A complex networks approach to find latent clusters of terrorist groups

Gian Maria Campedelli[1*] , Iain Cruickshank[2] and Kathleen M. Carley[2]

*Correspondence:
gianmaria.campedelli@unicatt.it
[1]Transcrime - Università Cattolica del Sacro Cuore, L.go Gemelli, 1, Milan, Italy
Full list of author information is available at the end of the article

Abstract

Given the extreme heterogeneity of actors and groups participating in terrorist actions, investigating and assessing their characteristics can be important to extract relevant information and enhance the knowledge on their behaviors. The present work will seek to achieve this goal via a complex networks approach. This approach will allow to find latent clusters of similar terror groups using information on their operational characteristics. Specifically, using open access data of terrorist attacks occurred worldwide from 1997 to 2016, we build a multi-partite network that includes terrorist groups and related information on tactics, weapons, targets, active regions. We propose a novel algorithm for cluster formation that expands our earlier work that solely used Gower's coefficient of similarity via the application of Von Neumann entropy for mode-weighting. This novel approach is compared with our previous Gower-based method and a heuristic clustering technique that only focuses on groups' ideologies. The comparative analysis demonstrates that the entropy-based approach tends to reliably reflect the structure of the data that naturally emerges from the baseline Gower-based method. Additionally, it provides interesting results in terms of behavioral and ideological characteristics of terrorist groups. We furthermore show that the ideology-based procedure tend to distort or hide existing patterns. Among the main statistical results, our work reveals that groups belonging to opposite ideologies can share very common behaviors and that Islamist/jihadist groups hold peculiar behavioral characteristics with respect to the others. Limitations and potential work directions are also discussed, introducing the idea of a dynamic entropy-based framework.

Keywords: Terrorism, Political violence, Community detection, Computational criminology, Von Neumann entropy, Gower's coefficient

Introduction

Complex networks have demonstrated their potential in many different domains. Approaches that rely on dynamic, multi-mode, multi-partite and meta-networks have been fruitful in shedding light on a wide variety of phenomena, including social ones (Barabási et al. 2002; Carley 2002; Szell et al. 2010; Centola 2010). In the last years, this process has indeed also touched areas as criminology, international security, and terrorism research (Cranmer et al. 2015; Berlusconi et al. 2016; Bx et al. 2015) .

This methodological shift has been facilitated by the increasing availability of open access data sets, the sensibility and interest of social scientists towards novel empirical approaches and the dramatic popularity of statistical software and data-science oriented languages. In spite of this shift, several critical points and pitfalls have been highlighted

by scholars regarding the actual results of scientific inquiry in the field of terrorism research. Sageman (2014), for instance, noted that the lacking collaboration between intelligence and academia led to a stagnation that is mainly motivated by the scarcity of rich, detailed and precise data on terrorist groups and events, which makes it difficult for researchers to develop models that are actually useful in reducing or assessing the terrorist threat.

In fact, Sageman argued that the intelligence community should be more willing to share crucial and rich data sets to the academia, in order to exploit their methodological rigour and capabilities. Recently, in an attempt to extensively review the field of terrorism research, Schuurman (2018) noted that many longstanding weaknesses and issues have been either completely or partially solved (e.g. scholars have expanded the range of data gathering techniques), while at the same time other issues are still in place. Among the others, scarcity of international and interdisciplinary collaborations and the high number of one-time contributors are preventing the field to develop in a more structured direction, therefore limiting the probability for high-impact and practical research.

In spite of these structural limitations, we seek to demonstrate the potential capabilities of complex networks to highlight hidden patterns within the terrorist global scenario, with the final aim to stimulate the debate on the application of novel methodological frameworks to research on terrorism. Hidden patterns could highlight operational similarities between groups that do not share any ideological background, peculiar attack-planning characteristics related to actors operating in certain areas, or even relevant behavioral differences between groups that fight for similar motivations but are settled in distinct regions. Our intuition is that there is first and foremost a need for advancing and experimenting novel methodological approaches that, in case of promising results, might be employed and applied to other contexts with more reliable data, allowing to draw more useful and solid conclusions. In fact, the unavailability of and search for better data should not stop the process of innovation within the field. This works relies on data retrieved from the Global Terrorism Database (GTD henceforth) on terrorist attacks occurred at global level from 1997 to 2016. The paper aims to propose a new algorithm for detecting latent clusters of terrorist groups expanding and extending the analytic approach we have provided in Campedelli et al. (2019): we demonstrate that, tested against our previous approach (which we will refer to as "baseline" throughout the paper) and a weak heuristic approach based on pure groups ideology, our novel algorithm confirms many results obtained with the baseline approach and also provides new interesting results on the hidden similarities across groups belonging to very different contexts and motivations.

The paper is organized as follows: the next section provides a review of network- and clustering-based approaches to the study of terrorism, trying to identify the main areas of application in which these methods have been experimented. Following, the "Data" section will thoroughly present the information contained in the data employed to conduct our analyses. The "Methodology" section will describe and explain the graph construction and algorithmic framework. The "Results" section will then describe relevant outcomes and patterns found after the experiments. Finally, "Discussion & future work" section will focus on the potential implications of this work, on its limits and on the possible directions that can be explored using this paper as a starting point.

Background

In recent years, one of the methodological frameworks that have been tested and have attracted the attention of both scholars and policymakers in the social sciences is network science, broadly intended (Borgatti et al. 2009). Network science has gained popularity in sociology (Granovetter 2018; Keuschnigg et al. 2018; Centola 2018), economics (Schweitzer et al. 2009; Hausmann and Hidalgo 2011; Vitali et al. 2011; Caccioli et al. 2018), political science (Hays et al. 2010; Ward et al. 2011; Gerber et al. 2013; Ribeiro et al. 2018) and criminology (Morselli 2009; Papachristos et al. 2015; Agreste et al. 2016; Calderoni et al. 2017; da Cunha and Gonçalves 2018).

This also applies to terrorism research. The first application of social network analysis to terrorism, the branch of network science that specifically seeks to map and study relation between human entities such as people or organizations, was the renowned paper by Krebs (2002). Krebs tried to understand the existing connections between hijackers and terrorists responsible for the 9/11 attacks using unstructured data retrieved from open access sources, as newspapers. Despite its limited sophistication, the study opened a path towards the study of terrorism under this new perspective. Following this strategy, other scholars have used relational data on individuals to reconstruct terrorist networks and investigate roles and key players (Koschade 2006; Brams et al. 2006; Belli et al. 2015).

Shifting from the pure physical and relational information gathered and structured to investigate the structure of groups, scholars have also tested and simulated the strength and resilience of terrorist networks. As an example, works in this area have put a strong emphasis on the application for intelligence purposes. They have relied on mathematical models focusing on network topology for either proposing methods for maximizing efficiency in disruption strategies (Carley 2006; Lindelauf et al. 2011; Eiselt 2018; Ren et al. 2019) or understand the most resilient topology structures to be learned from terrorism behavior and applied to other domains (e.g. infrastructure networks) (Gutfraind 2010). This interest towards increasingly complex questions regarding the nature and behavior of terrorist networks encourages scholars and scientists to integrate relational and topological information on networks with their spatial and temporal dynamics. Spatial and temporal dynamics are crucial when aiming at understanding the evolution of a certain entity or phenomenon, therefore several works have focused on these aspects using either synthetic-generated data or real-world information on existing networks (Moon and Carley 2007; Medina and Hepner 2011).

In the meanwhile, the revolution of social media has provided an unprecedented and massive amount of data to study the online social behaviour of people. As for the real physical world, individuals act criminally or violently also within the internet, and therefore researchers have started to be attracted by the potential consequences of criminal, and even terrorist, behaviors in the cyberspace. Indeed, a recent stream of research has focused on the detection of terrorist or radical behaviors retrieving network-information from social media platforms. Social media allow to go beyond pure relational information, integrating instead geographical, temporal features and many other profile attributes to infer patterns and dynamics of extremist users (Bouchard et al. 2014; Chatfield et al. 2015; Klausen 2015; Benigni et al. 2017).

The previous lines of research, though generally different in their data gathering techniques, modelling architectures and complexity scales, all mostly focus on mapping relations between individuals or, at most, organizations belonging to the same terrorist

sphere (e.g. the al Qaeda network). However, a very recent sub-domain explored the power of the complex network landscape when dealing with event data and abstract meta-networks of attack characteristics, with the aim to predict future terrorist behaviours in terms of target or weapon selection, targeted locations and employed tactics (Desmarais and Cranmer 2013; Tutun et al. 2017; Campedelli et al. 2018) or, more broadly, to highlight operational similarities between different terrorist organizations (Campedelli et al. 2019; Campedelli et al. 2019).

While network approaches for modelling terrorism have gained a certain degree of success and have tested and experimented techniques focusing on a variety of research questions, it is worth to note how this advancements have not been followed by the consequent combination of network science with unsupervised learning and, more specifically, cluster analysis. In one of the first attempts at using cluster analysis to group terrorist organizations, Chenoweth and Lowham (2007) used data on groups which targeted American citizens to explore alternative ways to conceive terrorist typologies. Qi et al. (2010) used both social network analysis and unsupervised learning to group extremist web pages using an hierarchical multi-membership clustering algorithm based on the similarity score of these pages. Finally, Lautenschlager et al. (2015) developed the Group Profiling Automation for Crime and Terrorism (GPACT) prototype that generates terrorist group profiling via a multi-step methodology that also includes clustering of terrorist events.

In light of this gap in research, following the intuition that network science may provide rich insights on the terror phenomenon, we modify our previous proposed methodology to test the performance of an automatic weighting scheme for the Gower's coefficient of similarity based on von Neumann's Entropy to preserve intrinsic qualities of the data that already emerge from our baseline approach.

Data

This work relies on data retrieved from the Global Terrorism Database (GTD) (LaFree and Dugan 2007; National consortium for the study of terrorism and responses to terrorism 2016). The GTD is the most comprehensive and detailed open access dataset on terrorist events at global scale, maintained by the START research center. Information are 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 (1) intentionality of the incident, (2) presence of violence (or immediate threat of violence) of the incident and (3) to the sub-national nature of terrorist actors.

The second level criteria are three and at least two of them must be 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, an additional filtering mechanism (variable *doubter*) controls for conflicting information or acts that may not be of exclusive terrorist nature (START 2017). For our analysis, we aggregated data (i.e., we did not separate by year or other time windows) from 1997 to 2016 on worldwide events and related perpetrators, excluding all the attacks which were of

doubtful terrorist nature.[1] This methodological choice led from 106,114 events to a total of 88,513. Furthermore, we have removed all the events plotted by "Unknown" actors. Considering the large amount of attacks with no identified perpetrator, we would have faced the risk of biased results. We have thus kept only attacks of clear terrorist nature with an identified author, accounting for a total of 41,456 events.

The multi-partite network which has been created and employed for our study relied on six main terrorist dimensions, namely: Events (N=41,456), Groups (N=1,493), Targets (N=22)[2], Weapons (N=13)[3], Tactics (N=9)[4] and operating Regions (N=12) [5]. These dimensions have been chosen because they represent the visible core of terrorist activity: the terror attack itself can indeed be represented by its perpetrator, the chosen target, the employed weapons and tactics and the geographic and political context in which it occurred. These variables are thus helpful in gathering a rich knowledge structure that will then be crucial for our methodology.

In addition to this information which represent the basis of this work, other variables extracted from the GTD and other sources have been employed to detect and assess behavioral patterns of terrorist groups belonging to the same clusters *ex post*. This information will include group-based attributes regarding terrorist activity such as ideology, success rate, suicide rate, fatality rate, casualty rate, multiplot rate, international rate and number of targeted countries. The ideology of each group has been mapped using existing information present in two open access data sets (Big Allied and Dangerous 1 and an extraction of Big Allied and Dangerous 2) when that information was available within those sources (Asal and Rethemeyer 2015), and by exception from other qualitative open access information sources.

This mapping led to include seven ideology categories: (i) Islamist/Jihadist groups, (ii) Far Left/Anarchist/Communist (FL), (iii) Far Right/Racist/Nazi (FR), iv) Ethno-Nationalist, (v) Other/Unknown, (vi) Religious (Islam excluded), (vii) Animal-rights/Environmentalist. A given group may belong to more than one category at a time (e.g.: the Popular Front for the Liberation of Palestine which contains at the same time elements of Marxism and Nationalism) (Table 1). It is worth to specify that these are labels that aim at giving context regarding the main motivations and ideological positions of the groups. This of course does not imply that enviromentalism, for instance, has to be associated with terrorism *per se*. These categories only mean that a given group that has plotted at least one attack included in the GTD had motivations and roots that can be matched with a given ideology. The same applies to left-wing or right-wing organizations: having a particular political position does not automatically qualify an existing entity (either a person or a group) as terrorist. However, there are diverted and extremist positions on both political sides that are tightly connected with groups and actors that have been responsible of terrorist acts.

---

[1]We have also excluded attacks from 1970 to 1996 because, as reported in the official GTD codebook, many variables on attacks occurred prior to 1997 were not available or sufficiently reliable.
[2]Targets list includes: Abortion Related, Government (General), Private Citizens & Property, Business, Religious Figures/Institutions, Police, Airports & Aircraft, Utilities, Educational Institution, Unknown, Journalists & Media, Government (Diplomatic), Other, Military, Telecommunication, Tourists, Terrorists/Non-State Militia, Transportation, NGO, Violent Political Party, Maritime, Food or Water Supply.
[3]Weapons list includes: Incendiary, Explosives/Bombs/Dynamite, Firearms, Unknown, Melee, Fake Weapons, Chemical, Other, Sabotage Equipment, Vehicle (not to include vehicle-borne explosives, i.e., car or truck bombs), Biological
[4]Tactics list includes: Facility/Infrastructure Attack, Bombing/Explosion, Armed Assault, Unknown, Assassination, Hostage Taking (Kidnapping), Unarmed Assault, Hijacking, Hostage Taking (Barricade Incident)
[5]Operating Region list covers the entire world and specifically includes: North America, South Asia, Middle East & North Africa, Sub-Saharan Africa, Western Europe, Eastern Europe, South America, Southeast Asia, East Asia, Central America & Caribbean, Australasia & Oceania, Central Asia

Table 1 Descriptive Statistics of Group Ideologies

Ideology	N groups	Share
Islamist/Jihadism	462	*0.31*
FL	271	*0.18*
Ethno/Nationalist	601	*0.40*
FR	50	*0.03*
Other/Unknown	206	*0.14*
Religion (No Islam)	45	*0.03*
Animal/Environmentalist	14	*0.01*
N of ideologies	**N groups**	**Share**
1	1347	*0.90*
2	136	*0.09*
3	10	*0.01*
Total	**1493**	*1.00*

The *success share* is given by the ratio between the successful attacks and the total number of events attributed to a given group. The *suicide share* maps the ratio of suicide attacks over the total number of events plotted by the same group. *Fatality* and *casualty* ratios are produced by the number of attacks with at least one dead victim (fatality) or one wounded victim (casualty) divided by the total number of events. The *international rate* is simply the ratio between attacks with some international features (e.g. logistic organization) and the total number of attacks. Finally, *multiplot share* quantifies the share of attacks that were part of a coordinated strategy (e.g. 9/11 case), out of total attacks. All these variables seek to enrich the knowledge associated to each group and to understand whether the identified clusters highlight certain patterned and eventually unexpected behaviors (Table 2).

Methodology

At the general level, the entropy-based approach that is presented and analyzed in this work is structured as follows: (i) calculation of the weights of each mode using the graph entropy of that mode; (ii) computation of the weighted Gower's Coefficient of Similarity between each of the terrorist groups using the entropy as the weight; (iii) extraction of the latent network from the pairwise Gower's Coefficient similarities and analyze its structural and intra-cluster properties. The detailed process is described in the following subsections. The entropy-based method will be then compared with the baseline model presented in (Campedelli et al. 2019) and with a heuristic method. The baseline model uses a simplified version of Gower's method. In this simplified version, no weights are

Table 2 Group-based Attributes on Terrorist Activity - Descriptive Statistics

	Mean	St.Dev.	Median	Min	Max
Events	27.76	205.81	2.00	1.00	5,634
Success Share	0.90	0.24	1.00	0.00	1.00
Suicide Share	0.03	0.14	0.00	0.00	1.00
Fatality Ratio	2.75	8.55	0.50	0.00	170
Casualty Ratio	7.88	23.34	1.67	0.00	385.29
Multiplot Share	0.14	0.28	0.00	0.00	1.00
International Share	0.29	0.41	0.00	0.00	1.00
N Targeted Countries	1.35	1.83	1.00	1.00	42

applied to the different modes and we only consider the natural structure of the data deriving from the affinity matrix that originates by the pairwise Gower's coefficient of similarity. In other words, instead of using the graph entropy of each mode as its weight, every mode is simply just given a weight of one. The heuristic method we use for comparison only uses groups' ideologies as the clustering criterion. In this method, we just use the dominant ideology of a particular terrorist group as its cluster label. So, if two groups share a dominant ideology, like Ethno/Nationalist, then they are in the same cluster.

Entropy-based gower's method for multi-partite data

Since the variables of the modes of Targets, Weapons, Tactics, and Regions form a many-to-many relationship with the terrorist groups, we first model this data as a multi-partite network (Fig. 1) with each partition joined to the terrorist groups; this is often referred to as a 'Star' structure with the partitions.

More specifically, we define:

$$\mathfrak{G}^N := \langle (V_1, V_2, \cdots, V_n), (E_{1,2}, E_{1,3}, \cdots, E_{m,n}), (W_{E1,2}, \cdots, W_{Em,n}) \rangle \tag{1}$$

as a multi-partite graph 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 $W \in \mathbb{R}_{\geq 0}$ and each mode in the multi-partite network is represented as $G_{m,n} := \langle (V_m, V_n), E_{m,n}, W_{E_{m,n}} \rangle$. With this data structure we then employ Gower's Coefficient of Similarity (Gower 1971) to place the groups in a latent space, whereby we can create a latent network of the groups and assign groups to clusters based upon the multi-partite network. In the latent network, the edges maps the similarity between group i and j, calculated using Gower's Similarity Coefficient defined as:

$$S_{ij} = \frac{\sum_{k=1}^n w_{ijk} S_{ij}^{(k)}}{\sum_{k=1}^K w_{ijk}} \tag{2}$$
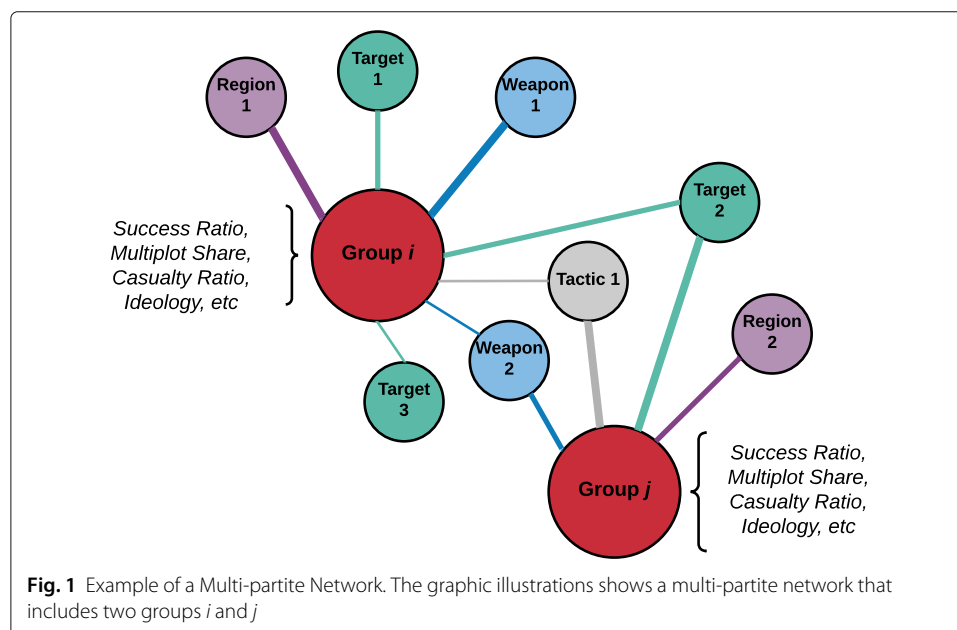
Fig. 1 Example of a Multi-partite Network. The graphic illustrations shows a multi-partite network that includes two groups *i* and *j*

where S_{ij} is the similarity between terrorist groups i and j on a variable (i.e. Targets, Weapons, etc.), k, and K is the total number of variables across all N modes, and w_{ijk} is the weight of the similarity between group i and group j for variable k. $S_{ij}^{(k)}$ is then dually defined as:

$$S_{ij}^{(k)} : \begin{cases} 1, \ if (x_{ik} = x_{jk}) \neq \emptyset \\ 0, \quad otherwise \end{cases} \tag{3}$$

if the variable, k, is categorical (to include binary) for node i and j's responses, x_{ik}, x_{jk}, and:

$$S_{ij}^{(k)} : \frac{|x_{ik} - x_{jk}|}{r_k} \tag{4}$$

where r_k is the range of x_k, if k is numerical. For each variable, k that is numerical the range is calculated as:

$$r^k = |max(x_k) - min(x_k)| \tag{5}$$

which means that the range is given by the absolute value of the maximum value of the variable k minus the minimum value of the variable k. Gower's coefficient of similarity provides a wide degree of flexibility as it can take various data types, like integer, binary, or continuous values and does so without the use of dummy variables. So, with this coefficient of similarity we are able to incorporate various means of describing terrorist groups, which can be of nearly any data type, and do so in such a way that keeps the original structure of each of the modes as bipartite networks intact.

Another advantage of Gower's Coefficient is the weighting term. As was noted at the beginning of this section, each of the possible variables used to find similarity between terrorist groups are not independent, but rather fall into various related modes. For example, if group i has operated in the Middle East, it is possible that it has also operated in North Africa or Southwest Asia. Furthermore, since the relationships within the mode are many to many (e.g. a terrorist groups can use many different weapons and vice-versa) each of these modes are a bipartite network. Thus, our data is modeled as a collection of bipartite networks, where there are relationships between the entities within each of the mode networks. So, to take advantage of this model of our data, we employed *network entropy* as the means to weight different modes for the weighting scheme in the Gower's Coefficient (Passerini and Severini 2008). Von Neumann's network entropy is a spectral measure originating from Gibbs entropy that has been applied to the quantum realm and that provides information on the complexity of a graph and on the amount of information that a network contains. In general, network entropy can be thought of as a measure of how heterogeneous a network is in terms of its connections (Passerini and Severini 2008; Feng et al. 2019). As such, network entropy has been used to characterize changes within dynamic graphs, as it is good as distinguishing different graph snapshots from each other (Ye et al. 2018). Furthermore, network entropy has also been used a means of distinguishing certain graphs from each other, with those graphs that have a higher entropy having more complex structures like subgroups (Passerini and Severini 2008; Ye et al. 2014). So, we similarly employ network entropy to distinguish between the different modes, which are bipartite networks, in such a way that those modes that have more heterogeneous structures — and as such are likely to be better for separating terrorist groups into clusters — are considered as more important. Since Gower's Coefficient allows for weighting

for exactly the purpose of emphasizing more important features in data, we can use network entropy with the weighting term in Gower's Coefficient to automatically emphasize more useful modes of our data. Following the derivations of the entropy of a network in Passerini and Severini (2008); Silva et al. (2015), we define the entropies of each of our modes as:

$$H^n = -\sum_{i=1}^{|V|} \frac{\tilde{\lambda}_i}{|V|} ln \frac{\tilde{\lambda}_i}{|V|} \tag{6}$$

where $\tilde{\lambda}_i$ are the eigenvalues of the normalized Laplacian of the graph of the particular mode. So, for $n \in N$ the normalized Laplacian is $\tilde{L}^n = D^{-\frac{1}{2}}(D-X^n)D^{-\frac{1}{2}}$, where $(D-X^n)$ is the unnormalized Laplacian and therefore X^n is the adjacency matrix from a particular mode and D is the degree matrix, which is created by:

$$D : \begin{cases} \sum_{j=1}^{V} X_{ij} & \text{if } D_{ii} \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

Following the findings on using network entropies to characterize heterogeneous graphs in Ye et al. (2018); Feng et al. (2019), we let those modes with higher entropy have more impact on the similarity measurements through Gower's coefficient. So, the weighting term in our Gower's coefficient is:

$$w_{ijk} = \sum_{n}^{N} \delta(k,n) \times H^n \tag{8}$$

where $\delta(k,n)$ is an indicator function that returns 1 if variable k is in mode n, and 0 otherwise. It should be noted that each variable within a mode will recieve the same weight. So, those modes which have a more heterogeneous structure, which should be better for producing structures like clusters, will have a higher weight in the comparison of the various terrorist groups.

Asymmetric kNN modularity graph construction

Having obtained pairwise similarities between all of the terrorist actors, we now move on to extracting a network from the data, which we refer to as the latent network. Following our work in Campedelli et al. (2019), we continue to use the kNN modularity maximization procedure proposed in Ruan (2009). At a high level, once similarities have been computed for each of the terrorist groups, the method iterates through various possible numbers of neighbors for each node, k, and selects that k which produces the most modular graph, relative to a null-model, random graph produced on the same similarities. Modularity in this case is the network modularity as described in Newman (2010):

$$mod(G) = \frac{1}{2m} \sum_{ij} [A_{ij} - \frac{deg(i) \times deg(j)}{2m} \delta(c_i, c_j)] \tag{9}$$

where m is the number of links in the network, and c are the cluster assignments of the nodes. A graph with high modularity is one which will have sub groups that have a lot of interconnections. Since it is known that random graphs can give rise to modular structures, we also compare this modularity value to the modularity value obtained from a random graph with the same number of vertices and edges, and the same similarities between the vertices (Ruan 2009). The general idea behind this method is that a

kNN that is higher in modularity, relative to a null-model, is better for detecting community structure in the underlying data used to make the network. It should be noted that this procedure applies only after a measure of similarity has been applied to the data. We have, however, modified the algorithm slightly to better suit our data. First, we use an asymetric kNN network. More precisely, for each point i, let $N_k(i)$ be the k nearest neighbors of i, then an asymmetric kNN network has links between two nodes i and j if $i \in N_k(j)$ OR $j \in N_k(i)$. Second, in the clustering step, we differ from the original algorithm proposed in Ruan (2009), as we use a faster method of modularity maximization of unimodal networks, the Louvain Method (Blondel et al. 2008), as opposed to the author's *QCut* algorithm. The psuedo-code of our implementation of network construction by kNN modularity maximization is detailed in Algorithm 1.

Algorithm 1 kNN Modularity Maximization Procedure

    input: Distance or Affinity Matrix, S, (n x n)

    output: Optimal k-Nearest Neighbor Network, G^*, and sub group assignments $C(G^*)$.

    for $i = 1 : \lfloor log_2(n) \rfloor$ **do**

        $k \leftarrow 2^i$

        $G_k \leftarrow kNN(S, k)$

        $C(G_k) \leftarrow Louvain(G_k)$

        $G_k^r \leftarrow randomize(G_k)$

        $C(G_k^r) \leftarrow Louvain(G_k^r)$

        $Modularity_k \leftarrow Modularity(C(G_k)) - Modularity(C(G_k^r))$

    end for

    $k^* \leftarrow argmax_k Modularity_k$

    $G^* \leftarrow kNN(s, k^*)$

    $C(G^*) \leftarrow Louvain(G^*)$

    return $C(G^*)$, G^*

In Algorithm 1, G_k is a particular kNN graph, where each vertex connects to exactly k of its nearest neighbors. The sub-step of $randomize(G_k)$ is to randomly re-wire all of the edges in G_k. This is equivalent to creating an Erdos-Renyi random graph that has the same number of edges and vertices as G_k. This step is performed in order to create a null-model of G_k, so that we can get a better idea of the strength of the modularity of the proposed G_k by comparing it to the modularity of its null counterpart, G_k^r. So, a good kNN should not just have high modularity, but also high modularity with respect to a randomized version of that kNN; the modular structures should not be just an artifact of the kNN's density or size. Finally, we return G^* which is that kNN which has the most modular structure. A Python implementation of the code will be available on the author's GitHub page with publication of this article.

Results

Comparing clustering assignments

To compare the similarity of clustering assignments of the three grouping procedures we have calculated the Adjusted Mutual Information (AMI), which is a modified version of

the ordinary mutual information adjusted for randomness, and it is calculated as

$$AMI(U, V) = \frac{MI(U, V) - E\{MI(U, V)\}}{max\{H(U), H(V)\} - E\{MI(U, V)\}} \tag{10}$$

where $E\{MI(U, V)\}$ is the expected mutual information between two random clusterings and $H(U)$ and $H(V)$ are the entropies associated to each partition U and V. This is a standard way to measure how different or similar are the outcomes of clustering procedures. The results are displayed in Fig. 2.

The results clearly show how, with respect to the baseline unweighted model, subgrouping using entropy weighting is far more similar than the ideology-based subgrouping. This, on one hand, suggests that relying only on this latter heuristic can extremely underestimate and distort the latent similarity that exist across groups when fully considering behavioral or operational variables. While ideology is certainly important for contextualizing a certain terrorist actor, the multi-partite original network includes information that are not captured by this method.

On the other hand, the entropy-based weighting is able to capture a relatively high portion of the information associated with the unweighted baseline model. While certainly introducing this data-driven discriminatory procedure affects the final clusters, our first conclusion is that this method is far more reliable if we want to preserve the original information structure of our data. Moreover, this algorithmic approach might provide a more solid tool to analysts and policymakers if they need to go beyond the original data, exploiting the richness of the original data itself.

Given these results, we now proceed to compare the baseline and the entropy-based clusterings more in depth, to understand whether, besides pure group assignments, they also share stable similarities in terms of behavioral and ideological features.

Unweighted vs entropy-based subgrouping: similarities and differences

Our algorithmic procedure to create the k-Nearest Neighbor network yielded two different graphs. Focusing on global characteristics of both networks, we can highlight how these graphs hold distinct structural and topological characteristics (Table 3).

The table above highlights how the k-NN procedure provided two different optimal k's for the two networks. The entropy-based network has a higher k, and this justifies the
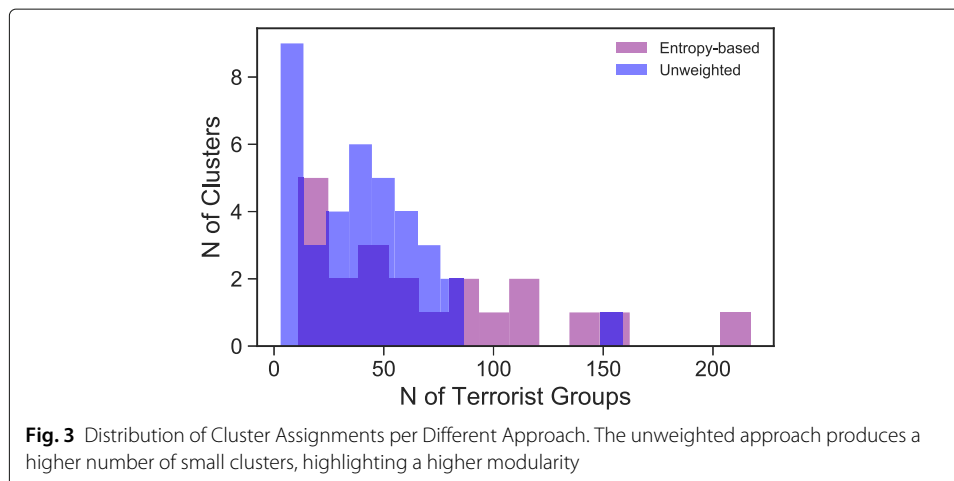

Fig. 2 Adjusted Mutual Information of three Subgrouping Schemes. The ideology-based heuristic approach produces subgroupings which are extremely different when compared with the baseline and the entropy-based approach
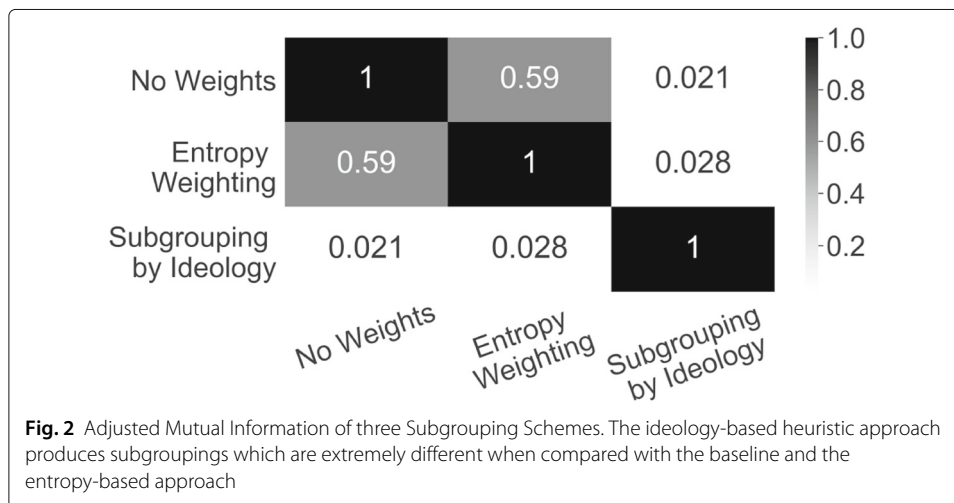
Table 3 Network structural and topological characteristics for both approaches

	Unweighted	Entropy-based
Best k	2	4
Modularity	0.805	0.390
N of Links	5,242	10,658
Bi-directional Link Count	2,621	5,329
Density	0.002	0.004
Clustering Coefficient	0.342	0.336
Betweenness Centralization	0.656	0.596
Eigenvector Centralization	0.826	0.703
Total Degree Centralization	0.106	0.257

higher number of links (both overall and bi-directional), and the higher network density. Nonetheless, the unweighted network proves to be higher in clustering coeficient, betweenness centralization and eigenvector centralization, while the entropy-based one yielded higher total degree centralization. With regard to modularity, the unweighted network notably performs a higher value, suggesting that the emerging clusters are more defined than the ones yielded by the entropy-based approach.

Connected to this aspect, and with regard to the actual subgroupings, is the fact that the entropy-based procedure produced less clusters (21) compared to the unweighted one (37). As Fig. 3 shows, the entropy-based approach produces a greater number of highly populated clusters, while in the unweighted case, a considerable amount of clusters includes a little number of groups (in fact, 25 clusters include less than 50 terrorist groups each). These figures further justify the different scores in term of modularity, since a higher number of smaller clusters is highly likely to indicate a higher degree of diversity in the network itself, as captured by modularity.

Focusing on the different node-level measure distributions, Fig. 4 displays histograms and 2D Kernel Density Estimations (KDE) of three selected metrics, namely Log Unscaled Total Degree Centrality, Log Betweenness Centrality and Clustering Coefficient. Total Degree Centrality and Betweenness Centrality have been transformed in log scale in order to provide more intuitive graphic results, since the original distributions are extremely left-skewed and the bivariate visualizations would have been extremely difficult to interpret.

Fig. 3 Distribution of Cluster Assignments per Different Approach. The unweighted approach produces a higher number of small clusters, highlighting a higher modularity

Fig. 4 Histograms (20 bins) and 2D Kernel Density Estimation of Log Unscaled Total Degree Centrality, Log Betweenness Centrality and Clustering Coefficient of Terrorist Groups across both approaches 2D Kernel Density Estimations are performed through regular grids made of 500 bins for each axis

Regarding Log Unscaled Total Degree Centrality, the histogram highlights how the groups in the Entropy-based approach generally have less connections that the ones in the Unweighted approach. The 2D KDE displays a strong concentration of data points in the bottom-left side of the graph, with more dense concentrations. With regards to Log Betweenness Centrality, the histogram shows relatively similar distributions for the considered approaches. The Unweighted one displays a greater number of groups with betweenness equal to zero ($\log(n) \approx -3$).[6] The 2D KDE displays a very high concentration of nodes on the top right of the plot, showing a positive correlation of log betweenness centrality across nodes for both approaches. However, it is worth noting that there is also an interesting small concentration of nodes that have very high value in the Entropy-based case but, conversely, very low ones in the Unweighted case. Conversely, in relation to clustering coefficient, more evident differences emerge when looking at the histogram. In fact, the Entropy-based approach shows a more concentrated distribution, while the Unweighted one highlights a very different behavior, with a considerably high number of extreme values, on both left and right side of the x-axis. However, these differences are mitigated in the KDE plot. Indeed, it shows the concentration of the majority

---

[6]When betweenness was equal to 0 we have transformed it to 0.001 in order to allow the value for log transformation. This transformation did not affect the results since it was performed only to provide intuitive visualizations and interpretable bivariate relations

of data points in the bottom left, almost indicating a linear relationship. In spite of this, it is wort noting that there exist a portion of groups which obtain very high levels of clustering coefficient in the Unweighted case, while their corresponding values in the Entropy-based approach are significantly lower. In light of the considerations on these detected differences in topology, structural and node-level measures of both networks and cluster formation, it is worth inspecting the types of groups in terms of operational and behavioral features and ideologies that are clustered together in both approaches.

The correlation results interestingly showcase that the majority of relations (either positive or negative) hold stably across both approaches, while only few have opposite directions from one approach to the other (Fig. 5). Notably, correlations on events (first columns of both plots) are generally very different. This may suggest that the raw number of events do not drive any consistent information flow regarding cluster assignments. This would indicate that there are other types of features that actually capture similarities or differences across terror groups, and that the latent data structure is independent from the individual frequency of attacks of actors. In terms of stable results, both approaches demonstrate how clusters with a high percentage of islamist or jihadist groups are associated with high levels of attack success, while this type of relation goes in the opposite direction for all the other ideologies. This marks a distinctive feature of jihadism or islamism as terror motivation. Expectedly, suicide attacks are also found to be positively correlated with jihadist or islamist ideology. Furthermore, Islamist/jihadist ideology is again the only one positively associated with high levels of both casualties and fatalities in both approaches, while other ideologies seem to be less lethal. In terms of the multiplot, which captures the extent to which a terrorist group is able to plot multiple coordinated attacks in the same day as part of a more complex logistic structure, FL groups, along with religious (non-Islamist) groups, are the only ideologies to display a stable, positive relation.
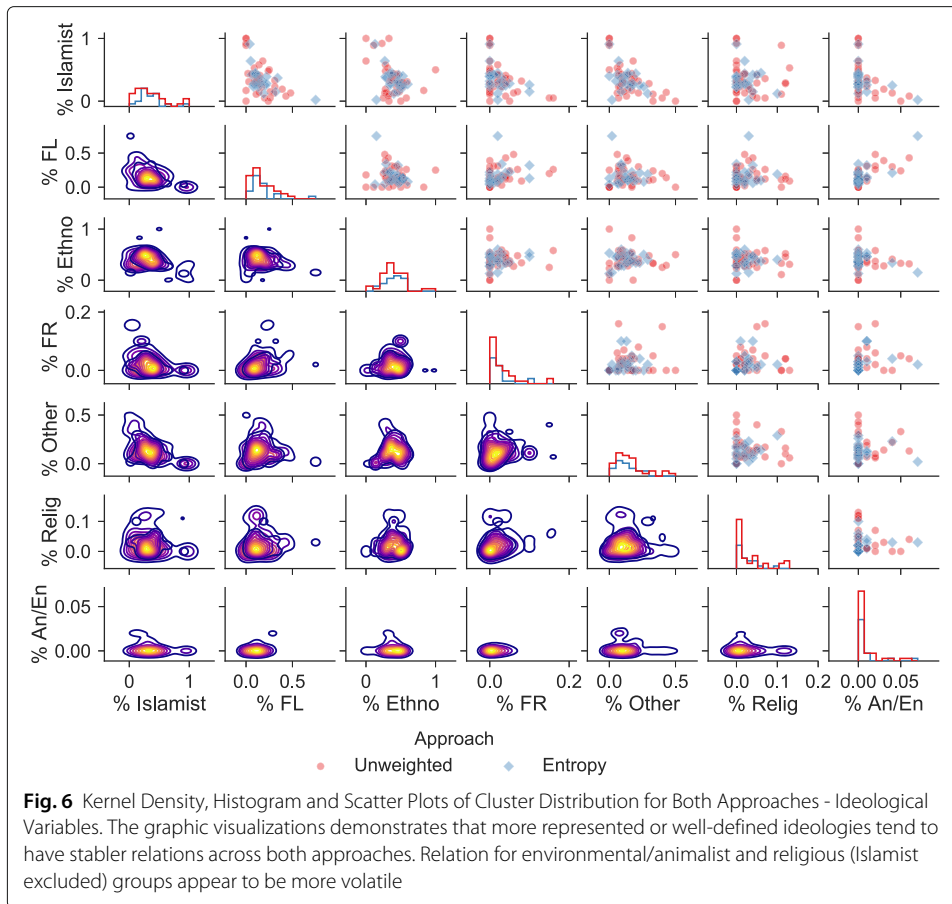
Clear results emerge when focusing on pairwise relations between ideologies (Fig. 6). Overall, the majority of relations are stable across both algorithmic approaches. Islamism has negative correlations in both cases with all the other ideologies, implying that groups belonging to or motivated by this ideology represent very distinguished entities in the global terrorist scenario. With regard to FL groups, they strongly share similar cluster assignments with enviromentalist and animalist groups and, surprisingly, they share similar assignments also with FR actors. This result suggest that, while these two ideologies are considered very different from one another and the motivations of groups belonging to these factions are extremely distant, from the operational point of view (namely, from the standpoint of employed weapons, hit targets, applied tactics and targeted regions) FL and FR terrorist groups are quite similar, and this result is corroborated by its stability across the two approaches. FL groups are not the only actors that share cluster assignments with FR terrorists. In fact, the analysis show that the higher the fraction of FR groups, the higher the number of religious (Islam excluded) and ethno/nationalist actors.

These type of relations might be expected, considering that many FR groups include elements of bigotry or radical religious views and that many nationalist actors generally rely upon fascist or far-right discourse and political behaviours. However, another surprising result is given by the negative relation between FL and ethno/nationalist groups, considering two factors: first, the positive stable correlation between FR and FL groups; second, the fact that radical leftist or communist ideologies are generally the other
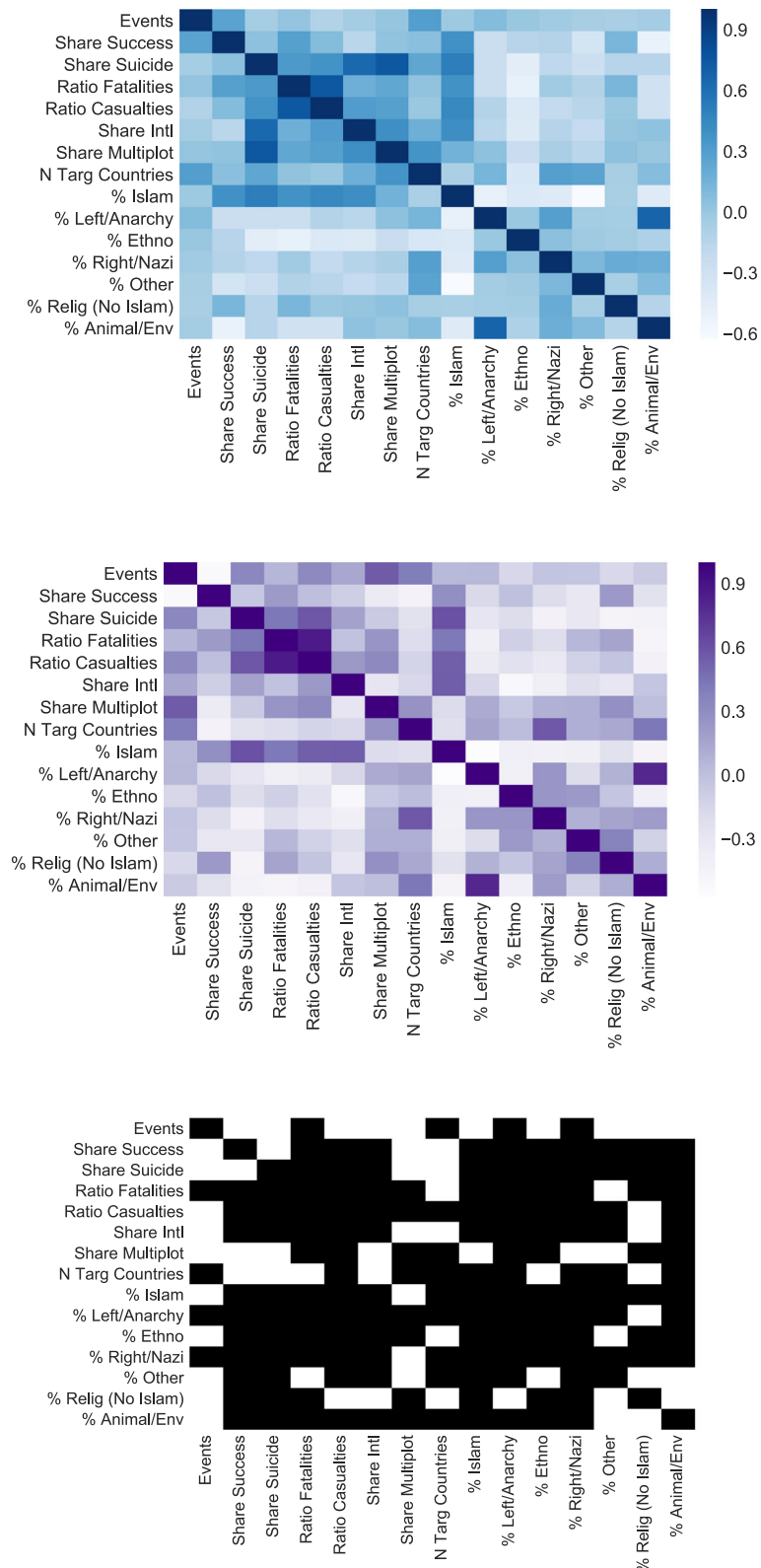
Fig. 5 Correlation among behavioral and ideological variables Unweighted (left) and Entropy-based (right) approaches. Correlation shows what bivariate relations stay stable across both approaches (namely, they maintain the same direction). Bottom plot highlights (in black) stable relations against non-stable (white)

Fig. 6 Kernel Density, Histogram and Scatter Plots of Cluster Distribution for Both Approaches - Ideological Variables. The graphic visualizations demonstrates that more represented or well-defined ideologies tend to have stabler relations across both approaches. Relation for environmental/animalist and religious (Islamist excluded) groups appear to be more volatile

opposite driving force of certain nationalist or independence-driven actors, such as the Euskadi Ta Askatasuna (ETA) in Spain. This means that, besides the potentially similar motivations and background, these two types of groups act and organize attacks that are generally dissimilar.

Potential directions: introducing a dynamic entropy-based approach

The comparative analysis across the two approaches demonstrated that the entropy-based approach preserves a relatively high amount of outcomes derived from the baseline model, also shedding light on additional mechanisms when focusing on the *ex*-post analysis of behavioral and ideological features. However, our experiment comes with a limitation that should be adressed in the future. In fact, our original multi-partite network spans across twenty years and includes the cross-sectional information on the whole universe of active groups without taking into account time in a dynamic fashion. Terrorism has undergone several changes in the last two decades: many new groups have appeared only recently, many have disappeared or have been dismantled, other actors have been active only for very short period. In general, over the years, the trend of active actors has not shown a stable behavior. As it is expected, this type of trend also regards the number of attacks and terrorist events (Fig. 7).

These changes and trends over the last twenty years may also be related to strategies and types of attack, besides mere frequency of attacks. This is an aspect which intersects

Fig. 7 Number of Active Groups and Terror Attacks per Year (1997-2016)
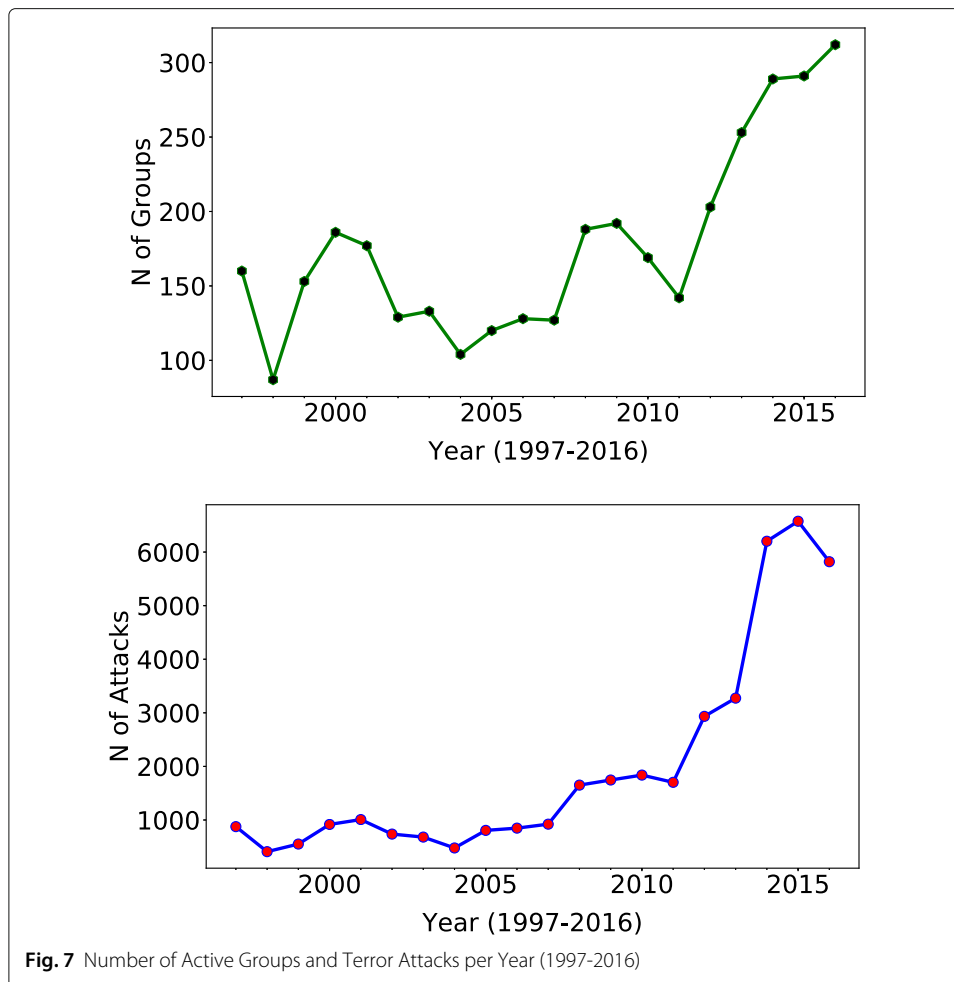
our analysis. Since our original multi-partite network stores all the information of the past twenty years, these changes and trends may be underestimated or even vanish altogether in our algorithmic procedure. As for the current data structure, groups that have plotted very few attacks in two distant years may be clustered together considering their similarity in the data space, however, it might be not useful for analysts or policy-makers to compare two groups that are too distant in time. For this reason, our intuition is to introduce a potential solution to this issue via a dynamic Entropy-based approach. Instead of constructing a single static multi-partite network, we build yearly multi-partite graphs to capture the variations in the entropy of each mode (Fig. 8).

As shown by the plot above, there exists clear variations in the entropies of each mode from which we have built our multi-partite network. The plot highlights how the Weapon mode generally follows completely different trends with respect to all the others. At the same time, the entropies of Tactic and Target modes display very similar behavior in the last four years of our considered timespan. Notably, there is an almost complete similarity of entropies for Region, Tactic and Target modes with respect to year 2001. This graph clearly suggests that time should be considered in the algorithmic procedure, since it is highly likely that sensible changes in the subgrouping outcomes will emerge. Besides the relevance of embedding temporal dynamics from the purely research point of view, this

Fig. 8 Entropy variation for each mode (1997-2016)
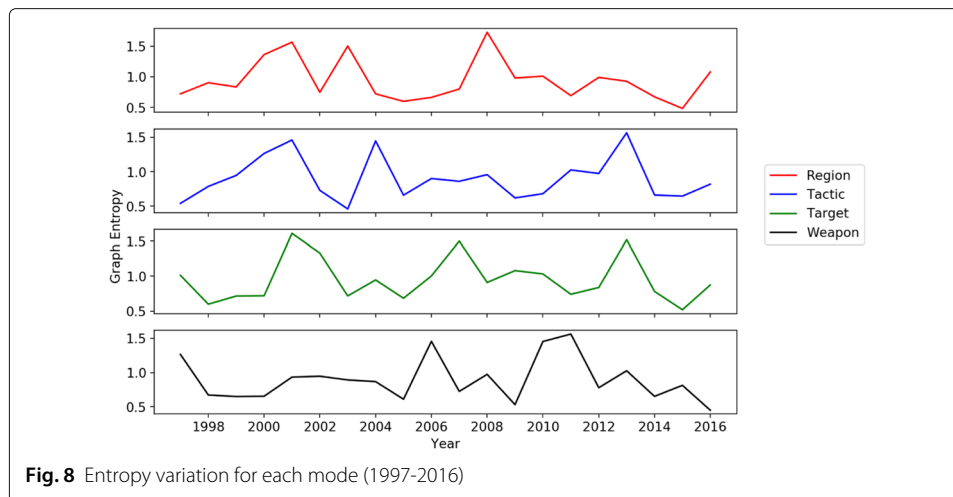
furthermore provides a richer tool for potential users and analysts interested in using the algorithm to detect, assess and study patterns in the data. In fact, since our model seeks to provide a practical framework that can be easily deployed for the aforementioned purposes, we feel that time-aware results are able to exclude all the non-active groups in a given year and would increase the usability of previous years information and its efficiency for real-time objectives oriented to intelligence profiling.

Discussion & future work

This work has presented a novel algorithmic framework for detecting latent clusters of similar terrorist groups via a complex network approach. We have created a multipartite network for the entire known population of terror groups active worldwide from 1997 to 2016, where modes were Weapons, Tactics, Targets, operating Regions, and proposed a novel clustering architecture expanded from (Campedelli et al. 2019). We have then compared our new entropy-based architecture with two alternative solutions: a weak-heuristic approach based on terror group clustering by ideology and our baseline unweighted approach. The entropy-based approach modifies the baseline approach simply weighting each mode by its graph entropy, in order to provide a data-driven approach that takes into account the relevance of a certain mode with respect to the others. The analysis has first demonstrated that subgrouping by ideology leads to cluster assignments very different to the ones obtained with our baseline method, where we let patterns emerge naturally from data with no a priori knowledge and, secondly, that the entropy-based and the baseline approaches have similar results both in terms of stability of cluster assignment for terrorist groups and behavioral and ideological intra-cluster association. Both approaches corroborated interesting findings that go beyond the pure methodological intent of this work. To investigate the meta-connections between groups resulting from our work, we have analyzed behavioral characteristics (e.g., share of successful attacks, international propensity, etc.) and we have also focused on the ideological background of each actor, retrieving this information from BAAD version 1 and 2 and other open access qualitative sources. Though labelling a group under few ideological categories may oversimplify certain complex components of terrorism, interesting relations emerged. Besides several expected patterns (e.g. Islamist/jihadist groups tend not

to be associated with groups belonging to other ideologies), the algorithm reveals other results that may shed light on terrorism in terms of research and policy. The clustering procedures highlighted a certain similarity between FL groups and FR groups, indicating that besides their divergent objectives and goals, these two types of groups share similar behaviors. Furthermore, FL groups on one side are often associated with animal-rights and environmentalist actors, suggesting that some overlapping in terms of motives and aims is also connected to similar methods and ways of acting. On the other side, FR groups tend to be clustered together with ethno/nationalist and religious groups, as it might be expected given that many FR groups hold nationalist or religion-related elements.

Overall, the entropy-based approach is a flexible tool for capturing the intrinsic and hidden knowledge included in the manifold via a data-driven procedure, rather than using subjective knowledge and weaker heuristics, which was one of the limitations of several experiments conducted in (Campedelli et al. 2019). In spite of the aforementioned results, our approach may suffer from the fact that the original multi-partite network does not take time into account. Indeed, the manifold includes the whole set of available data from 1997 and 2016. This might be interpreted as a limitation, especially considering that our work is inherently policy-oriented. The last twenty years have been susceptible of several dramatic changes in the ways terrorism manifests itself at global scale. On one side, they have seen the rise of islamist and jihadist terrorism not only in Africa and Middle East, but also in Western and Eastern Europe and countries of the North America. On the other side, politically motivated terrorism has showed shifts and different concentration over time and space. Furthermore, the considered time span is relatively long and therefore includes groups that may have been already disappeared and dismantled or even actors that have plotted one or very few single attacks, therefore constituting a sort of "noise" in the whole scenario. In light of this, we have opened the path for future work showing that, besides variations in the trends of active groups and actors and plotted attacks, there exist also significant variations in the entropies of each mode over time. Entropies change sensibly over-time and we have highlighted the presence of some similarities in these trends across certain modes (e.g. Tactic and Target), while others follow completely different behaviors (e.g. Weapon). For these reasons, future work should test the entropy-based setup within a dynamic framework. While considering all the groups and being able of compare still active groups with those that have been already dismantled or disrupted is certainly useful, we feel that controlling for the noise in the manifold and including only groups that are still part of the global terrorist scenario will provide more insights and will help policy-makers or analysts in understanding to what extent certain groups are similar or different compared to others. Additionally, leveraging upon the entropy-based structure automatically allows one to take into account the most relevant sources of information in real-time (e.g. modes): this type of setup, for instance, would be capable of highlighting anomalous behaviours or strategical behavioral evolution of certain terror groups. Future work will also seek to eventually exclude groups that are not strictly considerable as terrorists (e.g., Mexican drug cartels) although their actions are of terrorist nature: this operation would reduce potential noise and distortion of the results.

Finally, inherent limitations come from the data. While the GTD is certainly the most reliable and solid open access dataset freely available for research-purposes on terrorist events, its structure poses issues of missing data and level of detail of the information. Despite the fact that, as opposed to other criminal phenomena, terrorist attacks tracked

and recorded do not face the risk of underestimation (generally, every terrorist attack is reported by newspapers or media agencies), not all details on the attacks might be consistently retrieved and included in the dataset. This would therefore lead to a certain degree of bias or missing information regarding event characteristics, which are the core of our work. Another linked type of limitation is the risk of too generic information, especially for terror attacks occurred outside Europe and North America (which are actually the majority of the events). While our algorithmic framework has demonstrated a certain degree of potential using the GTD, the intent is to test it on more detailed databases in the future. Additionally, in our algorithmic framework, we do not consider any correlation between the different modes of the data. More specifically, it is possible that certain groups use certain weapons or tactics because of limited availability of alternative means and not, instead, as the product of a free choice. Unfortunately, we are not able to assess whether this is the case for the groups under analysis, but this potential explanation shall be kept in mind. Additionally, the ideology labelling process, though based on a scientifically recognized dataset, may oversimplify certain characteristics and motivations behind each group's actions. Reducing the complexity of the causes and motives behind the decision to resort to terrorism is challenging and attention should be payed not to provide distorted or biased interpretation of the results.

Author details
[1]Transcrime - Università Cattolica del Sacro Cuore, L.go Gemelli, 1, Milan, Italy . [2]School of Computer Science - Carnegie Mellon University, 5000 Forbes Avenue, 15213 Pittsburgh, PA, USA .

References

Agreste S, Catanese S, De Meo P, Ferrara E, Fiumara G (2016) Network structure and resilience of Mafia syndicates. Inf Sci 351:30–47. https://linkinghub.elsevier.com/retrieve/pii/S0020025516300925

Asal VH, Rethemeyer RK (2015) Big Allied and Dangerous Dataset Version 2. http://www.start.umd.edu/baad/database. Accessed 17 June 2019

Barabási A, Jeong H, Néda Z, Ravasz E, Schubert A, Vicsek T (2002) Evolution of the social network of scientific collaborations. Physica A: Stat Mech Appl 311(3-4):590–614. http://linkinghub.elsevier.com/retrieve/pii/S0378437102007367

Belli R, Freilich JD, Chermak SM, Boyd KA (2015) Exploring the crime-terror nexus in the United States: a social network analysis of a Hezbollah network involved in trade diversion. Dynamics of Asymmetric Conflict 8(3):263–281

Benigni MC, Joseph K, Carley KM (2017) Online extremism and the communities that sustain it: Detecting the ISIS supporting community on Twitter. PLOS ONE 12(12) e0181:405

Berlusconi G, Calderoni F, Parolini N, Verani M, Piccardi C (2016) Link Prediction in Criminal Networks: A Tool for Criminal Intelligence Analysis. PLOS ONE 11(4) e0154:244. https://dx.plos.org/10.1371/journal.pone.0154244

Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. J Stat Mech: Theory Exp 2008(10) P10:008

Borgatti SP, Mehra A, Brass DJ, Labianca G (2009) Network Analysis in the Social Sciences. Science 323(5916):892–895. https://science.sciencemag.org/content/323/5916/892

Bouchard M, Joffres K, Frank R (2014) Preliminary Analytical Considerations in Designing a Terrorism and Extremism Online Network Extractor(Mago VK, Dabbaghian V, eds.), Vol. 53. Springer International Publishing, Cham. http://link.springer.com/10.1007/978-3-319-01285-8_11

Brams SJ, Mutlu H, Ramirez SL (2006) Influence in Terrorist Networks: From Undirected to Directed Graphs. Stud Confl & Terrorism 29(7):703–718. http://www.tandfonline.com/doi/abs/10.1080/10576100600701982

Bx LI, Zhu JF, Wang SG (2015) Networks model of the East Turkistan terrorism. Phys A: Stat Mech Appl 419:479–486. https://linkinghub.elsevier.com/retrieve/pii/S0378437114008607

Caccioli F, Barucca P, Kobayashi T (2018) Network models of financial systemic risk: a review. J Comput Soc Sci 1(1):81–114. http://link.springer.com/10.1007/s42001-017-0008-3

Calderoni F, Brunetto D, Piccardi C (2017) Communities in criminal networks: A case study. Soc Netw 48:116–125. https://linkinghub.elsevier.com/retrieve/pii/S0378873316300363

Campedelli GM, Bartulovic M, Carley KM (2019) Pairwise similarity of jihadist groups in target and weapon transitions. J Comput Soc Sci. https://doi.org/10.1007/s42001-019-00046-8. Accessed 17 June 2019

Campedelli GM, Cruickshank I, Carley KM (2018) Complex Networks for Terrorist Target Prediction (Thomson R, Dancy C, Hyder A, Bisgin H, eds.), Vol. 10899. Springer International Publishing, Cham

Campedelli, GM, Cruickshank I, Carley KM (2019) Detecting Latent Terrorist Communities Testing a Gower's Similarity-Based Clustering Algorithm for Multi-partite Networks(Aiello LM, Cherifi C, Cherifi H, Lambiotte R, Lió P, Rocha LM, eds.), Vol. 812. Springer International Publishing, Cham. http://link.springer.com/10.1007/978-3-030-05411-3_24

Carley KM (2002) Computational organization science: A new frontier. Proc Natl Acad Sci 99(Supplement 3):7257–7262. http://www.pnas.org/cgi/doi/10.1073/pnas.082080599

Carley KM (2006) Destabilization of covert networks. Comput Math Org Theory 12(1):51–66

Centola D (2010) The Spread of Behavior in an Online Social Network Experiment. Science 329(5996):1194–1197. http://www.sciencemag.org/cgi/doi/10.1126/science.1185231

Centola D (2018) How Behavior Spreads: The Science of Complex Contagions. Princeton University Press, Princeton

Chatfield AT, Reddick CG, Brajawidagda U (2015) Tweeting propaganda, radicalization and recruitment: Islamic state supporters multi-sided twitter networks. In: Proceedings of the 16th Annual International Conference on Digital Government Research - dg.o '15. ACM Press, Phoenix, Arizona. pp 239–249. http://dl.acm.org/citation.cfm?doid=2X00000.

Chenoweth E, Lowham E (2007) On Classifying Terrorism: A Potential Contribution of Cluster Analysis for Academics and Policy-makers. Def Secur Anal 23(4):345–357

Cranmer SJ, Menninga EJ, Mucha PJ (2015) Kantian fractionalization predicts the conflict propensity of the international system. Proc Natl Acad Sci 112(38):11,812–11,816. http://www.pnas.org/lookup/doi/10.1073/pnas.1509423112

da Cunha BR, Gonçalves S (2018) Topology, robustness, and structural controllability of the Brazilian Federal Police criminal intelligence network. Appl Netw Sci 3(1):36. https://doi.org/10.1007/s41109-018-0092-1

Desmarais BA, Cranmer SJ (2013) Forecasting the locational dynamics of transnational terrorism: a network analytic approach. Secur Inf 2(1):8

Eiselt H (2018) Destabilization of terrorist networks. Chaos, Solitons & Fractals 108:111–118. https://linkinghub.elsevier.com/retrieve/pii/S0960077918300183

Feng X, Wei W, Zhang R, Wang J, Shi Y, Zheng Z (2019) Exploring the heterogeneity for node importance by von neumann entropy. Phys A: Stat Mech Appl 517:53–65. http://www.sciencedirect.com/science/article/pii/S0378437118314274

Gerber ER, Henry AD, Lubell M (2013) Political Homophily and Collaboration in Regional Planning Networks: POLITICAL HOMOPHILY. Am J Polit Sci 57(3):598–610. http://doi.wiley.com/10.1111/ajps.12011

Gower JC (1971) A General Coefficient of Similarity and Some of Its Properties. Biometrics 27(4):857–71

Granovetter M (2018) The Sociology of Economic Life. 3rd. Routledge. https://www.taylorfrancis.com/books/9780429494338. Accessed 17 June 2019

Gutfraind A (2010) Optimizing Topological Cascade Resilience Based on the Structure of Terrorist Networks. PLoS ONE 5(11):e13448. https://dx.plos.org/10.1371/journal.pone.0013448

Hausmann R, Hidalgo CA (2011) The network structure of economic output. J Econ Growth 16(4):309–342. https://doi.org/10.1007/s10887-011-9071-4

Hays JC, Kachi A, Franzese RJ (2010) A spatial model incorporating dynamic, endogenous network interdependence: A political science application. Stat Methodol 7(3):406–428. https://linkinghub.elsevier.com/retrieve/pii/S1572312709000975

Keuschnigg M, Lovsjö N, Hedström P (2018) Analytical sociology and computational social science. J Comput Soc Sci 1(1):3–14. http://link.springer.com/10.1007/s42001-017-0006-5

Klausen J (2015) Tweeting the Jihad : Social Media Networks of Western Foreign Fighters in Syria and Iraq. Stud Confl Terrorism 38(1):1–22

Koschade S (2006) A Social Network Analysis of Jemaah Islamiyah: The Applications to Counter-terrorism and Intelligence. Stud Confl Terrorism 29(6):559–575

Krebs V (2002) Mapping Networks of Terrorist Cells. Connections 24(3):43–52

LaFree G, Dugan L (2007) Introducing the Global Terrorism Database. Terrorism Polit Violence 19(2):181–204

Lautenschlager J, Ruvinsky A, Warfield I, Kettler B (2015) Group Profiling Automation for Crime and Terrorism (GPACT). Proc Manuf 3:3933–3940

Lindelauf R, Borm P, Hamers H (2011) Understanding Terrorist Network Topologies and Their Resilience Against Disruption(Wiil UK, ed.). Springer Vienna, Vienna. http://link.springer.com/10.1007/978-3-7091-0388-3_5

Medina RM, Hepner GF (2011) Advancing the Understanding of Sociospatial Dependencies in Terrorist Networks: Sociospatial Dependencies in Terrorist Networks. Trans GIS 15(5):577–597. http://doi.wiley.com/10.1111/j.1467-9671.2011.001281.x

Moon IC, Carley KM (2007) Modeling and Simulating Terrorist Networks in Social and Geospatial Dimensions. IEEE Intell Syst 22(5):40–49

Morselli C (2009) Inside Criminal Networks. Springer-Verlag, New York. https://www.springer.com/gp/book/9780387095257

National consortium for the study of terrorism and responses to terrorism (2016) Global Terrorism Database (Data file). https://www.start.umd.edu/gtd. Accessed 17 June 2019

Newman MEJ (2010) Networks: An Introduction. Oxford University Press, New York

Papachristos AV, Braga AA, Piza E, Grossman LS (2015) The Company You Keep? The Spillover Effects Of Gang Membership On Individual Gunshot Victimization In A Co-Offending Network: Gang Membership, Networks, & Victimization. Criminology 53(4):624–649. http://doi.wiley.com/10.1111/1745-9125.12091

Passerini F, Severini S (2008) The von Neumann entropy of networks. arXiv e-prints arXiv:0812.2597

Qi X, Christensen K, Duval R, Fuller E, Spahiu A, Wu Q, Zhang CQ (2010) A Hierarchical Algorithm for Clustering Extremist Web Pages. In: 2010 International Conference on Advances in Social Networks Analysis and Mining. IEEE, Odense. pp 458–463

Ren XL, Gleinig N, Helbing D, Antulov-Fantulin N (2019) Generalized network dismantling. Proc Natl Acad Sci 116(14):6554–6559. http://www.pnas.org/lookup/doi/10.1073/pnas.1806108116

Ribeiro HV, Alves LGA, Martins AF, Lenzi EK, Perc M (2018) The dynamical structure of political corruption networks. J Complex Netw 6(6):989–1003. https://academic.oup.com/comnet/article/6/6/989/4823561

Ruan J (2009) A Fully Automated Method for Discovering Community Structures in High Dimensional Data. In: 2009 IEEE International Conference on Data Mining (ICDM). pp 968–973

Sageman M (2014) The Stagnation in Terrorism Research. Terrorism Polit Violence 26(4):565–580. https://dx.doi.org/10.1080/09546553.2014.895649

Schuurman B (2018) Research on Terrorism, 2007–2016: A Review of Data, Methods, and Authorship. Terrorism Polit Violence:1–16. https://www.tandfonline.com/doi/full/10.1080/09546553.2018.1439023. Accessed 17 June 2019

Schweitzer F, Fagiolo G, Sornette D, Vega-Redondo F, Vespignani A, White DR (2009) Economic Networks: The New Challenges. Science 325(5939):422–425. https://www.sciencemag.org/lookup/doi/10.1126/science.1173644

Silva FN, Comin CH, Peron TKD, Rodrigues FA, Ye C, Wilson RC, Hancock E, Costa LDF (2015) Modular Dynamics of Financial Market Networks. arXiv e-prints arXiv:1501.05040

START (2017) GTD Codebook: Inclusion Criteria and Variables. Tech. rep. University of Maryland. College Park, MD

Szell M, Lambiotte R, Thurner S (2010) Multirelational organization of large-scale social networks in an online world. Proc Natl Acad Sci 107(31):13,636–13,641. https://www.pnas.org/cgi/doi/10.1073/pnas.1004008107

Tutun S, Khasawneh MT, Zhuang J (2017) New framework that uses patterns and relations to understand terrorist behaviors. Expert Syst Appl 78:358–375. https://linkinghub.elsevier.com/retrieve/pii/S0957417417301161

Vitali S, Glattfelder JB, Battiston S (2011) The Network of Global Corporate Control. PLoS ONE 6(10):e25995. https://dx.plos.org/10.1371/journal.pone.0025995

Ward MD, Stovel K, Sacks A (2011) Network Analysis and Political Science. Ann Rev Polit Sci 14(1):245–264. https://www.annualreviews.org/doi/10.1146/annurev.polisci.12.040907.115949

Ye C, Wilson RC, Comin CH, Costa LdF, Hancock ER (2014) Approximate von neumann entropy for directed graphs. Phys Rev E 89:052–804. https://link.aps.org/doi/10.1103/PhysRevE.89.052804

Ye C, Wilson RC, Rossi L, Torsello A, Hancock ER (2018) Thermodynamic analysis of time evolving networks. Entropy 20(10). https://www.mdpi.com/1099-4300/20/10/759

Publisher's Note