crimes in addition to terrorism, yet there are small fundamental distinguishing factors that make the more restrictive samples unsuitable for transfer knowledge. One of the most decisive aspects analyzed by the authors of our current research was the *crime motive*. When we take a closer look at what drives terrorism in parallel with crimes of similar consequences, we find a strong divide. In most types of crimes, the criminal seeks a personal benefit of some sort, however terrorists often times are serving an ideology and consider their acts as personal altruist sacrifice (Reeve 2020). In our initial research we sought a transfer knowledge methodology between areas of crime, yet due to this main defining reason as well as other subsequent ones, the lessons we learned in predictive policing about other types of crimes do not transfer to the study of terrorism.

Available Organized Crime Data. Our team performed research on several public databases that contain international events and we will refer to in this manuscript. We have investigated the GDELT Project data base that contains news data gathered daily since 1979; The University of Maryland START database for the Study of terrorism and response to terrorism; the National Security Research Division (RAND) database of worldwide terrorism incidents and particularly the RAND-MIPT terrorism incident database; the Center for International Security and Corporation (CISAC) from Stanford for mapping militants; and the World Bank database for statistics and to extract our relative distributions. Data is showing that the incidence of a serious event in society can vary from 0.01% in most countries, and go as high as 7% in the terrorism epicenters of the planet (based on World Data statistics corroborated with RANDT-MIPT terrorist data). **Figure 1** is a visual illustration of RAND-MIPT database of 13274 deadly events gathered from 1968-2007 (Clauset 2015). Based on this data, which is in line with other databases, the occurrence of an event such as 9/11 with 2,977 deaths and 6,000 injured it is truly a rare occurrence. We are seeing in **Figure 1 (c)** that most terrorist attacks do not have any casualty. The data shows over 12000 samples with zero deaths. Still, despite its rare incidence, over 50% of the people surveyed in the US alone, live in fear of being a victim of a terrorist activity (Ritchie et al. 2013).

Challenges in the Study of Terrorism. With scarce and geographically segregated data (Ritchie et al. 2013) we find most research in this area focusing on qualitative aspects of the terrorist activity. *Transfer Knowledge:* Several quantitative works conducted by universities and science foundations draw similarities, maps events, and correlations to other areas of science networks we have more knowledge of (Bagozzi et al. 2019), (**?**), (Chang, Chen, and Fomby 2017). Publications in the area of direct modeling and pattern recognition, that study terrorist networks, have focused on the dynamics of the structure and how well they respond to an adverse reaction (Li et al. 2015). The ABM is a known technique (Grimm et al. 2006) (Macy and Willer 2002), able to mimicking the interaction within the terrorist groups using simple rules, recognizing their non-traditional organizational pattern. Similarly (Tsvetovat and Carley) uses an ABM model due to observations scarcity and lack of insight. The authors look at the overall terrorist groups, their behavior and interactions with little differentiation between different agent roles. Other research turns to simulation to evaluate strategic options(**?**), model rare events behavior (Gilbert 2019) and gain some insight of how that system is going to behave within the variance of input parameters.

Complex systems can be difficult to not only predict, but even describe. In-depth work (Ilachinski 2012) focused on modeling the entire network of terrorist as self-adaptive societies, to help Intel analysts with a larger context. This in turn created an adaptive system, into a tool continuously feeding variety of sources. While it has very extensive formulations its available and reputability is highly restrictive. However the work in this domain shades light on a variety of circumstances even if it doesn't provide us a perfect prediction of how that system is going to behave (Brown et al. 2005). In consequence, terrorist behavior is a constant high visibility matter with no easy answers. In addition to *not having sufficient data* to study these events, we do *not have readily available mathematical models to produce synthetic data*, making **threat detection a hard, but vital problem**. The ability to use a set of simple rules to create synthetic data could in turn create a highly useful solution both for data augmentation and unclassified analysis of restrictive data sets.

Table 1: Hypothesis of Terrorist Traits

Role	Education	Marriage	Wealth	Religion Training
Perpetrator	low	single	poor	neutral
Leader	high	married	neutral	trained while young
Financier	neutral	married	wealthy	trained older

Our Contribution. In this paper we present *empirical evidence of a potential solution to producing synthetic data to synthetically model an unknown distribution for rare terrorist events*. We do this by creating synthetic societies that follow the profiles of different countries of the world. In this synthetic environment we create agents and rules. We allow time for the society to evolve and mature. Subsequently we compare features of synthetic distribution against reported terrorist activity within modeled geographic boundaries (a given country).

Subsequent *novelty of our work is in modeling terrorist interaction based on desegregated terrorist profiles*. Our model proposes a holistic societal interaction, as opposed to modeling only terrorist agent behavior. We recognize terrorist activity as only one small part of the whole societal interaction. In consequence we model the entire population within a geographical region.

Due to lack of data, the study of terrorist events using ABM simulation was investigated before. In this work we are combining the modeling with political science research to offer an interdisciplinary insight in the problem. The political research shades light on the importance of disaggregating terrorist profiles. We model this desegregation in a synthetic environment using mathematical distributions and computer simulations to create individual agents that interact in a resulting complex environment(Macal and North 2005). To support modeling, data is extracted from different geographic regions, with distinctive profiles.

The *terrorist profile desegregation* distinguishes various types of roles agents assume. To carry out a simulation of a modelled society in which terrorist activity exists, we follow research based evidence proven reliable in qualitative studies from social and political sciences(Perliger, Koehler-Derrick, and Pedahzur 2016) . Therefore, we desegregated the terrorist label based on their role within the organization such as **perpetrator**: the actor that carries out an terrorist attack, **group leader**: the mastermind behind the attack, and **financier**: has the ability of supplying funds so terrorist acts do not lack resources.

To model a holistic *society*, in addition to agents tasked with generating an attack we added other agents with different roles in preventing/stopping, or by-standing activities related to terrorism. Subsequently use simple rules to simulate terrorism recruiting, leading, and carrying the attack. These activities take place in a real society in addition to other agents' activities that have no connection with terrorism such as police and civilians. !

Models and algorithm description

Conceptual Profiles. Due to the various activities required to carry out a terrorist attack, terrorists need to perform different roles. Research in political science (Perliger, Koehler-Derrick, and Pedahzur 2016) depicts the gap between participation and violence in terrorist activities. The authors conclude a possible correlation between socio-economical status, marital status, religious education, and opportunity as defining factors for someone to turn to terrorist activities as well the role they would play within a terrorist organization. (Perliger, Koehler-Derrick, and Pedahzur 2016) connects the characteristic of terrorists to the role they might play. They note that wealth, education, marital status are key indicators for terrorists to become a perpetrator, leader, or financier. Concluding recorded events, they note that an individual could over time fill more than one of these roles. For clarity we attempt to distinguish the characteristics of these individuals in our simulation separately.

The perpetrator is key in carrying out the attack and in most cases their life is just one of the assumed sacrifices their role requires. These people are likely low educated, unemployed, are not married, local to the area(Perliger, Koehler-Derrick, and Pedahzur 2016). Their motivation could be terrorism as altruism (Reeve 2020) or for financial reasons to support their family (Krueger and Malečková 2003). For these reasons we will refer to the *perpetrator* as the individual who attempts to carry out the terrorist attack. We will create a profile for this actor with parameters as low income, low-medium education, not married, young, highly religious.

The second role would be of the "master mind" of the attack (Russell and Miller 1977). Normally these people carry a special persona and dedicate their lives to the mission. They are regarded as spiritual leaders. In Islamic terrorist groups, there is evidence that leaders are likely to be married, employed, and found religion later in life (Perliger, Koehler-Derrick, and Pedahzur 2016). We simulated agents for our models that we refer to as *leaders* their profile parameters are low income, medium-high education, married, middle aged, highly religious.

As the leader plans the attack the *financier* provides money or goods towards the attack. An individual who finances terrorist activities is more likely to be married, have religious training, and was introduced to religion at a younger age (Perliger, Koehler-Derrick, and Pedahzur 2016). Another distinct role we need for a successful attack is financing the operations. These actors in our modeling would have a medium -high income for the environment they are financing or residing in, they would have medium-high education, are mostly married, of various ages, and at times might display medium-high religious beliefs. We depict these characteristics in Table 1. to synthesise the connection between a certain terrorist profile and their role. In here the reader notices that we are setting agents' characteristics based on the known distributions. If however we are unsure, do not have supporting evidence, or there there are strong conflicting or non-converging opinion in the qualitative studies for a particular characteristic, this feature in the agent profile was not populated. We did not have conclusive data for the "Financier"'s Education-Status, the "Leader"'s Wealth-Level or the "Perpetrator"'s Religious-Training

Based on a hypothetical collection of threats we desegregate the terrorist profile in agents under three profiles (Section 5). We make theoretical assumptions of individual terrorist, which reproduce empirical data on terrorist attacks in qualitative analysis (Perliger, Koehler-Derrick, and Pedahzur 2016). As described above, individual terrorist are separated into three different roles (perpetrator, leader and financier). In addition to these agents that carry the terrorist acts, we will add police enforcement agents and general civilian agents.

The *police* agents have the role to prevent and stop the conflict, while the *civilians* can be recruited to become terrorist or denounce a plot they came in contact to.

At any point in time the simulation contains agents under five profiles *perpetrators, leaders, financiers, civilians*, and *police/police*.

Profile Implementation. Each agent has a serious of features(or parameters). The corroborated value of these fea-

Table 2: Table of Definitions

Parameter	**Description**
n	Total number of the agents in a simulation at **state *i***.
α	Agents. Total number of agents at **state *i*** present in in the synthetic environment: $\alpha_1^i, \alpha_2^i, \alpha_3^i, ...\alpha_i^i, ...\alpha_n^i$.
τ	Feature. Each agent has 9 features. The tensor of all features characterize an agent at a certain state. For an agent α at state *i* we write tensor: $\alpha^i = \{\tau_1, \tau_2, \tau_3, \tau_4, \tau_5, \tau_6^i, \tau_7^i, \tau_8^i, \tau_9^i\}$
m	Total number of states. An agent α in final state will look like $\alpha^m = \{\tau_1, \tau_2, \tau_3, \tau_4, \tau_5, \tau_6^m, \tau_7^m, \tau_8^m, \tau_9^m\}$
w	Agent feature values τ generated based on the characteristics of the modelled geographical region
μ	Continuous probability distribution/normal distribution mean
s	Scale parameter proportional with the standard deviation

tures (think of a tensor) determines **the state of the agent at a certain time**. Some of these features such as:*(i) education status*, *(ii) marital status*, *(iii) wealth level*, *(iv) religious training*, *(v) exposure to crime*(Table 3) remain constant over time. The set values of these features is created as a normal distribution that follows the characteristics of an actual society (i.e. countries, cities, states, a.s.o.) present within certain geographical boundaries.

In addition to these parameters that describe the incipient state of an agent and remain constant throughout the simulation, additional parameters are used to help them adapt and function in the environment. The value of these parameters change throughout the evolution of the model and they are: *(vi) crimes committed*, *(vii) predisposition towards police*, *(viii) predisposition towards terrorism*, *(ix) power level*(Table 3). All agents in the environment have these nine features. Based on the corroboration of their values agents can be assigned to a certain group described above(terrorist(financier, perpetrator, leader), civilian, or police). Their profile(denoted by a tensor of rank 9, Table 3) in the synthetic environment dictates the role they play. The agents are adaptive, therefore the value of the last four characteristics change over time. Subsequently one particular agent can play different roles in the course of a simulation. We refer to this as *different states of an agent*. These traits were chosen based on literature (Perliger, Koehler-Derrick, and Pedahzur 2016), as well as for functionality of the model.

Table 3: Hypothesis of Agent Traits

$AgentTraits$	$Civilian$	$police$	$Terrorist$
(i)*Education Status*	constant	constant	constant
(ii)*Marital Status*	constant	constant	constant
(iii)*Wealth Level*	constant	constant	constant
(iv)*Religious Training*	constant	constant	constant
(v)Exposure to Crime	constant	constant	constant
(vi)Crimes Committed	*variable*	*variable*	*variable*
(vii)Predisposition towards Police	*variable*	*variable*	*variable*
(viii)Predisposition towards Terrorism	*variable*	*variable*	*variable*
(ix)Power	*variable*	*variable*	*variable*

Notation. We initialize all the agents (α) with $\alpha_1, \alpha_2, \alpha_3....\alpha_n$ to represent agents in synthetic environment, represented as all the nodes in a graph. Therefore at any point in time *t* any agent α_i will have the sate α_i^t. For every state we will define the agent α as a tensor and each of its features τ, with values from $\tau_1, \tau_2, \tau_3, \tau_4, \tau_5, \tau_6, \tau_7, \tau_8, \tau_9$ for the agent features (i)-(ix) descried above.

In the synthetic environment certain traits remain constant while others change. We present the hypothesised fixed and variable traits in the synthetic environment in Table 2. The constant values that describe an agent do not change as a result of agent interaction of behaviors within this particular synthetic society. In consequence they were initialized with *high, low, or neutral* at initialization bare no change throughout modeling.

In addition to their role, the agents have different **states** as they evolve through the synthetic environment. The *state* of an agent is set at the inception of the synthetic environment and can change throughout the simulation. We will refer to *the state of an agent* with the superscript *1, 2, ..., n* and refer to it generically as *state t* (from "time" or "ticks"). The variable time as perceived by humans, is traditionally emulated in the agent based modeling environments in *ticks*. Each agent(represented as node) has a certain state t at every tick: for tick 1 will have state 1, tick 2 state 2, a.s.o. The value of the agent characteristics described in Table 2 might remain constant or change, depending on the interaction with other agents. If an interaction took place, an edge will be added to the graph, as the representation of an ideological connection between two agents(represented by nodes in the graph).

Agent Initialization: We can think of this simulation as an un-directed graph, where the nodes are the agents and the edges are the relationships they have. To initialize the society, we first create the agents (nodes). Each node represents an agent with its features, stored in a tensor as described above. Subsequently a random number of connections are created, to ensure a minimum of two connections for each agent(represented by a node in the simulation). This would translate in reality to affinity towards people in society (it could be family, or ideology based affinities). Subsequently we induce a percentage(¡10%) of small random highly connected clusters rather than a complete random graph, to simulate scattered groups of loyalty. Therefore at **any state *i*** the values of the features of an agent(represented by a node in our simulation) is a rank 9 tensor with values described by τ_i features (Table 3.) As the world advances (in ticks) the values of *crime-committed*, *predisposition toward police*, *predisposition towards terrorism* and *power* of the agents (nodes) can change their state following simple rules (Algorithm 2). We assume *normal distribution* when populating the initial state of the agents. We do *not* assume any *feature correlation* between the nine features used as agent traits. In consequence at the time of inception, our synthetic modelled world will have a normal distribution (Figure 4(a)), with no feature correlation for the agents. Consequently at any time, an agent will play a certain role as a perpetrator, leader, or financier. *To reflect this variation the tensor places 6 through 9 have a superscript indicating that their is set to potentially change as result of an interaction.* In Table 3 we describe the tensor values that remain constant or can change in the evolution of the environment.

If we are to describe all the states of an agent α_1 from initial iteration i=1 to final iteration i=m, with 1 being the initial state and m being the final state, we will write the following formulation:

$$\alpha_1^1 = \{\tau_1, \tau_2, \tau_3, \tau_4, \tau_5, \tau_6^1, \tau_7^1, \tau_8^1, \tau_9^1\}$$
$$\alpha_1^2 = \{\tau_1, \tau_2, \tau_3, \tau_4, \tau_5, \tau_6^2, \tau_7^2, \tau_8^2, \tau_9^2\}$$
$$.....$$
$$\alpha_1^m = \{\tau_1, \tau_2, \tau_3, \tau_4, \tau_5, \tau_6^m, \tau_7^m, \tau_8^m, \tau_9^m\}$$

Algorithm 1: Constructing the synthetic environment

Input : Agent feature decomposition (see section 3.2): $\alpha^i = \{\tau_1, \tau_2, \tau_3, \tau_4, \tau_5, \tau_6^i, \tau_7^i, \tau_8^i, \tau_9^i\}$. Establish geographical area indicators (see section 3.3); World size; Number of runs.
Output : Synthetic data
Method:
Step 1. Calculate probability density function based on the reported parameters in considered geographical area.
Weight := {education status, marital status, religion training, relative wealth, exposure to crime}.
Step 2. Call in function to populate the environment with variable trait values: exposure to crime, crimes committed, predisposition towards police, predisposition towards terrorism, power level.
Step 3. Environment setup: generate and balance agent features based on the established distributions for the overall synthetic environment and agent level per feature
Step 4. Ensure ideological links 1:1 and few highly connected groups
Step 5. Write update state after every tick.

Environment setup: We constructed an uniform environment as a continuous, unbounded space. We instruct an infinite plan topology consistent with ideological loyalty independent of space.

Interactions: In this environment interactions are limited to Agent-to-Agent interaction. They will be modeled in the *states i* of the agents. The connections between agents are not to be interpreted as physical proximity, but devotion, affinity, and doctrine. These ideological connections are modelled by randomly initiated edges and clusters of *loyally*.

The interaction outcome is modeled using a logistic distribution (1). In addition to creating the agents, we model interaction rules for agents to interconnect. We assume the

Algorithm 2: Adaptive Agents Interaction

Input : Agent feature decomposition (see section 3.2): $\alpha^i = \{\tau_1, \tau_2, \tau_3, \tau_4, \tau_5, \tau_6^i, \tau_7^i, \tau_8^i, \tau_9^i\}$; Establish geographical area indicators (see section 3.3); World size; Number of runs.
Output : Outcome of an interaction
Method:

```
if agent has a neutral predisposition then
    (a) collides w/terrorist, increases terrorist predisposition
    (b) collides w/ police, increases police predisposition
if agent meets threshold for predisposition towards police then
    (a) collides w/ terrorist:
    Stage 1: no action takes place. Subsequent interactions
      result in removal(arrest)
    (b) collides w/ police, both police increase power
    (c) collides w/ civilian, police predisposition increases
if agent meets threshold predisposition towards terrorism then
    if leader then
        collides w/ terrorist, leader power increases
        if power > threshold then
            attack ⟹ crimes committed increases
        collides w/ police, power decreases and
        if power < threshold then
            ⟹ leader removed
        collides w/ civilian, terrorist predisposition increases
          and
        if predisposition > threshold then
            ⟹ recruit
    if agent meets threshold financier then
        (a) collides w/terrorist, both increase power
        (b) collides w/ police, power decreases and
        if power drops under a set threshold then
            leader removed
        collides w/ civilian, terrorist predisposition increases
    else
        collides w/terrorist ⟶ terrorist power increases and
        if power > threshold then
            attack ⟹ committed crimes increases
        collides w/ police : power decreases and
        if power < threshold then
            terrorist removed
        collides w/ civilian, terrorist predisposition increases
```

distribution of attributes for each role and compare recorded events of terrorism after each iteration of the process.

Since we are going to model this society as a graph, the agents are going to be represented by nodes and affinity by un-oriented edges. For exemplification we would write the states of all agents at a random state *t*. Since here we conceptualize the entire environment, all nodes, their states, we will use additional subscripts to capture the uniqueness of a certain characteristic of a node at the global environment level. For exemplification agent α_1^t(agent one at time t) will have its own Predisposition towards Terrorism τ_{18}^t reflected at the global environment level as τ with a subscript $_1$ meaning it belongs to agent one followed by subscript $_8$ meaning it is the 8th value of the tensor, and superscript ^{t} showing it is captured at tick/time *t*:

$$\alpha_1^t = \{\tau_{11}, \tau_{12}, \tau_{13}, \tau_{14}, \tau_{15}, \tau_{16}^t, \tau_{17}^t, \tau_{18}^t, \tau_{19}^t\}$$
$$\alpha_2^t = \{\tau_{21}, \tau_{22}, \tau_{23}, \tau_{24}, \tau_{25}, \tau_{26}^t, \tau_{27}^t, \tau_{28}^t, \tau_{29}^t\}$$

$$\alpha_3^t = \{\tau_{31}, \tau_{32}, \tau_{33}, \tau_{34}, \tau_{35}, \tau_{36}^t, \tau_{37}^t, \tau_{38}^t, \tau_{39}^t\}$$
$$\alpha_4^t = \{\tau_{41}, \tau_{42}, \tau_{43}, \tau_{44}, \tau_{45}, \tau_{46}^t, \tau_{47}^t, \tau_{48}^t, \tau_{49}^t\}$$
$$\alpha_5^t = \{\tau_{51}, \tau_{52}, \tau_{53}, \tau_{54}, \tau_{55}, \tau_{56}^t, \tau_{57}^t, \tau_{58}^t, \tau_{59}^t\}$$
$$...$$
$$\alpha_i^t = \{\tau_{i1}, \tau_{i2}, \tau_{i3}, \tau_{i4}, \tau_{i5}, \tau_{i6}^t, \tau_{i7}^t, \tau_{i8}^t, \tau_{i9}^t\}$$
$$...$$
$$\alpha_n^t = \{\tau_{n1}, \tau_{n2}, \tau_{n3}, \tau_{n}, \tau_{n5}, \tau_{n6}^t, \tau_{n7}^t, \tau_{n8}^t, \tau_{n9}^t\}$$

To model the probability of a loyalty connection we assume a logistic distribution depending on the profile polarization of the agents and mean of the modelled distribution. We are using a cumulative distribution function to model the probability of success for each interaction as shown below.

$$F(x) = \frac{1}{1 + \exp\left(\frac{-(\mathrm{w}*\tau)-\mu)}{s}\right)} \tag{1}$$

Where we have: specific geographic rates(w), agent traits(τ), mean distribution(μ), and scale of the logistic distribution(s). The values for (w) and (s) are chosen based on the geographical recorded data for a certain region, where (μ) and (τ) are dynamically calculated for with each interaction, during the synthetic environment.

Modeling and Simulation

For the initial setup we used we consider the following recorded distributions of real data (Department 2006)(World Bank 2018).

- percent of married population aged 15 years and older
- population wealth index based on Poisson distribution for scaled relative wealth to geo-location
- religion distribution
- education level(decimal point)(e_1 no education, e_2 primary, e_3 secondary, e_4 tertiary education) where:

$$e_1 + e_2 + e_3 + e_4 = 1 \tag{2}$$

- crime - exposure and density for considered area.

For calibration we used Maryland terrorism database(UMD) (for the Study of Terrorism and to Terrorism (START). 2018). The limitations of this data-set were in recording terrorist activities as following: in order for an incident to be in the database, *the action needs to be intentional with violence or threat of violence with non-national actors involved*. It also must have two of the following: be outside of war activity, have an intended message for others beyond the victims, or had a declared a social, religious, political, or economic goal. The data used was from 2012 to 2017 as the method of collection prior to 2012 was significantly different.

To initialize the model we followed demographic indexes on Libya, France, and Pakistan. We chose these countries because we thought, corroborating all criteria we were interested created different profiles. In consequence if the country population 15 years or older are 63% married, we weighted the population distribution for *Marital Status* feature as 0.65. Educational data was found in the World Bank/Barro Lee data-set (World Bank 2018). Marital status was found through different sources based on the country (Moore ; Department 2006). Religious orientation was modeled using Pew Research Center data(Center 2010). Recorded crimes were pulled from UMD database(for the Study of Terrorism and to Terrorism (START). 2018).

Parameters, Interactions, and Outcome

Parameters are set on hypothesis for individual terrorist profile desegregation corroborated with mathematical formulation using input values gathered from recorded data. At every point in time the profiles of each agent were recorded as values in a matrix. Based on the value of *predisposition*(towards Police or Terrorism), the agent in the modelling could be identified as terrorist, civilian, or police(Figure 2(a),(b),(c)), where we notice (on x-axis) the predisposition towards terrorist or police as a value that evolves over the course of synthetic environment. Subsequently, the terrorists could be clustered in financier, perpetrator, or leader based on a combination of values described in Table 1. In consequence, the synthetic environment was configured and it is briefly described in Algorithm 1. *Interaction outcome* between two nodes(agents) with a high predisposition towards terrorism could constitute the bases of an attack or recruitment into a different role. The probability of success of such an action, will be modeled with a logistic distribution. If the probability from the logistic distribution is greater for a certain role, the individual is recruited to this role. Whenever a terrorist or police meet their own kind, they gain more power. However when a police individual meets a terrorist, nothing happens until they exceed the power threshold to enable a change in agents they are connected with. Once a leader has at least three underlings and has accumulated enough power, the leader can begin to plan an attack, if a connection with a financier profile of enough power is present. At this point the premises for an attack are present. Now if the police personal interacts with a terrorist, then the terrorist can be arrested. The the success of the attack comes from a logistic distribution based on the power, number of terrorist, and right probabilistic context from the other agents they are connected with. If for any reason an agent is removed from the environment, its place is replaced with a new agent with profile values matching the initial environment setup. A set of simple rules for outcome of an interaction is presented in Algorithm 2.

Recorded statistics of the geographical country are embedded into the model. For each simulation, the model runs through cycles of 1000, 2000, 3000, and 4000 iterations of interactions (ticks). Because we found a strong divide between the profiles after 4000 iterations, we did not run the environment further. To mimic ideological affinity random small highly connected clusters are formed when the society is initialized. Each iteration 10% of the population is chosen randomly to interact.

We extracted the data and consider the medians of these results. We assume that terrorist attacks are rare events, so we look for those which have a median of zero compared to those with higher medians. Those with higher medians are having a large number of attacks and people killed com-

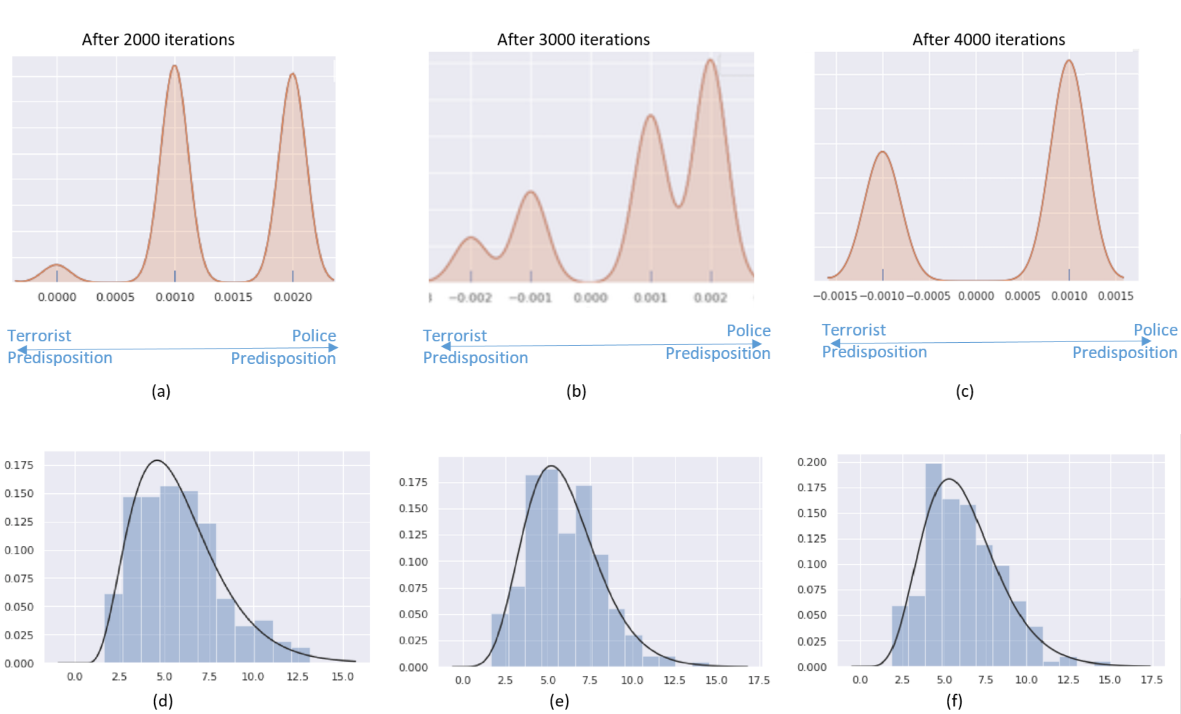

Figure 2: Synthetic environment: (a) society is highly positive, x-axis shows the polarity of the population towards terrorist or law enforcement support. We notice a high positive polarization towards the police and some neutrality (b) noticing some very negative attitudes (c) the society has a high divide towards police or terrorism (d) consecutively the attacks start to show a long tail towards the right (e) accentuated long tail to the right, with high variation in values (f)long tail rare events and variance, comparable to recorded data Figure 1(d).

pared to what the data suggests. We as well tried various methods of sampling when allowing the nodes to interact. We have noticed that the Latin-Hypercube sampling yield a better spread than our initial sampling. Figure 3 (a) (b) (c) are showing visualization of the data recorded for the *predisposition towards terrorism* as opposed to *predisposition towards police*. Predisposition towards terrorism would account for negative values on the x-axis, and positive values on this axis would show a favorable attitude, *predisposition towards police* Figure 2 (a) shows the predisposition after 2000 runs the population is highly in favor of the police, and there is some neutrality. In subsequent run, in Figure 3 (b) we show the predisposition distribution after 3000 runs. Here we are noticing that there is significantly higher *predisposition towards terrorism*. The population evolves very fast in the next runs into a strong divide of agent polarization Figure 2 (c). For this runs we are showing the types of actions that take place in the synthetic environment. At the beginning the environment is initialized with a normal distribution of profiles. Figure 3(d), (e), (f) show the model evolution over time. We can notice a right hand tail gets longer and longer, which models the terrorist attacks. Over time the events remain very rare in the synthetic environment.

Discussion

While this remains exploratory work at the intersection of political science and modeling, it takes the theoretical research of rare terrorist events, one step further in implementation of these concepts. The desegregation of terrorist profiles based on their role and motive, corroborated with the lack of proximity boundaries showed initial evidence of ability to simulate synthetic data pertinent to rare events, in the ensemble of a global society. Further investigations are needed into modeling interactions based on the socio-geographic standpoint. Despite modeling different countries, with weights to describe various geographical environment, our overall societal behavior did not change dramatically (therefore we did not show the different distributions), we advise towards stronger weights should be assigned to model diverse socio-geographic environments.

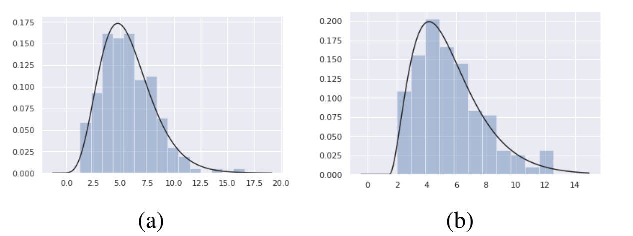

Figure 3: Synthetic environment (a) at 4100 ticks, (b) at 4200 ticks

The model is highly sensitive to variation based on the number of runs. Therefore we note that a more rigorous approach(Asmussen et al. 2000) is needed towards measuring the divergence in synthetic data outcome and hyper parameter tuning. To validate the results under different assumptions we ran the same environments using various modalities of sampling. We notice a slight predisposition preference for the weights in our initial hypothesis and chanced to employ a Latin Hyper-cube random sample. We compared the mean and variance distribution of data extracted from the synthetic distribution with the data from real events. For deadly events in the RAND-MIPT data-set we record low means(4.2) and very high variance (727). This matches our synthetic environment mean and variance just before 4000 ticks(mean 3.1, variance 748), when the population becomes strongly divided, and subsequently Figure 3(a) and (b) the terrorist attacks increase, raising the mean of deaths much higher than recorded distributions.

Conclusion

In comparison to other works in the area, that might have touched upon the subject and methodology, this light weight formulation using basic rules and profiles, matches the outcome of recorded distributions and validates the initial hypothesis. The density estimation, variance, and mean of rare modeled events (Figure 2 (e)) is in sync with the indexes recorded events from RAND-MIPT data (Figure 1(d)). We can in-turn, extract the percentage indicated by Poisson distribution for rare events (Clauset 2015) and use to augment data for traditional modeling. A future direction is identifying hyper-parameter tuning, identification of how variation plays a role in the outcome, as well as extraction of data to be fed to a machine learning algorithm and measure outcomes.

We thank Dr. Jerry Ballard and Dr. Brian Jalaian for assistance with presenting the methodology, Ms. Nicole Wayant, Megan Maloney, and Vijay, all from US Army for comments. Their assistance greatly improved the quality of the manuscript.

References

Asmussen, S.; Binswanger, K.; Højgaard, B.; et al. 2000. Rare events simulation for heavy-tailed distributions. *Bernoulli* 6(2):303–322.

Bagozzi, B. E.; Joo, M. M.; Kim, B.; and Mukherjee, B. 2019. A bayesian split population survival model. *Political analysis*.

Brown, D. G.; Riolo, R.; Robinson, D. T.; North, M.; and Rand, W. 2005. Spatial process and data models: Toward integration of agent-based models and gis. *Journal of Geographical Systems* 7(1):25–47.

Center, P. R. 2010. Table: Muslim population by country.

Chang, K.; Chen, R.; and Fomby, T. B. 2017. Prediction-based adaptive compositional model for seasonal time series analysis: Prediction-based adaptive compositional model for seasonal time series analysis. *Journal of Forecasting* 36(7).

Clauset, A. 2015. *Estimating the probability of rare events*.

Corporation, R. 2018. *Cost of Terrorism*.

Department, U. N. S. 2006. Population by marital status, age, sex, and urban/rural residence.

for the Study of Terrorism, N. C., and to Terrorism (START)., R. 2018. Global terrorism database [data file].

Galarnyk, M. 2019. Explaining the 68-95-99.7 rule for a normal distribution.

Gilbert, N. 2019. *Agent-based models*, volume 153. Sage Publications, Incorporated.

Grimm, V.; Berger, U.; Bastiansen, F.; Eliassen, S.; Ginot, V.; Giske, J.; Goss-Custard, J.; Grand, T.; Heinz, S. K.; Huse, G.; et al. 2006. A standard protocol for describing individual-based and agent-based models. *Ecological modelling* 198(1-2):115–126.

Ilachinski, A. 2012. Modelling insurgent and terrorist networks as self-organised complex adaptive systems. *International Journal of Parallel, Emergent and Distributed Systems* 27(1):45–77.

Krueger, A. B., and Malečková, J. 2003. Education, poverty and terrorism: Is there a causal connection? *Journal of Economic perspectives* 17(4):119–144.

Li, B.; Sun, D.; Zhu, R.; and Li, Z. 2015. Agent based modeling on organizational dynamics of terrorist network. *Discrete Dynamics in Nature and Society* 2015.

Macal, C. M., and North, M. J. 2005. Tutorial on agent-based modeling and simulation. In *Proceedings of the Winter Simulation Conference, 2005.*, 14–pp. IEEE.

Macy, M. W., and Willer, R. 2002. From factors to actors: Computational sociology and agent-based modeling. *Annual review of sociology* 28(1):143–166.

Moore, M. French marriage rate plunges as population, birth rate rise. *Seattle Times*.

Mullins, S. 2009. Parallels between crime and terrorism: A social psychological perspective. *Studies in Conflict & Terrorism* 32(9):811–830.

Perliger, A.; Koehler-Derrick, G.; and Pedahzur, A. 2016. The gap between participation and violence: Why we need to disaggregate terrorist 'profiles'. *International Studies Quarterly* 60(2):220–229.

Reeve, Z. 2020. Islamist terrorism as parochial altruism. *Terrorism and political violence* 32(1):38–56.

Ritchie, H.; Hasell, J.; Appel, C.; and Roser, M. 2013. Terrorism.

Russell, C. A., and Miller, B. H. 1977. Profile of a terrorist. *Studies in conflict & terrorism* 1(1):17–34.

Tsvetovat, M., and Carley, K. Knowing the enemy: A simulation of terrorist organizations and counter-terrorism strategies.

World Bank, W. D. I. 2018. Educational dataset barro-lee dataset.