# Detecting Latent Terrorist Communities Testing a Gower's Similarity-based Clustering Algorithm for Multi-Partite Networks

Gian Maria Campedelli<sup>1,2</sup>, Iain Cruickshank<sup>2</sup>, and Kathleen M. Carley<sup>2</sup>

<sup>1</sup> Università Cattolica del Sacro Cuore, L.go Gemelli 1, 20123, Milan, Italy gianmaria.campedelli@unicatt.it

<sup>2</sup> Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA, 15213, USA. icruicks@andrew.cmu.edu, kathleen.carley@cs.cmu.edu

Abstract. Finding hidden patterns represents a key task in terrorism research. In light of this, the present work seeks to test an innovative clustering algorithm designed for multi-partite networks to find communities of terrorist groups active worldwide from 1997 to 2016. This algorithm uses Gower's coefficient of similarity as the similarity measure to cluster perpetrators. Data include information on weapons, tactics, targets, and active regions. We show how different dimensional weighting schemes lead to different types of grouping, and we therefore concentrate on the outcomes of the unweighted algorithm to highlight interesting patterns naturally emerging from the data. We highlight that groups belonging to different ideologies actually share very common behaviors. Finally, future work directions are discussed.

**Keywords:** multi-partite networks, unsupervised learning, community detection, terrorism

# 1 Introduction

In recent years, a slow but constant shift towards multidisciplinary dialogue between academic fields (e.g. criminology, political science, sociology, statistics, computer science) has directed the attention of the scientific community towards terrorism as a quantitatively measurable social phenomenon. Data on terrorism often involves different types of information: events, organizations, perpetrators, and tactics to name the most prominent. Since terrorism is a multifaceted and extremely complex issue, each of these different dimensions helped and may still help in understanding specific dynamics. Despite the progress made, Sageman [20] highlighted how terrorism research faces a stagnation which is mainly caused by the still limited availability of primary source information that are kept private by governments, leading to speculations with little empirical grounding in academia. Additionally, Sageman claims that this lack of data has not only affected the solidity of academic results only but, also, the results achieved by the intelligence community which has the data but lacks methodological rigor.

#### 2 Campedelli et al.

Although Sageman points out a relevant factual problem, our belief is that still a lot can be done in the effort of making open access data meaningful. In light of this, this work seeks to employ public event data on terrorism events from 1997 to 2016 to derive dynamic meta-networks on terrorist groups and subsequently test the performance of a new algorithm for community detection in multi-partite networks using Gower's Similarity Coefficient. The aims are to analyze how different weighting specifications of the algorithm affect the latent communities<sup>3</sup> and understand whether, besides ideology, purely operational and behavioral patterns can shed light on terrorist groups, highlighting hidden communities that may include actors which are actually considered distant or separate those actors that are usually seen as very similar. Furthermore, community detection can provide insights that may be relevant for policy strategies: if clear and distinct profiles exist, then different counter strategies should be deployed. The article is organized as follows: in the next section it presents related work on both network analysis and terrorism and previous clustering approaches to the phenomenon. In the third section, it will describe the data source and structure that will be employed in our analysis. The methodology section will then explain the structure of the Gower's Similarity Coefficient-based algorithm for finding communities in multi-partite networks. Results will then be presented, highlighting relevant findings. In the discussion section, results are reviewed with a particular focus on future work directions.

## 2 Related Work

The availability of open access data (combined with the diffusion of statistical software or data science oriented programming languages) has contributed to a change of perspective towards quantitative and computational approaches to the study of terrorism. From an applied network science standpoint, network analysis for the study of terrorism has been employed in several specific subdomains. A classic application aims at understanding and highlighting internal dynamics and roles within terrorist groups [1, 11, 12]. Shifting from the relational information gathered and structured to investigate roles and key players, scholars have also tested and simulated the strength and resilience of terrorist networks. Some of the works in these subdomains relied on mathematics to reveal network topologies and possible effective strategies to destabilize terrorist networks [5]. Furthermore, researchers integrated spatial and temporal dynamics to simulate the evolution of these networks, relying on the conceptual assumption that time and space are relevant features when aiming to analyze the phenomenon from the evolutionary standpoint [16,17]. With the explosion of social media, the attention of researchers has been attracted to the possible consequences of criminal behavior in cyberspace. Indeed, a recent stream of research has started to focus on detecting terrorist and radical behaviors on these social media platforms. such as Twitter [2, 10]. Finally, in the fourth and last subdomain, researchers are trying to use event data to reconstruct multi-mode or multiplex networks in

<sup>&</sup>lt;sup>3</sup> Throughout the paper, communities and clusters are used as synonyms.

order to predict future terrorist attacks, locations, and tactics. This is the most recent and underdeveloped field in which network science is being applied to terrorism [4,7]. In another line of research, cluster analysis has not been extensively applied to the analysis of terrorism. In one of the first attempts at using cluster analysis to group terrorist organizations, Chenoweth and Lowham [6] used data on groups which targeted American citizens to explore alternative ways to conceive terrorist typologies. Qi et al. [18] 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 [14] 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.

### 3 Data

The data used in this work come from the Global Terrorism Database (henceforth GTD) [13]. GTD includes information on terrorist attacks from 1970 to 2016. Information on worldwide attacks is retrieved from open sources and each event is required to meet certain criteria to be included in the dataset and labelled as terrorist. Additionally, events which meet these criteria but have uncertainty as to whether they should be considered terrorist events are included in the dataset but mapped with the "doubter" variable. For our analysis, we used data from 1997 to 2016 on worldwide events (and related perpetrators), excluding all the attacks which were of doubtful terrorist nature. This methodological choice led from 106,114 events to a total of 88,513, and was the only pre-processing of the data performed. The meta-networks which have been created and employed for our study relied on six main terrorist dimensions, namely: Events (N=88,513), Groups (N=1,494), Targets (N=22), Weapons (N=13), Tactics (N=9) and operating Regions (N=9). Since a terrorist group can attack many different targets, use many different weapons, and operate in many different regions, and vice-versa, these data naturally form many-to-many relationships, and can therefore be easily modeled as networks. In addition to this information which represent the basis of this work, other variables extracted from the GTD will be 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, and by exception from other qualitative open access information sources. Finally, seven ideology categories were created: (i) Islamic/Jihadist groups, (ii) Left Wing/Anarchist, (iii) Right Wing/Racist, (iv) Ethno-Nationalist, (v) Other/Unknown, (vi) Religious (Islam excluded), (vii)

#### 4 Campedelli et al.

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). The success rate is given by the ratio between the successful attacks and the total number of events attributed to a given group. The suicide rate maps the ratio of suicide attacks over the total number of events plotted by the same group. Fatality and casualty rates are the ratios of attacks with at least one dead victim (fatality) or one wounded victim (casualty) person out of the total number of events. Finally, the international rate is simply the ratio between attacks with some international features (e.g. logistic organization) and the total number of attacks. All these variables seek to enrich the knowledge associated to each group and to understand whether the identified clusters highlight certain unexpected behaviors.

|                    | Mean  | St.Dev. | Median | Min  | Max    |
|--------------------|-------|---------|--------|------|--------|
| Events             | 60.24 | 1271.77 | 2.00   | 1.00 | 48,537 |
| Success Rate       | 0.90  | 0.24    | 1.00   | 0.00 | 1.00   |
| Suicide Rate       | 0.03  | 0.14    | 0.00   | 0.00 | 1.00   |
| Fatality Rate      | 2.75  | 8.55    | 0.50   | 0.00 | 170    |
| Casualty Rate      | 7.88  | 23.34   | 1.67   | 0.00 | 385.29 |
| Multiplot Rate     | 0.14  | 0.28    | 0.00   | 0.00 | 1.00   |
| International Rate | 0.29  | 0.41    | 0.00   | 0.00 | 1.00   |
| Targeted Countries | 1.46  | 4.57    | 1.00   | 1.00 | 163    |

Table 1. Group-based Attributes on Terrorist Activity - Descriptive Statistics

# 4 Methodology

Since the variables of Targets, Weapons, Tactics, and Regions form a manyto-many relationship with Groups, we first model this data as a multi-partite network with each partition joined to Groups. Indeed, we define:

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

as a multi-partite graph that cointains 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  $W \in \mathbb{Z}_{\geq 0}$  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 [8] to place the terror organizations<sup>4</sup> in a latent space, whereby we can create a latent network of the organizations and assign these organizations to clusters based upon the multi-partite

<sup>&</sup>lt;sup>4</sup> Throughout the work, Group, terror organization and organization are used as synonyms.

network. We use Gower's Similarity Coefficient defined as:

$$S_{ij} = \frac{\sum_{k=1}^{r} w_{ijk} S_{ij}^{(k)}}{\sum_{k=1}^{v} w_{ijk}}$$
(2)

where  $S_{ij}$  is the similarity between Groups *i* and *j* on a variable (i.e. Targets, Weapons, etc.), *k* and *v* is the total number of variables and  $w_{ijk}$  is the weight of the similarity between Group *i* and Group *j* for metric *k*.  $S_{ij}^{(k)}$  is then dually defined as:

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

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

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

where  $r_k$  is the range of  $x_k$ , if k is numerical. Since we are interested in how the different variables, or modes of the multi-partite network, affect the possible latent network and clusters of the terrorist organizations, the weighting term will take values of:

$$w_{ijk} = \frac{N}{\sum_{n}^{N} \delta\left(k, n'\right)} \tag{5}$$

where N is the total number of n modes (in this case, N=4), and  $\delta(k, n')$  is an indicator function that returns 1 if k is within one of the specified important nodes, n'. Thus, we can use this weighting term to explore what happens when we consider certain modes, like Region or Tactics as more important to cluster formation than others, which allows us to investigate whether certain analytic theories regarding terrorist Group similarities are present in the data. Following the completion of the weighted pairwise Gower's Coefficient of Similarity calculation we are then left with an  $N \times N$  affinity matrix that contains the pairwise similarities between each terrorist group and every other terrorist group. To cluster this affinity matrix into sub groups and create a latent network of the terrorist groups, we use k-NN network modularity maximization [19]. k-NN network modularity maximization takes an affinity or distance matrix and creates a k-NN graph where each node connects to its k nearest neighbors (Figure 1). We applied k-NN network modularity maximization because the matrix is fully connected, thus impeding the use of Louvain, and spectral clustering is proved to be more efficient for sparser graphs [15]. Then, this graph is clustered using the Louvain method of modularity maximization [3]. The algorithm iterates over certain values of k and then selects the corresponding latent network and sub group assignments that maximize the modularity of the latent network. The following is a quick visual depiction of the k-NN network modularity maximization procedure. In summary, our method for finding latent networks and groups of terrorist organizations, which allows for testing of analytic theories, is structured as follows:

- 6 Campedelli et al.
  - 1 Determine which modes (i.e. Tactics, Weapons, etc.) will be significant and construct the weighting matrix appropriately;
  - 2 Calculate the Gower's Coefficient of Similarity between each Group and every other Group, using the weighting matrix, to form an affinity matrix;
- 3 Find latent graphs and sub-groups in the affinity matrix using the k-NN network modularity maximization procedure;
- 4 Compare the different clustering outputs and latent networks for the different modal weighting schemes to better understand their impact on terrorist groups



Fig. 1. The k-NN Network Modularity Maximization Procedure

Based on this algorithm, we test the relation between the different groupings emerging from the different weighting processes. The comparison will be based on the scores of two popular metrics for evaluating cluster similarity: the Adjusted Rand Index (ARI) and the Adjusted Mutual Information (AMI). A total of six models have been run. The first one assigns the same weight to each dimension (i.e.: no weights), the second one weights Region as the most important dimension, the third does the same with Tactic, and so on. In the last model, we used the "Ideology" attribute, which is generally seen as the main discriminant between terrorist groups, to test if it provides clustering assignments similar to the other models. This procedure seeks to test two working hypotheses:

- H1: Weighting differently a given terrorist dimension would lead to results that are considerably variable across different weighting schemes;
- H2: Weighting by ideology would lead to extremely different results compared with the other schemes, eventually posing the risk of missing relevant hidden patterns and feature clusters that arise regardless of ideology itself. Our intuition is that operational and behavioral characteristics are more fitting in explaining clustering rather than relying on mere political or religious motivations.

# 5 Results

The models run with different weighting schemes yield interesting results for both working hypotheses. Firstly, considering H1, both ARI and AMI measures vary with different ranges when we decide to weight differently a particular mode. Indeed, the ARI ranges from a minumum of 0.13 to a maximum of 0.36, indicating that the highest pairwise similarity is given when comparing models where Tactics and Weapons are considered more important. The AMI, although with different absolute values (the range is between 0.33 and 0.55), confirms this same results.



Fig. 2. Adjusted Mutual Information of Different Weighting Schemes



Fig. 3. Adjusted Rand Index of Different Weighting Schemes

#### 8 Campedelli et al.

Therefore, we argue that applying different weights poses the risk of biasing outcomes that may naturally emerge using the algorithm treating all features as equal. Secondly, regarding H2, clustering on ideology provided communities that are completely different from all the other previous weighting schemes. As shown in Fig. 2 and Fig. 3, ideologically-driven clusters possess very little similarity (mostly around 0) in both AMI and ARI to our behavioral-based clustering. This confirms the hypothesis that, though most research and public debate usually distinguishes terror groups based on their ideologies and/or motivations, it may be more useful to look at operational and behavioral characteristics in order to find meaningful terrorist clusters. Investigating further, we now focus on the resulting clusters of the "No weights" model, analyzing the distribution across clusters of the variables described in the Data section, namely Success Ratio, Suicide Rate, Fatalities Rate, International Rate and the seven different mapped Ideologies. Our algorithm clustered the terrorist groups in the dataset in 37 distinct terrorist communities. The size of communities ranged from a minimum of 3 groups to a maximum of 159 groups per community (Figure 5). Table 2 provides the descriptive statistics of each of the variables employed for the expost evaluation across clusters.



Fig. 4. Pearson's Correlation Matrix of Variables and Ideologies across Groups

There are several significant results taken from this model. First of all, while Success Ratio is extremely high in each community (the average is 0.90), the community that mostly deviates from this value (i.e.: cluster 28, Success Ratio=0.66) includes a total of 21 terror groups and none of these groups is Islamic/Jihadist.

|                 | Succ | Suic | Fatal | Casual | Multi | Intl | Targ  |       |      |      |       |      | _    |      |
|-----------------|------|------|-------|--------|-------|------|-------|-------|------|------|-------|------|------|------|
| С               | Rate | Rate | Rate  | Rate   | Rate  | Rate | Count | Islam | Left | Nat  | Right | Rel  | Env  | Oth  |
| 0               | 0.88 | 0.06 | 4.25  | 14.61  | 0.09  | 0.29 | 1.19  | 0.37  | 0.14 | 0.36 | 0.04  | 0.01 | 0.01 | 0.19 |
| 1               | 0.92 | 0.17 | 4.49  | 20.77  | 0.13  | 0.55 | 1.36  | 0.64  | 0.18 | 0.00 | 0.00  | 0.00 | 0.00 | 0.18 |
| 2               | 0.96 | 0.04 | 1.92  | 4.66   | 0.12  | 0.19 | 3.73  | 0.23  | 0.31 | 0.44 | 0.01  | 0.01 | 0.01 | 0.08 |
| 3               | 0.92 | 0.06 | 1.65  | 5.88   | 0.06  | 0.16 | 1.04  | 0.33  | 0.31 | 0.37 | 0.02  | 0.00 | 0.02 | 0.10 |
| 4               | 0.90 | 0.03 | 4.39  | 19.59  | 0.16  | 0.27 | 1.12  | 0.42  | 0.12 | 0.38 | 0.02  | 0.05 | 0.00 | 0.14 |
| 5               | 0.97 | 0.07 | 3.12  | 6.10   | 0.18  | 0.25 | 1.02  | 0.44  | 0.02 | 0.39 | 0.00  | 0.02 | 0.00 | 0.17 |
| 6               | 0.86 | 0.03 | 2.46  | 4.57   | 0.13  | 0.17 | 1.07  | 0.29  | 0.10 | 0.54 | 0.00  | 0.05 | 0.00 | 0.17 |
| 7               | 0.96 | 0.03 | 1.92  | 6.95   | 0.11  | 0.34 | 1.88  | 0.28  | 0.10 | 0.48 | 0.00  | 0.00 | 0.00 | 0.30 |
| 8               | 0.93 | 0.01 | 2.37  | 8.38   | 0.08  | 0.27 | 1.08  | 0.23  | 0.17 | 0.52 | 0.03  | 0.03 | 0.00 | 0.08 |
| 9               | 0.89 | 0.03 | 2.45  | 8.06   | 0.17  | 0.21 | 1.19  | 0.41  | 0.15 | 0.44 | 0.04  | 0.01 | 0.00 | 0.07 |
| 10              | 0.74 | 0.11 | 2.32  | 5.96   | 0.16  | 0.35 | 1.05  | 0.31  | 0.03 | 0.59 | 0.03  | 0.00 | 0.00 | 0.10 |
| 11              | 0.95 | 0.03 | 1.01  | 2.74   | 0.11  | 0.31 | 1.91  | 0.43  | 0.06 | 0.45 | 0.02  | 0.02 | 0.00 | 0.09 |
| 12              | 0.96 | 0.02 | 2.57  | 5.45   | 0.20  | 0.33 | 1.23  | 0.29  | 0.12 | 0.46 | 0.04  | 0.12 | 0.00 | 0.10 |
| 13              | 0.93 | 0.07 | 2.68  | 9.91   | 0.19  | 0.19 | 1.09  | 0.53  | 0.09 | 0.31 | 0.00  | 0.13 | 0.00 | 0.06 |
| 14              | 0.88 | 0.01 | 1.10  | 5.20   | 0.16  | 0.40 | 2.03  | 0.19  | 0.39 | 0.28 | 0.00  | 0.03 | 0.03 | 0.14 |
| 15              | 0.99 | 0.00 | 2.37  | 4.78   | 0.08  | 0.21 | 1.11  | 0.15  | 0.26 | 0.48 | 0.07  | 0.04 | 0.00 | 0.11 |
| 16              | 0.95 | 0.06 | 2.93  | 7.16   | 0.19  | 0.36 | 1.68  | 0.58  | 0.14 | 0.28 | 0.02  | 0.00 | 0.00 | 0.08 |
| 17              | 0.89 | 0.02 | 1.57  | 7.03   | 0.19  | 0.59 | 1.28  | 0.28  | 0.16 | 0.52 | 0.04  | 0.12 | 0.00 | 0.16 |
| 18              | 0.87 | 0.01 | 1.74  | 4.30   | 0.18  | 0.43 | 1.40  | 0.05  | 0.26 | 0.58 | 0.16  | 0.07 | 0.02 | 0.07 |
| 19              | 0.89 | 0.01 | 0.72  | 1.85   | 0.20  | 0.21 | 1.29  | 0.13  | 0.48 | 0.27 | 0.06  | 0.02 | 0.02 | 0.06 |
| $\overline{20}$ | 1.00 | 0.04 | 2.16  | 6.34   | 0.00  | 0.71 | 1.08  | 0.92  | 0.00 | 0.23 | 0.00  | 0.00 | 0.00 | 0.00 |
| 21              | 0.90 | 0.06 | 6.53  | 18.50  | 0.21  | 0.28 | 1.03  | 0.48  | 0.24 | 0.28 | 0.00  | 0.00 | 0.00 | 0.07 |
| 22              | 0.95 | 0.03 | 3.07  | 6.30   | 0.09  | 0.23 | 1.17  | 0.30  | 0.11 | 0.52 | 0.02  | 0.02 | 0.00 | 0.10 |
| $\overline{23}$ | 0.96 | 0.00 | 4.79  | 7.12   | 0.16  | 0.12 | 1.06  | 0.11  | 0.11 | 0.33 | 0.00  | 0.11 | 0.00 | 0.33 |
| $\overline{24}$ | 0.86 | 0.02 | 2.14  | 4.46   | 0.15  | 0.36 | 1.76  | 0.36  | 0.12 | 0.38 | 0.05  | 0.07 | 0.00 | 0.17 |
| 25              | 0.78 | 0.04 | 1.31  | 4.93   | 0.19  | 0.41 | 1.39  | 0.13  | 0.30 | 0.41 | 0.00  | 0.00 | 0.04 | 0.22 |
| 26              | 0.95 | 0.01 | 7.03  | 11.19  | 0.11  | 0.27 | 1.72  | 0.32  | 0.09 | 0.34 | 0.08  | 0.01 | 0.01 | 0.22 |
| $\overline{27}$ | 0.83 | 0.00 | 1.17  | 4.60   | 0.24  | 0.38 | 1.46  | 0.08  | 0.43 | 0.32 | 0.04  | 0.04 | 0.06 | 0.13 |
| $\overline{28}$ | 0.66 | 0.00 | 0.83  | 1.32   | 0.02  | 0.30 | 1.14  | 0.00  | 0.24 | 0.33 | 0.05  | 0.00 | 0.05 | 0.33 |
| 29              | 0.90 | 0.09 | 4.80  | 13.40  | 0.06  | 0.58 | 1.22  | 0.89  | 0.00 | 0.11 | 0.00  | 0.11 | 0.00 | 0.00 |
| 30              | 1.00 | 0.00 | 1.67  | 3.00   | 0.00  | 0.00 | 1.00  | 1.00  | 0.00 | 0.33 | 0.00  | 0.00 | 0.00 | 0.00 |
| 31              | 0.88 | 0.01 | 2.21  | 5.52   | 0.09  | 0.04 | 2.35  | 0.05  | 0.20 | 0.25 | 0.15  | 0.05 | 0.00 | 0.40 |
| 32              | 0.91 | 0.00 | 2.29  | 10.29  | 0.29  | 0.14 | 2.00  | 0.14  | 0.14 | 0.29 | 0.00  | 0.00 | 0.00 | 0.43 |
| 33              | 0.80 | 0.00 | 0.54  | 11.38  | 0.14  | 0.26 | 1.00  | 0.50  | 0.25 | 1.00 | 0.00  | 0.00 | 0.00 | 0.00 |
| 34              | 0.92 | 0.78 | 4.80  | 14.72  | 0.63  | 1.00 | 2.00  | 1.00  | 0.00 | 0.00 | 0.00  | 0.00 | 0.00 | 0.00 |
| 35              | 0.75 | 0.00 | 0.00  | 3.00   | 0.00  | 0.50 | 1.00  | 0.00  | 0.00 | 0.50 | 0.00  | 0.00 | 0.00 | 0.50 |
| 36              | 1.00 | 0.00 | 0.02  | 0.03   | 0.15  | 0.00 | 1.00  | 0.17  | 0.00 | 0.83 | 0.00  | 0.00 | 0.00 | 0.17 |
| Μ               | 0.90 | 0.03 | 2.74  | 7.88   | 0.13  | 0.29 | 1.45  | 0.30  | 0.18 | 0.40 | 0.03  | 0.03 | 0.01 | 0.13 |

Table 2. Variables Distribution Across Resulting Clusters. ANOVA tests revealed statistically significant differences across groups for all variables (Prob>F=0.000)



Fig. 5. Number of Terrorist Groups in Each Cluster

There is only another cluster with no Islamic/Jihadist groups in it (cluster 35), and although it is a small community (only 4 groups in it), the Success Ratio is significantly low in that case too (0.75). This seems to indicate that, generally, Islamic/Jihadist groups tend to have good operational performance in their attacks. As further proof, it is worth to note that the clusters with the highest percentage of Islamic/Jihadist groups (namely, clusters 1, 16, 20, 30, 34) show Success Rate values always higher than the average, specifically in the range from 0.92 to 1.

Secondly, when focusing on the Fatalities Rate, data highlight that the clusters that yield higher values generally include several types of ideologies. Two major examples are cluster 0 (Fatalities Rate=4.25) and Cluster 4 (4.39). In the first case, out of a total of 159 terror groups, 37.11% are Islamic/Jihadist and 35.85% are Ethnic/Nationalist. In the second case, out of a total of 66 groups, 42.42% are Islamic/Jihadist and 37.88% are Nationalists. This demonstrates how the fact of being able to carry out particularly devastating attacks is not a feature that is specifically related to a single ideology.

Thirdly, Pearson correlation (Fig. 4) revealed interesting relations between ideologies and behavioral characteristics, looking at data from a more general perspective. Listing some: multiplot rates are extremely correlated with suicide and international operations. Additionally there are is no strong relationship between the degree of severity of an attack (fatality rate) and the extent to which this attack is internationally plotted or motivated. This result indicates that both domestic and transnational terrorism are able to inflict high-magnitude terror. Our further hypothesis is that, however, the distribution of highly fatal attacks per geopolitical terror type (i.e.: domestic or transnational) is not equally distributed across regions in the world. In fact, it may be that being transnational or domestic is not a discriminant feature *per se*, but it is intrinsically related

11

to the operating geographic context. Moreover, it is worth to note how there is a positive correlation between the shares of environmental/animalist groups and leftists/anarchist organizations: two ideological types of terrorism which are overlapping in some cases.

### 6 Discussion and Future Work

This work has applied a novel clustering algorithm for multi-partite networks based on Gower's coefficient of similarity to define latent communities of terror groups at a global scale, using data for the period 1997-2016. Besides the innovative application, this work has presented multiple models based on different weighting schemes and demonstrated how (1) weighting more specific features may lead to substantially different results and (2) giving more importance to ideology will ultimately hide common behavioral patterns that groups shared regardless of their motivations. Finally, this work has presented the results of the algorithm in the "No Weight" case, analyzing the most relevant outcomes in terms of attribute variables across the behaviorally-detected clusters. This exploratory application calls for future work. Specifically, further directions could involve the use of machine learning algorithms developing feature spaces adding contextual, operational, and temporal information to evaluate if it is possible to train a classifier that correctly predicts the cluster to which each group is assigned. This analysis would eventually confirm the meaningfulness of the resulting communities. Using a simple ex-post descriptive analysis, we have demonstrated that our algorithms yield outcomes that give a certain patterned structure to the data. Thus, a further investigation of the data using supervised learning may strongly corroborate our findings. Furthermore, our application seeks to be compared to other existing clustering algorithms designed for multimode matrices, e.g. the Infinite Relational Model [9], in order to evaluate the stability of the results when other methodological architectures are applied.

#### References

- Belli, R., Freilich, J.D., Chermak, S.M., Boyd, K.A.: Exploring the crimeterror 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 (2015). DOI 10.1080/17467586.2015.1104420
- Benigni, M.C., Joseph, K., Carley, K.M.: Online extremism and the communities that sustain it: Detecting the ISIS supporting community on Twitter. PLOS ONE 12(12), e0181,405 (2017). DOI 10.1371/journal.pone.0181405
- Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment 2008(10), P10,008 (2008). DOI 10.1088/1742-5468/2008/10/P10008
- Campedelli, G.M., Cruickshank, I., Carley, K.M.: Complex Networks for Terrorist Target Prediction. In: R. Thomson, C. Dancy, A. Hyder, H. Bisgin (eds.) Social, Cultural, and Behavioral Modeling, vol. 10899, pp. 348–353. Springer International Publishing, Cham (2018)

- 12 Campedelli et al.
- Carley, K.M.: Destabilization of covert networks. Computational and Mathematical Organization Theory 12(1), 51–66 (2006). DOI 10.1007/s10588-006-7083-y
- Chenoweth, E., Lowham, E.: On Classifying Terrorism: A Potential Contribution of Cluster Analysis for Academics and Policy-makers. Defense & Security Analysis 23(4), 345–357 (2007). DOI 10.1080/14751790701752402
- Desmarais, B.A., Cranmer, S.J.: Forecasting the locational dynamics of transnational terrorism: a network analytic approach. Security Informatics 2(1), 8 (2013). DOI 10.1186/2190-8532-2-8
- 8. Gower, J.C.: A General Coefficient of Similarity and Some of Its Properties 27
- 9. Kemp, C., Tenenbaum, J., Griffiths, T., Yamada, T., Ueda, L.: Learning systems of concepts with an infinite relational model. In: AAAI, vol. 3 (2006)
- Klausen, J.: Tweeting the *Jihad*: Social Media Networks of Western Foreign Fighters in Syria and Iraq. Studies in Conflict & Terrorism 38(1), 1–22 (2015). DOI 10.1080/1057610X.2014.974948
- Koschade, S.: A Social Network Analysis of Jemaah Islamiyah: The Applications to Counterterrorism and Intelligence. Studies in Conflict & Terrorism 29(6), 559–575 (2006). DOI 10.1080/10576100600798418
- 12. Krebs, V.: Mapping Networks of Terrorist Cells. Connections 24(3), 43–52 (2002)
- LaFree, G., Dugan, L.: Introducing the Global Terrorism Database. Terrorism and Political Violence 19(2), 181–204 (2007). DOI 10.1080/09546550701246817
- Lautenschlager, J., Ruvinsky, A., Warfield, I., Kettler, B.: Group Profiling Automation for Crime and Terrorism (GPACT). Proceedia Manufacturing 3, 3933–3940 (2015). DOI 10.1016/j.promfg.2015.07.922
- von Luxburg, U.: A Tutorial on Spectral Clustering. arXiv:0711.0189 [cs] (2007). URL http://arxiv.org/abs/0711.0189. ArXiv: 0711.0189
- Medina, R., Hepner, G.: Geospatial Analysis of Dynamic Terrorist Networks. In: I.A. Karawan, W. McCormack, S.E. Reynolds (eds.) Values and Violence, vol. 4, pp. 151–167. Springer Netherlands, Dordrecht (2009). DOI {10.1007/ 978-1-4020-8660-1\_10}
- Moon, I.C., Carley, K.M.: Modeling and Simulating Terrorist Networks in Social and Geospatial Dimensions. IEEE Intelligent Systems 22(5), 40–49 (2007). DOI 10.1109/MIS.2007.4338493
- Qi, X., Christensen, K., Duval, R., Fuller, E., Spahiu, A., Wu, Q., Zhang, C.Q.: A Hierarchical Algorithm for Clustering Extremist Web Pages. In: 2010 International Conference on Advances in Social Networks Analysis and Mining, pp. 458–463 (2010). DOI 10.1109/ASONAM.2010.81
- Ruan, J.: A Fully Automated Method for Discovering Community Structures in High Dimensional Data. pp. 968–973. IEEE (2009). DOI 10.1109/ICDM.2009.141
- Sageman, M.: The Stagnation in Terrorism Research. Terrorism and Political Violence 26(4), 565–580 (2014). DOI 10.1080/09546553.2014.895649. URL http: //dx.doi.org/10.1080/09546553.2014.895649