Future Generation Computer Systems 101 (2019) 983–992

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Understanding the composition and evolution of terrorist group networks: A rough set approach

Vincenzo Loia, Francesco Orciuoli *

DISA-MIS, Università degli Studi di Salerno, Via G. Paolo II, 132, 84084 Fisciano (SA), Italy

HIGHLIGHTS

- Eliciting information of terrorist groups' networks from past terrorist events.
- A novel similarity function based on boundary regions in three-way decisions.
- Approximate conceptualizations of terrorist groups' behaviors through Rough Sets.
- Analysis of temporal evolution of terrorist groups supporting counter-terrorism.

ARTICLE INFO

Article history:
Received 22 March 2019
Received in revised form 13 June 2019
Accepted 23 July 2019
Available online 26 July 2019

Keywords:
Terrorist group network
Rough set theory
Temporal evolution
Counter terrorism

ABSTRACT

Nowadays, many resources for counter-terrorism operations are available for researchers belonging to different areas. In particular, the START project provides the Global Terrorism Database (GTD) that can be analyzed in order to provide, for instance, prediction models. The main idea underlying this work is using the historical data provided by GTD, which offers information related to terrorist attacks perpetrated since 1970, in order to conceptualize the behaviors of terrorist groups in specific time intervals. Such conceptualizations are, subsequently, used to understand the similarity between terrorist groups and elicit relations to represent terrorists' networks. The above networks can be used to study the temporal evolutions of terrorist groups' behaviors by applying the approach in different time periods along the timeline and studying differences among the resulting networks. The approach is mainly based on Rough Set Theory and Three-way Decisions Theory and provides an original similarity function based on the definition of boundary regions.

1. Introduction and related works

Terrorism is a type of collective violence having direct impact on peace, normal routine of a country/community and security and it is also a way to generate fear in civilians using violence.[1] Terrorism is an evolving phenomenon, thus it is vital to provide counter-terrorism operators with tools for the prevention of it. In particular, counter-terrorism could be defined as a collection of activities that might include techniques, tactics or strategies carried out by either government, politicians, police department, business or military for the prevention of terrorism.[2] The main objective of this work is to define an approach aiming at eliciting knowledge on terrorist attack perpetrators by analyzing terror events along the timeline. The idea is to construct a sociogram, i.e., a network of perpetrators, where the nodes represent terrorist groups and the edges represent generic relations occurring between two groups. Such relations are elicited by using an original method based on similarity of terrorist groups and taking into account several dimensions related to terrorist groups' characteristics (e.g., behavior of the attacks, motivations for the attacks, claim modes). More in details, terrorist groups are conceptualized by means of rough sets [1] (in particular through the application of both lower and upper approximations) and the similarity measure of two groups is calculated by comparing, in general, their respective rough sets. Lastly, considering terrorist group networks along the timeline allows to study the temporal evolution of relations between such groups. With respect to the adopted data source, the aforementioned conceptualization is realized by considering the historical data provided by the Global Terrorism Database[3] (GTD) [2,3]. GTD is an open source database that includes terrorist activities or events information all around the world since 1970. This dataset[4] considers 10 main attributes

* Corresponding author.
E-mail address: forciuoli@unisa.it (F. Orciuoli).

1 https://www.slideshare.net/shaanyadav3/terrorismcauses-and-types.

2 https://en.wikipedia.org/wiki/Counter-terrorism.

3 National Consortium for the Study of Terrorism and Responses to Terrorism (START). (2018).

4 https://www.start.umd.edu/gtd/.

https://doi.org/10.1016/j.future.2019.07.049
0167-739X/

and many sub attributes, those 9 attributes are: (i) ID and date; (ii) incident information, (iii) incident location, (iv) attack information, (v) weapon information, (vi) target/victim information, (vii) perpetrator information, (viii) casualties and consequences, and (ix) additional information and sources. As stated before, the proposed approach can be put under the umbrella of computational frameworks for counterterrorism operators and analysts. In particular, the scientific literature offers several works focusing on methodological aspects like, for instance, in [4], where a decision support methodology for risk assessment and prediction of terrorism insurgency, based on artificial neural networks and analytic hierarchy process, has been proposed. Some of these paper deal with risk assessment to evaluate radicalization on Social Media like Twitter [5]. Other works focus on both attack behaviors and networks of terrorist groups. Among such works, the authors of [6] provide a framework based on two logic flows in the framework, the behavior recognition flow and the network predicting flow. The behavior recognition flow is designed to mine the behavior pattern from both the structural characteristics of the group network and the group behavior. In particular, the behavior recognition problem is modeled as a pattern classification problem. The network predicting flow, based on the wavelet transform theory, is employed to better understand and predict changes in terrorist group networks.

Moreover, the same database used in the present work has been already exploited in [7] where classification algorithms (e.g., KNN and SVM) are adopted for prediction tasks, and in [8], where a modified version of k-means clustering algorithm is evaluated against the above database of terrorist events. Additionally, multidisciplinary approaches have been used to understand the dynamics of terrorism events. In this context, the work [9] proposes a machine learning method to simulate the risk of terrorist attacks at a global scale based on multiple resources, long time series and globally distributed datasets.

With respect to the existing approaches, the proposed one focuses on providing information about the evolution of the behaviors of terrorist groups by constructing terrorist groups' networks and considering also how the relations with other groups change as the time goes on. The idea to model terrorist groups' networks has been already proposed in [10], where association rule mining and cosine similarity measure have been adopted to extract rules and compute similarities for building the social network model.

The remaining parts of the paper are structured as it follows. Section 2 provides background knowledge on rough set theory and three-way decision theory. The main approach is described in Section 3 (explaining the main steps of the proposed approach) and 5 (considering the time dimension in the approach). Section 4 provides a complete illustrative example to better clarify the approach. A case study is described in Section 6. Final remarks and future works are discussed in Section 7.

2. Rough set and three-way decisions theories

The rough set theory (RST) is a wellknown mathematical tool that is useful to deal with imprecise, inconsistent, incomplete information and knowledge [11]. The basic idea of the RST can be divided into two parts. The first part is to form concepts and rules through the classification of objects. The second part is to discovery knowledge through the classification of the equivalence relation and classification for the approximation of the target [12]. Moreover, the authors of the above cited work affirm that RST is a certain mathematical tool to solve uncertain problems. RST has become an important information processing tool in the field of intelligent information processing [13]. Let us start with fundamentals definitions of RST. First of all, being S an information system defined as the 4-tuple: $S = \langle U, R, V, f \rangle$, $R = C \cup D$, where U is a finite nonempty set of objects, R is a finite nonempty set of attributes, the subsets C and D are called condition attribute set and decision attribute set, respectively. $V = \bigcup_{a \in R} V_a$, where V_a is the set of values of attribute a, and $card(V_a) > 1$, and $f : R \rightarrow V$ is an information or a description function.

Definition 1 (*Indiscernible Relation*)**.** Given a subset of attribute set $B \subseteq R$, an indiscernible relation $IND(B)$ on the universe U can be defined as follows,

$$IND(B) = \left\{(x, y) \,|\, (x, y) \in U^2, \forall_{b \in B} (b(x) = b(y))\right\} \tag{1}$$

The equivalence relation is an indiscernible relation. And the equivalence class of an object x is denoted by $[x]_{IND(B)}$ and the pair $\left(U, [x]_{IND(B)}\right)$ is an *approximation space*. From this point $[x]_{IND(B)}$ will be indicated as $[x]$.

Definition 2 (*Upper and Lower Approximation Sets*)**.** Given an information system $S = \langle U, R, V, f \rangle$, for a subset $X \subseteq U$, its lower and upper approximation sets are defined , respectively, by:

$$\overline{apr}(X) = \{x \in U | [x] \cap X \neq \emptyset\}, \tag{2}$$

$$\underline{apr}(X) = \{x \in U | [x] \subseteq X\} \tag{3}$$

where $[x]$ denotes the equivalence class of x.

The rough membership for the element $x \in U$ is defined as follows:

$$\mu_X(x) = \frac{|X \cap [x]|}{|[x]|} \tag{4}$$

We can affirm that rough membership is a coefficient that describes inaccuracy for $x \in X$.

Furthermore, the Three-way Decisions Theory (3WDT) [14] has been introduced to divide the universe U into three disjoint regions: positive, negative and boundary. Such three regions are viewed, respectively, as the regions of acceptance, rejection, and noncommitment in a ternary classification [15] obtained by employing rough set theory [16]. More formally, it is possible to define:

$$\begin{aligned} POS(X) &= \underline{apr}(X), \\ BND(X) &= \overline{apr}(X) - \underline{apr}(X), \\ NEG &= U - \overline{apr}(X). \end{aligned} \tag{5}$$

If an object $x \in POS(X)$, then it belongs to target set X certainly. If an object $x \in BND(X)$, then it does not belong to target set X certainly. If an object $x \in NEG(X)$, then it cannot be determined whether the object x belongs to target set X or not.

3. Approach to elicit terrorist groups' networks

This section contains the description of the approach for modeling terrorist groups and use such conceptualization to elicit groups' networks. The workflow of the approach is depicted in Fig. 1.

In particular, the first step is to select the relevant features from the whole set of attributes (see Section 3.2) provided by the Global Terrorism Database (see Section 3.1). In the second step, the crisp conceptualizations of terrorist groups and their approximations are provided (see Sections 3.3 and 3.4). The preliminary actions for the second step are the definition of the time interval in which considering terrorist events and the selection of the terrorist groups for which analyzing the behavior. The third step, starting from the rough approximations provided by the second step and employing the three-way decisions theory, constructs

Fig. 1. Workflow of the overall approach.

the boundary regions for the above rough conceptualizations (see Section 3.5). In the fourth step, the similarity matrix, reporting the similarities between all pairs of considered terrorist groups, is constructed (see Section 3.7) by using the novel similarity function defined in Section 3.6. Lastly, the fifth step is committed to transform the similarity matrix into a network graph representing the generic relations among terrorist groups (see Section 3.7).

3.1. Global Terrorism Database

Global Terrorism Database (GTD) [17] is an open source database including information on terrorist events around the world from 1970 through 2016. It contains information about 170350 terrorist events, and a total of 135 attributes (features) for each event, including exact date, location, group, weapon, casualty and so on. The features are grouped in the following categories: IDs and dates, incident information, incident location, attack information, weapon information target/victim information, perpetrator information, casualties and consequences, and additional information and sources. The GTD defines a terrorist attack as the threatened or actual use of illegal force and violence by a non-state actor to attain a political, economic, religious, or social goal through fear, coercion, or intimidation.

3.2. Selecting features from GTD

The idea underlying the proposed approach is conceptualizing terrorist groups by using knowledge on the attacks they have perpetrated. Thus, it is needed to select a relevant subset of the features provided by the GTD. Such subset must be suitable to describe the behavior of the terrorist groups. In this work, it was not considered an automatic approach for feature selection but two behavioral aspects have been addressed: (i) *attack and operative modes*, and (ii) *target and motives*. In particular, the *attack and operative modes* aspect is defined by using following features:

- `attacktype1` provides information about the main method of the attack or the general tactics adopted and executed by the perpetrators during the event.
- `weaptype1` indicates the main type of the weapon used during the attack by the perpetrators.
- `suicide` informs us if there is the evidence of a suicide attack or not.
- `ishostkid` informs us whether or not the victims were taken hostage or kidnapped during the event.
- `ransom` indicates if the event involved a demand of monetary ransom.

Moreover, the *target and motives* aspect is defined by using the following features:

- `targtype1` specifies the type of the target for the current event.
- `INT_LOG` indicates if a perpetrator group crossed a border to carry out the attack.
- `INT_IDEO` indicates if a perpetrator group attacked a target of a different nationality.

Table 1
Sample value vector associated with features and interpretations.

Feature	Value	Interpretation
`attacktype1`	2	*Armed Assault*
`weaptype1`	6	*Explosives/Bombs/Dynamite*
`suicide`	0	*No suicide*
`targtype1`	3	*Police*
`INT_LOG`	0	*The nationality of the perpetrator group is the same of the location of the attack*
`INT_IDEO`	0	*Any and all nationalities of the perpetrator group are the same as the nationalities of the victims*
`ishostkid`	0	*No kidnapping*
`crit1`	1	*The incident meets criterion 1*[a]
`crit2`	1	*The incident meets criterion 2*[b]
`crit3`	1	*The incident meets criterion 3*[c]

[a]The act must be aimed at attaining a political, economic, religious, or social goal.
[b]There must be evidence of an intention to coerce, intimidate, or convey some other message to a larger audience (or audiences) than the immediate victims.
[c]The action must be outside the context of legitimate warfare activities.

- `crit1` informs us if the attack has a political, economic, religious, or social goal.
- `crit2` informs us if the attack has the intention to coerce, intimidate or publicize to larger audience.
- `crit3` informs us if the attack is outside the context of legitimate warfare activities.

Other two features could be used to specify the location of a given attack: `region` and `country`. For the aim of this work we will not use the above geographical features in order to avoid an excessive specialization.

The values corresponding to the aforementioned features represent the behavior expressed by a perpetrator in a specific terrorist event. We need to summarize the behaviors expressed by the same perpetrator in a given set of events in order to obtain the typical behavior of such perpetrator. Therefore, we have a view $U_{[t_1,t_2]}$ on GTD that represents the Universe of Discourse including all terrorist events in a time window of interest. The set C contains all relevant features we need to consider and each event e, belonging to the Universe, is represented by the vector of values corresponding to the aforementioned features. The sample vector $\langle 2, 6, 0, 3, 0, 0, 0, 1, 1, 1\rangle$ is reported in Table 1.

Potentially, the feature set could be increased by considering also additional aspects like *communication* (with features `claimed` and `claimmode`) and *impact* (with features `nkill`, `nwound` and `propextent`). Such features could be used in future works.

3.3. Conceptualizing the behavior of terrorist groups

The typical behavior of a terrorist group can be conceptualized by considering the set of events in the dataset that have been perpetrated by such a group. And, in particular, by considering the behavior expressed during these events. Such behavior can be obtained by studying the values corresponding to the features selected before (represented as shown in Section 3.2). The set of the events perpetrated by a group in a given time window represents a crisp conceptualization of the behavior of such a group.

More in details, being $U_{[t_1,t_2]}$ the Universe of Discourse including data of all terrorist attacks tracked in the time window $[t_1, t_2]$, C the set of features discussed in Section 3.2 used to describe the events in the Universe (C is the set of condition attributes), IND_C the indiscernibility function defined on C and g the name of the terrorist group to conceptualize. Additionally, consider a further set of attributes $D = \{gname\}$ (decision attributes), such that $A = C \cup D$. Moreover, $V_{[t_1,t_2]} = \{e \in U_{[t_1,t_2]} | gname(e) = g\}$ is the set of all events in the time window $[t_1, t_2]$ associated (in the GTD) to the terrorist group g. Therefore, $V_{[t_1,t_2]}$ conceptualizes, in a crisp way, the behavior of the group g in the time window $[t_1, t_2]$.

This crisp conceptualization is not completely satisfactory for our aims. In fact, the original data (in GTD) could be imprecise (e.g., a given attack is associated to the wrong group — an error occurs in the value for the `gname` attribute) or it is possible to have outliers (atypical behaviors) in different attacks of the same group. For this motivations, a rough conceptualization is needed.

A rough conceptualization can be constructed by applying the rough sets operators over the set $V_{[t_1,t_2]}$ in order to obtain the rough set $(\underline{V}_{[t_1,t_2]}, \overline{V}_{[t_1,t_2]})$. With respect to the definitions of Section 2, $\overline{V}_{[t_1,t_2]} = \overline{apr}\left(V_{[t_1,t_2]}\right)$ and $\underline{V}_{[t_1,t_2]} = \underline{apr}\left(V_{[t_1,t_2]}\right)$.

Other types (variants, extensions and so on) of rough sets could be evaluated for the task described in this section [18–23] and [24]. Some criteria to select a suitable rough set approach are described in the work [25].

3.4. Interpreting the rough set for a terrorist group

Given the constructed rough set $(\underline{V}_{[t_1,t_2]}, \overline{V}_{[t_1,t_2]})$, its first component $\underline{V}_{[t_1,t_2]}$ is the lower approximation and includes the attacks that certainly have been perpetrated by the group g. In other terms, the lower approximation is a conceptualization of the characterizing behavior of the terrorist group g. The second component $\overline{V}_{[t_1,t_2]}$ is the upper approximation and includes the attacks that could be have perpetrated by the group g plus the attacks that certainly have been perpetrated by such a group. In other terms, the upper approximation represents both the characterizing behavior of g and behaviors that can be possibly associated to g (e.g., a behavior adopted also by other terrorist groups).

3.5. Applying the three-way decisions theory

Once a terrorist group's behavior is summarized by means a rough set, it is possible to apply the well-know Three-way Decisions Theory to classify the attacks (of a given terrorist group) in three regions: positive, negative and boundary. The boundary region contains the ambiguous events (coming from uncertain data) and should be further investigated in order to discover more knowledge on the similarities between the behaviors of different groups. Fig. 2 shows the conceptualization of the behavior of the terrorist group g (this is obtained by filtering the terrorist attacks with respect to the attribute `gname`), the approximation of such a concept and the construction of the three regions.

More in details, the three regions could be defined as:

$$POS = \underline{V}_{[t_1,t_2]} \tag{6}$$

$$BND = \overline{V}_{[t_1,t_2]} - \underline{V}_{[t_1,t_2]} \tag{7}$$

$$NEG = U_{[t_1,t_2]} - BND \tag{8}$$

The interpretation for the three regions, constructed for the terrorist group g, is the following one:

- The POS region includes all events certainly perpetrated by the considered terrorist group and that characterize its behavior;
- The NEG region includes all events that surely have not been perpetrated by the considered terrorist group and that cannot characterize the behavior of such a group;
- The BND region includes all events that could or could not be been perpetrated by the considered terrorist group and that possibly express its behavior.

It is interesting to discuss on the composition of the boundary region. The BND region for the generic terrorist group g contains events representing terrorist attacks having one of the following characteristics:

- they express behaviors of g that are not specific for g but are acted also by other terrorist groups,
- they express behaviors of other groups h that are also acted by g.

Therefore, the BND region for a group g represents the gray zone of the behaviors of g that could be associated also to other terrorist groups. Thus, it is possible to consider such a BND region as one of the existing sources useful to identify similarities among different terrorist groups.

3.6. Definition of the similarity function

In the proposed approach the similarity among two terrorist groups is calculated by starting from the approximations obtained in Sections 3.3 and 3.4. Thus, a similarity function, exploiting the above approximations is needed. Let us start from a number of similarity functions applicable to sets as described in [26], where the authors introduce the general form of a similarity function applicable between sets:

$$F(A, B) = \frac{\psi_1\left(|A \cap B|\right)}{\psi_2\left(|A|, |B|, |A \cup B|\right)} \tag{9}$$

where $A, B \subset \Omega$, ψ_1 is a strictly increasing function and ψ_2 is an increasing function of three variables. Moreover, the function F, for all $A, B \subset \Omega$, has the following characteristics:

$$0 \leq F(A, B) \leq 1 \tag{10}$$

$$F(A, B) = 1 \Longleftrightarrow A = B \tag{11}$$

$$F(A, B) = 0 \Longleftrightarrow A \cap B = \emptyset \tag{12}$$

If the denominator of F is constant then F is strictly

increasing with $|A \cap B|$ (13)

Measures based on F and satisfying requirements (10)–(13) are called strong similarity measures. Furthermore, it is possible to define the requirements for weak similarity functions by replacing the requirement (11) with the following one:

$$F(A, B) = 1 \Longrightarrow A \subset B \text{ or } B \subset A \tag{14}$$

In literature, a number of strong similarity measures have been recognized. An incomplete list of such function is reported in Table 2.

Furthermore, it is also possible to generalize Eq. (9) for ordinary sets (e.g., fuzzy sets, rough sets) as it follows:

$$F^*(A, B) = \frac{\psi_1\left(\sum_{x \in A \cap B} \min\left(P_A(x), P_B(x)\right)\right)}{\psi_2\left(\sum_{x \in A} P_A(x), \sum_{x \in B} P_B(x), \sum_{x \in A \cup B} \max\left(P_A(x), P_B(x)\right)\right)} \tag{15}$$

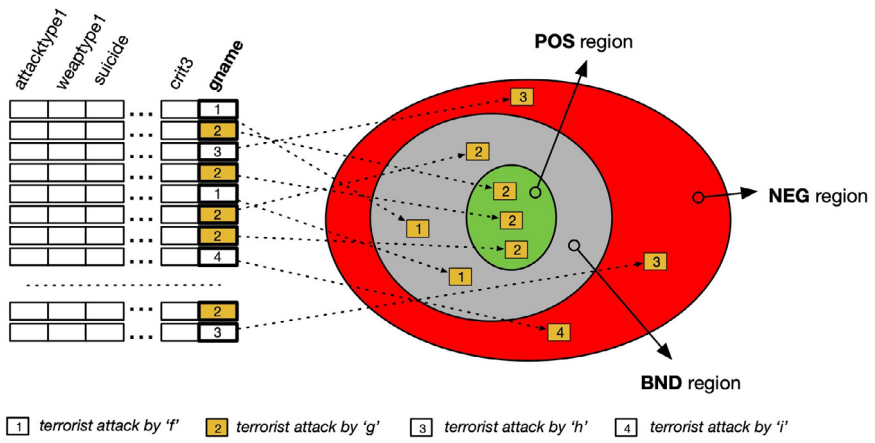

Fig. 2. Diagram representing the three regions by which the dataset has been partitioned.

Table 2
Wellknown strong similarity measures between sets.

Name	Equation
Jaccard's index	$J(A,B) = \frac{\lvert A\cap B\rvert}{\lvert A\cup B\rvert}$
Dice's index	$D(A,B) = \frac{2\lvert A\cap B\rvert}{\lvert A\rvert+\lvert B\rvert}$
Cosine function	$C(A,B) = \frac{\lvert A\cap B\rvert}{\sqrt{\lvert A\rvert\lvert B\rvert}}$
Measure N	$N(A,B) = \sqrt{2}\frac{\lvert A\cap B\rvert}{\sqrt{\lvert A\rvert^2+\lvert B\rvert^2}}$
Overlap measure 1	$O_1(A,B) = \frac{\lvert A\cap B\rvert}{\min(\lvert A\rvert,\lvert B\rvert)}$
Overlap measure 2	$O_2(A,B) = \frac{\lvert A\cap B\rvert}{\max(\lvert A\rvert,\lvert B\rvert)}$

where $P(x)$ is a membership function that ranges in the interval $[0, 1]$. It is possible to prove that F^* (and functions based on it) is a strong (or weak) similarity function for ordinary sets with the same requirements of F, i.e., (10), (11) (or (14) in the case of weak similarity), (12) and (13).

Basing on (15), the authors of [27] introduced a strong similarity measure between two rough sets. More in details the above measure is defined by the function:

$$Sim(A,B) = \begin{cases} 1, & \text{if } A = B = \emptyset, \\ \frac{\sum_{i=1}^{n}\min\{a_i,b_i\}}{\sum_{i=1}^{n}\max\{a_i,b_i\}}, & \text{otherwise} \end{cases} \tag{16}$$

where $A, B \subseteq U$, $U = \{u_1, u_2, \ldots, u_n\}$ is the universe of discourse and $\forall u_i \in U$, $a_i = \mu_A^I(u_i)$ is the rough membership of u_i in A and $b_i = \mu_B^I(u_i)$ is the rough membership of u_i in B. Then, by applying Eq. (4) to A and B, it is possible to obtain $A' = \{\mu_A^I(u_1)/u_1, \mu_A^I(u_2)/u_2, \ldots, \mu_A^I(u_n)/u_n\}$ and $B' = \{\mu_B^I(u_1)/u_1, \mu_B^I(u_2)/u_2, \ldots, \mu_B^I(u_n)/u_n\}$.

Taking care of the considerations reported in Section 3.5 it is possible to consider only the BND regions of two terrorist groups in order to measure the similarity between them. In fact, it is convenient to exclude the POS regions given that they include only behavior that strongly characterizes the behavior of a group. In other terms, such characterizing behavior cannot be recognized in the POS region of another terrorist group. Despite the POS regions, BND regions include those behaviors that are shared by different groups. Therefore, BND regions are interesting to be investigated for searching the aforementioned similarities. In order to calculate such a similarity measure it is needed to employ a suitable function. Among the similarity functions presented above, the *Overlap Measure 1* seems to be suitable given that it is not affected by the set cardinality.

Let G_1 and G_2 be two rough sets representing the conceptualized behaviors of two terrorist groups and, after applying the formulas reported in (6)–(8) on such rough sets, let BND_{G_1} and BND_{G_2} the boundary regions of the first and the second terrorist group respectively. Now it is possible to define the following similarity function based on the *Overlap Measure 1*:

$$Sim_{TG}(G_1, G_2) = \frac{\left|BND_{G_1} \cap BND_{G_2}\right|}{\min\left(\left|BND_{G_1}\right|, \left|BND_{G_2}\right|\right)} \tag{17}$$

The interpretation of the function Sim_{TG}, that ranges in the interval $[0, 1]$, is that two different terrorist groups are more similar if there exist more terrorist attacks following patterns (behaviors) that could be been perpetrated by both the groups.

3.7. Building the network

In order to construct a network of possible relations starting from a set of terrorist groups $g_1, g_2, \ldots, g_n$, it is needed to calculate $Sim_{TG}(G_1, G_2)$ for each $(G_1, G_2) \in \Gamma \times \Gamma$, where Γ is the set of rough sets representing conceptualizations of all terrorist groups $g_1, g_2, \ldots, g_n$, i.e., G_i is the rough set corresponding to the concept g_i. Using the calculated similarity measures it is possible to build a similarity matrix:

$$M = \begin{bmatrix} s_{1,1} & s_{1,2} & \ldots & s_{1,n} \\ s_{2,1} & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ s_{n,1} & \ldots & \ldots & s_{n,n}, \end{bmatrix} \tag{18}$$

where, for all $i, j = 1 \ldots n$, the similarity value between g_i and g_j is calculated by using $s_{i,j} = Sim_{TG}(G_i, G_j)$, where G_i is the rough sets obtained starting from the concept g_i and G_j is the rough set calculated on the concept g_j.

The last step is constructing the terrorist groups' network by using the similarity matrix and also considering a thresholding operation. The *network building algorithm* is pretty simple. Assume $W = (V, E)$, where W is the terrorist groups' network, V is the set of all nodes of the network and $E \in V \times V$ is the set of all edges of such a network. The pseudocode of the algorithm is reported as Algorithm 1:

Take care to the constant γ that is a threshold. The value γ is used to avoid that an edge between two nodes is put into the network even if such two nodes represent terrorist groups that are not strongly related.

4. Illustrative example

In order to clarify the concepts described above, in this section an illustrative example is proposed. Such example adopts a subset of the features described in Section 3.2 and contains synthetic

Algorithm 1 Terrorist Groups' Network Building

$V \leftarrow \{g_1, \ldots, g_n\}$
$E \leftarrow \emptyset$
$W \leftarrow (V, E)$
for $i = 1 \ldots m$ **do**
  for $j = 1 \ldots m$ **do**
    if $s_{i,j} \geq \gamma$ and $i \neq j$ and $(g_j, g_i) \notin E$ **then**
      $E \leftarrow (g_i, g_j)$
    end if
  end for
end for

Table 3
Information table for the first illustrative example.

attackID	attacktype1	weaptype1	targtype1	gname
0	1	1	1	0
1	1	1	1	0
2	1	1	0	0
3	0	1	0	0
4	0	0	1	1
5	0	1	0	1
6	0	0	0	2
7	0	0	0	2
8	1	0	2	2
9	0	1	0	2
10	1	1	0	1
11	0	0	1	1
12	0	0	1	1
13	1	1	2	0
14	1	0	2	1

data that are reported in Table 3. The rows of the table contain information related to terrorist attacks. In particular, attributes `attacktype1`, `weaptype1` and `targtype1` have been used as condition attributes, whilst `gname` has been used as decision attribute. Lastly, `attackID` is the identifier of the event.

In this example the objective is conceptualizing the behavior of the terrorist group identified with `gname=0` in the information table. The first operation to execute is to define the concept that is going to be approximated. Such concept is: $X = \{0, 1, 2, 3, 13\}$. The second operation to execute is to calculate lower and upper approximations of the concept X:

$$\underline{X} = \{0, 1, 13\}, \tag{19}$$

$$\overline{X} = \{0, 1, 2, 3, 5, 9, 10, 13\}, \tag{20}$$

Lastly, the third operation to execute is calculating the three regions (POS, BND, NEG) by means of the Three-way Decision Theory as described in Section 3.5:

$$POS = \{0, 1, 13\} \tag{21}$$

$$BND = \{2, 3, 5, 9, 10\} \tag{22}$$

$$NEG = \{4, 6, 7, 8, 11, 12, 14\} \tag{23}$$

It is possible to interpret the true regions by observing the values for the considered attack features from the information table. The positive region asserts that the characterizing behavior of group 0 is summarized by attacks of type 1 to target 1 by using the weapon 1 or attacks of type 1 to target 2 by using the weapon 0. Moreover, the boundary region asserts that other behaviors could be associated to group 0. More in details, such behaviors have been acted by group 0 but also by other different terrorist groups like, for instance, 1 and 2.

This section provides an illustrative example in order to clarify the approach to build the terrorist groups' network.

Table 4 shows a set of terrorist attacks perpetrated by four different groups, namely 0, 1, 2 and 3. After the conceptualization step accomplished by using the approach described in

Table 4
Information table for the second illustrative example.

attack ID	attacktype1	weaptype1	targtype1	gname
0	1	1	1	0
1	1	1	1	0
2	1	1	0	0
3	0	1	0	0
4	0	0	1	1
5	0	1	0	1
6	0	0	0	2
7	0	0	0	2
8	1	0	2	2
9	0	1	0	2
10	1	1	0	1
11	0	0	1	1
12	0	0	1	1
13	1	1	2	0
14	1	0	2	1
15	2	2	2	3
16	2	2	2	3

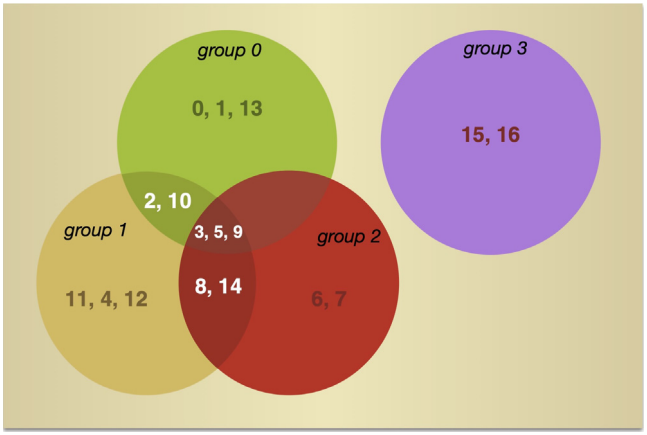

Fig. 3. Terrorist groups' network for the second illustrative example.

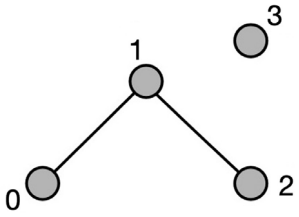

Fig. 4. Terrorist groups' network for the second illustrative example.

Sections 3.3–3.5 we obtain the representations of the above four terrorist groups' behaviors as reported in Table 5.

The BND regions for the four groups are depicted in the diagram reported in Fig. 3.

After applying the formula (17), the following similarity matrix is obtained:

$$M = \begin{bmatrix} 1.0 & 0.5 & 0.0 & 1.0 \\ 0.5 & 1.0 & 1.0 & 0.0 \\ 0.0 & 1.0 & 1.0 & 0.0 \\ 1.0 & 0.0 & 0.0 & 1.0 \end{bmatrix} \tag{24}$$

The similarity matrix M has to be processed by the Algorithm 1, assuming $\gamma = 0.6$ in order to obtain the terrorist groups' network reported in Fig. 4.

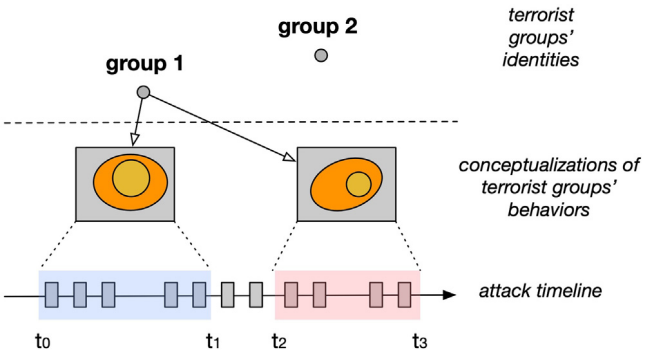

Fig. 5. Different conceptualizations for the behaviors of the same terrorist group.

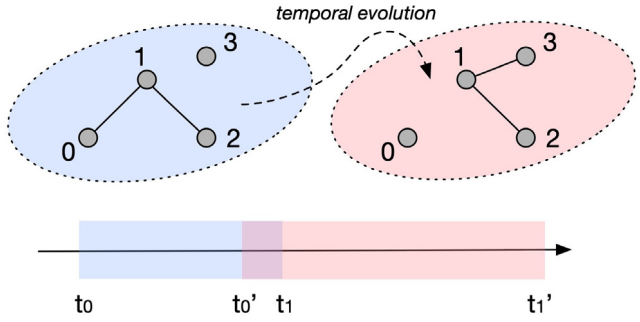

Fig. 6. Temporal evolution of terrorist groups' networks.

5. Analyzing the temporal evolution of the network

Some additional considerations must be done with respect to the temporal evolution of terrorist groups' behaviors. First of all, the proposed approach depends on the used dataset, therefore it also depends on the attacks included into such dataset. Moreover, if we choose the dataset by chronologically considering all the attacks and select only a specific time window we obtain

Table 5
Results for the conceptualization of groups in the information Table 4.

group	POS	BND
0	{0, 1, 13}	{2, 3, 5, 9, 10}
1	{11, 4, 12}	{2, 3, 5, 8, 9, 10, 14}
2	{6, 7}	{3, 5, 8, 9, 14}
3	{15, 16}	{}

that the representation of groups' behaviors is affected by the characteristics of the attacks perpetrated in such a window. This situation is graphically represented in Fig. 5, where it is possible to observe that *group1* is conceptualized in two different ways given that two different time intervals ($[t_0, t_1]$ and $[t_2, t_3]$), and thus two different attack sets, have been considered in the conceptualization process. Secondly, if we select two time intervals $[t_0, t_1]$ and $[t'_0, t'_1]$ where $t_0 < t'_0$ and $t_1 < t'_1$, extract the attacks corresponding to such windows from the whole dataset and construct the conceptualizations for the same group g, it is possible to analyze the temporal evolution of the behavior of this terrorist group. Thirdly, if we construct the terrorist group's networks for two different time intervals, characterized as done above, it is possible to analyze the temporal evolution of the relationships among groups ([28] and [29]). In particular, we can observe the evolution reported in Fig. 6 where two networks have been constructed by using two time intervals. It is possible to note that the relation between groups 0 and 1 disappeared and the relation between 1 and 3 appeared. Moreover, the relation between 1 and 2 appears in both the first and the second interval. Temporal evolutions of relations are due to possible changes in groups' behaviors along the timeline. In fact, groups maintaining their behaviors tend to maintain also their relational states, groups that change their behaviors tend to modify the relations they are involved in. Lastly, it is possible that groups that change their behaviors maintain the relations among them because changing their behavior accordingly.

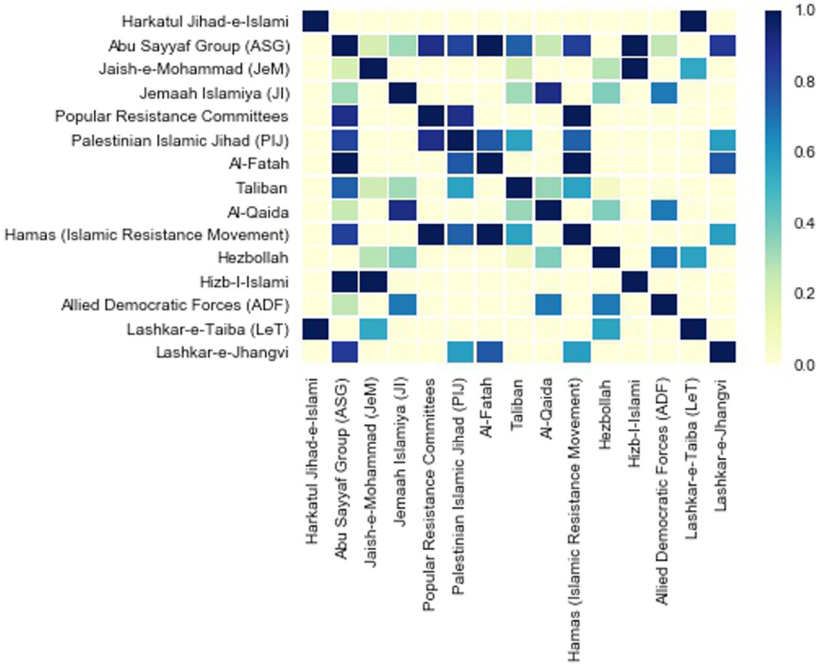

Fig. 7. Similarity matrix for interval [2000–2002].

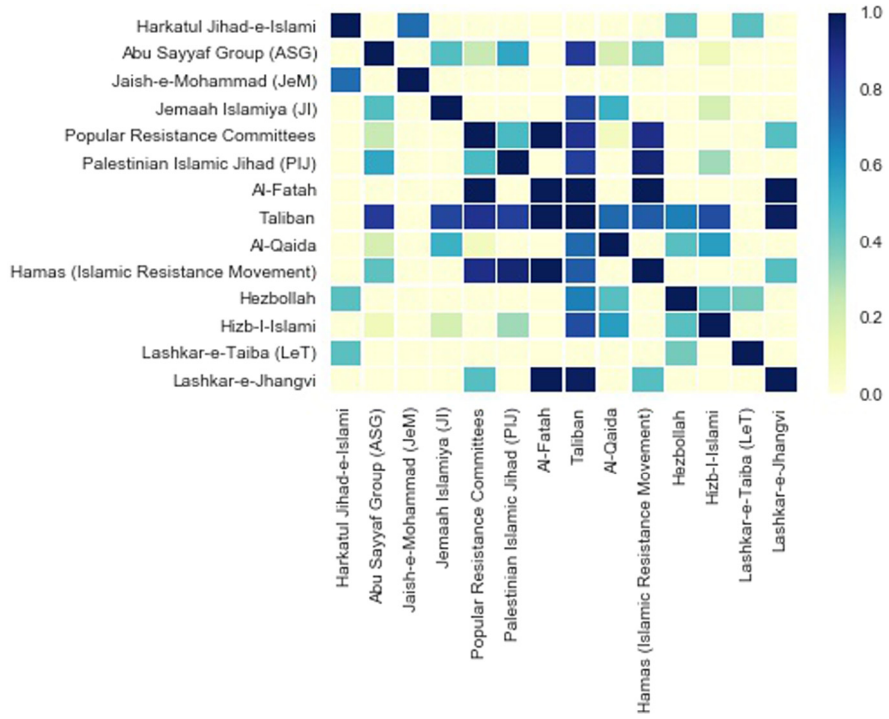

Fig. 8. Similarity matrix for interval [2003–2005].

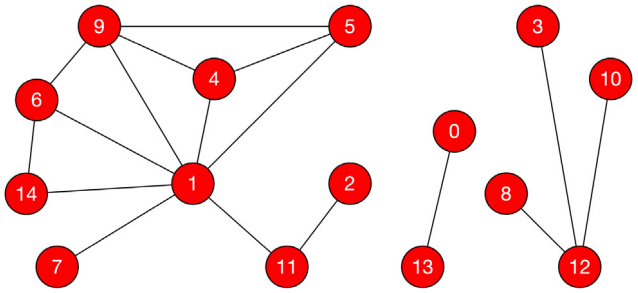

Fig. 9. Terrorists' network in the interval [2000, 2002].

6. Case study

In order to evaluate the proposed approach we provide a case study focused on the analysis of terrorist attacks from 2000 to 2011 included in the GTD. More in details, the case study foresees the construction of terrorist groups' networks in four time intervals and the assessment of the results by comparing them to networks created by experts and retrieved from the *Big, Allied and Dangerous (BAAD)*[5] [30] online platform that offers relationship information and social network data on 50 of the most notorious terrorist organizations in the world since 1998, with additional network information on more than 100 organizations.

6.1. Selected data and parameters

In order to create the information tables to start the conceptualization process, it was needed to identify the time intervals to consider and filter data (from the GTD) with respect to such intervals. In particular, the considered time intervals are: (1) [2000, 2002], (2) [2003, 2005], (3) [2006, 2008], and (4) [2009, 2011]. Moreover, the features used for the case study are 11 (`attacktype1`, `weaptype1`, `suicide`, `targtype1`, `INT_LOG`, `INT_IDEO`, `ishostkid`, `crit1`, `crit2`, `crit3`, `gname`), where the first ten are condition attributes and the last one is a decision attribute. More in details, the first interval induces an information table of 5052 terrorist attacks, the second one induces an information table of 4433 attacks, the third interval consists of 10793 attacks and the fourth interval contains 14612 attacks. Lastly, the threshold γ was set equal to 0.6.

6.2. Results and evaluation

In this section, the similarity matrices and networks for all intervals will be presented and discussed with respect to the expert knowledge coming from the BAAD database. In particular, we have considered the following terrorist groups: (0) Harkatul Jihad-e-Islami, (1) Abu Sayyaf Group (ASG), (2) Jaish-e-Mohammad (JeM), (3) Jemaah Islamiya (JI), (4) Popular Resistance Committees (PRC), (5) Palestinian Islamic Jihad (PIJ), (6) Al-Fatah, (7) Taliban, (8) Al-Qaida, (9) Hamas (Islamic Resistance Movement), (10) Hezbollah, (11) Hizb-I-Islami, (12) Allied Democratic Forces (ADF), (13) Lashkar-e-Taiba (LeT), (14) Lashkar-e-Jhangvi. The group selection has been accomplished by considering the groups that are more active in the time window [2000, 2011]. From such set, we have extracted the terrorist groups that are hubs for the star networks, provided by the BAAD database, which constitute the ground truth in the evaluation process. The similarity matrices, for the first two time intervals (indicated above and calculated by using the approach proposed in this work) are represented in Figs. 7 and 8 as heatmaps, where it is already possible to note an evolution of the similarity measures among the same terrorist groups.

In particular, it emerges that the similarity between Palestinian Islamic Jihad (PIJ) and Hamas (Islamic Resistance Movement) becomes stronger in the second interval (with respect

[5] Funded through the Department of Homeland Security's Science and Technology Directorate's Office of University Programs, the underlying BAAD database - https://www.start.umd.edu/baad/database - was created and is maintained by the Project on Violent Conflict at the University at Albany's Rockefeller College of Public Affairs and Policy.

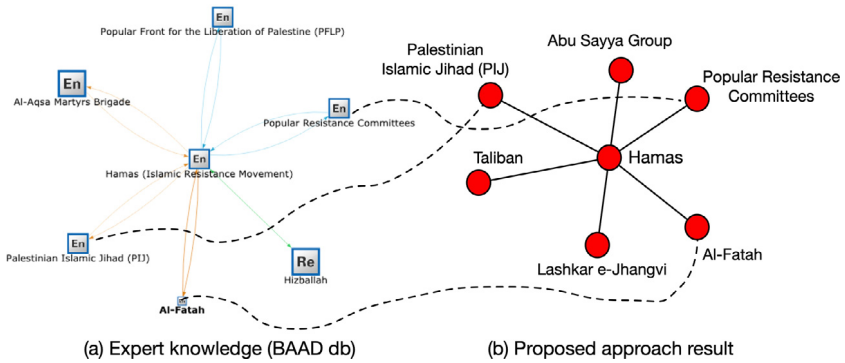

Fig. 10. Hamas network in the interval 2002 from expert knowledge (BAAD database).

Table 6
Precision, Recall and F-Measure for evaluating results.

Terrorist group	Interval	Precision	Recall	F-Meas.
Hamas	2000–2002	0.5	0.75	0.6
	2003–2005	0.5	0.75	0.6
	2006–2008	0.5	0.75	0.6
	2009–2011	#	#	#
PIJ	2000–2002	0.33	0.67	0.44
	2003–2005	0.4	0.67	0.5
	2006–2008	0.4	1	0.57
	2009–2011	0.25	0.5	0.33
PRC	2000–2002	0.33	1	0.5
	2003–2005	0.4	1	0.57
	2006–2008	0.4	1	0.57
	2009–2011	1	0.5	0.67

to the first one) and the similarity between Popular Resistance Committees and Palestinian Islamic Jihad (PIJ) does not reach the threshold in the first interval but appears in both the second interval and the third interval. Moreover, it is interesting to observe that the similarity between Palestinian Islamic Jihad (PIJ) and Al-Fatah is evident in the first interval but decreases along the timeline and disappears in the second and third intervals. All these insights are fully confirmed by the considered expert knowledge.

The second step is the construction of the terrorist groups' network starting from the similarity matrices, one for each matrix. The network constructed for the interval [2000, 2002] is reported in Fig. 9, where the node IDs are associated to terrorist groups' names at the beginning of this section.

In order to evaluate the results of the proposed approach, we have compared the constructed terrorist groups' networks to the one obtained from the BAAD database (see example in Fig. 10 for the Hamas terrorist group). Such comparison has been realized, for all the considered time intervals, by calculating *precision*, *recall* and *F-measure*. The set of relevant terrorists' groups (taken by the expert knowledge) is represented by *KTG* and the set of retrieved terrorists' groups (obtained by using the proposed approach) is represented by *ATG*. Thus, it is possible to define:

$$precision = \frac{|KTG \cap ATG|}{|ATG|}, \tag{25}$$

$$recall = \frac{|KTG \cap ATG|}{|KTG|}, \tag{26}$$

$$F\text{-}measure = 2 \times \frac{precision \times recall}{precision + recall}. \tag{27}$$

The approach offers promising results in terms of precision, recall and F-measure. In particular, for the whole set of groups that have been analyzed, the constructed networks for around the 75 percent of the above groups provides *F-measure* values equal or greater than 0.5. Table 6 partially reports the results of the last phase of the evaluation activity. In particular, in the table, it is possible to find the results for terrorist groups present in at least three of four time intervals. Missing results are mostly due to the fact that the BAAD database does not offer for every year the information about networks for all terrorist groups. This is motivated by a lack of knowledge with respect to specific groups in specific periods.

7. Conclusions

The paper provides an original approach to elicit terrorist group networks from a database of terrorist events by using a similarity function based on rough set theory. The approach has been described and illustrated by providing several illustrative examples. Considerations on the time dimension have been also provided. Furthermore, the proposed approach has been demonstrated and evaluated by using a Python implementation of rough set operators (realized by the authors) and comparing the results to the expert knowledge obtained by the resources provided by the BAAD database (START project). Given the promising results, the approach could have the potential to be included in the analysis frameworks adopted by counter-terrorism operators. Limitations of the results provided by this work are: (i) the lack of a method to automatically deal with time intervals, (ii) the lack of a method to automatically analyze the temporal evolution of terrorist groups' networks, and (iii) the incapacity to determine the semantics of relations in the elicited networks. All the above aspects could be investigated and, in particular, the authors have already planned several enhancements for the proposed approach: (i) a systematic extraction of relevant features from the original dataset including 135 features (new relevant features could emerge and improve the expert-based approach adopted in the present work), (ii) the definition of a semantic similarity function for terrorists' groups (it could be interesting to inject semantics in the similarity function in order to consider pieces of existing and possibly incomplete knowledge to improve the approach), (iii) the evaluation of several alternatives to traditional rough sets, e.g., tolerance rough sets, neighborhood rough sets, for conceptualizing the behaviors of terrorists' groups, and (iv) the definition of an approach for the temporal evolution analysis of terrorist groups' networks.

Declaration of competing interest

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

References

[1] Z. Pawlak, A. Skowron, Rough sets and Boolean reasoning, Inf. Sci. 177 (1) (2007) 41–73.
[2] Global Terrorism Database, 2018. URL https://www.start.umd.edu/gtd.
[3] J. Lee, Exploring global terrorism data: a web-based visualization of temporal data, ACM Crossroads 15 (2) (2008) 7–14.
[4] A. Kengpol, P. Neungrit, A decision support methodology with risk assessment on prediction of terrorism insurgency distribution range radius and elapsing time: An empirical case study in thailand, Comput. Ind. Eng. 75 (2014) 55–67, http://dx.doi.org/10.1016/j.cie.2014.06.003, URL http://www.sciencedirect.com/science/article/pii/S0360835214001831.
[5] R. Lara-Cabrera, A. Gonzalez-Pardo, D. Camacho, Statistical analysis of risk assessment factors and metrics to evaluate radicalisation in twitter, Future Gener. Comput. Syst. 93 (2019) 971–978, http://dx.doi.org/10.1016/j.future.2017.10.046, URL http://www.sciencedirect.com/science/article/pii/S0167739X17308348.
[6] Z. Li, D. Sun, B. Li, Z. Li, A. Li, Terrorist group behavior prediction by wavelet transform-based pattern recognition, Discrete Dyn. Nat. Soc. 2018 (2018).
[7] G.M. Tolan, O.S. Soliman, An experimental study of classification algorithms for terrorism prediction, Int. J. Knowl. Eng. 1 (2) (2015) 107–112.
[8] S.B. Salem, S. Naouali, M. Sallami, A computational cost-effective clustering algorithm in multidimensional space using the manhattan metric: Application to the global terrorism database, World Acad. Sci. Eng. Technol. Int. J.Comput. Electr. Autom. Control Inf. Eng. 11 (6) (2017) 703–708.
[9] F. Ding, Q. Ge, D. Jiang, J. Fu, M. Hao, Understanding the dynamics of terrorism events with multiple-discipline datasets and machine learning approach, PLoS One 12 (6) (2017) e0179057.
[10] J. Górecki, K. Slaninová, V. Snášel, Visual investigation of similarities in global terrorism database by means of synthetic social networks, in: 2011 International Conference on Computational Aspects of Social Networks (CASoN), IEEE, 2011, pp. 255–260.
[11] Z. Pawlak, Rough sets, Int. J. Comput. Inf. Sci. 11 (5) (1982) 341–356, http://dx.doi.org/10.1007/BF01001956, URL https://doi.org/10.1007/BF01001956.
[12] Q. Zhang, Q. Xie, G. Wang, A survey on rough set theory and its applications, CAAI Trans. Intell. Technol. 1 (4) (2016) 323–333, http://dx.doi.org/10.1016/j.trit.2016.11.001, URL http://www.sciencedirect.com/science/article/pii/S2468232216300786.
[13] Z. Pawlak, Rough sets and intelligent data analysis, Inform. Sci. 147 (1) (2002) 1–12, http://dx.doi.org/10.1016/S0020-0255(02)00197-4, URL http://www.sciencedirect.com/science/article/pii/S0020025502001974.
[14] Y. Yao, An outline of a theory of three-way decisions, in: J. Yao, Y. Yang, R. Słowiński, S. Greco, H. Li, S. Mitra, L. Polkowski (Eds.), Rough Sets and Current Trends in Computing, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 1–17.
[15] Y. Yao, Three-way decisions with probabilistic rough sets, Inform. Sci. 180 (3) (2010) 341–353, http://dx.doi.org/10.1016/j.ins.2009.09.021, URL http://www.sciencedirect.com/science/article/pii/S0020025509004253.
[16] Y. Yao, Rough sets and three-way decisions, in: D. Ciucci, G. Wang, S. Mitra, W.-Z. Wu (Eds.), Rough Sets and Knowledge Technology, Springer International Publishing, Cham, 2015, pp. 62–73.
[17] G. LaFree, L. Dugan, Introducing the global terrorism database, Terror. Political Violence 19 (2) (2007) 181–204, http://dx.doi.org/10.1080/09546550701246817.
[18] Q. Hu, D. Yu, J. Liu, C. Wu, Neighborhood rough set based heterogeneous feature subset selection, Inf. Sci. 178 (18) (2008) 3577–3594.
[19] Y. Yao, Relational interpretations of neighborhood operators and rough set approximation operators, Inf. Sci. 111 (1–4) (1998) 239–259.
[20] X.R. Zhao, B.Q. Hu, Fuzzy probabilistic rough sets and their corresponding three-way decisions, Knowl.-Based Syst. 91 (2016) 126–142.
[21] S. Greco, B. Matarazzo, R. Słowiński, Dominance-based rough set approach as a proper way of handling graduality in rough set theory, in: Transactions on Rough Sets VII, Springer, 2007, pp. 36–52.
[22] H. Li, X. Zhou, Risk decision making based on decision-theoretic rough set: a three-way view decision model, Int. J. Comput. Intell. Syst. 4 (1) (2011) 1–11.
[23] Y. Yao, Decision-theoretic rough set models, in: International Conference on Rough Sets and Knowledge Technology, Springer, 2007, pp. 1–12.
[24] J. Meng, P. Wang, X. Wang, T.Y. Lin, Rule induction for tolerance relation-based rough sets, in: 2011 IEEE International Conference on Granular Computing, IEEE, 2011, pp. 445–450.
[25] J.P. Herbert, J. Yao, Criteria for choosing a rough set model, Comput. Math. Appl. 57 (6) (2009) 908–918.
[26] L. Egghe, C. Michel, Construction of weak and strong similarity measures for ordered sets of documents using fuzzy set techniques, Inf. Process. Manage. 39 (5) (2003) 771–807.
[27] K. Qin, Z. Song, Y. Xu, Soft rough sets based on similarity measures, in: T. Li, H.S. Nguyen, G. Wang, J. Grzymala-Busse, R. Janicki, A.E. Hassanien, H. Yu (Eds.), Rough Sets and Knowledge Technology, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 40–48.
[28] Y.-R. Lin, Y. Chi, S. Zhu, H. Sundaram, B.L. Tseng, Analyzing communities and their evolutions in dynamic social networks, ACM Trans. Knowl. Discov. Data 3 (2) (2009) 8.
[29] Y.-R. Lin, Y. Chi, S. Zhu, H. Sundaram, B.L. Tseng, Facetnet: a framework for analyzing communities and their evolutions in dynamic networks, in: Proceedings of the 17th International Conference on World Wide Web, ACM, 2008, pp. 685–694.
[30] V.H. Asal, R.K. Rethemeyer, Big Allied and Dangerous Dataset Version 2, 2015, URL http://www.start.umd.edu/baad/database.

Vincenzo Loia received the master's degree in computer science from the University of Salerno, Fisciano, Italy, in 1985, and the Ph.D. degree in computer science from the University of Paris 6, Paris, France, in 1989. He is a Full Professor of computer science with the University of Salerno. Professor Loia is the Coeditor-in-Chief of Soft Computing journal and the Editor-in-Chief of Ambient Intelligence and Humanized Computing journal. He serves as an Editor for 14 other international journals.

Francesco Orciuoli received the master's degree cum laude in computer science from the University of Salerno, Fisciano, Italy. He is an Associate Professor in computer science with the University of Salerno. He is currently focusing his research activities on semantic technologies, computational intelligence and situation awareness.