Expert Systems With Applications 78 (2017) 358–375

Contents lists available at ScienceDirect

Expert Systems With Applications

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

# New framework that uses patterns and relations to understand terrorist behaviors☆

Salih Tutun[a,b,1,2,*], Mohammad T. Khasawneh[b], Jun Zhuang[c]

[a] Turkish Military Academy, Defense Sciences Institute, Ankara, Turkey
[b] Department of Systems Science and Industrial Engineering, Binghamton University, The State University of New York, Binghamton, New York
[c] Department of Industrial and Systems Engineering, University at Buffalo, The State University of New York, Buffalo, New York

ARTICLE INFO

*Article history:*
Received 6 September 2016
Revised 19 January 2017
Accepted 16 February 2017
Available online 17 February 2017

*Keywords:*
Link formation
Feature selection
Adaptive optimization
Networks
Decision making
Homeland security

ABSTRACT

Terrorism is a complex phenomenon with high uncertainties in user strategy. The uncertain nature of terrorism is a main challenge in the design of counter-terrorism policy. Government agencies (e.g., CIA, FBI, NSA, etc.) cannot always use social media and telecommunications to capture the intentions of terrorists because terrorists are very careful in the use of these environments to plan and prepare attacks. To address this issue, this research aims to propose a new framework by defining the useful patterns of suicide attacks to analyze the terrorist activity patterns and relations, to understand behaviors and their future moves, and finally to prevent potential terrorist attacks. In the framework, a new network model is formed, and the structure of the relations is analyzed to infer knowledge about terrorist attacks. More specifically, an Evolutionary Simulating Annealing Lasso Logistic Regression (ESALLOR) model is proposed to select key features for similarity function. Subsequently, a new weighted heterogeneous similarity function is proposed to estimate the relationships among attacks. Moreover, a graph-based outbreak detection is proposed to define hazardous places for the outbreak of violence. Experimental results demonstrate the effectiveness of our framework with high accuracy (more than 90% accuracy) for finding patterns when compared with that of actual terrorism events in 2014 and 2015. In conclusion, by using this intelligent framework, governments can understand automatically how terrorism will impact future events, and governments can control terrorists' behaviors and tactics to reduce the risk of future events.



## 1. Introduction

### 1.1. Background and motivations

The amount of crisis and chaos have increased across the world due to terrorist groups that use more complex tactics and strategies that cannot be easily recognized (see in Fig. 1). In particular, suicide terrorism has become the hardest attack type for counter-terrorism, and it is an easy and cheap usage by terrorist for attractive attacks. Counter-terrorism agencies (e.g., CIA, FBI, NSA) use social media and telecommunication to capture intentions of terrorists. However, terrorists have become more careful in using social media for planning and preparing of attacks. A constant problem for intelligence services is understanding terrorism when there is no clue (Byman & Shapiro, 2014). In this regard, governments need to use new intelligence systems that discern patterns for future threats.

Predicting suicide attacks, which encompasses high uncertainty, is almost impossible. The uncertain nature of terrorism is the main challenge in the design of counter-terrorism policy. Instead of a prediction, using proper protection reduces uncertainty in the prevention of attacks (Jackson & Frelinger, 2009). In general, the challenge in protecting society from terrorism is being able to correctly identify associated activity patterns with the given information such as incident type, attack type, and weapon type. Intelligence gathering can reduce uncertainty for terrorism. Recently, innovative intelligent approaches have been widely used to analyze the terrorist activity patterns, to predict their future moves, and finally to deter potential terrorist attacks.

For innovative intelligent approaches, Knowledge Discovery from Databases (KDD) techniques can play a significant role to improve counter-terrorism and crime-fighting capabilities of intelligence and security agencies/organizations (Chen, 2011). These techniques can refer to potentially useful knowledge (previously unknown) for data. These techniques deliver convenient, easy and

☆ Fully documented templates are available in the elsarticle package on CTAN.

* Corresponding author.
*E-mail addresses:* stutun1@binghamton.edu (S. Tutun), mkhasawn@binghamton.edu (M.T. Khasawneh), jzhuang@buffalo.edu (J. Zhuang).

[1] Funded for PhD Education by Turkish Military Academy.

[2] Ph.D. Candidate in Binghamton University, The State University of New York.

http://dx.doi.org/10.1016/j.eswa.2017.02.029
0957-4174/

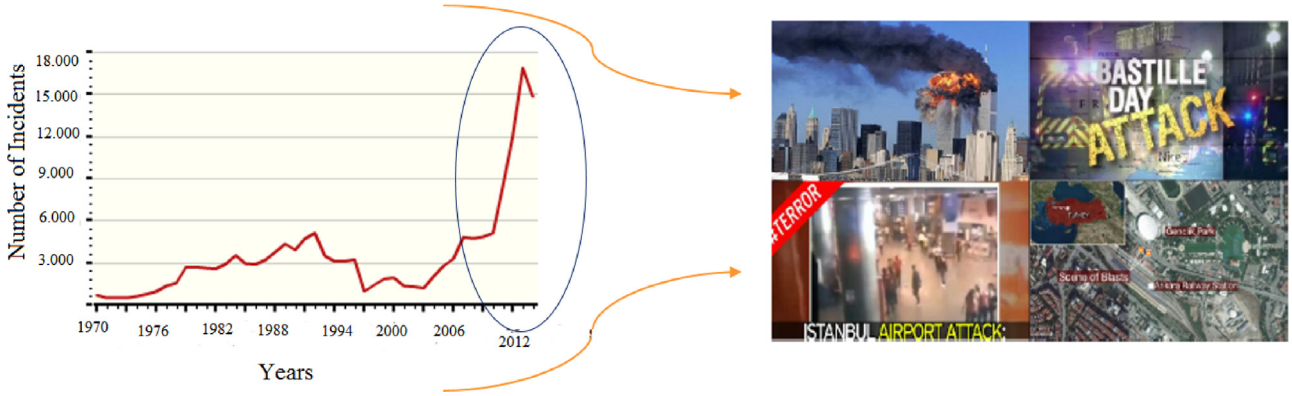


**Fig. 1.** Recent suicide bombing attacks in the world.

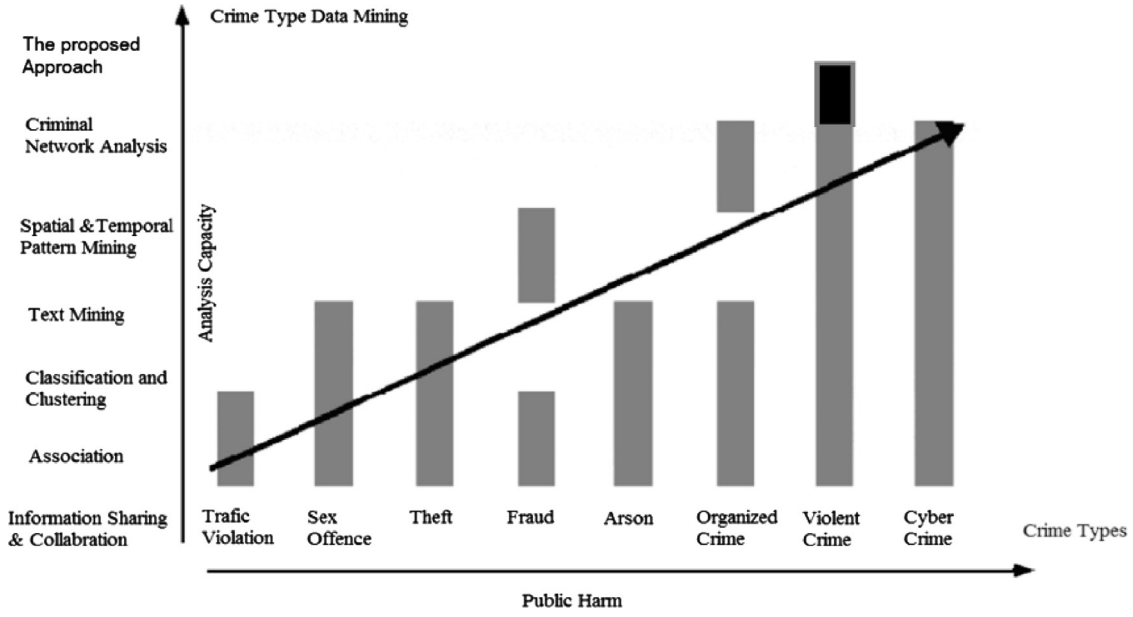


**Fig. 2.** Analysis capability of knowledge discovery techniques for crime types.

practical exploration of large data for organizations. However, understanding terrorism is a challenging issue due to the high uncertainty in terrorist strategies and tactics. In this context, graph theoretic approaches have become increasingly a key technique that allows the capture of complex interactions (Netzer et al., 2012). Due to the changing nature of terrorism concerns, a new type intelligence is demanded by counter-terrorism using network analysis. In Fig. 2, existing techniques are presented for security concerns based on information sharing and analysis capability. For this reason, we focus on a network analysis, which has a high analysis capability for violent crime (terrorism) being a significant public harm and information sharing (Chen, 2011).

Network models have been proven to be a valuable tool for understanding terrorism (Bohannon, 2009). However, most researchers have focused on understanding the behavior of individual terrorists by modeling their relationship with each other within a particular terrorist group, hoping that such information can provide insights about the leader of that specific terrorist group. In this research, we focus on the spatial and tactical relationship of different attacks rather than the connection of individuals within a terrorist group. Furthermore, current network-based approaches in literature concentrate on prosecution instead of prevention (Krebs, 2002; Xu & Chen, 2005). However, the ultimate goal of counter-terrorism agencies is preventing terrorist attacks. Hence, we incorporate data mining techniques with network analysis methods to prevent future suicide attacks in Iraq, as a case study.

In this paper, a new framework is proposed to understand the behavior of suicide attacks. The results (patterns and relations) could potentially help with the understanding of suicide attacks and enable law enforcement agencies to propose reactive strategies. The expected results will enable policy makers to develop precise global and/or local counter-terrorism policies. From the standpoint of governments, the overall goals of making counter-terrorism policy are to prevent terrorist attacks, and thereby to reduce financial, human, and political losses.

### 1.2. Contributions of this study

The network of terrorist attacks is modeled in the framework to prevent future attacks. Moreover, feature selection and similarity function are first proposed as a framework to find relations that will help construct a network for events. The primary contributions of this study are as follows:

This study covers an implementation of a network of terrorist attacks for intelligence analysis. It is an early attempt to identify meaningful patterns in suicide attacks using network models. More specifically, a new Evolution Stimulating Annealing Lasso (least absolute shrinkage and selection operator) Logistic Regression

(ESALLOR) is proposed to define the best features for similarity function, and to find the important key features (e.g., how important the features are shown for successful and interactive terrorist events). Afterward, a new heterogeneous similarity function is also proposed to estimate relationships among interactive events. Moreover, a graph-based outbreak detection is proposed to define important risk areas for terrorism by using spatial characteristics of the past events.

As a secondary level contribution, the proposed approach is different than that of other approaches because it is interested in terrorist events instead of terrorists as decision makers. The data is collected by combining various databases and sources and preprocessed to define certain attacks. The proposed framework also provides animation capability to assist in analyzing dynamic data, as well as showing the importance of detecting suicide bombing attacks in Iraq as a case study.

The rest of the paper is organized as follows. In Section 1, based on the existing literature, network analysis with link formation, feature selection, and outbreak detection are presented to define relations of terrorism events. In Section 2, the methods used in the new proposed approaches are explained briefly. Section 3 explains how to make data collection and analysis for terrorism. Section 4 describes in detail how the proposed methodology is used to understand suicide terrorist attacks in Iraq. In Section 5, the new patterns are defined for the years 2003–2013. The proposed framework is discussed to understand the suicide attacks based on finding patterns with sensitivity analysis (for the years 2014 and 2015). Finally, Section 6 shows the improvement in modeling terrorism and the contribution of the paper as a conclusion.

## 2. Literature review

Crime and tendency to violence continue to remain severe threats to the entire world with highly complex criminal activities (Hassani, Huang, Silva, & Ghodsi, 2016). Crime is not limited to the streets now because the use of the Internet causes a tendency to violence and more sophisticated behaviors in the modern age. In this section, as seen in Fig. 2, eight different criminal categories show how local enforcement and international enforcement are classified for crime types (Chang, Chung, Chen, & Chou, 2003). Also, the related data mining techniques are presented to show the effective tool for crime types by uncovering hidden patterns for investigation and preventing crime and violence by both government and private institutions. Therefore, the most important and dangerous crimes are defined as terrorist (violent) attacks, and it is the hardest crime type to capture behaviors because strategies and tactics are changed dynamically by terrorists and governments. Thus, terrorism is the most complex crime, and when the terrorism is prevented, other crimes can be prevented easily (Chen et al., 2004). As an event, there is no research to understand how terrorist attacks interact each other for future strategies because there is no dataset shows relations of events. However, there are some studies about crime events, and we searched them and improved for understanding terrorism.

Based on the behaviors of the crime and terror, use of technology is the most powerful tool to organize and make complex crime behaviors by terrorists. The need for new and efficient methods becomes increasingly important to understand complex behaviors (Kanellis, 2006). Data mining and social network analysis are described as powerful tools to prevent and understand these behaviors (Thongtae & Srisuk, 2008). Several applications have been made in recent years to help the investigations by using data mining (Nath, 2006) and social network analysis (Xu & Chen, 2005). The main challenge facing all intelligence-gathering agencies is accurately and influentially analyzing the criminal data. Thus, researchers classify existing data mining techniques into six classes: information sharing and entity extraction, association mining, classification, clustering, spatial and temporal pattern mining, and network analysis. Researchers have studied these data mining techniques for security applications (Chen, 2011), as seen in Fig. 2.

For information sharing, researchers have used the statistics-based (concept space algorithm) to associate persons, organizations, vehicles, etc. (Hauck, Atabakhsb, Ongvasith, Gupta, & Chen, 2002). Named entity extraction is also used to extract valuable information from text data (e.g., name, address, location, time, etc.,) to understand connections of persons for catching criminals (Arulanandam, Savarimuthu, & Purvis, 2014). With developed technologies, since the early 1970s, several studies have been performed as crime data mining to reduce crimes and terrorism using association, classification, and clustering methods (Perry, Berrebi, Brown, Hollywood, & Jaycocks, 2013). Cluster analysis is utilized to detect crime hot spots by automatically identifying associations from existing crime data (Agarwal, Nagpal, & Sehgal, 2013; Kalaikumaran, Karthik et al., 2012). Association is used to link crime and terrorist events and provide an informative association for discovering patterns (Usha & Rameshkumar, 2014). The researchers also used classification techniques for detection of terrorism and criminal activities, and prediction of crime and terrorist hot spots (Choi, Ko, Kim, & Kim, 2014). However, these methods are not adequate for understanding complex relationships for crimes, network analysis is used to define similar transactions, and the interaction measurements to show relationships and connections of terrorists and criminals (Chiu, Ku, Lie, & Chen, 2011; Prakash & Surendran, 2013; Sparrow, 1991). Network analysis has been used for application of fraud and criminal cases for a long time (Chen, 2011; Wang, Zhe, Kang, Wang, & Chen, 2008). After the attack on September 11, the use of network analysis for terrorism increased significantly.

For understanding terrorism, according to United States General Accounting Office report (Senate, 2004), there were at least 52 agencies planning to use data mining in 2005. There were 199 efforts, which are 29 projects about identifying terrorist activities or patterns (Fienberg, 2005). The most important one was the Total Information Awareness (TIA) program (called Manhattan project) started by the Defense Advanced Research Program (DARPA) in DARPA's Information Awareness Office (IAO), which was founded in 2002. After the September 11, attack, the TIA started to integrate information technologies (e.g., data mining) to detect potential terrorists. After 2003, TIA changed the name to Terrorist Information Program (Fienberg, 2005). Hence, this organization is focused on individuals to prevent future attacks by finding their transactions, sponsorship, and support. Moreover, another very important program is the Multi-state Anti-Terrorism Information Exchange System (MATRIX) (Clarke, 1988) that has the capability to store, analyze and exchange terrorism-related data in MATRIX database. This database has personal information about individuals. By using MATRIX and TIA systems, analysts can extract multiple links to identify terrorists or criminals. These two systems are focused on identification of terrorist profiling for potential terrorists (Clarke, 1988). A major problem with these systems is that there is no privacy for individuals, and they just focused on innocent civilians instead of the behaviors of terrorists.

For the advantages and disadvantages of the methodologies in the literature (see Table 1), the researchers mostly used clustering techniques for crime mining (Perry et al., 2013). However, this method is limited in the use of significant amounts of data due to high computational intensity (Chen et al., 2004). Classification is mostly used to predict crime trends to reduce the necessary time to identify criminal people (De Vel, Anderson, Corney, & Mohay, 2001). However, the pre-defined classification scheme is needed to use crime data mining. It also needs to have adequate training and testing data for effective classification. At the same time, these techniques are inadequate to capture complex interactions

**Table 1**
Data mining techniques for crime and violence.

| Data Mining Techniques and References | Key Techniques | Advantages and Disadvantages | Purpose |
|---|---|---|---|
| Entity Extraction (Arulanandam et al., 2014) | Named Entity Extraction (lexical lookup, rule-based, machine learning and hand-crafted rules), Natural Language Processing | It is just used to extract valuable information from text data (e.g., name, address, location, time, etc.,) to understand connections of persons for catching criminals. | Extract valuable information especially from unstructured text data (i.e., Person, Address, Location, Time, Vehicle, Nationality, Phone, Gender and Race, Crime Type, Personal Property). |
| Cluster Analysis (Hauck et al., 2002; Nath, 2006), (Kalaikumaran et al., 2012), (Agarwal et al., 2013) | GIS (Geographic Information System), Self-Organizing Map, Hierarchical Clustering Technique, Partitioning Clustering Technique, Co-occurrence Analysis, K-Means Clustering | It is limited in the use of significant amounts of data due to high computational intensity | Detect crime hot spots; automatically identify associations from existing crime data and weight relationships to detect the strongest association among all possible pairs of crime related entities |
| Association Rule (Usha & Rameshkumar, 2014) | Apriori Algorithm, Association Rule Mining, Distributed High Order Text Mining, Temporal Association Rule | It is used to link crime and terrorist events and provide an informative association for discovering crime patterns. | Link crime incidents, provide informative association, discover crime patterns |
| Classification (Chen et al., 2004), (Kalaikumaran et al., 2012), (De Vel et al., 2001) | Iterative Dichotomiser 3, C4.5, STAGE, CART, Hunt's Algorithm, Deceptive Theory | It is mostly used to predict crime trends to reduce the necessary time to identify criminal people. However, the pre-defined classification scheme is needed to use crime data mining. It also needs to have adequate training and testing data for effective classification. At the same time, these techniques are inadequate to capture complex interactions and terrorism. | Efficient detection of specific criminal activities among large-sized data sets; Categorize crime data; Predict crime hot spots |
| Network Analysis (Sparrow, 1991), (Chiu et al., 2011), (Prakash & Surendran, 2013) | K-core, Core/periphery Ratio; Measure of Centrality, Closeness and Betweenness; Center Weights Algorithm; Borgatti's Key Player Approach; | It is used to describe the roles and interactions of nodes in a network. Researchers have used this technique to demonstrate a criminal's role and the flow of information. It is used to visualize criminal networks. However, the researchers focused on individuals (people) instead of individuals' behaviors. | Provide analyses of functions, structures and the interaction measurements, detect relationships and connections of criminal individuals, identify key members and interaction patterns |

and terrorism. For this reason, social network analysis is used to describe the roles and interactions of nodes in a network (Coffman & Marcus, 2004). Researchers have used this technique to demonstrate a criminal's role and the flow of information. It is used to visualize criminal networks. However, this technique is used more commonly for prosecution than for prevention in literature (Akgun, Kandakoglu, & Ozok, 2010; Wang et al., 2008). Some researchers also investigate and study criminal and terrorist networks, but mostly focus on terrorist networks as decision makers in order to identify the leaders using network analysis (Krebs, 2002; Xu & Chen, 2005). Furthermore, the researchers have not focused on quantitative research because terrorism data (relational data) are not available for the public to view due to security issues (Telesca & Lovallo, 2006). However, the researchers focused on individuals (people) instead of individuals' behaviors. In order to analyze terrorism events, the researchers need to use intelligence methods by using new approaches (e.g., feature selection and similarity function, etc.) to look for relations. A new framework can be proposed to understand and detect terrorist activities. It can capture complex relations and interactions to understand terrorist' tactics and behaviors by finding relations (similarities) among terrorist attacks. It is also robust because the ESALLOR model is proposed to select and define the significance of the features. Finally, it is related to events instead of people to define behaviors of terrorists.

Therefore, because terrorists are more careful with their transactions, sponsorship, and support, methodologies in the literature could be inadequate to understand terrorist activities. If terrorists use technologies and social media carefully, we can extract suspicious activities based on the past events. Network analysis is the most powerful approach for understanding terrorism and criminal activities because it captures complex interactions. However, researchers are focused on people and their relations. The framework can easily understand interactions and uncontrolled learning by terrorists for future attacks. Governments can understand what are the popular tactics to make a provision for defined locations. Before describing the details of our framework, we briefly review some related works, introduce existing terrorism network models, feature selection methods, and discuss similarity functions. Therefore, we can use feature selection, similarity function, and network analysis to propose a new framework to understand the behaviors of terrorist events.

### 2.1. Feature selection methods

Since the early1970ʼs, several studies on dimension reduction were performed using various methods (Moradi & Rostami, 2015). High dimensional data are a big challenge for pattern recognition approaches. In literature, there are different methods to reduce dimensionality, such as Principle Component Analysis (PCA) and Independent Component Analysis (ICA) (Hyvärinen & Oja, 2000; Song, Guo, & Mei, 2010). However, these approaches capture only features that are of significant variance, and they only work for numerical features. Therefore, they perform ineffectively for a subset of features that are informative for supervised and unsupervised learning approaches (Zhang & Hancock, 2011). Moreover, feature selection is also used to reduce dimensions of data in the literature. Existing feature selection methods are explained briefly with five categories such as filter, wrapper, embedded, hybrid and graph-based methods in recent years.

Some researchers used a filter approach as an interesting statistical analysis without learning algorithm for feature sets. This can be fast because the learning algorithm is not used for selection. Some researchers used specific measures, such as Fisher score (Gu, Li, & Han, 2012), information gain (Yu & Liu, 2003),

**Table 2**
Dynamic terrorism dataset collection phases by collection institution. Note: PRIVATE is private collection by authors.

| Dates of attacks | PGIS | CETIS | ISVG | START | RAND-MIPT | PRIVATE |
|---|---|---|---|---|---|---|
| 1/1/1970 – 12/31/1997 | X | | | X | X | |
| 1/1/1998 – 3/31/2008 | | X | | X | X | |
| 4/1/2008 – 10/31/2011 | | | X | X | | |
| 11/1/2011 – 12/31/2015 | | | | X | | X |

gain ratio (Mitchell, 1997), and Laplacian score (He, Cai, & Niyogi, 2005) to rank the importance of these features. However, they ignore future dependency that can reduce performance because each feature is examined separately with these measures. To fill this gap, researchers are focusing on multivariate filter methods by grouping to explain their dependency (Ferreira & Figueiredo, 2012).

For the wrapper approach, each subset is appraised by using black-box learning algorithms, such as ANN (artificial neural network) and GA (genetic algorithm), that are able to find optimal features for high prediction accuracy (Chandrashekar & Sahin, 2014; Moradi & Rostami, 2015). However, these methods need high computational time and could provide incorrect results with a large number of features due to learning algorithms that are used in the evaluation of subsets. Generally speaking, in the literature, sequential feature selection, and heuristic search are used for wrapper approaches (Moradi & Rostami, 2015; Zorarpacı & Özel, 2016). In hybrid approaches, the researchers took advantage of filter and wrapper approaches by improving accuracy and computational running time. Moreover, feature selection could be the combination of a search technique to select the key features (Gunasundari, Janakiraman, & Meenambal, 2016; Hsu, Hsieh, & Lu, 2011). In order to use all methods to select the best features with high accuracy, embedded approaches are used (Archibald & Fann, 2007). In the embedded approach, feature selection is joined as a piece of the training process (Archibald & Fann, 2007). This approach is a catch-all group of methods that uses feature selection as model construction process to improve accuracy in selecting the best feature subsets. This approach is the best approach for catching the best accuracy. The new model is proposed by improving this approach for robust results. Moreover, it is defined to select the best features, and weighted the features. In this paper, the new ESALLOR model is proposed to find the optimal feature set with feature weights that are used in the similarity function.

### 2.2. Similarity function

There are various books for clustering analysis that discuss similarity between categorical and numerical features (Kaufman & Rousseeuw, 2009). For continuous features, Minskowski (e.g., Manhattan and Euclidean) distance (Kaufman & Rousseeuw, 2009) is usually used to calculate the distance between two points. For categorical features, these methods do not work because categorical features cannot be ordered (Boriah, Chandola, & Kumar, 2008). The researchers used the overlap measure in the literature (Stanfill & Waltz, 1986). This measure to find similarity between two points assigns a similarity as 1 (if the points are similar) or 0 (if the points are not similar). Moreover, some researchers improved this measure for categorical data and proposed a heterogeneous distance function (Wilson & Martinez, 1997). However, they just combined overlap measure and Minskowski distance. Therefore, similarity learning can be used to form relationships between nodes (Scholz, 2010). In order to form relationships between nodes, we need to use effective similarity function for categorical and numerical variables. We proposed the new weighted heterogeneous similarity function by adding weights (from the ESALLOR model) for each feature and combining distance measures with the probability of frequency measure. Hence, this robust similarity function can be used to define relationships among events.

## 3. Data collection and analysis

In general, one challenge is to identify associated activity patterns. The reason for this challenge is that it lacks terrorist data in depth due to confidentiality, and some data just do not exist. Due to technologies with high levels of capability, a considerable amount of data about terrorist activities were acquired and released for counter-terrorism research purposes.

In this section, a description of the data, as well as the data analysis behind it, will be provided. Our study is based on a sample of terrorist attacks in the field of counter-terrorism. The data sets are combined from five different databases, as seen in Table 2. The GTD (Global Terrorism Databases) by the Study of Terrorism And Responses to Terrorism (START) is used as the original platform of this research (START, 2015). Other four databases are added to improve the quality of this terrorism data set.

The data are collected with historical incidents of domestic and international terrorism. As seen in Table 2, the first phase of data (between the years 1970 and 1997) were collected by Pinkerton Global Intelligence Service (PGIS). Afterward, data between the years 1998 and 2008 were collected by the Center for Terrorism and Intelligence Studies (CETIS) with START (START, 2015). Then, data between the years 2008 and 2011 were collected by the institute for the Study of Violent Groups (ISVG) (START, 2015). For recent years, data between the years 2011 and 2015 were collected by START. At the same time, data from RAND Database of Worldwide Terrorism Incidents (RDWTI) were used to improve the collected data (Division, 2016). Therefore, we have over 140,000 incidents and approximately 75 features for each incident after preprocessing.

Data preprocessing is made by cleaning missing values and non-terrorist attacks to strengthen precision in the data. Subsequently, we analyzed the collected data to understand which locations have a challenging problem in recent years. As seen in Fig. 3, suicide attacks that use bombs have increased more recently in Iraq than compared to that of previous years. In recent years, the most frequent attacks are the bombing by suicide attacks in Iraq. At the same time, terrorist groups have become more successful, as seen in Fig. 4. One of our objectives for this research is focused on this problem so that future attacks can be prevented from occurring. Data between the years 2003 and 2013 are used to model the network of terrorist attacks. Finally, data for the years of 2014 and 2015 are used for testing of our framework.

## 4. Methodology

In this section, a new (terrorism) network that can be used to prevent future threats is discussed in detail. Moreover, a new model (ESALLOR) is explained to select the subset of relevant features for similarity function. Then, a new heterogeneous similarity function is proposed to define links between nodes (attacks). In addition, we propose a graph-based outbreak detection for defining

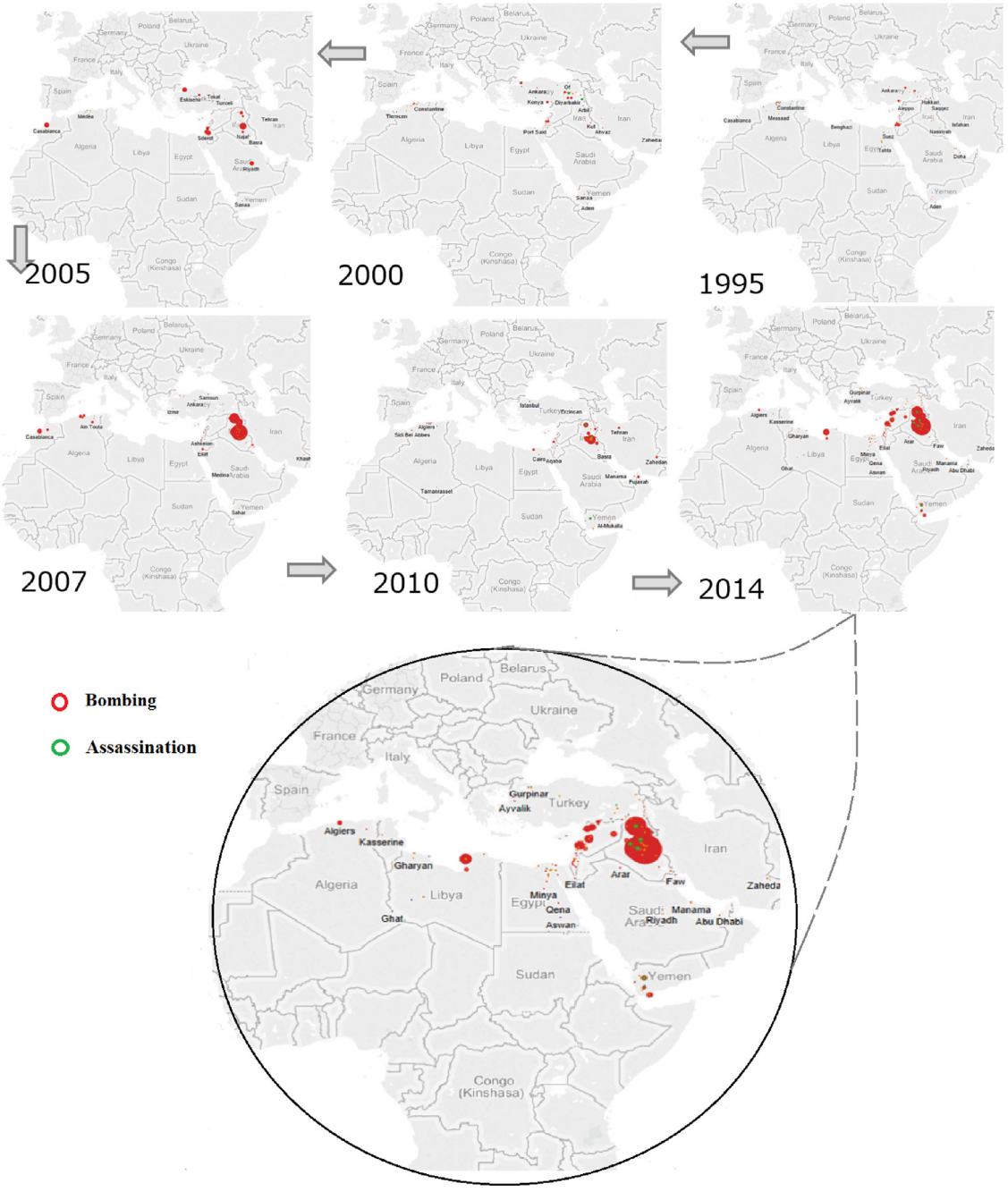


**Fig. 3.** Data analysis for suicide attacks in the world. Note: Red circles are bombings, and green circles are assassinations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

outbreak locations of terrorist attacks. Finally, patterns for suicide attacks are obtained by the proposed approaches to prevent future threats, as seen in Fig. 5.

### 4.1. The proposed methodology for the selecting of key features

Feature selection is critical to identifying a subset of features, which includes information clustering and classification. It aims to reduce dimensionality for the future space and to expedite and reduce the cost for the learning algorithm (Zhang & Hancock, 2012). The new model is offered to select key futures by using evolutionary strategy and adaptive simulated annealing with Lasso logistic regression.

#### 4.1.1. Hybrid meta-heuristic approach

In the model, simulated annealing (SA) is a random search technique (Kirkpatrick, 1984) and a single-based optimization approach. The base of the idea was first presented by Metropolis in 1953. Afterward, Kirkpatrick (1984) offered a simulation search model by using the annealing approach to get an optimal solution. The algorithm mimics the annealing process in materials physics as metals freeze and cool into a crystalline state with minimum energy level by using bigger crystal sizes to decrease defects. The efficiency of the algorithm for optimization depends on the control of temperature and cooling schedule. Moreover, in order to move to new solutions, the algorithm uses random walk, which describes the movement of the algorithm by searching randomly for the current solution to a neighborhood solution in order

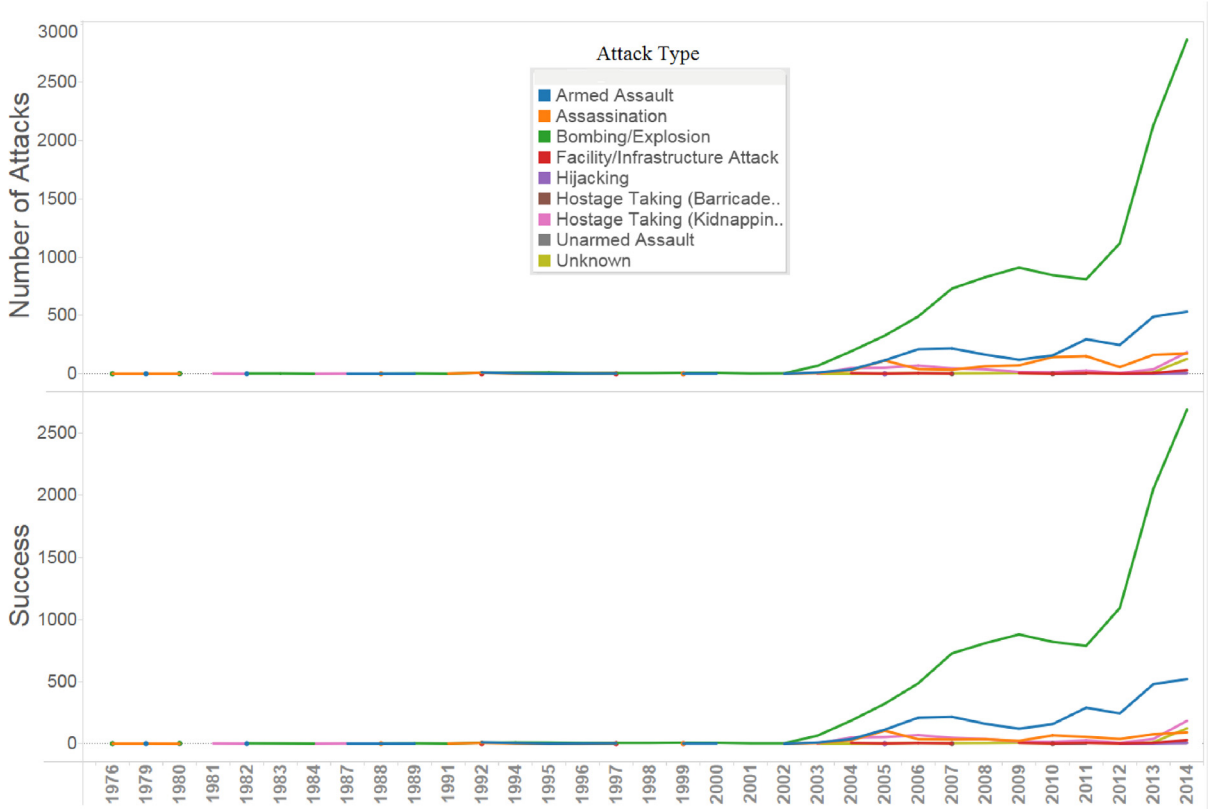


**Fig. 4.** Attack types and success of attack types in Iraq. Note: Comparing how the attacks are successful and how they increased in recent years.

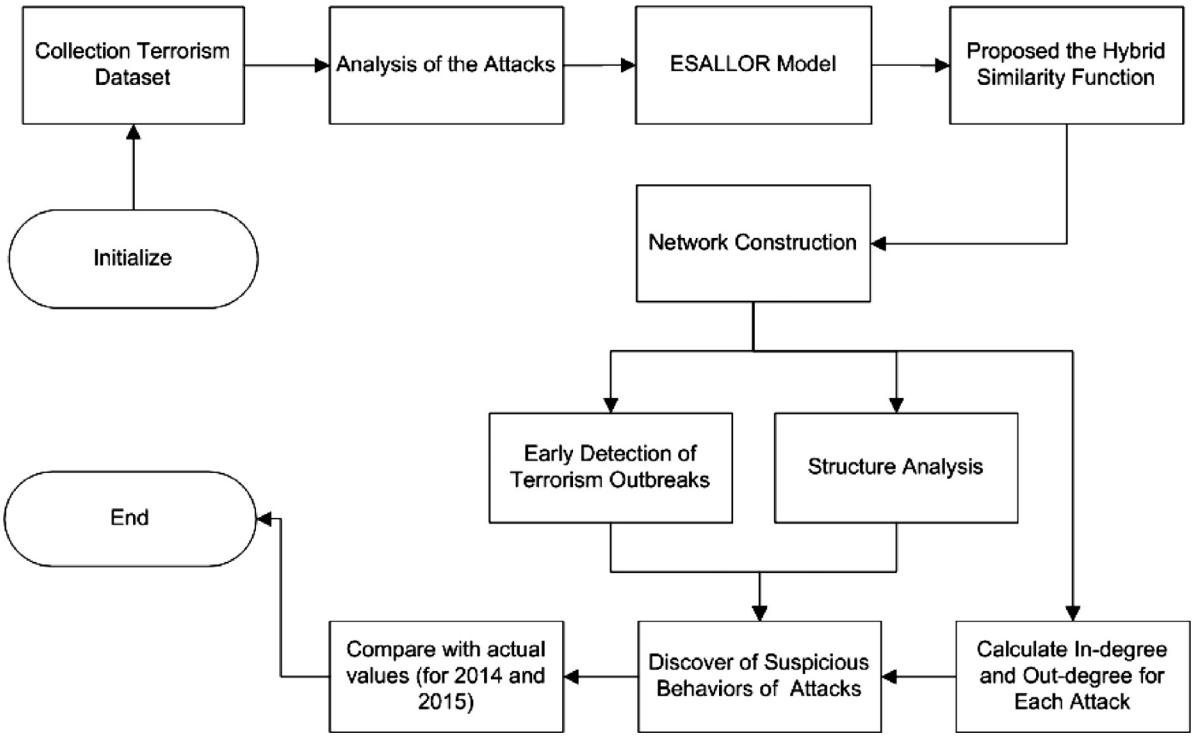


**Fig. 5.** The flow chart of the new framework for discovering of suspicious behaviors.

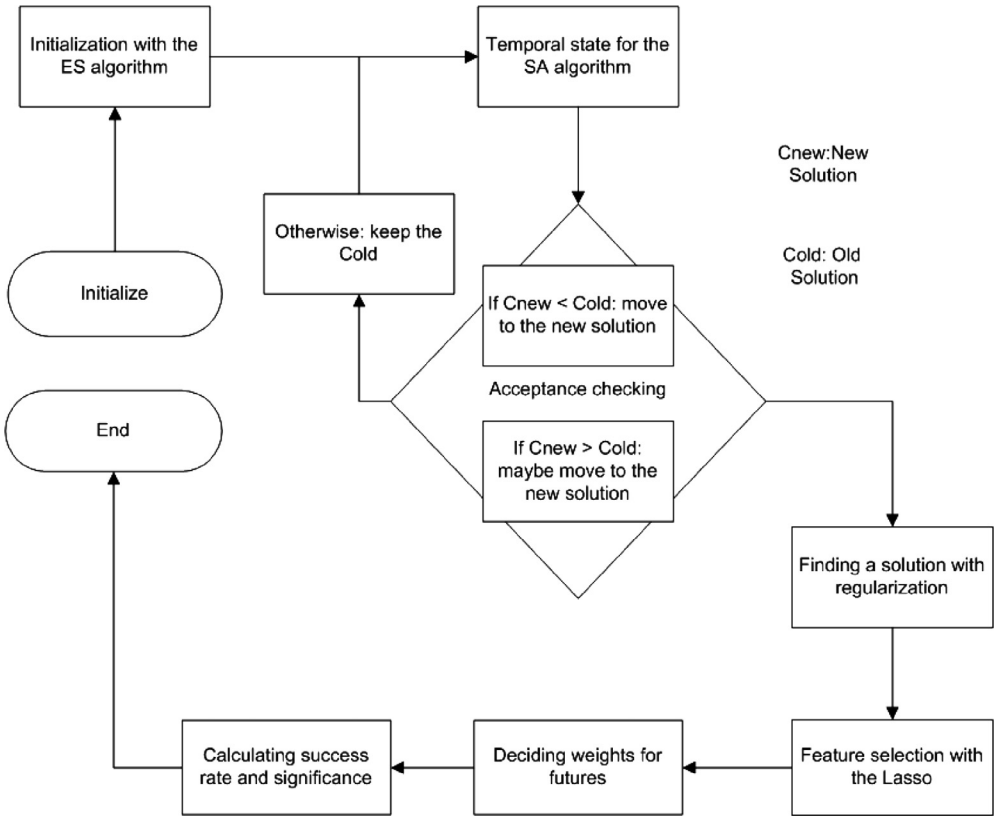


**Fig. 6.** The ESALLOR model for selecting key features.

to explore the optimal feasible solution (Kirkpatrick, 1984). In addition, the temperature is reheated when the new solution is not suitable for movement, and the method is made adaptive to prevent premature convergence.

In Evolutionary Strategy (ES), new solutions (called children) are compared with old solutions (called parents). ES is used to find a good initial solution for the simulated annealing method because it is a population-based algorithm, which can search out more solutions for the global optimum in large search areas. At the same time, these methods can cause over-fitting because a meta-heuristic approach is used. In order to eliminate this situation, regularization is added to the objective function. Lasso is used for regularization of coefficients to find the optimal regression model by optimizing parameters with a hybrid based on the ES and SA algorithms.

The proposed model, as seen in Fig. 6, is described in the following steps:

**Step 1:** Initialization with the ES algorithm that sets the boundaries of parameters. The initial values of the parameters are then generated for the model. If the ES has better offspring as parents, the standard deviation for movement to a new solution is decreased for the adaptive model.

**Step 2:** Temporal state for the SA algorithm that makes a random move to change the current system state by using the optimal initial parameters for the ES.

**Step 3:** Acceptance checking that looks at the following equations to understand whether there is acceptance or rejection of the temporal state. If there is rejection, the temperature is reheated as the adaptive model in the SA.

⋆The temporal state is accepted if the energy of the new solution is greater than the energy of the old solution and $p$, which is a random number, is less than $P$, which is the accepted rate with the new solution as $0 \leq p \leq 1$.

⋆The temporal state is accepted if the energy of the new solution $\leq$ the energy of the old solution.

⋆The temporal state is otherwise rejected.

**Step 4:** Finding a solution with regularization that finds the optimal solution by comparing all solutions. The algorithm with regularization (the Lasso) is also checked for over-training by comparing testing and training errors.

**Step 5:** Feature selection with the Lasso: The algorithm uses the Lasso regularization to improve the subset of features for analysis.

**Step 6:** Deciding weights ($\beta$) for futures: These values give the importance of features to use in the proposed similarity function.

**Step 7:** Calculating success rate (Eq. (1)) of attacks as output: Successful rate is defined to select attractive attacks. If the success rate for the event is less than 0.2 or more than 0.8, events are defined as attractive.

*4.1.2. Coupling evolutionary strategy with simulated annealing*

In the proposed model, there are absolute values that challenge the calculation of the parameters (weights) in the formulation, as seen in Eq. (2). In order to solve this problem, the hybrid meta-heuristic approach is used to optimize the parameters. ES is used to find initial solutions for decision variables (coefficient of the model) by giving initial ranges. Thereafter, by using the SA based on a single solution, the algorithm searches the neighborhood of the initial solution because a random walk is used for the next solution. It moves to new solutions for decision variables by using a normal random number. This means that the algorithm might get stuck unless it has a good initial solution. For instance, as seen in Fig. 9, it begins to find solutions from $S_0$ to $S_3$. After arriving at $S_3$, the algorithm tends to accept this point as the optimal solution for decision variables, even though it is only a local optimum.

The algorithm needs to search in a global way to find the optimum solution. Thus, the ES algorithm can find good (close to

the optimal) initial solutions that can be used in the SA algorithm. When we began with these solutions, the SA algorithm found the optimal solution by looking in the neighborhood of initial solutions.

#### 4.1.3. Formulation of the new model

In the literature, researchers use linear and quadratic regressions to obtain a new model for modeling. However, when they use a meta-heuristic approach for the training of the model, researchers need to consider over-training (Tutun, Chou, & Canıyılmaz, 2015). In the model, the Lasso regularization is used to prevent over-training. They are regression methods that involve penalizing the absolute and square size of the regression coefficients. In the formulation, the model is first decided by using logistic regression to estimate the probability of success for events, as seen in Eq. (1).

$$F_1 = \frac{1}{1 + e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2+\beta_3 x_3+\ldots+\beta_d x_d)}} \tag{1}$$

Lasso is used for the ESALLOR model, as seen in Eq. (2) for regularization of coefficients. It is also used as an objective function (Eq. (2)) in order to optimize the coefficients (decision variables).

For a given value of $\lambda > 0$,

$$min_{\beta_0,\beta_1,\beta_2,\beta_3,\ldots,\beta_d}\left(\frac{1}{2N}\sum_{i=1}^{N}(Y_i - (F_1))^2 + \lambda\sum_{j=1}^{d} \mid \beta_j \mid\right) \tag{2}$$

Where $d$ is the number of features used the in model, $Y_i$ is the response of the event $i$, $F_i$ is the estimated value of output, $N$ is the number of events, $x_i$ is data, a vector of d values at event $i$, $\lambda$ is a nonnegative regularization parameter, the parameters $\beta_0$ and $\beta$ are scalar and d-vector, respectively. As $\lambda$ increases, the number of nonzero components of $\beta$ decreases (Tutun et al., 2015).

### 4.2. A weighted heterogeneous similarity function

This paper explores the opportunities for the application of network analytic techniques to make provisions before terrorist attacks. In order to form links between nodes, similarity function can be used to measure similarities (relations). However, computing categorical data similarity is not straightforward because there is no clear ordering among categorical variables. A new data-driven heterogeneous similarity function is proposed to solve this problem.

For an overlap measure between categorical data, we define the notations as categorical data set $D$ that contains $N$ objects. This data set has $d$ categorical features and continuous features where $F_h$ denotes the $h$th feature, as seen below in the matrix. Let the feature $F_h$ take $n_h$ values in the data set $D$.

$$\begin{bmatrix} & \textit{Features ID} & F_1 & F_2 & F_3 & \cdots & F_d \\ & n_1 & x_{11} & x_{12} & x_{13} & \cdots & x_{1d} \\ & n_2 & x_{21} & x_{22} & x_{23} & \cdots & x_{2d} \\ N & n_3 & x_{31} & x_{32} & x_{33} & \cdots & x_{3d} \\ & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ & n_N & x_{N1} & x_{N2} & x_{N3} & \cdots & x_{Nd} \\ & \textit{Frequency} & f_1(x) & f_2(x) & f_3(x) & \cdots & f_n(x) \end{bmatrix}$$

The notations are as follows: The frequency of values is defined that the number of times that feature $F_h$ takes the value x in the $D$ data set (Note: if $x \notin F_h$, $f_h(x) = 0$), and $P_h(x)$: The sample probability of feature $F_h$ takes value $x$ in $D$ data set, as seen in above.

$$P_h(x) = \frac{f_h(x)}{N} \quad \forall h = (1, 2, 3, \ldots, d) \tag{3}$$

The importance of features (weights) is found by using Eq. (4). Features influence is checked by the success rate(as a class). After finding weights (see Eq. (2)), the weights are normalized to use in Eq. (5). Therefore, $w_h$ is used (from Eq. (2)) to give importance of feature.

Similarity value between X and Y belonging to the data set $D$ is calculated as follows:

$$F_{Linear} = \frac{1}{1 + e^{-(\beta_0+\sum_{h=1}^{d}\beta_h x_h)}} \tag{4}$$

If $F_{Linear}$ is greater than 0.8 or $F_{Linear}$ less than 0.2, incident is decided as attractive event. Otherwise, it is non-attractive event. If the past event is attractive for future events. $b_h$ and $n_h$ are decided to estimate similarity (interaction) for events.

$$b_h = \begin{cases} P_h(x) & if \;\; X_h = Y_h \; as \; categorical \; features \\ 0 & otherwise \end{cases}$$

$$n_h = \begin{cases} (X_h/Y_h) & if \; X_h < Y_h \;\; as \; continuous \; features \\ (Y_h/X_h) & otherwise \end{cases}$$

$$S(X, Y) = \sum_{h=1}^{d} \beta_h \; (\sqrt{(b_h)^2} \; or \; (n_h)) \tag{5}$$

where $S(X, Y)$ is the similarity between two events. This value is used to define relations between events in networks. For instance, after the September 11, 2001 attack, terrorist groups started to use the tactics of this attack. In order to understand how they used tactics, we calculate the similarities of attacks.In the above matrix, we have features ($F$) for each terrorist event ($n$). Between $n_1$ event and $n_2$ event, we can calculate similarity, and understand how $n_2$ event is interacted by $n_1$ event as an attractive event. Based on the ESALLOR model, $\beta$ values each feature $F_{Linear}$ are defined to show how important the features and attractive events like September 11, attack are, as seen in Eq. 4. Afterward, for calculating the similarity between two events, if the $F_1$ is the categorical feature, we compared $n_1$ event for $F_1$ and $n_2$ event for $F_1$. When they have the same values, we used Eq. 3 to calculate the probability of the feature $F_1$. Otherwise, the value is defined as zero, which means that there are no interactions for $F_1$. If the $F_1$ is the continuous feature, we calculate the ratio $(X_h/Y_h)$. For example, for the number of killed, we can have 10 people and 20 people for two events. Thus, the ratio is $10/20 = 0.5$, and 50% similarity between the two events. Therefore, for each feature, after calculating similarity, we can calculate total similarity with $\beta$ values, as seen in Eq. 5. $S(X, Y)$ values between 0 and 1, and is calculated to show the similarity between the $n_1$ event and $n_2$ event. It can also use relations between nodes (events) to construct the network for understanding complex interactions among terrorist attacks.

### 4.3. Network inference

The analysis of terrorist attacks indicates that both the evolutionary nature of terrorism and the adaptation of the tactics for it are recognized for terrorist attacks (Chenoweth & Lowham, 2007). Terrorist leaders in attacks tend to emulate the behavior of other terrorist leaders and learn from their mistakes and successes. It means that tactics and strategies are spreading as a contagious disease for future events. When this spread of tactics is captured, the future behavior of events can be understood for reactive strategies. Hence, correctly identifying activity patterns associated with different terrorist groups and predicting the success rate of their attacks enables us to reduce the effectiveness of terrorist attacks.

In our research, the network is defined as the structure that consists of many individuals (terrorist attacks as events) that have relations. It can be denoted as a matrix $G = (V, E)$ be a directed

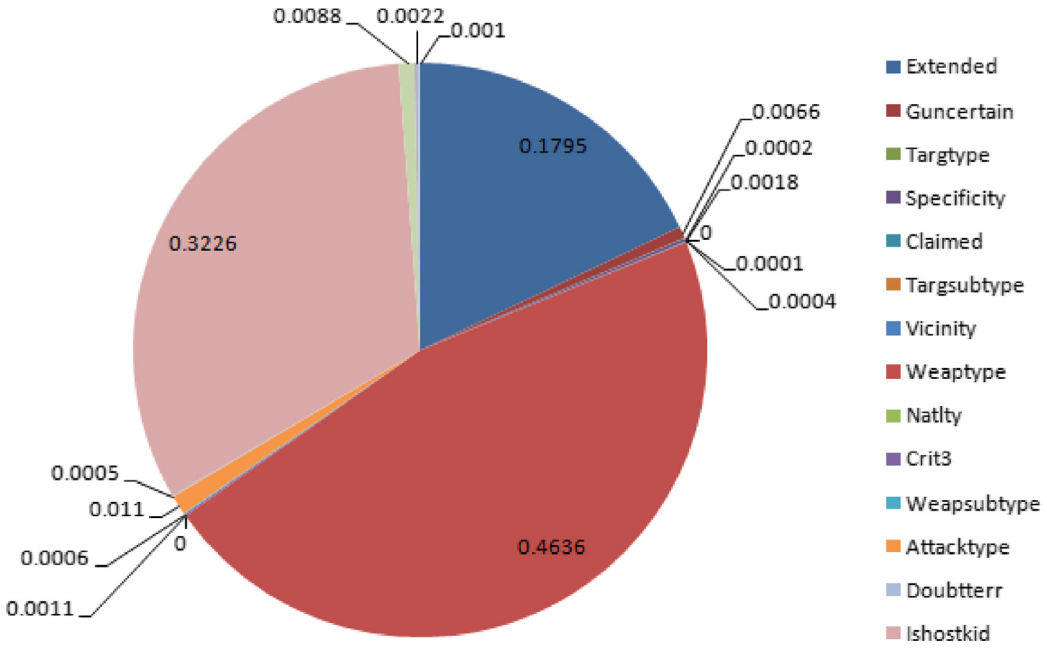


**Fig. 7.** Importance of the features for successful terrorist attacks.

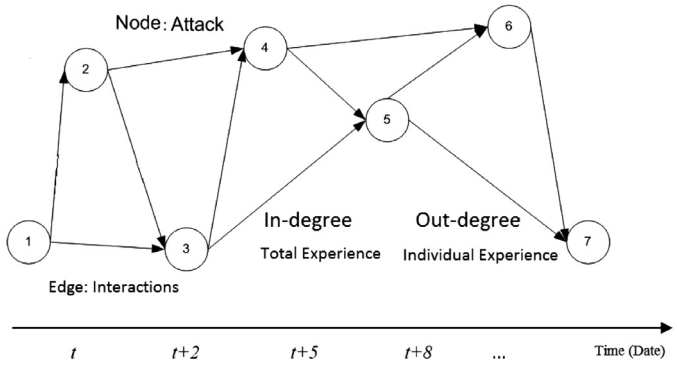


**Fig. 8.** Example of contagious learning from past attacks as a directed graph.

graph, where $V$ is the vertex set and $E$ is the edge set. Then matrix G is defined, as seen in Eq. 6.

$$G = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1j} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2j} & \dots & a_{2n} \\ & & \vdots & & \vdots & \\ a_{i1} & a_{i2} & \dots & a_{ij} & \dots & a_{in} \\ & & \vdots & & \vdots & \\ a_{n1} & a_{n2} & \dots & a_{nj} & \dots & a_{nn} \end{pmatrix} \tag{6}$$

If node $i$ is linked to node $j$, then $a_{ij} = 1$, else $a_{ij} = 0$. The matrix also could have a weight between 0 and 1 to show the level of relation (Li, Zhu, & Wang, 2015). The network is defined as a directed graph because terrorist leaders are learning tactics from mistakes and successes of past attacks, as seen in Fig. 8. A directed graph is constructed by estimating the in-degree ($k_i$) of nodes, which is the number of incoming edges, and with the out-degree ($k_o$) of nodes, which is the number of outgoing edges, as seen in Eq. (7) and Fig. 8 (Barabási & Frangos, 2014).

$$deg(i) = k_i = \sum_{j=1}^{m} a(i, j) \tag{7}$$

Spatial network and tactic-based network are used to analyze the relation of attacks, as seen in Fig. 10. We extract the best relations among terrorist attacks by combining these networks. First of all, spatial features such as city, state, vicinity, etc., are used as models for terrorist attacks to show how interactions spread. We select important locations to quickly make provision for terrorism. Therefore, outbreak detection for terrorism can be modeled similar to that of selecting nodes (e.g., attack locations) in a network to detect spread of information. Important locations of attacks are selected to monitor, so that any terrorism outbreak can be detected early, when very few violent attacks have occurred (Leskovec et al., 2007). In the following section, the proposed similarity function is used to calculate interactions among the events by using tactics such as weapon type, attack type, target type, etc.

#### 4.3.1. Structure analysis of the networks

The centrality of the network is calculated to make structure analysis. Popularity and information sharing of events are decided by using degree centrality, betweenness centrality, and closeness centrality.

Degree Centrality: Degree centrality is calculated to draw topology of terrorist attacks and features, as seen in Eq. (8) (Freeman, 1979; Li et al., 2015). This index helps to identify the most popular attacks in the network (Sayama, 2015).

$$deg(i) = C_{dc}(i) = \sum_{j=1, j\neq i}^{m} a(j, h) \tag{8}$$

Where m is the total number of attacks, and $a(j, h)$ is a binary variable that asserts whether a relation occurs between node $i$ and $h$.

Closeness centrality: The closeness centrality of node $h$ is identified as the sum of its distance to other nodes. Therefore, the central node has the smaller distance from others. This centrality measures the most efficient attack to collect information from the all terrorist groups (Alvarez-Hamelin, Dall'Asta, Barrat, & Vespignani, 2005; Li et al., 2015; Sayama, 2015).

$$C_{cc}(h) = \sum_{i=1, i\neq h}^{m} y(j, h) \tag{9}$$

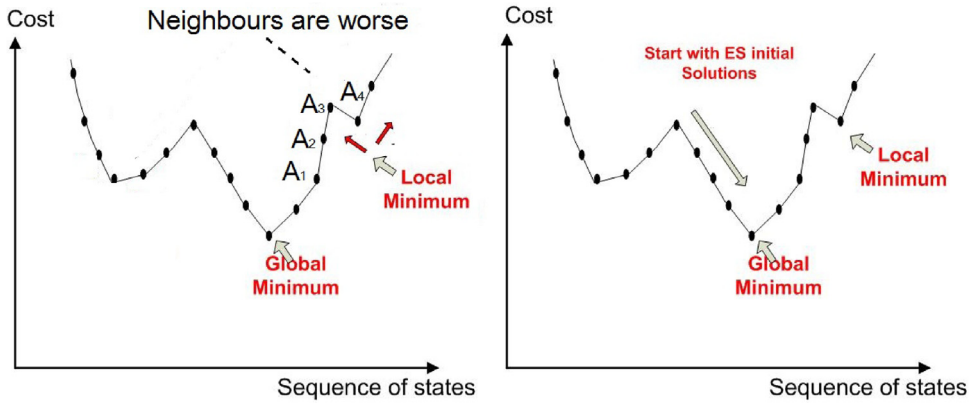


**Fig. 9.** Coupling the ES and the SA to explain how to prevent getting stuck in a local optimum.

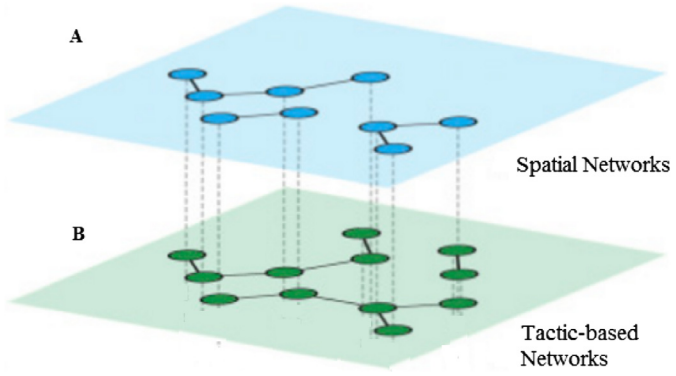


**Fig. 10.** Example for schematic representation of analyzed networks.

where $y(j, h)$ is the length of shortest path that connects nodes $j$ and $h$.

Betweenness Centrality: This measure indicates that the number of shortest paths from all nodes to all others passes through that particular node. This index describes the node that has most relations in a network. A gatekeeper of information flow can be found by using this centrality. The most powerful attack to control the information flow within terrorist groups can be found by using this centrality (Freeman, 1979; Sayama, 2015).

$$C_{bc}(h) = \sum_{j=1}^{m} \sum_{i=1}^{m} b_{ji}(h) \tag{10}$$

where $b_{ji}(h)$ is a binary variable that asserts whether a relation occurs between nodes $j$ and $h$.

## 5. Experimental results and discussion

In this section, patterns and relations are extracted by using collected data. The experimental results are presented by using the collected data and the proposed approaches. Our experiments include two parts. First, the key features are selected with their weights for similarity function. Moreover, the proposed similarity function is used to define popularity and outliers to understand how terrorist groups will attack in the future. Second, we show the finding patterns to make provision for future terrorist attacks by using calculating relations. In the discussion part of this section, the proposed framework is shown so that it can be used by the government for intelligence analysis. We discuss how the results can be used to find patterns for predicting future terrorist attacks.

### 5.1. Defining importance of the key features

The best features are selected and weighted because they have different effects on the success of events. In order to calculate well relations amongst the events, feature weights need to be considered for those of similarity functions. By calculating attractions for each attack, the weights are calculated for each feature. The proposed (ESALLOR) model is able to detect the relevant features. Using this robust model, the weights of the features are calculated to show the importance of the features, as seen in Table 3. Afterward, these weights are used to calculate the probability of success rate, as seen in Eq. (1). As seen in Table 3 and Fig. 7, the feature weights are found to define the importance of features for using in the proposed similarity function. Weapon type, kidnapping (ishostkid), attack type (attacktype) and extension of attack (extended) are very significant for terrorists become successful. Other features became nearly zero due to the fact that the ESALLOR model was used. The government needs to control these features to prevent successful attacks.

Furthermore, because Lasso regularization is used in the model, feature selection and weighting are utilized to improve similarity function. Based on the success rate, an attractive event is decided if this value is less than 0.2 as a bad experience, and more than 0.8 as good experience. After defining the best features, relations among events are calculated by using the proposed similarity function.

### 5.2. Defining the popular terrorist behaviors

In this section, by using popular (the most similar) behaviors, we define behaviors for future attacks in the defined locations. For example, for those near the city of Baghdad, we can find what are the popular attacks and tactics. In the popularity-based behaviors, terrorism outbreaks are defined, and structure analysis is made to understand the popularity of future attacks.

#### 5.2.1. Early defining terrorism outbreaks

For this section, spatial patterns are found by using network-based outbreak detection in Iraq. Outbreak detection is modeled as selecting nodes (e.g., attack location) in a network. As seen in Fig. 11, the attack's behavior comes from Baghdad in Iraq. Baghdad has a high degree in the network and is related to attacks that have a low degree. Future attacks can be controlled and prevented when this behavior in Baghdad is controlled.

Moreover, after using tactic-based networks and spatial networks, as seen in Fig. 10, structure analysis is conducted to understand popular behaviors in the combined network. Based on the degree centrality in the network, Event 17, Event 21, Event 90, and

**Table 3**
Importance of the features used in similarity function.

| Feature name | Weight | Feature name | Weight | Feature name | Weight |
|---|---|---|---|---|---|
| Extended | 0.1795 | Guncertain | 0.0066 | Targtype | 0.0002 |
| Specificity | 0.0018 | Claimed | 0.0000 | Targsubtype | 0.0001 |
| Vicinity | 0.0004 | Weaptype | 0.4636 | Natlty | 0.0000 |
| Crit3 | 0.0011 | Weapsubtype | 0.0006 | Attacktype | 0.0110 |
| Doubtterr | 0.0005 | Ishostkid | 0.3226 | Property | 0.0088 |
| Multiple | 0.0022 | Int_Any | 0.0010 | **Total** | **1.0000** |

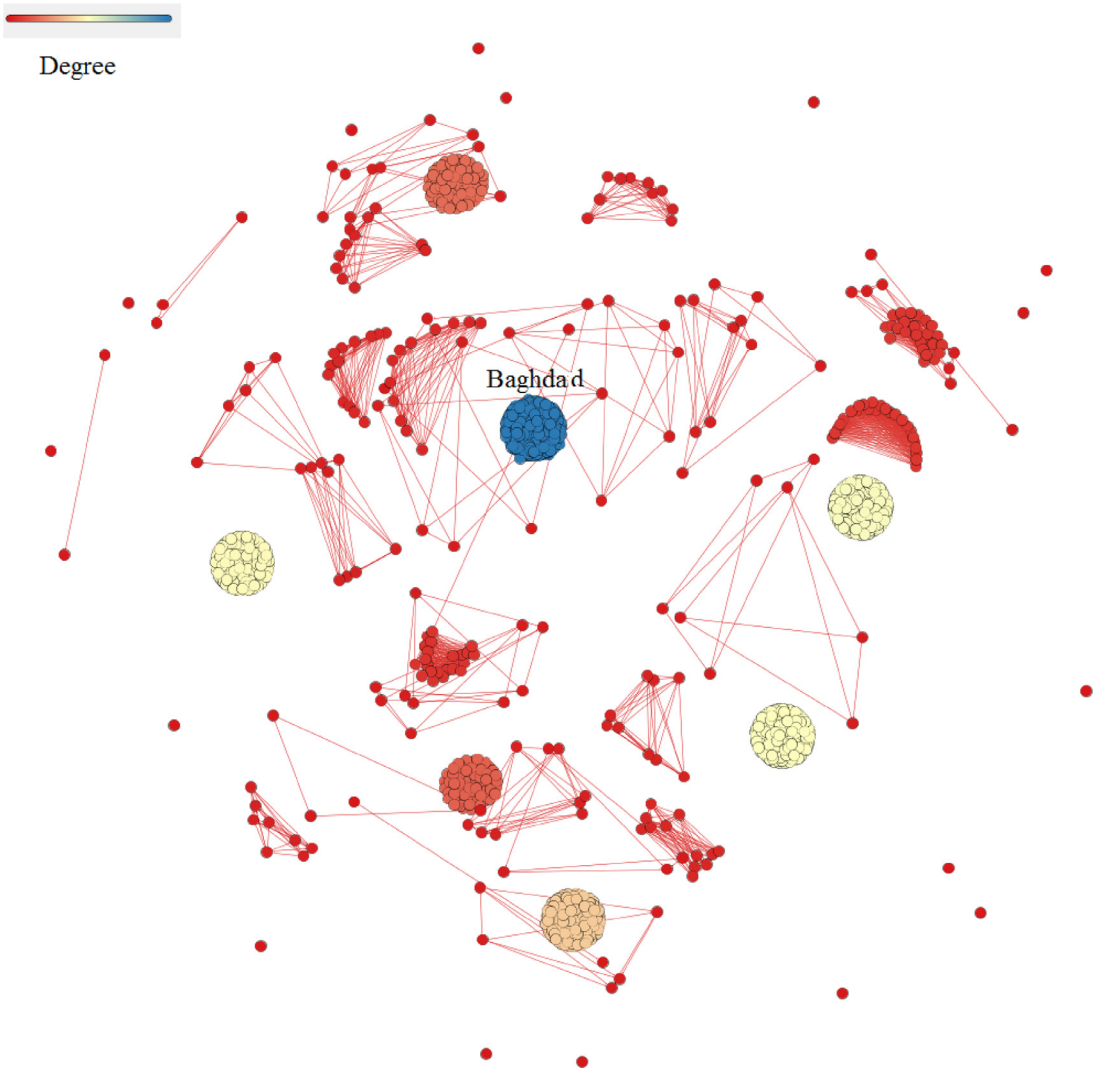


**Fig. 11.** Outbreak detection to define dangerous areas in Iraq.

Event 108, which are the most popular events used in the past, as seen in Fig. 12, there is general behavior (called popularity) found because hub nodes are similar in the network. As patterns, the general behavior happened in order in the years of 2005, 2007, 2008 and 2013. The attacks occurred in a city/village/town as the center. They did not use multiple attacks. They focused on the government as a target type. They used suicide bombing with a vehicle or carried bodily by a human being. The attacks resulted in property damage. They did not use hostages for attacks. They killed and wounded few people. It means that they tried to kill important people to attract an audience. They used suicide bombing as an assassination approach because it is more successful.

In the second place, the shortest path is used to measure network characteristics by using betweenness centrality and closeness centrality. Based on the closeness centrality, the most efficient attacks are found to collect the information from all terrorist groups, as seen in Fig. 14. Event 96, and Event 114 are important as the influence of future events. At the same time, a gatekeeper of information flow is found by using betweenness centrality. Event 20, Event 24, Event 23, Event 19, Event 77, and Event 81 are the most powerful attacks that control the information flow, as seen in Fig. 13.

### 5.3. Discovering patterns from the past terrorist attacks

Generally speaking, terrorist groups are learning from past attacks with interactions, and they do not know how to learn tactics. When the interactions are captured, tactics and evaluation of the

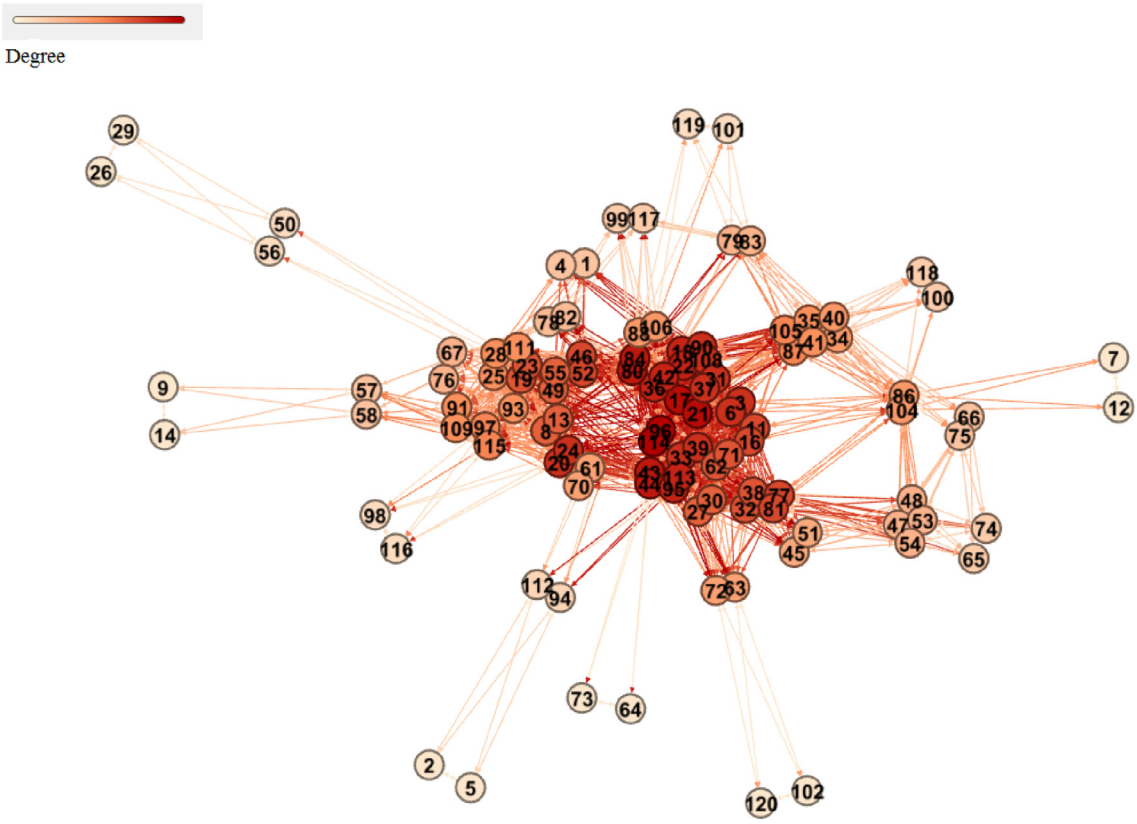


**Fig. 12.** Degree centrality for suicide attacks.

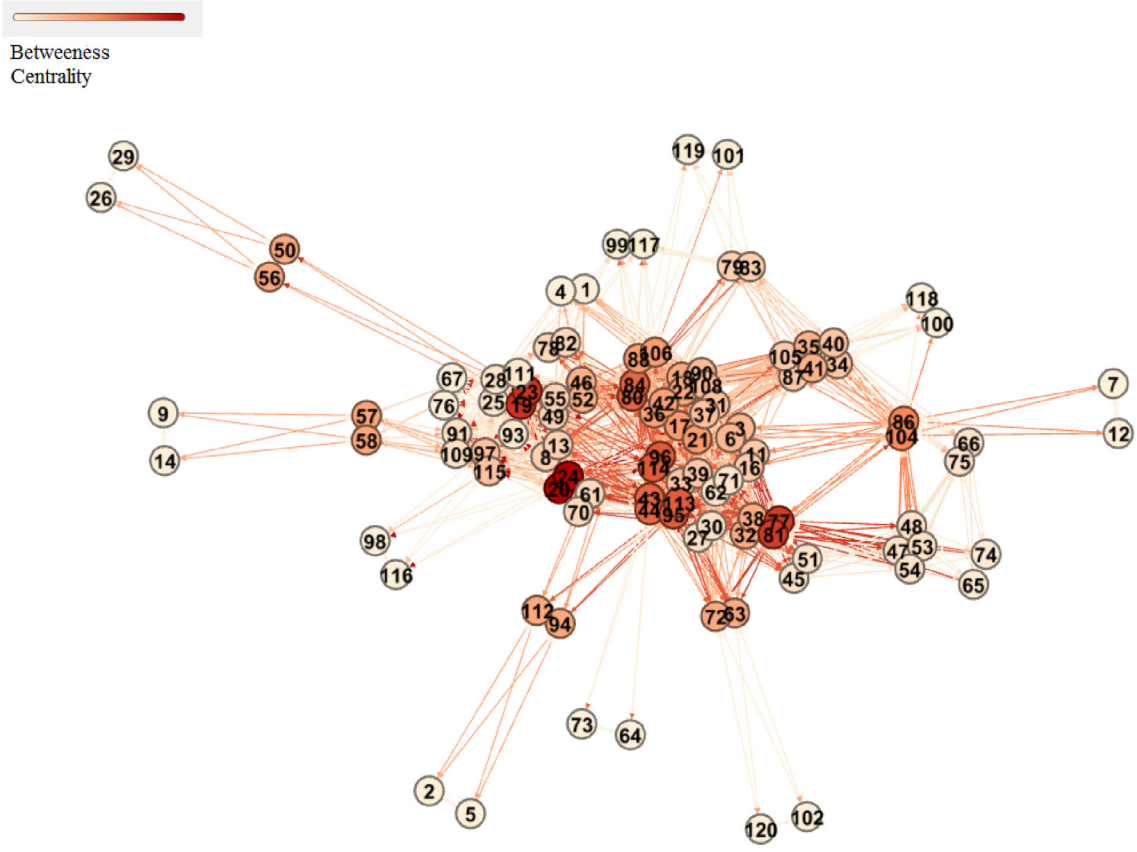


**Fig. 13.** Betweenness centrality for suicide attacks.

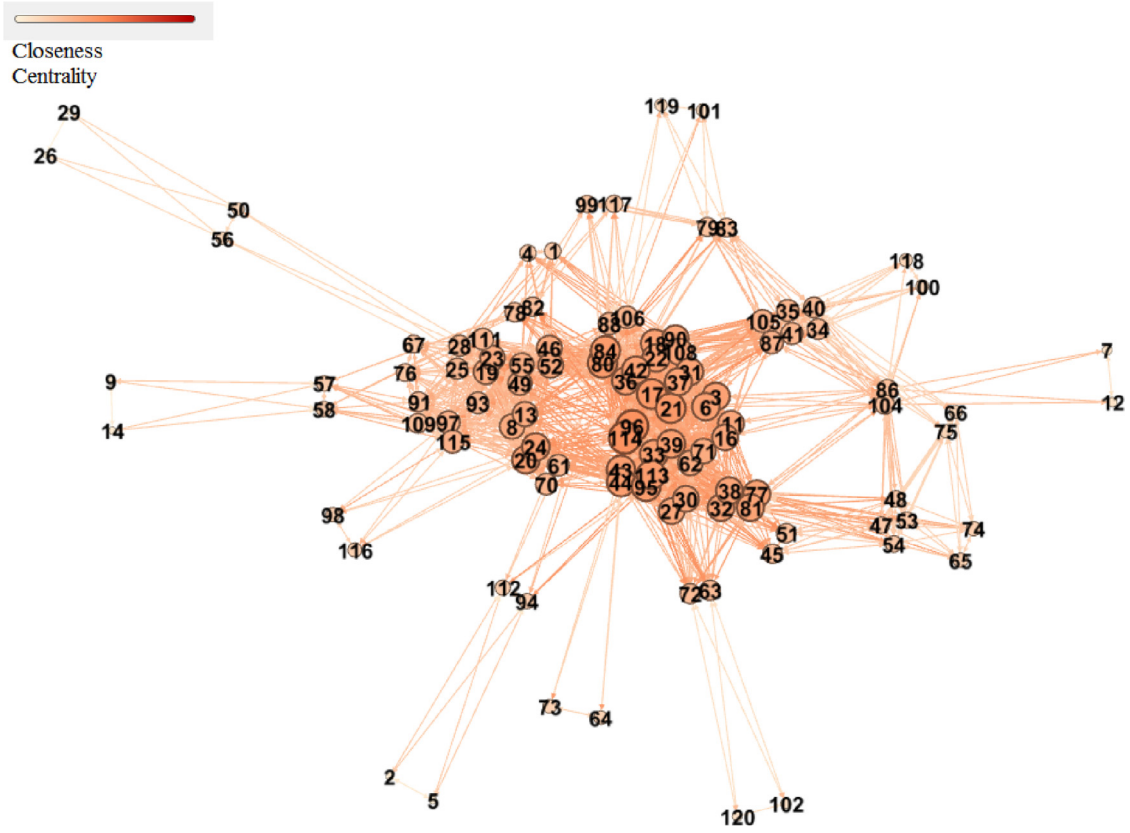


**Fig. 14.** Closeness centrality for suicide attacks.

**Table 4**
Suicide attacks being the highest in-degree and out-degree dynamically. Note: numbers of nodes show the attacks. Note: Highlighted nodes are determined for future behaviors.

| Years | In-degree | Out-degree |
|---|---|---|
| 2003–2005 | 52-46-43-44-95-96... | 6-3-17-21-20-24-16-11-8-13... |
| 2004–2006 | 46-52-43-44-55-80-39... | 17-21-22-18-20-24-11-16-13-8... |
| 2005–2007 | 43-44-80-84-90-108-62-71-114... | 17-21-22-18-20-24... |
| 2006–2008 | 43-44-80-84-62-95-71-96... | 31-36-33-37-42-39-43-44... |
| 2007–2009 | 80-84-90-108... | 31-32-33-36-42-43-44-38-39... |
| 2008–2010 | 81-77... | 43–44... |
| 2009–2011 | 77-81-108-91... | 46-49-52-55... |
| 2010–2012 | 96-114-113-95-81-77... | 80-62-71... |
| 2011–2013 | 114-96-113... | 80-62-63-71... |
| **2013–2015** | 80-84-90-108... | |
| **2015–2017** | 77-81-108-91... | |

tactics can be defined by governments. Networks are constructed by calculating in-degree and out-degree values for every two years, as seen in Fig. 15. In-degree is used to show total experience from past attacks. Out-degree is used to show individual experience of each attack. The framework shows the behaviors used by terrorists in past events as well as explains how the framework understands behaviors that will be used for future attacks.

The results of analyzing the in-degree and out-degree properties of the constructed network (shown in Table 4) demonstrate that the tactics of attacks with high out-degree are repeated after some period, which indicates a pattern of repetition. For example, the behaviors of attacks with high out-degree during the years 2010–2013 are repeated in the 2007–2009 period as in-degree. This demonstrates that terrorists use individual experiences for future attacks. After following the patterns, the sensitivity analysis was used for the years 2013–2015 that demonstrated the same patterns. Fig. 16 shows the identified patterns for the 2013–2015 period and 2015–2017 period. In order to show that finding patterns is useful, the attacks for the years of 2014 and 2015 are compared with these patterns.

### 5.4. Sensitivity analysis

To validate the understanding capabilities of our framework, we used the identified pattern in the suicide attack network to predict the features of attacks in 2014 and 2015. As seen in Table 5, comparing the results with existing data shows that our proposed method was able to successfully predict most of the characteristics of suicide attacks with more than 90% accuracy. The only exception was the extent of the property damage, which was accurate in 60% of the attacks. These results support the previous findings that terrorists tend to emulate the behavior of other terrorist groups and learn from their mistakes and successes (Chenoweth & Lowham, 2007).

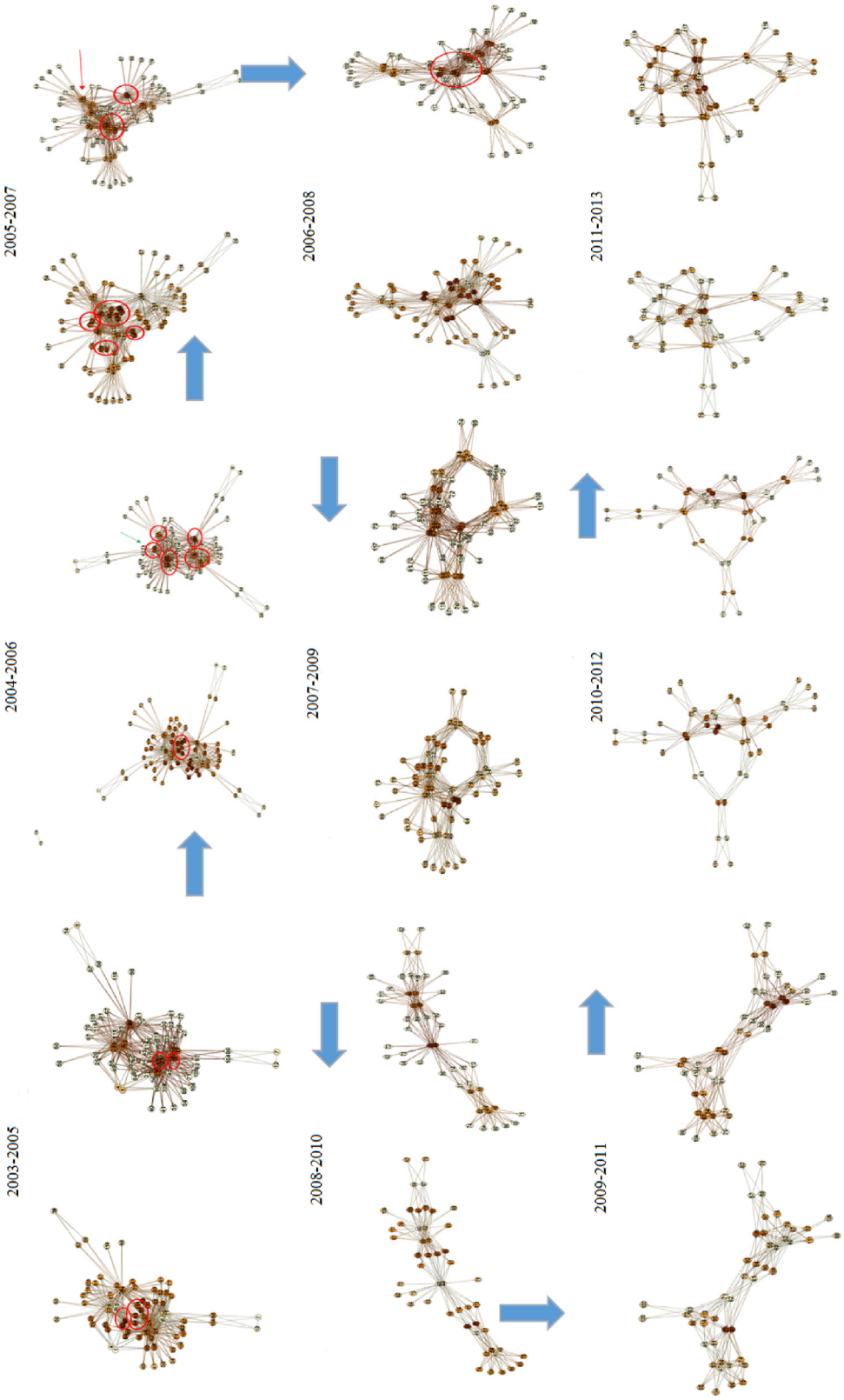


**Fig. 15.** The patterns based on in-degree and out-degree.

For the summary, as seen in Fig. 11, we defined that attractive terrorist tactics spread from Baghdad to all Iraq, middle east, and the entire world, respectively. When the attacks in Baghdad are prevented, the attacks in Iraq can be prevented, and we can understand the behaviors of terrorist groups before attacks happen in specific locations. Afterward, the most popular attacks are defined to understand how terrorist groups will attack by learning tactics from past attacks. Furthermore, by using the total experience (in-degree) and individual experience (out-degree), the framework understood what are the terrorist' behaviors for near

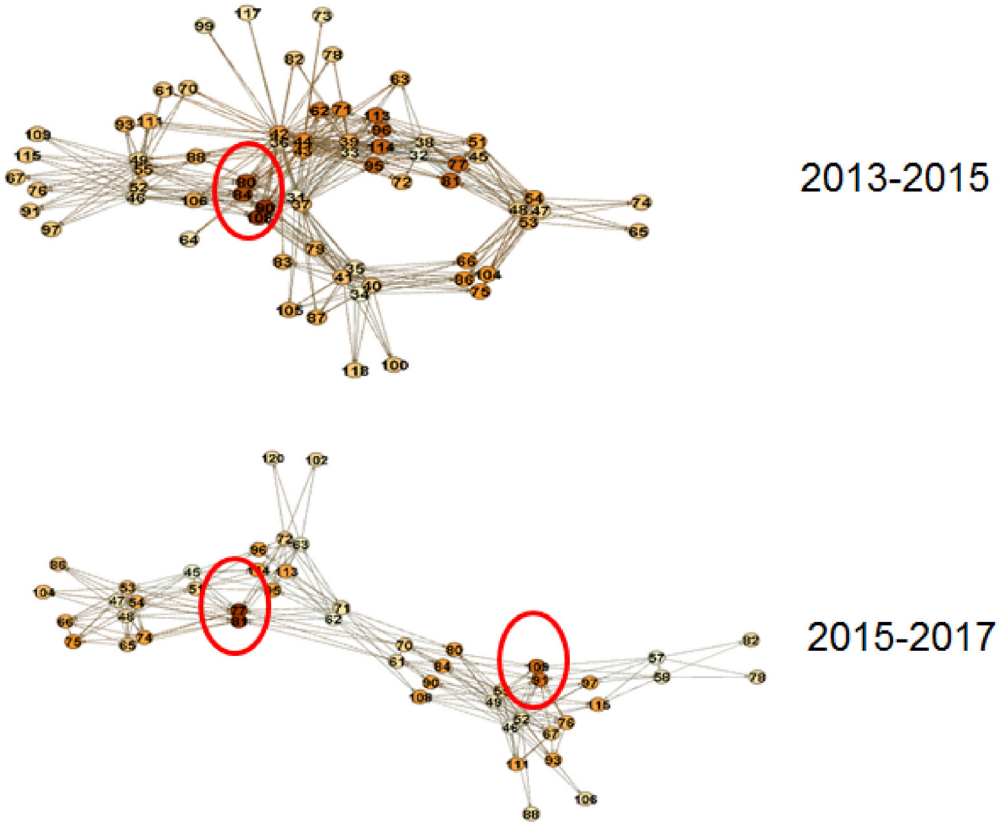


**Fig. 16.** The krebs2002mappingfinding patterns with networks analysis.

**Table 5**
Comparing finding patterns and occurred behaviors in the attacks for the years of 2014 and 2015.

| Features | Predicted feature value | Accuracy of occurrence |
|---|---|---|
| Geocoding specificity | Attacks will occur in city/village/town that attacked before | 96% |
| Weapon type | Explosive bombs and dynamite | 96% |
| Group name | ISIS and Al-Qaede | 75% |
| Multiple | Attacks are not multiple | 89% |
| Vicinity | Attacks will occur in city centers | 96% |
| Extented incident | The duration of attacks is less than 24 h | 100% |
| Goals of attacks | Attacks has political, economic, religious with larger audiences | 96% |
| Target type | Government (Politician, political party movement, head of state, government personnel, government building) | 96% |
| International-domestic | Attacks are domestic and unknown | 100% |
| Hostages-kidnaping | No kidnapping or hostages | 96% |
| Total number of fatalities | number of killing is less than 10 killing and 20 wounded | 96% |
| Extent of property damage | Minor (likely 1 million dollars) | 60% |

Baghdad. For the sensitivity analysis, we compared finding patterns and actual attacks for the years 2014 and 2015. Finally, we had more than 90% accuracy for most of the tactics. As an expert system, when the location is given, the framework gives the tactics to understand suicide terrorism, and proposes reactive strategies for counter-terrorism.

## 6. Conclusions and future works

Nowadays, preventing threats before they happen is extremely important in counter-terrorism. Governments need to understand how terrorist groups behave in terrorist attacks. In addition, as evidenced in Iraq, suicide attacks using bombs by terrorists were very successful because the government lacked the necessary understanding of terrorism and had insufficient provision. Counter-terrorism officials need to guide the government in developing better defense strategies to combat terrorist' tactics.

The primary conclusion of this paper is that the terrorism network of suicide attacks in Iraq is first modeled to prevent future threats. To form relations, the proposed ESALLOR model removes irrelevant and redundant features of terrorism events, and a new weighted heterogeneous similarity function is proposed to form links among suicide attacks, which comprise the most harmful and operative attack type. For finding dangerous areas in Iraq, outbreak detection for terrorism is modeled by looking at locations of attacks for terrorism.

In light of the results and discussion presented up to now, the proposed model shows that the results will work in the future due to the use of proposed feature selection and similarity function. New intelligent framework including feature selection and similarity function is investigated using supervised/unsupervised machine-learning techniques. The patterns and relations are extracted to improve counter-terrorism. We show that attacks can be prevented by learning the general behavior of attacks with

sensitivity analysis. The results on terrorism data in Iraq show that we can understand behaviors for suicide attacks by using finding patterns.

In future work, dynamic large networks could be used to discover the patterns as big data project for future events. Moreover, people could study a unified approach that applies pattern classification techniques to the proposed network topology to improve detection accuracy. Based on the proposed network, pattern recognition methods would be used to detect terrorism events and their terrorist group. Also, conditional probability can be used to understand which event will cause for the future events. At the same time, the framework can implement for another application areas if they can have interactions among observations (e.g., criminal events, credit card approval, etc.).

In conclusion, the results would enable policy makers to develop precise global and/or local counter-terrorism policies. The government can deter terrorist threats by using this intelligent framework. Governments can understand how terrorism will impact future events, and governments can control terrorist behaviors to reduce the risk of future events. Furthermore, this information can be useful for law enforcement agencies to propose reactive strategies.

## Acknowledgments

This research is supported by the Turkish Military Academy (TMA) and the Watson Institute Systems Excellence (WISE) at State University of New York at Binghamton. The authors wish to thank the TMA and the WISE for their support in the research. Moreover, this research (topic) was mentioned in more than fifteen news (e.g., http://www.eurekalert.org/pub_releases/2016-06/bu-nfu062816.php), and it was also used as a frontier cover in the Industrial and Systems Engineering (ISE) magazine (e.g., http://www.iienet2.org/ISEmagazine/details.aspx?id=42218). Also, we would like to thank the anonymous referees for valuable comments that have improved the quality of the paper.

## References

Agarwal, J., Nagpal, R., & Sehgal, R. (2013). Crime analysis using k-means clustering. *International Journal of Computer Applications, 83*(4).

Akgun, I., Kandakoglu, A., & Ozok, A. F. (2010). Fuzzy integrated vulnerability assessment model for critical facilities in combating the terrorism. *Expert Systems with Applications, 37*(5), 3561–3573.

Alvarez-Hamelin, J. I., Dall'Asta, L., Barrat, A., & Vespignani, A. (2005). Large scale networks fingerprinting and visualization using the k-core decomposition. In *Advances in neural information processing systems* (pp. 41–50).

Archibald, R., & Fann, G. (2007). Feature selection and classification of hyperspectral images with support vector machines. *Geoscience and Remote Sensing Letters, IEEE, 4*(4), 674–677.

Arulanandam, R., Savarimuthu, B. T. R., & Purvis, M. A. (2014). Extracting crime information from online newspaper articles. In *Proceedings of the second australasian web conference-volume 155* (pp. 31–38). Australian Computer Society, Inc.

Barabási, A.-L., & Frangos, J. (2014). *Linked: The new science of networks science of networks*. Basic Books.

Bohannon, J. (2009). Counterterrorism'S new tool:Metanetworkanalysis. *Science, 325*(5939), 409–411.

Boriah, S., Chandola, V., & Kumar, V. (2008). Similarity measures for categorical data: A comparative evaluation. *Red, 30*(2), 3.

Byman, D., & Shapiro, J. (2014). We shouldn't stop terrorists from tweeting. *The Washington Post, 9*.

Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering, 40*(1), 16–28.

Chang, W., Chung, W., Chen, H., & Chou, S. (2003). An international perspective on fighting cybercrime. In *International conference on intelligence and security informatics* (pp. 379–384). Springer.

Chen, H. (2011). *Dark web: Exploring and data mining the dark side of the web*: 30. Springer Science & Business Media.

Chen, H., Chung, W., Xu, J. J., Wang, G., Qin, Y., & Chau, M. (2004). Crime data mining: A general framework and some examples. *Computer, 37*(4), 50–56.

Chenoweth, E., & Lowham, E. (2007). On classifying terrorism: A potential contribution of cluster analysis for academics and policy-makers. *Defence & Security Analysis, 23*(4), 345–357.

Chiu, C., Ku, Y., Lie, T., & Chen, Y. (2011). Internet auction fraud detection using social network analysis and classification tree approaches. *International Journal of Electronic Commerce, 15*(3), 123–147.

Choi, D., Ko, B., Kim, H., & Kim, P. (2014). Text analysis for detecting terrorism-related articles on the web. *Journal of Network and Computer Applications, 38*, 16–21.

Clarke, R. (1988). Information technology and dataveillance. *Communications of the ACM, 31*(5), 498–512.

Coffman, T. R., & Marcus, S. E. (2004). Pattern classification in social network analysis: a case study. In *Aerospace conference, 2004. proceedings. 2004 ieee: 5* (pp. 3162–3175). IEEE.

De Vel, O., Anderson, A., Corney, M., & Mohay, G. (2001). Mining e-mail content for author identification forensics. *ACM Sigmod Record, 30*(4), 55–64.

Division, N. S. R. (2016). Rand database of worldwide terrorism incidents (rdwti). Retrieved from http://www.rand.org/nsrd/projects/terrorism-incidents.html.

Ferreira, A. J., & Figueiredo, M. A. (2012). An unsupervised approach to feature discretization and selection. *Pattern Recognition, 45*(9), 3048–3060.

Fienberg, S. E. (2005). Homeland insecurity: Datamining, terrorism detection, and confidentiality. *Bulletin of the International Statistical Institute*. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.486.8936&rep=rep1&type=pdf Accessed 15.02.17.

Freeman, L. C. (1979). Centrality in social networks conceptual clarification. *Social Networks, 1*(3), 215–239.

Gu, Q., Li, Z., & Han, J. (2012). Generalized fisher score for feature selection. arXiv preprint arXiv:1202.3725.

Gunasundari, S., Janakiraman, S., & Meenambal, S. (2016). Velocity bounded boolean particle swarm optimization for improved feature selection in liver and kidney disease diagnosis. *Expert Systems with Applications, 56*, 28–47.

Hassani, H., Huang, X., Silva, E. S., & Ghodsi, M. (2016). A review of data mining applications in crime. *Statistical Analysis and Data Mining: The ASA Data Science Journal, 9*(3), 139–154.

Hauck, R. V., Atabakhsb, H., Ongvasith, P., Gupta, H., & Chen, H. (2002). Using coplink to analyze criminal-justice data. *Computer, 35*(3), 30–37.

He, X., Cai, D., & Niyogi, P. (2005). Laplacian score for feature selection. In *Advances in neural information processing systems* (pp. 507–514).

Hsu, H.-H., Hsieh, C.-W., & Lu, M.-D. (2011). Hybrid feature selection by combining filters and wrappers. *Expert Systems with Applications, 38*(7), 8144–8150.

Hyvärinen, A., & Oja, E. (2000). Independent component analysis: Algorithms and applications. *Neural Networks, 13*(4), 411–430.

Jackson, B. A., & Frelinger, D. R. (2009). Understanding Why Terrorist Operations Succeed or Fail. *Technical Report*. DTIC Document.

Kalaikumaran, T., Karthik, S., et al. (2012). Criminals and crime hotspot detection using data mining algorithms: Clustering and classification. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 1*(10), pp–225.

Kanellis, P. (2006). *Digital crime and forensic science in cyberspace*. IGI Global.

Kaufman, L., & Rousseeuw, P. J. (2009). *Finding groups in data: An introduction to cluster analysis*: 344. John Wiley & Sons.

Kirkpatrick, S. (1984). Optimization by simulated annealing: Quantitative studies. *Journal of statistical physics, 34*(5–6), 975–986.

Krebs, V. E. (2002). Mapping networks of terrorist cells. *Connections, 24*(3), 43–52.

Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., VanBriesen, J., & Glance, N. (2007). Cost-effective outbreak detection in networks. In *Proceedings of the 13th acm sigkdd international conference on knowledge discovery and data mining* (pp. 420–429). ACM.

Li, B.-x., Zhu, J.-f., & Wang, S.-g. (2015). Networks model of the east Turkistan terrorism. *Physica A: Statistical Mechanics and its Applications, 419*, 479–486.

Mitchell, T. M. (1997). Machine learning. *McGraw-Hill series in computer science*. Boston (Mass.), Burr Ridge (Ill.), Dubuque (Iowa): McGraw-Hill. URL http://opac.inria.fr/record=b1093076.

Moradi, P., & Rostami, M. (2015). A graph theoretic approach for unsupervised feature selection. *Engineering Applications of Artificial Intelligence, 44*, 33–45.

Nath, S. V. (2006). Crime pattern detection using data mining. In *Web intelligence and intelligent agent technology workshops, 2006. wi-iat 2006 workshops. 2006 ieee/wic/acm international conference on* (pp. 41–44). IEEE.

Netzer, M., Kugler, K. G., Müller, L. A., Weinberger, K. M., Graber, A., Baumgartner, C., & Dehmer, M. (2012). A network-based feature selection approach to identify metabolic signatures in disease. *Journal of Theoretical Biology, 310*, 216–222.

Perry, W. L., Berrebi, C., Brown, R. A., Hollywood, J., & Jaycocks, A. (2013). *Predicting suicide attacks: Integrating spatial, temporal, and social features of terrorist attack targets*. Rand Corporation.

Prakash, D., & Surendran, S. (2013). Detection and analysis of hidden activities in social networks. *International Journal of Computer Applications, 77*(16).

Sayama, H. (2015). Introduction to the modeling and analysis of complex systems. *Open SUNY textbooks*. Milne Library, State University of New York at Geneseo.

Scholz, M. (2010). Node similarity as a basic principle behind connectivity in complex networks. arXiv preprint arXiv:1010.0803.

Senate, U. S. (2004). Federal Efforts Cover a Wide Range of Uses. *Technical Report*. GAO-04-548 2–3 (May 2004), online at http://www.gao.gov/new.items/d04548.pdf (visited Jan 12, 2008).

Song, F., Guo, Z., & Mei, D. (2010). Feature selection using principal component analysis. In *System science, engineering design and manufacturing informatization (ICSEM), 2010 international conference on: 1* (pp. 27–30). IEEE.

Sparrow, M. K. (1991). The application of network analysis to criminal intelligence: An assessment of the prospects. *Social Networks, 13*(3), 251–274.

Stanfill, C., & Waltz, D. (1986). Toward memory-based reasoning. *Communications of the ACM, 29*(12), 1213–1228.

START (2015). National consortium for the study of terrorism and responses to terrorism (start), global terrorism database. Retrieved from http://www.start.umd.edu/gtd.

Telesca, L., & Lovallo, M. (2006). Are global terrorist attacks time-correlated? *Physica A: Statistical Mechanics and its Applications, 362*(2), 480–484.

Thongtae, P., & Srisuk, S. (2008). An analysis of data mining applications in crime domain. In *Computer and information technology workshops, 2008. cit workshops 2008. ieee 8th international conference on* (pp. 122–126). IEEE.

Tutun, S., Chou, C.-A., & Canıyılmaz, E. (2015). A new forecasting framework for volatile behavior in net electricity consumption: A case study in turkey. *Energy, 93*, 2406–2422.

Usha, D., & Rameshkumar, K. (2014). A complete survey on application of frequent pattern mining and association rule mining on crime pattern mining. *International Journal of Advances in Computer Science and Technology, 3*(4).

Wang, S., Zhe, Z., Kang, Y., Wang, H., & Chen, X. (2008). An ontology for causal relationships between news and financial instruments. *Expert Systems with Applications, 35*(3), 569–580.

Wilson, D. R., & Martinez, T. R. (1997). Improved heterogeneous distance functions. *Journal of Artificial Intelligence Research, 6*, 1–34.

Xu, J. J., & Chen, H. (2005). Crimenet explorer: A framework for criminal network knowledge discovery. *ACM Transactions on Information Systems (TOIS), 23*(2), 201–226.

Yu, L., & Liu, H. (2003). Feature selection for high-dimensional data: A fast correlation-based filter solution. In *ICML: 3* (pp. 856–863).

Zhang, Z., & Hancock, E. R. (2011). A graph-based approach to feature selection. In *Graph-based representations in pattern recognition* (pp. 205–214). Springer.

Zhang, Z., & Hancock, E. R. (2012). Localized graph-based feature selection for clustering. In *Image analysis and recognition* (pp. 1–10). Springer.

Zorarpacı, E., & Özel, S. A. (2016). A hybrid approach of differential evolution and artificial bee colony for feature selection. *Expert Systems with Applications, 62*, 91–103.