UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI FISICA E ASTRONOMIA
GALILEO GALILEI

CORSO DI LAUREA IN FISICA

TESI DI LAUREA

Emergent Patterns in Global Terrorism

RELATORE: **Samir Suweis**

LAUREANDO: **Nicola Nicodemo**

ANNO ACCADEMICO 2016-2017

CONTENTS

LIST OF FIGURES

1 INTRODUCTION

Traditionally, *statistical mechanics* is the branch of physics that applies probability theory to the study of the thermodynamic behavior of systems composed of a large number of particles. In particular, it relates the microscopic properties of atoms and molecules to the macroscopic properties of materials.

However, statistical physics has also proven to be a fruitful framework to describe phenomena outside the realm of traditional physics. In an analogue way to Particle physics, the interactions of individuals as elementary units in social structures give rise to collective phenomena that can be studied by statistics ([1], [2]). In fact, human societies are characterized by a variety of global regularities [3].

In recent years, the idea of studying society within a framework of statistical physics has transformed to a concrete research effort, involving an increasing number of physicists. At the basis of this change is the availability of large databases and the appearance of new social phenomena that can be studied sistematically, because they are mostly related to the Internet [1].

In this work, we benefit from the large availability of records of terrorist attacks to introduce a systematic approach to the study of terrorism. Our objective is to point out regular patterns observed in insurgent activity. The dynamics of terrorist organizations is not utterly unpredictable.

Despite the large number of variables in their complex ecology, some basic mechanisms can be identified for the formation of insurgent aggregates [4], [5]. A model is proposed to relate to the macroscopical distribution of terrorist attacks.

Further investigation may eventually lead to a consistent theory for the dynamics of terrorism and support advancement in counter-terrorism strategies.

2 THE SIZE OF WARS

2.1 RICHARDSON'S STATISTICS OF DEADLY QUARRELS

The English physicist and mathematician Lewis Fry Richardson was a pioneer in studying social phenomena - e.g. violent conflicts - with the methodology and the systematicity typically used in natural sciences. In one of his paper dated 1948 - *Variation of the Frequency of Fatal Quarrels With Magnitude* [6] - he has showed that the size distribution of casualties in the wars recorded between 1816 and 1945 follows approximately a power-law distribution, with scaling exponent $\alpha \approx 1.50$. In other words, there is no characteristic dimension for the number of victims of a war.

Further investigation by the Swedish physicist and researcher in International conflict, Lars-Erik Cederman, extended these findings to more recent interstate wars (1820-1997), confirming [7] the power-law distributions of war casualties, with scaling exponent $\alpha = 1.41$.

We can replicate these results by analysing data from The Correlates of War Project Database (see Sec. 2.6). Fig. 1 shows the distribution of interstate war casualties recorded between 1820 and 2003, including more recent Invasions of Iraq and Afghanistan. Instead of relying on direct frequency counts, our calculations center on the *cumulative relative frequencies* $N(S > s)$ of war sizes: where S is the number of dead caused by a war (or war size, or *severity* of a war). This is used to calculate the probability $P(S > s)$ that a war has higher severity than s. Consequently, whereas for small conflicts the likelihood of more severe conflicts occurring is close to one, this probability approaches zero for very large events - because larger conflicts are unlikely to happen.

In formal terms, it can be argued that the cumulative probability scales as a power law:

$$P(S > s) = Cs^{-\alpha+1} \tag{1}$$

where C and α are positive numbers.

Our fit of empirical data from wars between 1820 and 2003 (see Sec. 2.6) returns the value $\alpha = 1.41$ for the scaling exponent, which is in good agreement with the results from literature ([6], [7]).

2.2 ANALYSIS OF DATA

For a comprehensive description of how the power law model has been fit to data, see Appendix A. Data analysis is performed by

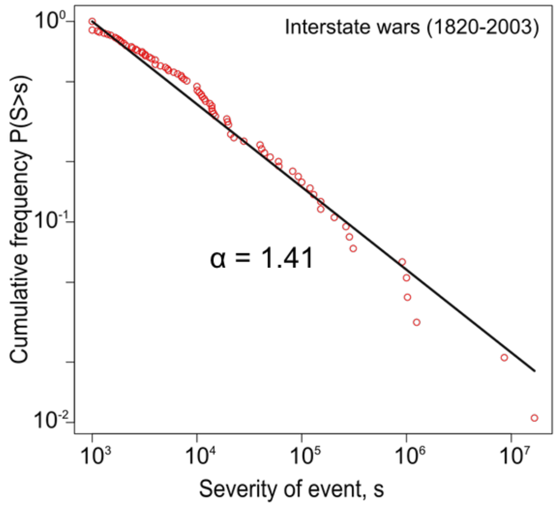


Figure 1: Cumulative frequency distribution of the severity of Interstate wars between 1820 and 2003. The scaling parameter of the power law model is $\alpha = 1.41$.

means of Excel, R and C++ programming languages. In particular, the fit of power law is performed using poweRlaw library for R, which uses a maximum likelihood estimate for the scaling parameter and a Kolmogorov-Smirnov statistics algorithm for the p-value (see Appendix A for more information). In the text, graphs of the Cumulative Frequency Distribution (CDF) are represented: so that the scaling parameter α is not the actual value of the slope - this being $\alpha - 1$.

p-values and other parameters of the power-law fit model are reported in tables or throughout the text.

2.3 GLOBAL TERRORISM: TOWARDS A NEW TYPE OF CONFLICT

The concepts and empirical data introduced so far pose two important questions: the statistics of conflicts makes some regularities manifest, which are similar in some aspects to those scientists use to study in other natural or human phenomena as diverse as earthquakes, biological extinctions, epidemics, fires, traffic jams, city growth, market and business performances, and, indeed wars [8]. On the other hand it is possible, and indeed necessary, to come up with a model which explains the underlying mechanism generating those regularities.

In this paragraph we introduce and analyze empirical data concerning terrorist events recorded from early Eighties to present. We want to show that the number of casualties follows a power-law distribution also in this case. Furthermore, the abundance of detailed data about thousands of individual events from tens of countries all over the world - with diverse social and economic background - let us observe a good variety of interesting patterns.

We are use to believe that 9/11 facts have fundamentally changed the nature of warfare. However, the discussion about the emergence of a new type of war had interested academics even before. Analystis such as Mary Kaldor, Thomas Hammes and William S. Lind were among the first to discern *new wars* emerging. They contributed to create the concept of *fourth-generation warfare* for the purpose of an argument for the changing of the face of war. A framework was created for understanding modern warfare, in which four generations are accounted for [9].

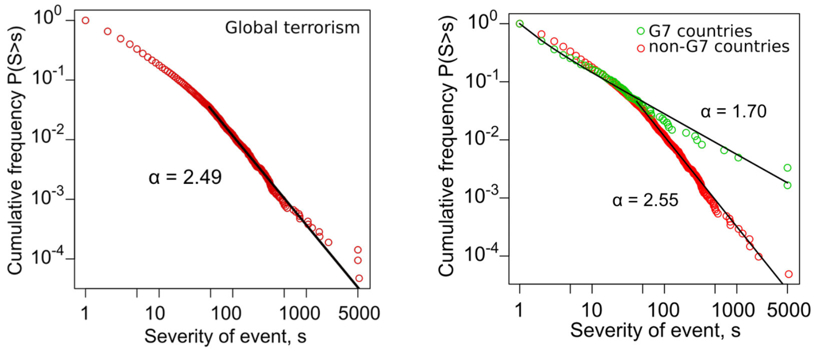
Whereas first-generation warfare was fought with mass manpower, the invention of machine guns changed the way of conducing battles. Still, the second-generation warfare is inspired to a culture of order, where plans and orders are carefully executed. World War II introduces an element of stealth and surprise to bypass enemy's defence: it is third-generation warfare and it is commonly known as *blitzkrieg*.

Fourth-generation warefare marks the end of the monopoly of the state on war. It is the case of insurgency, involving politics, combatants and civilians. This definition covers numerous well-known cases including conflicts in Iraq, Afghanistan, Israel-Palestine, terrorist organizations such as Al-Qaeda, the civil war in Syria or the Egyptian crisis. It includes elements of terrorism, insurgency or guerrilla tactics, decentralization, trans-national perspective, psychological influence and propaganda, especially through media manipulation and a spread out network of communication. These elements are important when it comes to recognize the patterns in empirical data and to formulate a theoretical model that explains them.

Fig. 2 shows our empirical findings for event sizes. In particular, we have collected terrorist attacks in the period from 1968 to 2009 (RDWTI RAND Database, see Sec. 2.6) and show their cumulative distribution function with respect to their severity (the number of casualties, which - in this case - is the sum of deaths and injuries).

The regular scaling in the upper tail of the distribution demonstrates the existence of a global pattern in the frequency statistics which involves events orders of magnitudes larger than average attacks: thus, very severe events must not be thought as anomalies produced by accidental circumstances. Furthermore, it has to be noticed that the scaling behavior exists, although significant changes have occurred in politics, geography and technology during recent decades. We can make a point that scaling behavior is a universal property of terrorism.

Nonetheless, scaling properties (i.e. the scaling parameter) may still be affected by environmental features. Specifically, we will examine how geographical distribution, industrial development and access to different weapons, produce variations in these properties.



(a) Casualties for global terrorism

(b) Casualties for terrorism in G7 and non-G7 countries

Figure 2: Cumulative frequency distribution of the severity of terrorist attacks occurring globally, and then separating G7-countries (Canada, France, Germany, Italy, Japan, Uk, Usa) and non-G7 countries. Scaling exponents for differently industrialized countries: all countries (2.49), G7 countries (1.70), non-G7 countries (2.55)

In the first place, we use the method of maximum likelihood (see Appendix for more detailed discussion on the methodology) to estimate the scaling parameter α of the power-law model from data. We obtain $\alpha = 2.49$ which is in good agreement with values published by Clauset et al. [5].

We decompose this distribution, by separating events occurring in the G7 countries (Canada, France, Germany, Italy, Japan, Uk, Usa) from the rest of the world (non-G7 countries). See Fig. 2-b. Scaling parameters are $\alpha_{G7} = 1.70$ and $\alpha_{rest} = 2.55$ respectively.

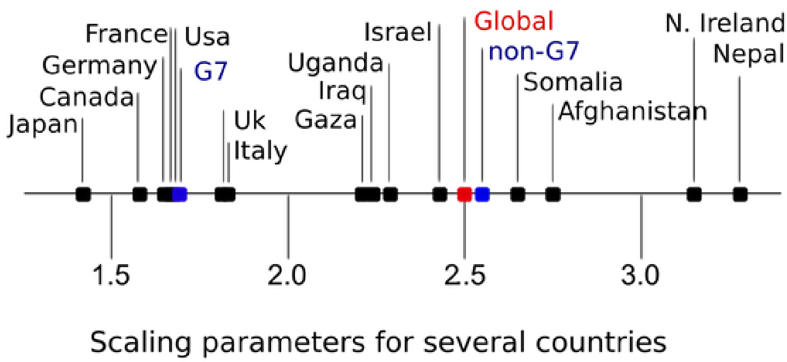


Figure 3: Scaling parameters for several countries. Two clusters emerge corresponding to values for G7 ($\alpha = 1.70$) and non-G7 ($\alpha = 2.55$) countries.

Research by Clauset, Young and Gleditsch reaches a similar result [5]. They consider the 30 nations belonging to the Organization for Economic Co-operation and Development (OECD). According to their study, 11% of the considered events took place in OECD countries. The remaining 89% occured throughout the rest of the world. However, the events inside OECD are mostly distributed in eight countries: Turkey, France, Spain, Germany, Usa, Greece, Italy and Uk (all of them are G7 countries with the exception of Turkey and Greece). Furthermore, 89.2% of the most severe events took place in these eight nations. In other words, terrorist attacks are slightly less likely to occurr in OECD area, but they tend to be more severe then in non-industrialized countries. Industrialization seems to have an impact on the frequency distribution of the events. But availability of certain technology itself, i.e. access to a certain kind of weapons, does not explain it all. Other political or social factors may be at the basis of the mechanisms producing different scaling parameters for different contexts.

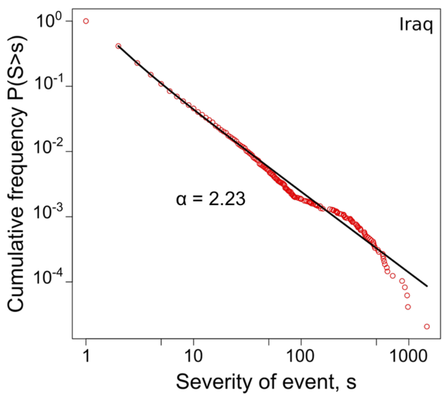
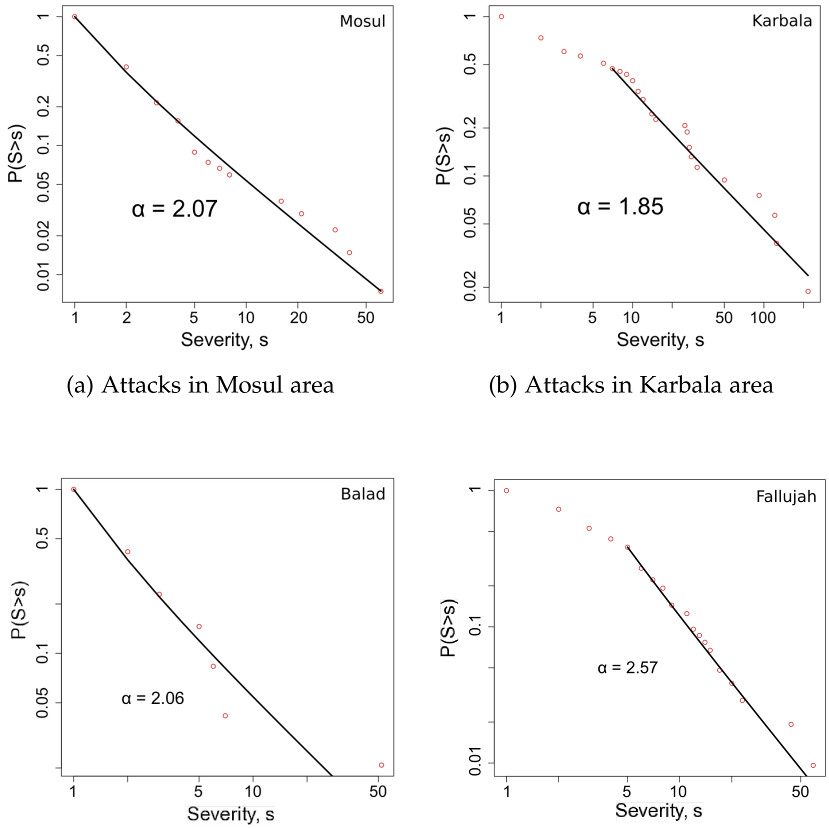


Figure 4: Cumulative frequency distribution of the severity of terrorist attacks in Iraq. Scaling parameter of the power-law model $\alpha = 2.23$

Johnson proposes that power-law scaling is not only a universal feature of global terrorism: the same behavior appears when considering conflicts in single nations [10]. We partition data by the geographical location and estimate the scaling parameter for each nation considered (see Fig. 3). Although the scaling parameters may vary from one country to the other, due to the peculiarities of the individual conflict that we have introduced before, we can distinguish two clusters around $\alpha = 1.7$ and $\alpha = 2.5$, which are the values for G7 countries terrorism and for non-G7 countries terrorism.

Fig. 4 shows the distribution of fatalities in terrorist attacks carried out in Iraq. The power-law model that fit data scales with parameter $\alpha = 2.23$. The same statistics continue to hold when considering restricted geographical areas of Iraq, suggesting that the severity of attacks carried out by a single terrorist group may also be distributed



(a) Attacks in Mosul area

(b) Attacks in Karbala area

(c) Attacks in Balad area

(d) Attacks in Fallujah area

Figure 5: Cumulative frequency distribution of the severity of terrorist attacks in some areas of Iraq (period 2003-2004). Attacks carried out locally - likely by a single terrorist group - follow approximately the same statistics of global terrorism.

as a power law. Fig. 5 shows the distributions for particular areas of Iraq: Mosul, Karbala, Balad, Fallujah.

2.4 POWER-LAW MODEL FIT

In the following table a list of the parameters of the power-law model fit for various data sets. N_{tail} is the number of events in the tail of the power law distribution; α is the scaling parameter; $x_{\min}$ is the lower bound of the distribution; p-value is calculated with KS statistics (see Appendix A).

Figure	Data set	N_{tail}	α	$x_{\min}$	p-value
Fig. 1	Interstate wars	95	1.41	1000	0.73
Fig. 2-a	Global terrorism	743	2.49	48	0.89
Fig. 2-b	G7 countries terrorism	605	1.70	1	0.13
Fig. 2-b	non-G7 countries terrorism	682	2.55	50	0.42
Fig. 4	Iraq terrorism	20020	2.23	2	0.11
Fig. 5-a	Mosul area terrorism	135	2.07	1	0.13
Fig. 5-b	Karbala area terrorism	25	1.85	7	0.12
Fig. 5-c	Balad area terrorism	48	2.06	1	0.43
Fig. 5-d	Fallujah area terrorism	40	2.57	5	0.83

2.5 ONLINE ECOLOGY OF INSURGENT AGGREGATES

Studying the internal dynamics of terrorist organizations is crucial to build up a model which can explain not only the universal pattern, but that also allows to understand the peculiarities of the conflicts. However, this cannot be done with real-world terrorist organizations operating on the field, due to a lack of empirical data. One particular case that can certainly be studied and analyzed in detail is that of online aggregates. Increased connectivity provided by the Internet facilitates the formation of real-world groups and the spread of pieces of propaganda that may inspire radicalized individuals to carry out violent attacks. This is the case of recent events in Europe directed by ISIS, but is also the case of less violent insurgent activities such as mass protests.

Johnson et al. have identified online aggregates supporting ISIS-related activity and, for comparison, online civil protestors across multiple countries (in particular Brazil) [4]. They have collected data to create a picture of how these aggregates develop online. In particular, pro-ISIS propaganda develop through self-organized online aggregates, which consist of groups of followers in online communities (e.g. on Facebook, Twitter, VKontakte or similar social networks). These groups are characterized by a strong social heterogeneity and no hierarchical structure. The lack of group leaders suggests that aggregation dynamics may be driven by self-organization. Online environment is also populated by predatory entities (i.e. police, hackers, social network moderators) seeking to shut down such aggregates activity. Hence, once the group is identified (this is usually done by manually analizing the content related to terrorist propaganda), one can analyze how the number of followers varies throughout its lifecycle.

According to Johnson et al. research, the distribution of pro-ISIS aggregates of size s follows a power law $s^{-\alpha}$ with exponent $\alpha = 2.33$. Moreover, the frequency of severe attacks perpetrated by ISIS also follows the characteristic power law distribution with $\alpha = 2.44$.

Online support is likely a condition for real-world action, and indeed proliferation of supporting aggregates can be an indicator of the conditions becoming favorable for an attack or a protest to take place. It has been observed that periods of relatively slow change, during which aggregates appear and disapper sporadically, are interrupted by an abrupt divergence in the rate at which new aggregates are formed. Such peak coincides with unexpected onset of real-world events (see for example Kobane, September 18, 2014 and Brazil mass protests June 11, 2013). This suggests that automated control of online activity of such aggregates can warn about the oncoming escalation and draw attention to the suspect activity.

Whereas an algorithm able to predict and thwart the outburst of terrorist attacks supported by online activity may not be here yet, one can benefit from the study of online aggregates dynamics.

2.6 DATA SOURCES FOR WAR AND TERRORISM EVENTS

In section 2.1, data from The Correlates Of War Project were used, restricting to interstate wars. In section 2.3, data from RDWTI RAND Database of Worldwide Terrorism Incidents have been used. In particular, data about terrorism in Iraq have been drawn from the Iraq Body Count Project.

Further information about databases can be found on the respective web pages:

`http://www.correlatesofwar.org`
`http://www.rand.org/nsrd/projects/terrorism-incidents.html`
`https://www.iraqbodycount.org`

3 A THEORETICAL MODEL

As discussed in Chapter 2 the analysis of empirical data suggests that - also for terrorist attacks - the distribution of the number of deaths or casualties follows a power law. Further investigation shows that the frequency and severity of events within individual conflicts exhibits the same power law statistics - although these vary depending on the characteristics of the specific conflict, indicating similarities in the underlying dynamic of the conflicts.

However, the question about which mechanism generates these behaviour still lacks a well-supported explanation. An accurate answer to this question - supported by a scientific and systematic approach to the subject, and in coordination with traditional studies - may shed a light on the internal dynamics of terrorist organizations and contribute to the development of novel policies and strategies for counter-terrorism.

Two explanations have been proposed for the origin of the observed pattern in the frequency of severe terrorist events: particularly, a competitive model by Clauset, Young and Gleditsch [5], and an adversarial aggregates model by Johnson et al. [4].

3.1 A COMPETITIVE MODEL

The model proposed by Clauset, Young and Gleditsch as an explanation to the origin of the scale invariance for terrorism, is based on a stochastic competitive process between state and non-state actors [5]. It is a variation of the model described by Reed and Hughes [11] and applied to a wide rage of phenomena from biology to economics to internet ecology: the basic idea is that if a process that grows exponentially is killed *randomly*, the distribution of the killed state follows a power law (in one or both tails).

For the purpose of our discussion, consider a terrorist who is planning an attack. The severity of an attack is given by two main contributions: 1) We make the idealization that the potential severity of an attack grows exponentially with time: in other words, attacks which require more amount of planning can potentially produce more casualties. The potential severity of an event is thus expressed as $p(t) \propto e^{kt}$, with $k > 0$.

2) On the other hand, the severity of a real event decades exponentially with time. Indeed, as time passes, there is less chance that the planning process succeedes, as the actors may be incarcerated or

killed by police or spontaneously abandon their plan. Thus, severity can also be expressed as $x \propto e^{-\lambda t}$, with $\lambda > 0$.

One can easily derive the distribution of the real event severities $p(x)$:

$$p(x) = p(t)\frac{dt}{dx} \propto x^{-\alpha}, \text{ where } \alpha = (1 + k/\lambda). \tag{2}$$

The two exponents of this model, k and λ, can be adjusted as parameters depending on the nature of the competitive process: e.g. they can be interpreted as the capabilities of state actors (i.e. police) and non-state actors (i.e. terrorist) of succeeding in their own tasks, respectively. In the likely hypotesis that both actors develop roughly equal capabilities (with state having maybe a slight advantage), the ratio $k/\lambda \geq 1$, giving the exponent $\alpha \geq 2$, in agreement with empirical data.

Irrespective of the actual value of the resulting exponent - which primarily depends on the features of the individual conflict that cannot be included explicitely in the equations, this model makes at least two good points: a simple and stochastic competitive process can explain the power law behaviour based on a few assumption which, despite the idealization, fit welll with the current understanding of terrorism; on the other hand, it produces a scaling exponent which varies depending on the model parameters: this may be necessary to explain the different scaling behaviours in empirical data for different industrialization level of the country or different use of weapon.

Still, further investigation about internal dynamics of the terrorist organization and about they ways of interaction with other state actors is needed in order to understand the nature of the model adjustable parameters.

3.2 THE ASSUMPTIONS OF JOHNSON'S MODEL

The mechanism proposed by Johnson et al. is a self-organized critical model of the internal dynamics of a terrorist organization [4]. It adapts particularly well to the dynamics of the online aggregates we have introduced in section 2.5. A process of aggregation and disintegration of terrorist cells produces a dynamic equilibrium that is characterized by a power-law distribution in the sizes of cells (and, by assumption of the severity of their attacks). The scaling exponent can be calculated exactly and is found to be $\alpha = 5/2$.

The model is based on four assumptions about the interaction between terrorist cells that compose a modern terrorist organization. It makes no other assumption about the relationship with other cells or antagonists, or with respect to the specific features of the conflict, or the type (strategy and weapons used) and the site of the attack.

Despite the simplicity and generalization of the assumptions - which consequently poses limits to a detailed analysis of the internal dynamics of the organization (that lack, however, a systematic study of empirical data), they permit us to solve the model mathematically.

In addition to explain the value of the scaling exponent, the model let us make qualitative considerations about the phenomenology of the conflict, and the basic underlying mechanism.

The assumptions of the model are:

- The total number N of radicalized individuals is $N >> 1$ and constant in time. These individuals can form cells of size $k = 1, 2, 3, 4...$ Let n_k denote the number of cells of size k.
- Terrorist cells undergo a process of aggregation (or *coalescence*), in which two cells merge to form a larger cell: the probability for unit of time that two cells of sizes i and j merge to form a cell of size $i + j$ is proportional to their sizes: $v_{coal}(ij)$
- Terrorist cells fall apart due to a process of disintegration (or *fragmentation*), in which a cell splits into single individuals: the probability of this occurring is proportional to its size: $v_{frag}k$ (This dependence can be generalized with no consequences on the result).
- Cells can produce an attack with probability independent of their dimension or their age or the number of attacks already launched; however, if the attack is produced, its severity $v(k)$ is directly proportional to the cell's size: $v(k) \propto k$.

This mechanism can be summed up by the equation:

$$\frac{\partial n_k(t)}{\partial t} = \frac{1}{2}\frac{v_{coal}}{N^2}\sum_{i,j=1}^{\infty} ijn_in_j - \frac{v_{coal}kn_k(t)}{N^2}\sum_{j=1}^{\infty} kn_j(t) - \frac{v_{frag}kn_k(t)}{N} \tag{3}$$

where $i + j = k$.

The first term represents an increase in the number n_k as a result of the merging of two aggregates of sizes i and j into a single cell of size $k = i + j$. The second term describes the decrease in n_k due to the merging of a cell of size k with another aggregate. Finally, the last term represents a decrease in n_k due to the fragmentation of an aggregate of size k.

3.3 A GENERALIZATION OF THE MODEL

We can operate a slight generalization of this model, as indicated by Clauset and Wiegel [12], to make the assumptions less strict and to extend the specific model to a family of such models. In particular, we let the probability of two cells merging depend on their sizes, i and j, in the form $C_0(ij)^a$ - where the parameters are $C > 0$ and $a \geq 0$. Furthermore we loosen the restrictions on the explicit form of the probability for a cell to be shut down, which is now just $b(k)$ instead of the direct proportionality.

The equation of the generalized model is:

$$\frac{\partial n_k(t)}{\partial t} = \frac{1}{2}C_0 \sum_{i,j=1}^{\infty} i^a j^a n_i n_j - C_0 k^a n_k \sum_{j=1}^{\infty} j^a n_j - b(k) n_k \tag{4}$$

where $i + j = k$.

Notice that the second summation in the right hand includes the term with $j = k$. To be precise, the number of combinations between some cell of size k and some of size j is $n_k n_l$ for $k \neq l$ and $\frac{1}{2}n_k(n_k - 1)$ for $k = l$. However, in the limit that $N \gg 1$, we shall find that also $n_k \gg 1$. In this case we can make the approximation $\frac{1}{2}n_k(n_k - 1) \cong \frac{1}{2}n_k^2$. Moreover, each combination of two cells of the same size k leads to the decrease of n_k by two: the loss term is then proportional to n_k^2.

3.4 ANALYTICAL SOLUTION OF THE MODEL

We analyse the model in the steady-state behavior. That is, we operate the limit $\lim_{t\to\infty} n_k(t)$ and denote as n_k^* the new variable in the steady-state limit. Eq. 4 simplifies to:

$$\frac{1}{2}C_0 \sum_{i,j} i^a j^a n_i^* n_j^* \delta_{i+j,k} = C_0 k^a n_k^* \sum_j j^a n_j^* + b(k) n_k. \tag{5}$$

for $k = 2, 3, 4...$

Notice that we don't consider the equation for $\partial n_1/\partial t$.

A way to solve Eq. 5 is by using the generating functions:

$$f(z) = \sum_{k=1}^{\infty} k^a n_k^* z^k; \qquad g(z) = \sum_{k=2}^{\infty} b(k) n_k^* z^k, \quad (b(1) = 0). \tag{6}$$

We multiply both sides of Eq. 5 by z^k and sum over k from 2 to ∞ and obtain:

$$\frac{1}{2}C_0 f(z) f(z) = C_0 f(1)[f(z) - n_1^* z] + g(z) \tag{7}$$

Eq. 7 can be solved easily for fixed z and$N \to \infty$. Assuming that the frequencies n_k^* are proportional to N, we just need to determine the leading orders of the various terms. Hence, eq. 7 can be replaced by the new equation:

$$\frac{1}{2}f^2(z) - f(1)f(z) + f(1)n_1^* z = 0 \tag{8}$$

which has the solution

$$f(z) = f(1) - \sqrt{f^2(1) - 2f(1)n_1^* z}. \tag{9}$$

We calculate $f(1)$ for $z = 1$: $f(1) = 2n_1^*$. Thus:

$$f(z) = 2n_1^*[1 - \sqrt{1-z}]. \tag{10}$$

Based on the definition of $f(z)$ given in Eq. 6, the term $k^a n_k^*$ can be found as the coefficient of z^k in the power series expansion of Eq. 10:

$$f(z) = 2n_1^* \left(\frac{1}{2}z + \frac{1}{8}z^2 + \frac{3}{48}z^3 + ... \right) \tag{11}$$

For generic k, we use Cauchy's theorem, which gives the contour integral:

$$k^a n_k^* = i\frac{n_1^*}{\pi} \oint_C z^{-k-1}\sqrt{1-z}dz \tag{12}$$

where the contour C encircles the complex z-plane for $1 \leq z < \infty$. If we choose z to be approximately 1, we can make the following substitution:

$$z = 1 + \chi, \qquad z^{-k-1} \cong e^{-(k+1)\chi}. \tag{13}$$

It yelds:

$$k^a n_k^* \cong \frac{2}{\pi} n_1^* \int_0^\infty \sqrt{\chi} e^{-(k+1)} d\chi = \frac{1}{\sqrt{\pi}} n_1^* (k+1)^{-3/2} \tag{14}$$

Hence, the number n_k^* of cells consisting of $k \gg 1$ individuals, in the steady-state, is given by the power law:

$$n_k^* \cong \frac{1}{\sqrt{\pi}} n_1^* k^{-a-3/2} \tag{15}$$

Because of the assumption that the severity of a terrorist attack is proportional to the size of the attacking cell, the probability p_k that an attack will produce k casualties is:

$$p_k \propto k^{-\alpha} \tag{16}$$

for $k \gg 1$, where the scaling exponent α is:

$$\alpha = a + 3/2. \tag{17}$$

In order to recover the results of Johnson's model, we set the parameter $a = 1$ as in the assumptions. It yelds the prediction $\alpha = 5/2$, in agreement with predictions from empirical data.

3.5 ON THE SIZE OF THE LARGEST CELL

In addition to the prediction of the scaling parameter $\alpha = 5/2$, this model makes another important prediction: the form of the function describing the fragmentation probability $b(k)$ determines how cells grow: in particular, it determines the largest possible size k_0 for a cell and whether it is possible that all terrorists merge in a single group of size N or not.

To show this relationship between $b(k)$ and k_0, let us consider the steady-equation for n_1^* that we haven't taken into account yet.

From previous assumptions:

$$\sum_{k=2}^{k_0} kb(k)n_k^* = C_0 n_1^* \sum_{l=1}^{\infty} l^a n_l^*. \tag{18}$$

With the same formalism used before, one finds that right side equals $C_0 n_1^* f(1)$ and that $f(1) = 2n_1^*$. Thus:

$$\sum_{k=2}^{k_0} kb(k)n_k^* = 2C_0(n_1^*)^2. \tag{19}$$

For the case $a = 1$, we can write:

$$N = \sum_{k=1}^{\infty} kn_k^* = f(1). \tag{20}$$

Hence $N = f(1) = 2n_1^*$. One finds that

$$n_1^* = \frac{1}{2}N. \tag{21}$$

Eq. 19 can be rewritten in the form

$$\sum_{k=2}^{k_0} kb(k)n_k^* = \frac{1}{2}C_0 N^2. \tag{22}$$

For explicit calculations, let $b(k)$ have the rather general form

$$b(k) = B_0 k^b \tag{23}$$

where b is some number near the unity, $1/2 < b < 3/2$.

Using the result of Eq. 15 we obtain (in the asymptotic limit when $k \gg 1$):

$$\sum_{k=2}^{k_0} kb(k)n_k^* \cong \frac{B_0 N}{2\sqrt{\pi}} \sum_{k=2}^{k_0} k^{b-3/2}. \tag{24}$$

The series on the right-hand can be approximated by an integral:

$$\sum_{k=2}^{k_0} k^{b-3/2} \cong \int_2^{k_0} k^{b-3/2} dk \cong \left(\frac{1}{b-1/2}\right) k_0^{b-1/2}. \tag{25}$$

With these results, Eq. 22 can now be written as:

$$\frac{B_0 N}{2\sqrt{\pi}} \left(\frac{1}{b-1/2}\right) k_0^{b-1/2} = \frac{1}{2} C_0 N, \tag{26}$$

which gives us an expression for the largest size possible, k_0:

$$k_0 = \left[\frac{C_0}{B_0}\left(b - \frac{1}{2}\right)\sqrt{\pi}N\right]^{1/(b-1/2)}. \tag{27}$$

In our case, we made the assumption that the probability of fragmentation is directly proportional to the size of the cell, hence $b = 1$ and $b(k) = B_0 k$. Then:

$$k_0 = \left[\frac{C_0}{B_0}\frac{\sqrt{\pi}}{2}N\right]^2. \tag{28}$$

3.6 NUMERICAL SIMULATION

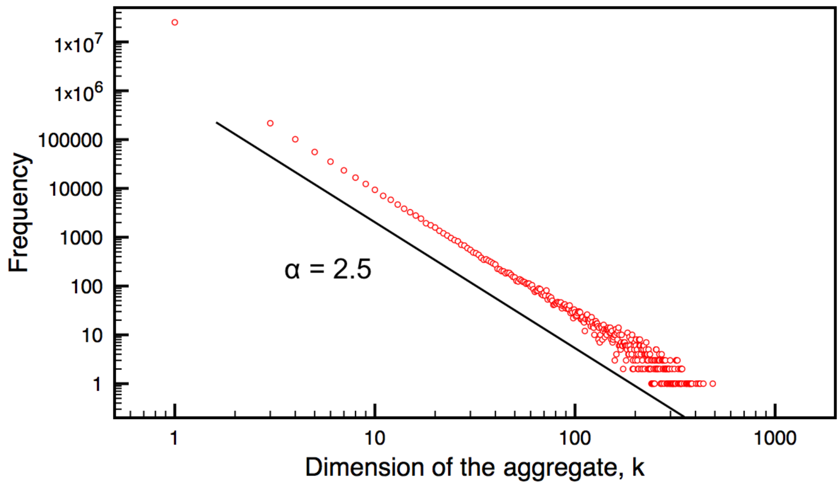


Figure 6:
Output of the numerical simulation: the number of cells of size k scales as a power law with scaling exponent $\alpha \sim 2.5$. The black line, with slope $\alpha = 2.5$ is a guide for the eye, not a fit.

We shall now run a numerical simulation which reproduces the mechanism described in Johnson's model (section 3.2). The simulation is coded in C++ and is based on the Johnson's et al. reference.

Let us have a pool of N aggregates, each denoted by a label $i, j, k, \ldots$, and let their size be denoted by $s_i, s_j, s_k, \ldots$. Initially, each aggregate is composed by just one individual such that the sum of their sizes equals N. This is the number of terrorists which will be dynamically distributed among aggregates as a result of the coalescence-fragmentation process. At each istant in time, a particular aggregate is selected randomly from the pool, but with a probability proportional to its size, and it is decided whether it will undergo coalescence or fragmentation process.

In the case of coalescence, another aggregate is picked randomly but witch chance to be selected proportional to its size. This is possible, due to the fact that in the array of a constant number N of elements, an aggregate with size s_i occurs s_i times, so that it has proportionally more chance to be picked. In fact, coalescence process consists in updating the sizes of the both merged aggregates to the total size of the two, and assigning them the same label, so that the same aggregate appears more times in the array. On the other hand, when the selected aggregate of size s_i is decided to undergo fragmentation, it is broken up into s_i aggregates of size equal to 1. The program update the size of the s_i identic elements in the array to 1 and give them different labels to indicate that know terrorist belongs to different cells (lone terrorists indeed).

This process is repeated for a number `tmax` of times. After the initial 2000 cycles, which account for the transient period before the system enters the steady-state behavior, a counter annotates the number n_s of aggregates of size $s = 1, 2, ..., N$ that have appeared at each step.

Fig. 6 shows the output of the simulation.

3.7 ON THE VALIDITY OF THE ASSUMPTIONS

Whether this model will prove useful or not in the long-term, it depends on how accurate it is in internal dynamics of modern terrorist organizations. There is a question about how strong is the dependence of the prediction - the power-law distribution - on the particular assumptions of the model. Further research would show how those assumptions can be relaxed or even eliminated while still producing the scaling behavior. Efforts can be done toward describing a more generalized and realistic mechanism.

First, we should ask how likely is that the probability of merging two cells is proportional to the product of their size. We may argue that a new individual can spontaneously decide to adhere to a group, e.g. based on similarity of interests or political ideology. Indeed, the model can be generalized by attaching to each group or individual a vector defining their peculiar character. Coalescence-fragmentation may then be depending on similarity of characters. At this time, however, we can simplify this mechanism by assuming that cells merge in order to increase their power, primarily. Thus, increasing size.

In fact, the geography of terrorist organizations in our model may be more variegated, if we allow interactions between cells and strategies to survive the shut down to be more complex (as happens indeed in online aggregates, ref.).

The assumption that the severity of attacks is directly proportional to the cell's size is more difficult to validate experimentally. One may think, in first approximation, that the capacity of launching a very severe attack is roughly proportional to the quantity of resources that the terrorist group disposes and to the capacity of coordinating the planning and the logistics - which are likely proportional to the size of the group itself.

Otherwise, one may show that a similar law-like pattern in empirical data holds when considering attacks carried by a single terrorist organization. In other words, the frequency of attacks with size s perpetrated by a specific group may scale as a power law with a roughly similar scaling parameter. The product of the two power laws with same exponents will yeld a power law pattern with same exponent as predicted by the model. Emphasis would shift to the dynamics of the attacked launched by a single group: Clauset's toy model could be helpful in this case.

4

CONCLUSIONS

Chapter 2 of this work showcases a series of regular patterns found in empirical data for conflicts. Following the path traced by Richardson, we use adequate statistical methods to find that the frequency distribution of fatal terrorist attacks scales as a power law, with scaling parameter $\alpha = 2.49$. Events of high severity, are not anomalies; they are rather produced by the same mechanism behind small but frequent attacks. Hence, large events occur with a probability orders of magnitude higher than we would expect from a normal distribution. Scaling behavior is not only a universal characteristic of global terrorism. We have separated events per geographical location. National conflicts show the same power law distribution. Different conflicts exhibit different scaling parameters, but two clusters emerge corresponding to terrorism in industrialized countries (i.e. G7 countries, having $\alpha = 1.70$) and in non-industrialized countries (i.e. non-G7 countries, having $\alpha = 2.55$). This scaling parameter can be thought as a "*measure*" of the type of conflict, thus reflecting the characteristics of the underlying mechanism. Guerrilla conflicts have higher parameter, while more traditional warefare tends to have lower parameter.

The scaling behavior depends on the internal dynamics of the terrorist organizations. Whereas it is generally difficult to model internal terrorist groups activity, due to lack of information and empirical data, one case that can be studied in detail is the *online ecology of insurgent aggregates*. These are groups of radicalized individuals participating in online communities supporting terrorist propaganda and playing a role in real-world activities. Research conducted by Johnson et al. on online pro-Isis aggregates shows that the number of aggregates of size s scales as a power law $s^{-\alpha}$ with $\alpha = 2.33$.

Chapter 3 introduces a model of the internal dynamics for terrorist organizations, which particularly fits the dynamics of online aggregates. It is based on a process of coalescence and fragmentation, where radicalized individuals or groups merge to create larger aggregates and are randomly shut down proportionally to their size. Solving the master equation for this mechanism in the steady state yelds a power law distribution in the number of cells with size s. Based on the assumptions of Johnson's model, the scaling parameter of this distribution is $5/2$, in agreement with experimental data.

Ultimately, we have depicted a framework in which terrorism may be studied sistematically and predictions about the dynamics of insurgent organizations can be done based on empirical data. It also poses the basis for a more consistent modelization of terrorism.

A FITTING POWER LAW TO EMPIRICAL DATA

Here we shortly describe the method, proposed by Clauset et al. [8], that we use to validate the power law distribution of empirical data and to estimate the scaling exponent and other parameters.

A first simplistic approach to determine whether a set of recorded data follows a power law distribution would be to construct a histogram representing the frequency of the measured quantity x, and plot it on doubly logarithmic axes. Doing so, one obtains a distribution that approximately follows a straight line, whose absolute slope gives the value of the scaling parameter α. Indeed:

$$\ln p(x) = \alpha \ln x + \text{const.} \tag{29}$$

Slope is typically estimated by performing least-squares linear regression on the modified data. Unfortunately, this method introduces systematic errors. The extimation of the scaling parameter is not trusty and we are not guaranteed that data really follow a power law.

A generally accurate method for estimating the parameters of a power law distribution follows. Furthermore, a way of testing the power-law hypotesis is discussed.

A.1 EXTIMATING THE SCALING PARAMETER

Let x represent the quantity we are interested in. A power-law distribution is represented by the *probability density* $p(x)$ such that:

$$p(x)dx = \Pr(x \leq X < x + dx) = Cx^{-\alpha}dx \tag{30}$$

where X is the observed value (in a continuous distribution) and C a normalization constant. As the equation 30 diverges as $x \to 0$ there must be some lower-bound x_{min} to the power-law behaviour.

Provided that $\alpha > 1$, the normalization condition yelds:

$$p(x) = \frac{\alpha - 1}{x_{\text{min}}} \left(\frac{x}{x_{\text{min}}} \right)^{-\alpha}. \tag{31}$$

The value of α for the power-law model that best fits our data maximizes the *likelihood* probability function:

$$p(x|\alpha) = \prod_{i=1}^{n} \frac{\alpha - 1}{x_{\text{min}}} \left(\frac{x_i}{x_{\text{min}}} \right)^{-\alpha} \tag{32}$$

or its logarithm $\mathcal{L}$:

$$\mathcal{L} = \ln p(x|\alpha) = \ln \prod_{i=1}^{n} \frac{\alpha - 1}{x_{\text{min}}} \left(\frac{x_i}{x_{\text{min}}} \right)^{-\alpha}. \tag{33}$$

where x_i, with $i = 1, 2, 3..., n$, are the observed values of x such that $x_i \geq x_{\text{min}}$. Setting $\partial\mathcal{L}/\partial\alpha = 0$ and solving for α we obtain the *maximum likelihood estimate* (MLE) for the scaling parameter:

$$\alpha = 1 + n\left[\sum_{i=1}^{n} \ln \frac{x_i}{x_{\text{min}}}\right]^{-1}. \tag{34}$$

A.2 ESTIMATING x_{min}

In order to extimate the value of the scaling parameter α, one should determine the lower bound x_{min} of the distribution. Athough this can be done by visually analyzing the points of the database, a more objective approach is desirable.

The method that we introduce here - but that is not the only - is based on the idea that the chosen x_{min} makes the measured data and the best-fit power law model as similar as possible above x_{min}. The Kolmogorov-Smirnov statistics, or KS statistics, is used to quantify the distance between two probability distributions:

$$D = \max_{x \geq x_{\text{min}}} |S(x) - P(x)|. \tag{35}$$

Here $S(x)$ is the *cumulative distribution function* of the data with $x \geq x_{\text{min}}$, and $P(x)$ is the CDF of the power-law model that best fits data for $x \geq x_{\text{min}}$. Hence, the best estimate of x_{min} minimizes the distance D.

A.3 TESTING THE POWER-LAW HYPOTESIS

The methods for estimating the scaling parameter and the lower bound does not determine whether the power law is actually a good model for empirical data. Indeed, regardless of the true distribution from which data are drawn, one can always fit a power law.

An approach to the question is to perform a *goodness-of-fit test*, returning a *p-value* that quantifies how plausible is the hypotesis of a power-law model. The basic idea is to use Kolmogorov-Smirnov statistics to compute either the distance of the distribution of empirical data from the hypothesized model, and the distance of synthetic data sets drawn from the same model. The *p-value* is defined to be the fraction of synthetic data sets whose distance from the hypothesized model is larger than the distance of empirical data.

If p is close to 1, then the difference between empirical data and the model can be attributed to statistical fluctuations; otherwise, if p is small, the model is not a good fit to the data. Clauset et al. quantify a treshold for p to rule out the power-law hypothesis: if $p \leq 0.1$ the model agrees poorly with data and the hypothesis can be rejected.

BIBLIOGRAPHY

[1] C. Castellano, S. Fortunato, and V. Loreto. "Statistical physics of social dynamics." In: *Rev. Mod. Phys.* 81.591 (2009).

[2] R. Albert and A. L. Barabási. "Statistical mechanics of complex networks." In: *Rev. Mod. Phys.* 74.47 (2002).

[3] M. Buchanan. *The social atom*. London, UK: Marshall Cavendish Business, 2007.

[4] N.F. Johnson et al. "New online ecology of adversarial aggregates: ISIS and beyond." In: *Science* 352.6292 (2016), pp. 1459–1463.

[5] A. Clauset, M. Young, and K.S. Gleidtsch. "On the Frequency of Severe Terrorist Events." In: *The Journal of Conflict Resolution* 51.1 (2007), pp. 58–87.

[6] L. F. Richardson. "Variation of the Frequency of Fatal Quarrels with Magnitude." In: *Journal of the American Statistical Association* 43.244 (1948), pp. 523–546.

[7] L.E. Cederman. "Modeling the Size of Wars: From Billiard Balls to Sandpiles." In: *American Political Science Review* 97.1 (2003), pp. 135–150.

[8] Aaron Clauset, Cosma Rohilla Shalizi, M. E. J. Newman A. Clauset, C. R. Shalizi, and M. E. J. Newman. "Power-Law Distributions in Empirical Data." In: *SIAM Rev.* 51.4 (2009), pp. 661–703.

[9] W.S. Lind, K. Nightengale, J.F. Schmitt, J. W. Sutton, and G. I. Wilson. "The Changing Face of War: Into the Fourth Generation." In: *Marine Corps Gazette* (1989), pp. 22–26.

[10] N.F. Johnson, M. Spagat, J. Restrepo, J. Bohórquez, N. Suárez, E. Restrepo, and R. Zarama. "From old wars to new wars and global terrorism." In: *eprint arXiv:physics/0506213* (2005).

[11] W. J. Reed and B. D. Hughes. "From gene families and genera to incomes and internet file sizes: why power laws are so common in nature." In: *Phys Rev E Stat Nonlin Soft Matter Phys.* 66.6 (2002).

[12] A. Clauset and F. W. Wiegel. "A generalized aggregation-disintegration model for the frequency of severe terrorist attacks." In: *The Journal of Conflict Resolution* 54.1 (2010), pp. 179–197.