Physica A 545 (2020) 123769

Contents lists available at ScienceDirect

Physica A

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

# Identifying influential nodes of global terrorism network: A comparison for skeleton network extraction

Kanokwan Malang [a], Shuliang Wang [a,*], Aniwat Phaphuangwittayakul [b], Yuanyuan Lv [a], Hanning Yuan [a], Xiuzhen Zhang [c]

[a] School of Computer Science & Technology, Beijing Institute of Technology, No.5 South Zhong-Guan-Cun Street, Beijing, 100081, PR China

[b] International College of Digital Innovation, Chiang Mai University, No. 239 Nimmanhaemin Road, Suthep, Muang, Chiang Mai 50200, Thailand

[c] School of Science (CSIT), RMIT University, GPO Box 2476, Melbourne 3001, Australia

ARTICLE INFO

*Article history:*
Received 9 March 2019
Received in revised form 20 November 2019
Available online 13 December 2019

*Keywords:*
Backbone network
Node evaluation
Global terrorism
Influential nodes
Network extraction
Complex network

ABSTRACT

The inherent structure and substantial information on global terrorism network are often understood by identifying influential nodes. Recently, novel node identification methods are developed from different perspectives. Each of them has trade-offs and strengths. However, the algorithms for exploring the key influential nodes have been adopted unevenly in light of network extraction research. A set of nodes that is more favorable to define the core network structure is unclear. In this paper, we, therefore, present a comparative study of node identification methods over the global terrorism network. The new insight each method contributes to identifying key influential nodes and core network structure is investigated. Six comparative methods are verified by the SIR model and monotonicity index. We further elaborate on experimental analysis by applying the critical nodes from each method to extract the skeleton network. All extracted skeletons are eventually compared with the original network in terms of node correlation and network structural-equivalence. Thus, the comparison and results not only used to reflect the potential of different methods to a particular network structure but also guide us to select a method that works best for extracting the skeleton network of real-world global terrorism.



## 1. Introduction

Recently, the integrative technological domain has been bringing some opportunities to enable node identification methods and complex network analysis for understanding the global terrorism network [1]. In particular, the vital insight on terrorism incidents can be derived from the investigation of influential nodes i.e., terrorists or targets [2]. Identifying the influential nodes of the global terrorism network can minimize the loss of living life and properties. For example, if we know the most vulnerable target nodes, counterterrorism could have a reasonable warning or prevention for a possible attack. On the other hand, if we know the deadliest terrorist nodes and their collaborators in advance, a government can monitor suspects and against terrorists from committing a terror attack together.

* Corresponding author.
*E-mail addresses:* kanokwan.malang@yahoo.com (K. Malang), slwang2011@bit.edu.cn (S. Wang), aniwat.ph@cmu.ac.th (A. Phaphuangwittayakul), lvyy@boe.com.cn (Y. Lv), yhn6@bit.edu.cn (H. Yuan), xiuzhen.zhang@rmit.edu.au (X. Zhang).

https://doi.org/10.1016/j.physa.2019.123769
0378-4371/

Over the past decade, centrality measures are proposed by many scholars to explain the influence power of the nodes in terrorist-to-terrorist networks [3]. Different centralities exhibit specific interpretations [4]. The most basic measure is Degree centrality which considers the direct influence from adjacent edges to each node. High degree value tells us the terrorist occupying a central position associated with the most potent node [5]. Betweenness centrality implies the bridge node or terrorist who is a key in-between two communities, while closeness centrality refers to the terrorist who good at network access because it is closer to other terrorists [6]. Although centrality measures can be beneficial for identifying the central nodes or hubs in general, the most central nodes do not necessary to carry the utmost leadership potential [7].

Besides, degree prestige is used to explore the structural features between terrorist organizations [8]. Prestige is usually applied in social network analysis (SNA) and applicable for only directed networks. The high prestige nodes refer to the terrorist who received many connections but initiates only a few relations. Research from [9] provides a new method called Influence Index which applied in the 9/11 terrorist network. For the same purpose, the idea of network efficiency was introduced [10]. By deactivating the node one by one, the algorithm can detect the most critical terrorist nodes that play an essential role in network functionalities. The higher influence value of the nodes is thus defined by the more massive drop in network efficiency caused by node removal. Unfortunately, this measure required high computational demand and inapplicable for the large-scale network. When nodes in the global terrorism network are corresponding to terrorism events, identifying the influential nodes could be done by measuring the severity of an event [11], e.g., the number of deaths and wounded. Although this is a simple and straightforward indicator to justify the node influence, the influence from neighborhoods and topology information are omitted.

Many studies on identifying the influential nodes have been conducted for network disruption. As mentioned in [12], the potential area of terrorist network can point out the role of leader as one factor indicating network vulnerability. Identifying the central nodes can essentially starve terrorist operations. Research from [2] and [13] show that removing high influential terrorist nodes can stop the terrorist activities, and the network ultimately broke down. One can see that a set of highly influential nodes has a high impact on network structure. Thereby, a fertile knowledge discovered from node identification is not limited to reduce the risk or effect of any attack that would happen in the future. Rather than that, it is deliberately carried out the feasible disruption plans and intensify the destruction of terrorist organizations.

Although countless studies were proposed, evaluating the nodes in the global terrorism network continually face serious difficulties. Since large-scale network typically contains noise connections that lead to frustration in understanding the actual role of the nodes. Analytically frivolous edges (e.g., attackers' actions, terrorist collaborations) may bring about biases in identifying the influential nodes. Furthermore, the emergence of terror nodes shares an evolving resemblance to other entities for social relationships [14]. Nodes in the global terrorism network are heterogeneous, and their relationship is hidden and complex. For these reasons, it requires specific node identification methods that are sensitive to network topology and efficient on big network data analysis.

In the network extraction context, a set of influential nodes that is advantageous to define the core network structure is unclear. Existing node identification methods mainly focus on discriminating the influence capability of a single node regardless of the group's ability to preserve key topology and execute the network functions. Researchers assumed that nodes with higher influence value or nodes whose ranking positions are more correlated with the real-spreading process; they are more important [15,16]. However, a group of influential nodes with an excellent combination might be significant for network structure. We believed that the influential nodes should not only be described by the capability of a node itself to influence or be influenced by other nodes, but they should also be a good representative for the whole network. Therefore, the degree to which the node identification method offers a set of influential nodes that preserves unique information and works best for a core network structure should be explored.

Another main reason is that most of the efficient node identification methods have been discussed and applied in various fields. For instance, social networks [17], biology [18], transportation [19], and WWW [20]. Nevertheless, the adoption of novel techniques developed in other scientific domains is limited in grounds of terrorism. It is truthfully that a few comparative research has been introduced to criminal and terrorist networks [4,21]. However, there is no systematic comparison relevant to the node's ability to controlling and influencing the overall structure of the global terrorism network.

Based on the reasons above, it is often helpful to compare the performance of existing node identification methods. There are two main research objectives. (1) We present a comparative analysis of six node identification methods over the global terrorism network. The capability of different methods on identifying the influential nodes is verified against the epidemic spreading process. (2) We examine the ability of influential nodes to produce a significant result in the process of skeleton network extraction. The comparison in this aspect aims at ensuring the quality of a set of nodes that influence the core network structure. It is inspired by the assumption that a small number of influential nodes preserving in a skeleton network can sufficiently describe ground-truth information and core structure of the original network. Besides, we use our findings to confirm the practical usefulness of influential node identification methods extensively, focusing on which method works best under network extraction circumstance and for what the extraction's objective.

The remainder of this paper is organized as follows. Section 2, we briefly review six node identification methods that exhibit potential in complex network analysis. In Section 3, we describe the details of global terrorism datasets, network construction process, and evaluation framework. In Section 4, we discuss the deployment of different methods, including the empirical results on node ranking correlation and structural-equivalence between the skeleton network and the original network. Section 5 is research conclusions that summarize our findings, limitations, and possible directions for future work.

## 2. Influential node identification methods

In this section, we review six efficient node identification methods, including Topology Potential algorithm, Topology Potential with mass parameter identified by k-shell centrality, Topology Potential with mass parameter identified by PageRank algorithm, Gravity Centrality Index, Influence Capability, and Multi-Attributes Centrality. The selection of these comparative methods is because they are innovative and have great potential in several contexts but lack of interests in the global terrorism network.

### 2.1. Topology potential

Topology potential (TP) was introduced in light of the data field [22,23]. TP is an indicator to reflect the ability of a node that depends on the differential position of a node over the network space. The importance of TP nodes can be explained by the sense that they adjacent to other important nodes [24,25]. On the ground of the Gaussian potential function [26], TP of any node $v_i$ is defined as

$$\varphi(v_i) = \sum_{j \in N} \left( m_j * exp^{-\left(\frac{d_{ij}}{\sigma}\right)^2} \right), \tag{1}$$

where $\varphi(v_i)$ is the topology potential of node $v_i$, for $v_i \in V$. $d_{ij}$ represents the shortest distance between a pair of nodes. $\sigma$ is a factor used to control the influence region which is generally calculated by a minimum entropy [22]. $m_j(\geq 0)$ stands for the mass of node $v_i$ $(i = 1, 2, \ldots, n)$ affected by the strength of the data field of node $v_j$ and is normalized to 1.

TP is considered as a pivotal technique to measure the importance value of the nodes. The application of TP covers the analysis methods in different domains, such as social networks [27,28], bioinformatics [29,30], and information search area [31]. TP is not only verified in the node evaluation domain but also has been proved by lots of evidence in community detection [32]. For example, the study from [33] improves the traditional spectral clustering algorithm using the appliance of TP. Their proposed algorithm can solve two significant problems, i.e., inefficient structural information and difficulty in deriving community number from ladder distribution. In [34], the TP-based community detection is integrated with the collaborative filtering recommendation in social networks. The more recent work from [35] proposed a method called DOCET by taking advantage of TP to solve some limitations in overlapping community detection, dynamic community identification, and community evolution analysis. This method can fully exploit the characteristics of topology potential field and shows its performance beyond most of the state-of-the-arts. TP nodes may contain different influential ability which can reflect meaningful knowledge from raw data to some extent. Applying TP may help illuminate our research on identifying the most influential nodes from the global terrorism network with vibrant network structural information.

### 2.2. Topology potential with mass identified by k-shell centrality

According to the TP formula, $m_j$ is an important parameter that affects the value of a node. Most research on applying TP adopts $m_j = 1$, representing the same influence for all nodes. In real-world networks, however, nodes may have different qualities and varying abilities to connect with others. Measuring the TP nodes with equal mass was criticized as unreasonable. In [31], the authors employed the entropy mapping method based on $f$ function corresponding to each node quality. The study from [36] also shows that $m_j$ can be used for describing an intrinsic property of the node. Therefore, TP could be induced by the diverse mass parameter.

In a previous study [37], a new node identification method called TPKS was introduced to find the skeleton nodes which are assumed to be responsible for the entire network information. The $m_j$ of TP is explored by various centrality indices. k-shell centrality (KS) [38] is one indicator used to define $m_j$ concerning the given Eq. (1). Their intensive experiments show that TP with $m_j = KS_{v_j}$outperforms other centralities and the classical TP in both social networks and syntactic networks. Specifically say that TP nodes with higher mass defined by KS, the more apparent of the node's influence and its ability.

### 2.3. Topology potential with mass identified by PageRank

As stated in 2.1, TP is not only applied for evaluating the node influence, but it is also putting forward to enhance the accuracy of community detection algorithms. Wang et al. 2014 [39] brought the PageRank algorithm [40] to improve $m_j$ in TP. The authors argue that when suppose $m_j = PageRank$, the nodes can reflect their actual value corresponding to different interaction abilities. We named this algorithm as TPPR. Similar work from [41], TPPR with the node location analysis is applied to select the seed node for community structure. It is proved that the mass parameter from PageRank produces a more reasonable distribution on topology potential field, leading to a more precise result in community detection. For this reason, the TPPR is also included in this comparative study.

### 2.4. Gravity Centrality Index

Gravity Centrality Index (GCI) [42] is one of the efficient methods that was presented under the light of k-shell centrality. The algorithm is inspired by the well-known Newton's law of gravity. In doing so, the mass parameter in the gravity formula is replaced by the k-shell value of two connected nodes, and the shortest path distance can be viewed as the distance among them.

The effectiveness of GCI is compared with different methods from various aspects and on different networks. GCI provides great potential beyond other methods when applied in scientist collaboration networks, yeast protein-to-protein networks, communication networks, and internet networks. This method also shows comparatively good performance in social networks, as described in [36]. The authors treated all tested networks as undirected and unweighted networks. However, none of the empirical studies is extended to explore the functioning of GCI on directed, weighted, and bipartite networks.

### 2.5. Influence capability

We can observe that the novel node identification methods based on k-shell centrality have aroused more attention among researchers. Another technique on this route is Influence Capability (IC) [43]. This method was proposed to fully consider the influence capability of a node determined by two attributes; node position and neighborhood information. Firstly, the IC utilizes the iteration number in the k-shell decomposition process to describe the node position information. Based on the decomposition process of k-shell, nodes are more important when they are peeled at a higher number of iterations. These nodes are closer to the network center and their position are crucial. Secondly, the neighborhood information respects the local influence capability of a node and can be calculated by the total degree of two-steps neighbors. The contribution of these two attributes relies on weight different information obtained by the entropy weighting method, rather than considering different attributes as equally important.

### 2.6. Multi-attributes centrality

Similarly, research from [44] introduced a new technique, namely Multi-Attributes Centrality (MAC). MAC relied on both node location information and neighbor information. This method takes advantage of the mechanism in k-shell decomposition to define node location and improve the resolution of the algorithm. Whereas, the neighbor information can be obtained by counting the number of nodes that can be reached by a particular node within two-steps.

Although the fundamental concept of IC and MAC is similar. However, we noticed that the given definitions and criteria for calculation in both node position information and neighborhood effect are different. Therefore, it is useful to bring these two methods to our comparative study as the methods might reveal the specific characteristic that contributes to identifying influential nodes and core network structure more clearly.

## 3. Experimental settings

### 3.1. Network data

A dataset from the Global Terrorism Database (GTD) has been used for our study. GTD is retrieved and maintained by the National Consortium for the Study of Terrorism and Responses to Terrorism (START) [45]. The dataset is publicly available and can be downloaded from https://www.start.umd.edu/gtd/contact/. The data are collected from several publicity sources that include media articles, electronic news, books, and journal, as well as legal documents. GTD consists of incidents that occurred around the world from 1970 to 2017. However, the incidents from 1993 are not included in our study due to there are more than 80% of attacks that are unreliable [46]. The entire GTD dataset contains nearly 182,000 observations with 135 attributes such as perpetrator information, target or victim information, and the number of casualties and consequences.

GTD is a large-scale dataset that is required a cleaning process before a network construction and investigation. The incidents recorded in the GTD are often ambiguous between terrorism, crime, or political violence. To make a precise determination, we deleted all incidents that START suggested to be unclear. In case that the incidents were carried out by "Unknown Terrorist" and "Unknown Target", they were also removed from the dataset. Our work does not include the attacks that were attempted but were ultimately unsuccessful. We eliminate all incidents that targeted to multination. This deletion is done for specifically remain network modularity to a particular nation. Also, all attacks occurred in the international sector, and the attacks with unavailable geolocation were filtered out. The data cleaning process results with 70,113 observations remaining in the dataset.

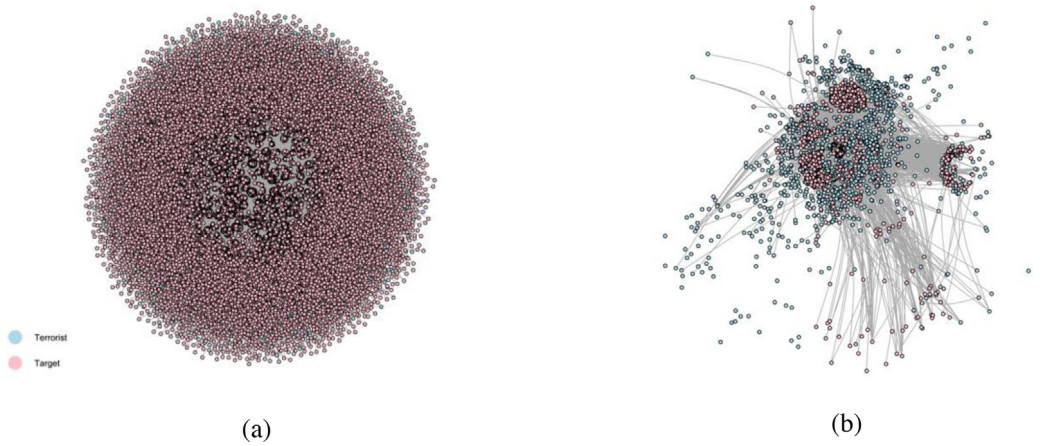


**Fig. 1.** The graphical representation of the global terrorism network. (a) corresponds to a network of the entire period 1970–2017, (b) is the zoom-in view of this network that exhibits the hairball network structure.

### *3.2. Network construction*

After cleaning the dataset, we can then construct the global terrorism network. The subject of this network is the terrorists and targets that can be represented by nodes. Edges illustrate the relationships between two nodes and indicate the involvement of two classes of nodes in a terrorism incident. Given $\mathcal{B}(U, V, E)$ as a bipartite and directed network with two disjoint node-sets $V$ and $U$, where $V$ is the primary node-set and $U$ is the secondary node-set. Such that $V \cap U = \emptyset$ and $E \subseteq V \times U$. Nodes are connected by a set of edge $E$ that signifies a violent action from terrorist node $v \in V$ to target node $u \in U$. Accordingly, the connection between nodes in the same class (e.g., terrorist-to-terrorist, target-to-target) is not allowed. Note that evaluating the influence value of the nodes in a bipartite (two-mode) network is still not well defined. The empirical node identification methods make no allowance for quantifying node's influential values that are specific to different classes of nodes. Therefore, we treat all nodes as though they are in the same class.

The constructed network expresses a more substantial number of target nodes than terrorist nodes. Along with 70,133 edges, the entire network comprises 34,199 nodes, while the largest connected component covers 31,461 nodes and 37,056 edges. This component encompasses nearly 92% of the overall nodes in the network. Fig. 1 confirms that the global terrorism network exhibits a hairball structure where its real topological structure is hidden. There are some of the periphery nodes around the network. Those can be considered as isolated nodes with a single or a few connections.

### *3.3. Framework for comparison*

In this study, we employ several evaluation techniques to compare the performance of different node identification methods. We have divided all evaluation metrics into two classes. First, the metrics used to verify the accuracy of a particular method in identifying the influence of individual nodes. The metrics integrated into this class include the node's influence capability by the SIR model, the node's ranking correlation by Kendall' tau coefficient correlation, and the node's difference measure by Monotonicity value. Second, the metrics used to confirm the significance of a node-set to the core network structure. The principle evaluations in this class are performed over the extracted skeleton networks. The metrics include the measurement of Spearman's rank-order correlation and the network structural-equivalence by QAP.

We conduct the implementation of all topology potential-based methods within Python programming language in Google Colab environment. Meanwhile, we process the GCI, IC, and MAC and do all types of comparisons using R platform in Mac OS. The SIR simulations are performed using Matlab, while data visualization is generated by Rstudio.

## 4. Results and discussion

### *4.1. Basic characteristics of the global terrorism network*

In an attempt to uncover the specific network structure, we divide the original network into ten subnetworks. Each subnetwork contains all terrorism incidents within five years (3 years for the last timeframe). Table 1 summarizes several basic network characteristics of the entire global terrorism network and all subnetworks. The size of subnetworks ranges from hundreds to ten-thousand nodes and edges, while their average shortest path length is identical ($d = 1$).

All subnetworks presented in Fig. 2 are independently drawn for better visualization. It is observed that a majority of the nodes are connected, while a few periphery nodes remain in all subnetworks. Accordingly, removing key nodes which

**Table 1**
The topological features of the entire global terrorism network and all subnetworks. $n$ and $m$ are the numbers of nodes and edges. $\langle k \rangle$ and $k_{max}$ are average degree and maximum degree. $d$ is the average shortest path length. *dense* and $c$ represent network density and the global clustering coefficient, while $g$ is the number of communities. $\beta_{th}$ denotes the spreading threshold of each subnetwork and $\beta$ is the infection probability of the SIR process.

| GTD Network | $n$ | $m$ | $\langle k \rangle$ | $k_{\max}$ | $d$ | *dense* | $c$ | $g$ | $\beta_{th}$ | $\beta$ |
|---|---|---|---|---|---|---|---|---|---|---|
| GTD: Original | 34 199 | 70 113 | 4.1003 | 5246 | 1 | 5.99E−05 | 0.3302 | 900 | 0.0014 | 0.01 |
| GTD: 1970–1974 | 896 | 1599 | 3.5692 | 216 | 1 | 1.99E−03 | 0.3031 | 70 | 0.0179 | 0.02 |
| GTD: 1975–1979 | 2102 | 3643 | 3.4662 | 354 | 1 | 8.25E−04 | 0.3019 | 175 | 0.0144 | 0.02 |
| GTD: 1980–1984 | 3317 | 5942 | 3.5828 | 1083 | 1 | 5.40E−04 | 0.1793 | 218 | 0.0052 | 0.01 |
| GTD: 1985–1989 | 4186 | 8531 | 4.0760 | 1574 | 1 | 4.87E−04 | 0.4957 | 205 | 0.0035 | 0.01 |
| GTD: 1990–1994 | 4299 | 7825 | 3.6404 | 958 | 1 | 4.23E−04 | 0.3680 | 338 | 0.0062 | 0.01 |
| GTD: 1995–1999 | 2751 | 3457 | 2.5133 | 223 | 1 | 4.57E−04 | 0.2322 | 324 | 0.0178 | 0.02 |
| GTD: 2000–2004 | 2881 | 3382 | 2.3478 | 236 | 1 | 4.08E−04 | 0.4716 | 234 | 0.0210 | 0.03 |
| GTD: 2005–2009 | 4000 | 5212 | 2.6060 | 741 | 1 | 3.26E−04 | 0.4601 | 214 | 0.0064 | 0.01 |
| GTD: 2010–1014 | 8930 | 14 457 | 3.2379 | 2170 | 1 | 1.81E−04 | 0.4415 | 287 | 0.0023 | 0.01 |
| GTD: 2015–2017 | 9915 | 16 065 | 3.2405 | 2926 | 1 | 1.63E−04 | 0.2800 | 293 | 0.0016 | 0.01 |

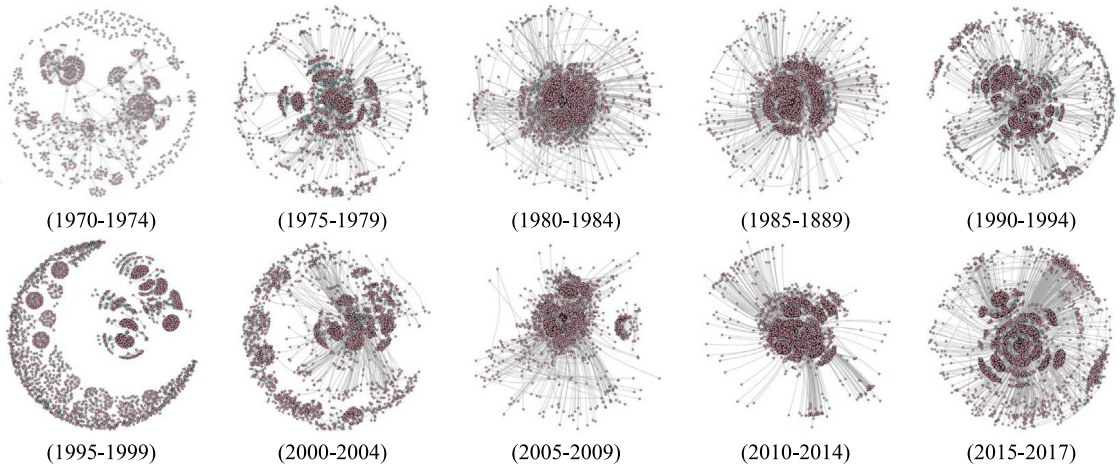


**Fig. 2.** The global terrorism networks in different timeframes.

is responsible for hub entities could easily affect other parts of the network. Different networks reveal the distinctive densely connected subgroups. Nodes of subnetworks from the earliest timeframes are loosely connected and carry a comparatively low degree. However, the situation in the more recent timeframes is different. Subnetworks of 2010–2014 and 2015–2017 represent the more centralized nature and contain tremendous edges, nearly twice times as high as the number recorded in other timeframes. A vast number of nodes and edges leads network highly complex and is more challenging to determine the actual node influence value. Leastwise, partitioning a terrorism network into small pieces should allow us to effortlessly investigate the evolution of the network and explore their characteristics with the time involved.

At this point, we study the cumulative degree distribution. Where $p(x)$ denotes probability density function, $k$ is a degree of the nodes. A simple measure of the degree distribution can give us a partial view of the network structure and discriminate the type of the network. There are three types of node degree, including total-degree, incoming-degree, and outgoing-degree. The cumulative degree distribution provided in Fig. 3(a) shows all-degree of the nodes exceeds the maximum value over 5000 degrees. Whereas the highest in-degree is only 714, and it is much lower than the total amount of outgoing degree. When the node degree increase, all-degree, and out-degree distributions become similar, this situation implies the node's outgoing connectivity highly induces the overall degree distribution of the network. All types of degree distribution follow a linear downward sloping trend. From this information, the global terrorism network probably is a scale-free network that is characterized by the power-law. The distribution in the graph means that there is only a small number of actors who have a large number of connections, while the majority of them carry a small number of links.

Another essential characteristic of the underlying network is the clustering coefficient. It worth to remind that some of the network analyses cannot be applied directly to a bipartite network. Hence, we apply the one-mode projection method from *tnet* [47] to reveal the clustering coefficient distribution. Fig. 3(b) illustrates the local clustering coefficient that is plotted against the degree $k$. As can be seen, the node clustering value decreases as the node degree increases. The distribution entails that the extremely dense sub-graphs contain nodes with a low degree, and hub nodes play a role in connecting these sub-graphs. The characteristic of the scale-free makes terrorism network resilient to a random

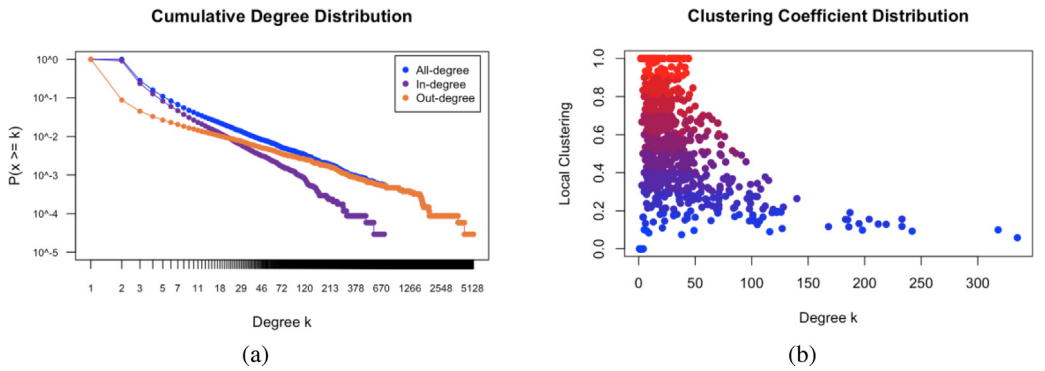


**Fig. 3.** The cumulative degree distribution (a) and clustering coefficient distribution (b) of the entire global terrorism network.

**Table 2**
The node ranking correlation of TP, TPKS, TPPR, GCI, IC, and MAC compared with the SIR model. $\tau(\cdot)$ is Kendall's tau correlation coefficient of the corresponding node identification methods.

| GTD Network | $\tau(TP)$ | $\tau(TPKS)$ | $\tau(TPPR)$ | $\tau(GCI)$ | $\tau(IC)$ | $\tau(MAC)$ |
|---|---|---|---|---|---|---|
| GTD: Original | 0.8780 | 0.6632 | 0.2664 | 0.5106 | 0.3976 | 0.3734 |
| GTD: 1970–1974 | 0.7938 | 0.5594 | 0.2884 | 0.5719 | 0.3768 | 0.3424 |
| GTD: 1975–1979 | 0.8264 | 0.6470 | 0.2754 | 0.5760 | 0.4005 | 0.3237 |
| GTD: 1980–1984 | 0.8125 | 0.5719 | 0.2986 | 0.4759 | 0.3370 | 0.3114 |
| GTD: 1985–1989 | 0.8291 | 0.5374 | 0.2696 | 0.4839 | 0.3613 | 0.3510 |
| GTD: 1990–1994 | 0.8850 | 0.6723 | 0.2859 | 0.5802 | 0.4576 | 0.4396 |
| GTD: 1995–1999 | 0.9644 | 0.8679 | 0.2658 | 0.5606 | 0.4507 | 0.4266 |
| GTD: 2000–2004 | 0.9502 | 0.8333 | 0.2108 | 0.4796 | 0.3651 | 0.3512 |
| GTD: 2005–2009 | 0.9535 | 0.8160 | 0.2741 | 0.4024 | 0.3375 | 0.3221 |
| GTD: 2010–1014 | 0.8616 | 0.5584 | 0.2335 | 0.4738 | 0.4272 | 0.4187 |
| GTD: 2015–2017 | 0.9034 | 0.5686 | 0.2615 | 0.5019 | 0.4618 | 0.4475 |

attack. As emphasized in [48], randomly attacking tends to target a node in the majority that has a lower influence. When removing the boundary nodes, it has no or less damage to the entire network's functions. Nonetheless, directly targeting only a small portion of known hubs or the relatively high influential nodes could affect the whole network structure to break-up and collapse. It could be confirmed that identifying the high influential nodes is a significant task that offers an excellent opportunity to destabilizing the global terrorism network.

### *4.2. Node ranking correlation*

To our knowledge, the practical node identification method should generate the influential nodes that are correlated with the real spreading process as much as possible. The Susceptible–Infected–Recovered (SIR) model is an epidemic spreading process that is commonly used [49,50]. Accordingly, we employ the SIR model to estimate the influence value of the nodes. Before performing the SIR, we first compute the epidemic threshold $\beta_{th}$. The $\beta_{th}$ is given as $\beta_{th} \sim \langle k \rangle / \langle k^2 \rangle$, where $\langle k \rangle$ and $\langle k^2 \rangle$ stand for the average degree and the 2nd-moment average degree. Afterward, we set the infection probability $\beta$ slightly larger than $\beta_{th}$. Both $\beta_{th}$ and $\beta$ showed in Table 1 are calculated from the largest connected components. SIR usually requires a number of simulations $\approx$ 1000 for more reliable results. However, applying this model over the large-scale network is time-consuming and exceeds the capability of our resources. We, therefore, run the SIR simulations for 100 times for each subnetwork. Finally, the influence of a node could be derived from the average number of recovered nodes at the end of the spreading process.

Consequently, we adopt Kendall's $\tau$ correlation coefficient [51,52], to measure the extent to which a ranking of the influential nodes from each method is consistent with the SIR model. Kendall's $\tau$ correlation value resulting in the range between $+1$ to $-1$. $\tau = 1$ means the ranking of the two methods is similar, while a negative score signifies entirely different among the two rankings. The larger $\tau$ value, the more accuracy the node identification method. The results of $\tau$ from different methods are varied. We collect all the results from different subnetworks as listed in Table 2. Note that we fail to derive the ground-truth information of the entire original network from the SIR simulations. Therefore, we represent $\tau$ of the original network by average correlation value from all subnetworks.

Table 2 manifests that TP, TPKS, TPPR, GCI, IC, and MAC are all positively correlate with the influence capability of the nodes evaluated by the SIR model. TP achieves the strongest correlations for all subnetworks, thus leads to the highest $\tau$ in the entire original network. The TPKS performs a slightly lower correlation than the classical TP. The performance of both TP and TPKS reflects high accuracy in capturing the influence value of individual nodes. As compared to other methods, the

correlation of TPPR is weaker. This because in these specific networks, PageRank itself always disagreed with the node's importance ranking from SIR. The negative correlation of PageRank thus directly affects the value of TPPR. Meanwhile, the $\tau$ result of IC and MAC are closely correlated, and both are comparatively lower than GCI. It shows that the influential nodes identified by GCI, IC, and MAC have moderate correlations with the ranking from the real spreading process.

In most situations of identifying the node's influence, we are more interested in the nodes which subsist at the top of the ranking. Thus, we rank all nodes according to their influence value in decreasing order. For those with the same value, their relative order is random. We then select the top three influential nodes from different methods to represent their effective results. For this comparison, the top nodes identified by SIR are illustrated as ground-truth information for all subnetworks. Meanwhile, the degree centrality is used as a comparison criterion for the original network. As shown in Table 3, nodes which are categorized into the top 3 most influential nodes from TP, TPKS, and TPPR are almost identical. They are also similar to the ranking from SIR. All high influential nodes from these three methods are in terrorist class while nodes in the target class are not even included in the top 30.

On the contrary, the ranking obtained by GCI, IC, and MAC completely deviates from the ground-truth. While SIR and other methods pinpoint at the nodes in the terrorist class, it is interesting that the most influential nodes discovered by IC and MAC are target nodes. This might due to IC and MAC determine the influential nodes regarding neighborhood concentrations. The high value of target nodes is collected from the influence of the critical terrorists who attacked them before.

Considering the subnetwork of 2015–2017, the terrorist nodes which carry the most influence value are Taliban, ISIL, and Boko Haram. Taliban is a Sunni Islamic group and political movement which emerges in Afghanistan. This terrorist group has recently expanded its territorial control at least 11% of the country and contest another 29 percent of Afghanistan's 398 provinces [53]. ISIL is the terrorist organization based in Iraq and Syria in which accounts for more than 2600 deaths in 2017. This terrorist is also the main driver for terrorist attacks in Europe and other developed countries. Boko Haram is operating primarily in northeastern Nigeria, also active in Chad, Niger, and Cameroon.

The top influential nodes discovered by TP, TPKS, and TPPR are consistent with the statistical reports by the Global Terrorism Index (GTI). As stated by GTI 2018 [54], Taliban, ISIL, and Boko Haram responsible for the most deaths in 2017. Furthermore, the top five countries most affected by global terrorism are ranked from Iraq, Afghanistan, Nigeria, Pakistan, and Syria, the countries in which these three terrorist groups sought to create Islamist state and control the political movements. This information gives evidence in the utility of the TP-based methods to deal with the influence capability of the nodes over the large-scale global terrorism network.

Table 4 mainly reveals the outcome of different methods over the subnetwork of 2015–2017. The information listed in this table confirmed that TP, TPKS, and TPPR contain the most similar set of top influential nodes in which correlates to the SIR. However, the outcomes produced by the GCI, IC, and MAC unveil more fluctuation. From this experiment, we also found that although the overall $\tau$ of TPPR is lowest, TPPR nodes at the top of the ranking are well-matched to the ranking from SIR, TP, and TPKS. Through this meaning, we argue that TPPR is precisely by providing accurate nodes at the top position, but not at the tail of the ranking order.

### 4.3. Node difference measure

As mentioned in [42], the excellent node identification method should provide a high resolution in its ranking result. Hence, we perform the Monotonicity index $M$ to evaluate the ability of each method in distinguishing the difference between the nodes. When $M$ of the one ranking vector equal to 1, the node importance ranking method is perfectly monotonic due to each node is classified into the different index values. $M$ which is closer to 0, means all nodes have a more identical rank.

In Table 5, MAC provides a perfect monotonicity value, which means the method can classify all nodes in the network into different ranks. TP, TPKS, TPPR, and IC are also good at distinguishing the node's difference. Except in the case of GCI in which the method provides much lower monotonicity than others. The reason is that the calculation of GCI relies on the k-shell value that quite coarse-grained [42]. Based on k-shell, a large number of nodes in the network are assigned to the same shell (rank) and have the same value. Therefore, the influence capability of the nodes cannot be wholly distinguished by the GCI method. Whereas the contributions of k-shell decomposition, when combined with TP, IC, and MAC, achieved a precise performance. This because the methods do not apply k-shell value directly. Instead, k-shell is integrated for the mass parameter under control of TP and plays a role in resource iteration for IC and MAC.

### 4.4. Skeleton network extraction

Aside from the significance of the individual nodes, we now investigate how variation in node identification methods relates to differences in the core structure of the network. It is of fundamental importance to determine which nodes are the critical target nodes indicating the fragile area and which terrorist nodes are responsible for the deadliest groups. Such that counterterrorism and government can put more concern on these critical nodes and increase awareness sessions to local outreach activities. However, in the case of terrorist actors, it is still difficult to identify the nodes that are highly active presence and create a strong impact within recent years. This difficulty is because many terrorists have regional affiliates. They work together as a partnership or cooperate under the same command. One to emphasize is that not only highly influential nodes are significant, but a set of connected nodes that efficiently combine their actions and efforts is also significant. In this section, we, therefore, demonstrate the achievement of a set of nodes identified by each method in capturing the core network structure called "skeleton network".

**Table 3**
The top 3 influential nodes of different subnetworks measured by six comparative methods. The number in parenthesis signifies the node lift (+1) or drop (−) in its ranking as compared to the result from SIR.

| Network | Rank | TP | TPKS | TPPR | GCI | IC | MAC | SIR |
|---|---|---|---|---|---|---|---|---|
| Original | 1 | Taliban | Taliban | Taliban | ETA | Business San Miguel | Government Santa Rosa | Taliban |
| | 2 | ISIL | ISIL | ISIL | FMLN | Government Santa Rosa | Business San Miguel | ISIL |
| | 3 | NPA | SL | NPA | Armenian Secret Army for the Liberation of Armenia | Police San Carlos | Citizens Dagu | SL |
| 1970–1974 | 1 | IRA | IRA | IRA | Black Sept.(+2) | Business New York City(+95) | Business New York City(+95) | IRA |
| | 2 | Left-Wing Militants | Left-Wing Militants | Left-Wing Militants | Weathermen (+6) | Government Washington(+97) | Citizens Belfast(+95) | Left-Wing Militants |
| | 3 | Black Sept. | Black Sept. | Black Sept. | JDL(+10) | Black September | Government Washington(+96) | Black Sept. |
| 1975–1979 | 1 | ETA | ETA | ETA | GRAPO(+11) | Government Madrid(+242) | Government Madrid(+242) | ETA |
| | 2 | IRA | IRA | IRA | Armenian Secret Army for the Liberation of Armenia(+9) | ETA(−1) | Citizens Bilbao(+242) | IRA |
| | 3 | FPL | FPL | Red Brigades(+3) | Palestinians(+2) | Citizens Belfast(+243) | Business Paris(+242) | FPL |
| 1980–1984 | 1 | SL | SL | SL | M-19(+5) | Business San Miguel(+244) | Business San Miguel(+244) | SL |
| | 2 | FMLN | FMLN | FMLN | Left-Wing Terrorists(+22) | Citizens San Miguel(+244) | Citizens San Miguel(+244) | FMLN |
| | 3 | ETA | ETA | ETA | Left-Wing Guerrillas(+26) | Police San Miguel(+245) | Citizens La Paz(+244) | ETA |
| 1985–1989 | 1 | SL | SL | SL | Terrorists(+31) | Citizens Concepcion(+188) | Citizens Concepcion(+188) | SL |
| | 2 | FMLN | ELN(+3) | FMLN | Death Squad(+13) | Citizens San Gerardo(+188) | Citizens San Gerardo(+188) | FMLN |
| | 3 | Sikh Extremists(+1) | FMLN(−1) | Sikh Extremists(+1) | FMLN(−1) | Business Lima(+189) | Government Santa Rosa(+188) | JVP |

(*continued on next page*)

**Table 3** (*continued*).

| Network | Rank | TP | TPKS | TPPR | GCI | IC | MAC | SIR |
|---|---|---|---|---|---|---|---|---|
| 1990–1994 | 1 | SL | SL | SL | Muslim Militants(+8) | Utilities Ayacucho(+209) | Utilities Ayacucho(+209) | SL |
| | 2 | PKK | PKK | PKK | Death Squad(+25) | Business Medellin(+217) | Business Lima(+209) | PKK |
| | 3 | Sikh Extremists | Palestini-ans | Palestinians | ELN(+2) | Police San Carlos(+212) | Citizens San Miguel(+209) | Palestinians |
| 1995–1999 | 1 | FARC | FARC | FARC | Colombian Patriotic Resistance(+76) | Government Medellin(+100) | Military Medellin(+99) | FARC |
| | 2 | LTTE | LTTE | LTTE | ELN(+2) | Military Medellin(+98) | Government Medellin(+99) | LTTE |
| | 3 | ELN(+1) | ELN(+1) | ELN(+1) | Islamist extremists(+2) | Government Uramita(+102) | Military Bogota(+99) | PKK |
| 2000–2004 | 1 | FARC | FARC | FARC | Gunmen(+9) | Citizens Bogota(+87) | Citizens Bogota(+87) | FARC |
| | 2 | Chechen Rebels(+1) | Chechen Rebels(+1) | Chechen Rebels(+1) | Rebels(+22) | Citizens San Pablo(+87) | Citizens San Pablo(+87) | MILF |
| | 3 | MILF(−1) | MILF(−1) | MILF(−1) | Muslim extremists(+12) | Police Cali(+89) | Business Medellin(+87) | Chechen Rebels |
| 2005–2009 | 1 | Taliban | Taliban | Taliban | TTP(+3) | Citizens Miran Shah(+100) | Citizens Miran Shah(+100) | Taliban |
| | 2 | CPI-Maoist | CPI-Maoist | CPI-Maoist | Muslim extremists(+43) | Citizens Bajaur district(+100) | Citizens Bajaur district(+100) | CPI-Maoist |
| | 3 | TTP(+1) | TTP(+1) | LTTE(+2146) | Al-Qaida(+38) | Citizens Mohmand district(+100) | Citizens Mohmand district(+100) | NPA |
| 2010–1014 | 1 | Taliban | Taliban | Taliban | Gunmen(+36) | Citizens Karma(+176) | Citizens South Waziristan district(+104) | Taliban |
| | 2 | CPI-Maoist | CPI-Maoist | CPI-Maoist | Muslim extremists(+9) | Citizens South Waziristan district(+103) | Citizens Bajaur district(+104) | CPI-Maoist |
| | 3 | Boko Haram(+1) | ISIL | Boko Haram(+1) | Lashkar-e-Jhangvi(+41) | Citizens Bajaur district(+103) | Educational Institution Safi(+104) | ISIL |
| 2015–2017 | 1 | Taliban | Taliban | Taliban | ISIL(+1) | Citizens Shawa(+96) | Citizens Shawa(+96) | Taliban |
| | 2 | ISIL | ISIL | ISIL | Muslim extremists(+10) | Military Basra(+124) | Military Mohmand district(+109) | ISIL |
| | 3 | Boko Haram | Boko Haram | Boko Haram | BLA(+45) | Government Tuulo Barwaqo(+96) | Military South Waziristan district(+109) | Boko Haram |

**Table 4**
The ranking results from six comparative methods over the subnetwork of 2015–2017.

| SIR | | TP | | TPKS | | TPPR | | GCI | | IC | | MAC | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ID | Value | ID | Value | ID | Value | ID | Value | ID | Value | ID | Value | ID | Value |
| 3 | 13.8 | 3 | 16.77 | 3 | 40.55 | 3 | 3.82E−03 | 1 | 8135 | 5903 | 2836.59 | 5903 | 1324.42 |
| 1 | 9.5 | 1 | 12.88 | 1 | 31.42 | 1 | 2.93E−03 | 35 | 7554 | 8280 | 2686.46 | 5724 | 1263.11 |
| 18 | 6.8 | 18 | 7.58 | 18 | 16.58 | 18 | 1.59E−03 | 42 | 5635 | 3292 | 2686.39 | 2754 | 1263.08 |
| 19 | 5.7 | 22 | 7.09 | 22 | 15.40 | 22 | 1.48E−03 | 111 | 5404 | 2896 | 2559.37 | 2003 | 1244.40 |
| 8 | 5 | 19 | 6.09 | 19 | 13.06 | 19 | 1.23E−03 | 20 | 5224 | 1629 | 2515.04 | 2209 | 1242.37 |
| 22 | 4.9 | 8 | 5.53 | 8 | 11.61 | 8 | 1.10E−03 | 17 | 4970 | 2416 | 2512.23 | 902 | 1240.93 |
| 39 | 4.8 | 39 | 4.63 | 39 | 9.40 | 39 | 8.91E−04 | 32 | 4853 | 2222 | 2512.12 | 749 | 1240.92 |
| 6 | 3.7 | 13 | 3.66 | 13 | 7.31 | 13 | 6.56E−04 | 7 | 4552 | 9151 | 2512.12 | 2014 | 1240.32 |
| 12 | 3.5 | 12 | 3.39 | 6 | 6.22 | 12 | 5.80E−04 | 18 | 4444 | 1325 | 2484.38 | 643 | 1240.31 |
| 63 | 3.4 | 6 | 3.36 | 12 | 6.00 | 6 | 5.74E−04 | 13 | 4424 | 924 | 2438.50 | 4047 | 1240.11 |
| 13 | 3.2 | 20 | 2.83 | 111 | 5.75 | 20 | 4.69E−04 | 112 | 4260 | 1 | 2438.14 | 1401 | 1240.10 |
| 35 | 3 | 111 | 2.77 | 35 | 5.64 | 35 | 4.60E−04 | 149 | 4255 | 2745 | 2434.57 | 6160 | 1239.90 |
| 111 | 2.4 | 35 | 2.64 | 20 | 5.41 | 111 | 4.47E−04 | 86 | 4227 | 4978 | 2417.83 | 1366 | 1239.89 |
| 31 | 2.4 | 63 | 2.47 | 63 | 3.96 | 63 | 3.72E−04 | 2 | 4211 | 7108 | 2409.09 | 1029 | 1239.88 |
| 20 | 2.1 | 31 | 2.39 | 7 | 3.60 | 31 | 3.38E−04 | 369 | 4203 | 576 | 2338.93 | 6796 | 1239.69 |

**Table 5**
The ability of six comparative methods to distinguish the node's difference. $M(\cdot)$ is the monotonicity index of the corresponding node identification methods.

| | $M(TP)$ | $M(TPKS)$ | $M(TPPR)$ | $M(GCI)$ | $M(IC)$ | $M(MAC)$ |
|---|---|---|---|---|---|---|
| GTD: Original | 0.9647 | 0.9780 | 0.9944 | 0.6082 | 0.9788 | 1 |

The emergence of research on network extraction has given new relevance to study the significance of the nodes to the entire network structure. Generally, a skeleton network is an outcome of the skeletonizing process after filtering some meaningless data from the complex network [55]. Skeleton network extraction considers only network key elements and aims to preserve the essential information of the original network. In this paper, the skeleton of the global terrorism network is defined as "a representative of the original network that contains highly influential nodes and edges, in which sufficiently maintain the whole network functions". As a result, the extracted skeleton networks can be used to explain behavioral patterns of the original network immediately. We suppose that the involvement of key actors identified by different node identification methods can allow us to discover distinctive skeleton structures.

In previous work [56], the network extraction model was applied over internet networks and citation networks. This model takes advantage of the hub nodes measured by TP to extract the skeleton network at different network granularities. The algorithm ranked all TP nodes in decreasing order and selects hub nodes that satisfied parameter $\alpha$. The initial set of hub nodes is denoted by source, while any isolated node in the source is considered as an island subnet. Consequently, the algorithm finds the connections between two subnets based on the shortest distance and join edges to the source. Intuitively, the parameter $\alpha$ is the only parameter of this algorithm which directly affects the number of preserved nodes and edges in the skeleton network.

In this paper, we build on that work by re-defining the value of hub nodes regarding the results from the six node identification methods. The skeleton network extraction algorithm is only applied to the entire network (original network). It is remarkable that the global terrorism network is bipartite and contains disconnected components. Over this specific network, the algorithm is limited in uncovering the shortest path among nodes. To solve this problem, we put our consideration for all formerly edges connecting to the hub nodes. All edges between two nodes are then simplified and mapped into a single one. We set the cut-off $\alpha$ parameter equal to 0.8, where the algorithm can retain around 20% of top nodes from the original network. Thereby, a skeleton network is reconstructed by encompassing with the smallest possible subset of highly influential nodes and their relevant edges. A small set of nodes in each skeleton network is believed to preserve the organization structure and major functions as similar to the original one. The more similar the skeleton network to the original network, as well as the reasonable network size and less complexity than others, the more important a set of nodes to the network's core structure. The extracted skeleton networks based on each of the six methods are called as TP-skeleton, TPKS-skeleton, TPPR-skeleton, GCI-skeleton, IC-skeleton, and MAC-skeleton hereafter. The representations of skeleton networks are displayed in Fig. 4.

The degree distribution of skeleton networks is illustrated in Fig. 5. As similar to the original network, node degree distributions from different skeleton networks are characterized by a wide variety and long-tailed distribution that reveals the heterogeneous topology in actor's activities. The distribution of the IC-skeleton and MAC-skeleton strays from the usual paths of the original network and other skeleton networks. This because the majority of their preserved nodes are different, most of which hold lower degree centrality. In observing the cumulative clustering coefficient in Table 6, we found that all nodes in different skeleton networks tend to cluster together at an almost similar degree, and consistent with the cluster behavior of the original network. The IC-skeleton provides much differ with the highest clustering

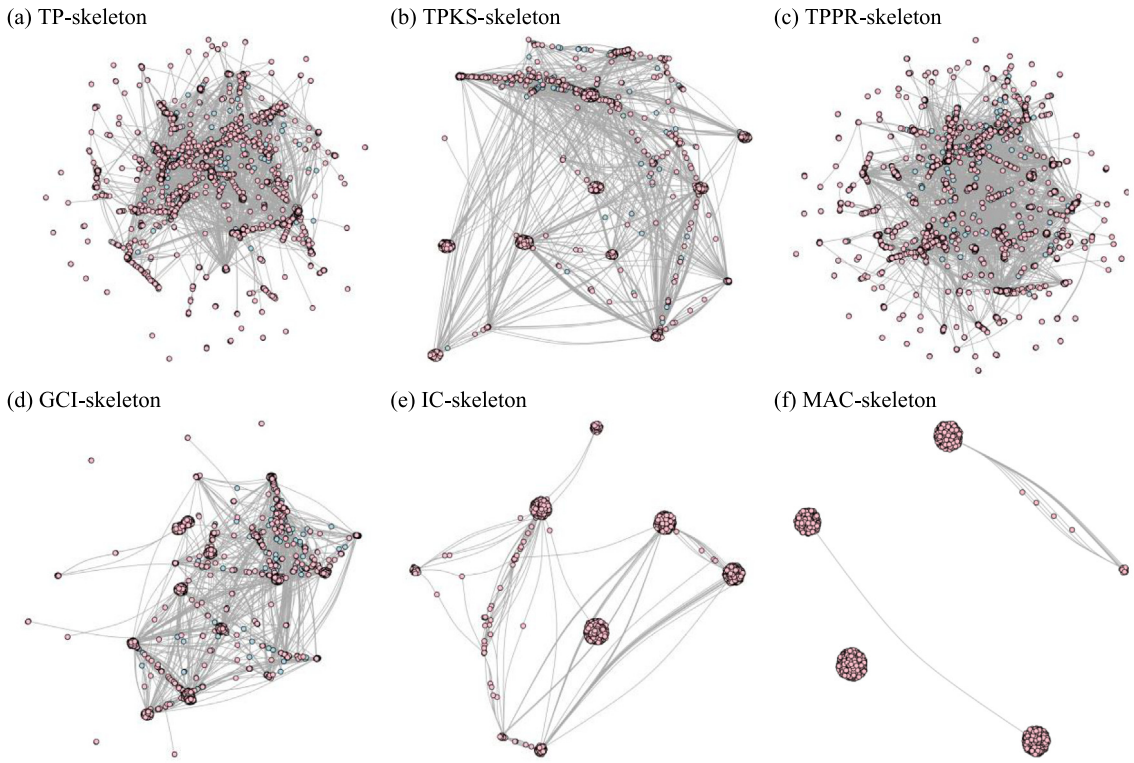


**Fig. 4.** The graphical representation of the extracted skeleton networks from different methods. (a) TP-skeleton, (b) TPKS-skeleton, (c) TPPR-skeleton, (d) GCI-skeleton, (e) IC-skeleton, and (f) MAC-skeleton.

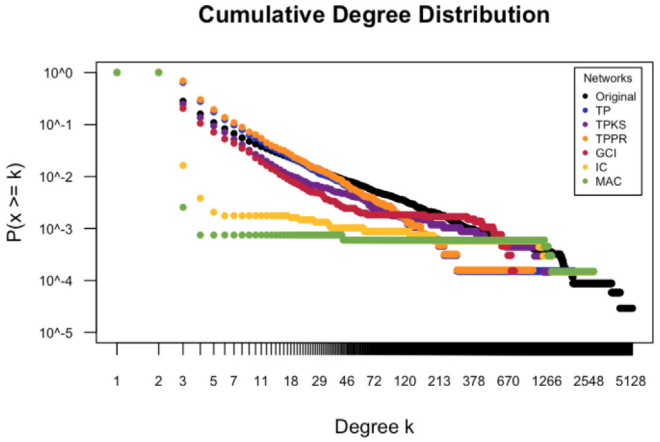


**Fig. 5.** The cumulative degree distribution of different skeleton networks.

coefficient. Unfortunately, the clustering coefficient of MAC-skeleton could not be derived since the skeleton network does not contain fully connected cliques that commonly used to calculate the clustering coefficient over the one-mode projected network [47].

The results in Table 6 manifests that all extracted skeleton networks can capture the topological features of the original network. Comparing with other skeleton networks, the TPPR-skeleton contains fewer nodes while its average degree is highest. It could refer to almost nodes in TPPR-skeleton have relatively high degree centrality. Although the TPKS, GCI, IC, and MAC preserve a larger number of nodes in the skeleton, their fraction of edges is much reduced than that of the TP-skeleton and TPPR-skeleton. Thus, the network structure and connectivity of the TPKS-skeleton, GCI-skeleton, IC-skeleton, and MAC-skeleton are less complicated.

**Table 6**
The topological features of different skeleton networks. $n$ and $m$ are the numbers of nodes and edges. $\langle k \rangle$ and $k_{max}$ are the average degree and the maximum degree. $d$ is the average shortest path length. *dense* and $c$ represent network density and the global clustering coefficient, while $g$ is the number of communities that exists in the network.

| Network | $n$ | $m$ | $\langle k \rangle$ | $k_{\max}$ | $d$ | *dense* | $c$ | $g$ |
|---|---|---|---|---|---|---|---|---|
| Original | 34 199 | 70 113 | 4.1003 | 5246 | 1 | 5.99E−05 | 0.3302 | 900 |
| TP-skeleton | 6610 | 12 190 | 3.6884 | 2002 | 1 | 2.79E−04 | 0.4291 | 24 |
| TPKS-skeleton | 6824 | 9849 | 2.8866 | 1372 | 1 | 2.12E−04 | 0.5886 | 1 |
| TPPR-skeleton | 6318 | 11 890 | 3.7638 | 1002 | 1 | 2.98E−04 | 0.3407 | 32 |
| GCI-skeleton | 6577 | 8646 | 2.6292 | 737 | 1 | 2.00E−04 | 0.3695 | 4 |
| IC-skeleton | 6827 | 6931 | 2.0305 | 2752 | 1 | 1.49E−04 | 0.8098 | 1 |
| MAC-skeleton | 6668 | 6675 | 2.0021 | 2752 | 1 | 1.50E−04 | NA | 3 |

**Table 7**
The QAP correlation and reduction rate of skeleton networks. Terrorist*, and Edges* represent the number of preserved nodes and edges from the one-mode projected networks. $\%R_N$, and $\%R_E$ are the percentage of network reduction from nodes and edges, respectively.

| Skeleton network | Terrorist* | Edges* | $\%R_N$ | $\%R_E$ | QAP |
|---|---|---|---|---|---|
| TP-skeleton | 1202 | 10 606 | 96.4852 | 84.8729 | 1 |
| TPKS-skeleton | 347 | 4117 | 98.9854 | 94.1281 | 0.9726 |
| TPPR-skeleton | 1902 | 17 119 | 94.4384 | 75.5837 | 1 |
| GCI-skeleton | 644 | 7394 | 98.1169 | 89.4542 | 0.9568 |
| IC-skeleton | 12 | 29 | 99.9649 | 99.9586 | 1 |
| MAC-skeleton | 5 | 2 | 99.9854 | 99.9971 | 1 |

### 4.5. Network structural correlation

We further evaluate the structural-equivalence of the skeleton networks and the original network by using the Quadratic Assignment Procedure (QAP). QAP is a non-parametric, permutation-based test that generates a similar network's structure by taking the matrix for the observed network and reshuffling its rows and columns [57]. In literature, quantifying QAP is only applicable to the one-mode network. The bipartite projection is again required for our investigations. As this step is completed, we can obtain the one-mode projected network of each skeleton structure. Generally speaking, the correlation between two network structures (i.e., the original network and skeleton network) could be obtained over the one-mode projected networks. To ensure the accuracy of the permutation process, we set 1000 iterations for this experiment. In doing so, the QAP should correctly reflect the network's structural correlation between the original network and different skeleton networks.

Table 7 accounts for the QAP correlation and a proportion of node reduction and edge reduction. The detailed information also includes a number of preserved nodes and edges in the one-mode projected networks. The results show that the structure of different skeleton networks after the projection and the structure of the original network are highly correlated. The TP-skeleton, TPPR-skeleton, IC-skeleton, and MAC-skeleton control the original network structure entirely (QAP $= 1$). It implies that the sets of core nodes identified by TP, TPPR, IC, and MAC are significant, each of which is the right combination for the whole network structure. By focusing on the TP and TPPR, however, both of these two techniques lead the skeletonizing process to conserves a large fraction of terrorist nodes and edges.

On the contrary, IC-skeleton and MAC-skeleton reduce excessively network sizes. The projected networks contain a too small number of key terrorist nodes and edges, which might fail in preserving the network's flows, essential information, and network functions. Interestingly, TPKS-skeleton and GCI-skeleton compressed both size and complexity reasonably; not too large or too small of preserved nodes and edges. At the same time, their network structures maintain a comparatively high similarity to the original network (QAP $= 0.9726$, and $0.9568$). Therefore, TPKS-skeleton and GCI-skeleton outperform other skeleton networks in this aspect.

To explore the core node-set, we list the top 20 influential nodes of all skeleton networks in Table 8. Different node identification methods can achieve a different set of highly influential nodes indicating skeleton nodes. The skeleton nodes might refer to terrorist partners who share hidden relationships across the groups and are the main drivers for entire global terrorism. The IC-skeleton and MAC-skeleton preserves a tiny set of terrorist nodes. There are only 12 key terrorists and 5 key terrorists in IC-skeleton and MAC-skeleton, respectively.

In-depth analysis of the significance of core nodes retained in the skeleton network, we adopt Spearman'$\rho$ rank-order correlation [58] to verify the strength of association between the influential nodes from the original network and ones from the skeleton network. Spearman'$\rho$ is calculated based on the ranked value and does not carry any assumption about the distribution of the data.

The traditional correlation coefficient is defined as a paired observation. Thus, it is not possible to calculate $\rho$ of networks which have different sizes and contain unpaired nodes. To find the $\rho$ between two node sets (i.e., nodes in the extracted network and nodes in the original network), we design the correlation measure using two techniques. First, we directly remove unpaired observations from the original network before computing $\rho$. Generally speaking, nodes of

**Table 8**
The top 20 influential nodes of different skeleton networks.

| Rank | TP-skeleton | TPKS-skeleton | TPPR-skeleton | GCI-skeleton | IC-skeleton | MAC-skeleton |
|---|---|---|---|---|---|---|
| 1 | Taliban | Taliban | Taliban | ETA | Taliban | Taliban |
| 2 | ISIL | ISIL | ISIL | FMLN | ISIL | ISIL |
| 3 | NPA | SL | NPA | Armenian Secret Army for the Liberation of Armenia | FARC | NPA |
| 4 | CPI-Maoist | FARC | CPI-Maoist | NPA | PKK | SL |
| 5 | SL | Muslim extremists | SL | PKK | SL | CPI-Maoist |
| 6 | FARC | ELN | FARC | MIR (Chile) | NPA | – |
| 7 | Boko Haram | Al-Qaida in Iraq | Maoists | Anarchists | Armenian Secret Army for the Liberation of Armenia | – |
| 8 | Maoists | IRA | Boko Haram | FPMR | Palestinians | – |
| 9 | PKK | NPA | PKK | FARC | IRA | – |
| 10 | Al-Shabaab | ETA | Al-Shabaab | Muslim extremists | ETA | – |
| 11 | ELN | Boko Haram | ELN | SL | CPI-Maoist | – |
| 12 | ETA | Al-Shabaab | ETA | GRAPO | Muslim extremists | – |
| 13 | TTP | CPI-Maoist | TTP | Gunmen | – | – |
| 14 | LTTE | MRTA | LTTE | EPL | – | – |
| 15 | FMLN | FMLN | FMLN | Black September | – | – |
| 16 | IRA | PKK | IRA | CGSB | – | – |
| 17 | Fulani extremists | Maoists | Muslim extremists | Students | – | – |
| 18 | Muslim extremists | ISI | Palestinians | FLNC | – | – |
| 19 | Sikh extremists | M-19 | Sikh extremists | M-19 | – | – |
| 20 | Palestinians | Gunmen | Fulani extremists | Left-wing extremists | – | – |

**Table 9**
The Spearman correlation coefficient $\rho(\cdot)$ of the corresponding skeleton networks. $corr_{rand}$ represents the correlation value obtained by randomly selected nodes, $corr_{cut}$ is the correlation value by removing unpaired nodes in the original network before comparing with the skeleton network.

| | $\rho(TP)$ | $\rho(TPKS)$ | $\rho(TPPR)$ | $\rho(GCI)$ | $\rho(IC)$ | $\rho(MAC)$ |
|---|---|---|---|---|---|---|
| $corr_{rand}$ | 0.6665 | 0.7455 | 0.4857 | 0.0182 | 0.0275 | 0.0199 |
| $corr_{cut}$ | 0.9973 | 0.9996 | 0.9948 | 1.0000 | 0.9999 | 1.0000 |
| Average | 0.8319 | 0.8726 | 0.7402 | 0.5091 | 0.5137 | 0.5099 |

the original network that does not appear in the skeleton network are filtered out. The influential nodes contained in two networks are then considered into pairwise matching. In the second technique, we adopted an idea of samples at random and representative samples, as was suggested in [59]. We set a number of selected nodes in the original network as equal to the number of nodes in the skeleton network. After that, we run the randomization test 1000 times for each method. The mean value from a random set is eventually used to reflect the real node's correlation between the skeleton network and the original network. The $\rho$ value can be interpreted between $+1$ and $-1$. A perfect degree of relationship between the two ranking lists is indicated by 1. As the $\rho$ goes towards 0, the correlation is weaker, while the – sign indicates anti-relationship between two node-sets.

The information in Table 9 demonstrates that the ranking correlation of the skeleton nodes from all methods is highly agreed with the ranking in the original network, specifically in the case of $corr_{cut}$. Nodes in GCI-skeleton and MAC-skeleton held the perfect $\rho$ for $corr_{cut}$. However, their results in $corr_{rand}$ are conversely low. A similar situation could be seen for IC-skeleton. Meanwhile, the TP-based skeleton networks include TP-skeleton, TPKS-skeleton, and TPPR-skeleton perform high $\rho$ in both $corr_{cut}$ and $corr_{rand}$ and leads average $\rho$ also higher. It is worth to remind that a skeleton network that has higher $\rho$ means the ranking positions of all nodes in the skeleton network are more similar to their formally ranked position in the original network. It implies that the certain node identification techniques are efficient as the techniques can provide a set of skeleton nodes equivalent to the node important ranking of the original network.

## 5. Conclusion

Identifying the influential nodes in the complex network is theoretical and practical significance. In this work, we compare the performance of six node identification methods over the global terrorism network. Our research believes that not only a node itself is essential, but a group of nodes who works as a team and provides a great impact on the entire network structure is also significant. For this reason, a comparative study of different methods, focusing on identifying the significant nodes and a core node-set, can help network analyst choosing the most relevant technologies that can

**Table 10**
The summary of strength and weakness of the six node identification methods, which is extended to the context of skeleton network extraction.

| Method | Strength | Weakness |
|---|---|---|
| TP | Simplicity,<br>Relatively high node's correlation with the real spreading process,<br>High network structure similarity | Omits global properties,<br>Preserves large number of key terrorist and edges in the skeleton |
| TPKS | Relatively high node's correlation with the real spreading process,<br>Relatively high ability to distinguish the node's difference,<br>Preserves a small number of key terrorists and sufficient edges in the skeleton,<br>High node's correlation with the original network | Omits global properties,<br>Ignores periphery nodes |
| TPPR | Measure both local and global properties of the nodes,<br>High ability to distinguish the node's difference,<br>High network structure similarity | Nodes at the tail of the ranking deviate from the real spreading process,<br>Preserves large number of key terrorists and edges in the skeleton |
| GCI | Preserves a small number of key terrorists and sufficient edges in the skeleton | Deviate ranking for top nodes as compared to the real spreading process,<br>Deviate ranking for top nodes in the skeleton,<br>Low ability to distinguish the node's difference |
| IC | Relatively high ability to distinguish the node's difference,<br>High network structure similarity | Deviates ranking for top nodes as compared to the real spreading process,<br>Preserves too small number of key terrorists and edges in the skeleton |
| MAC | High ability to distinguish the node's difference,<br>High network structure similarity | Deviate ranking for top nodes as compared to the real spreading process,<br>Preserves too small number of key terrorists and edges in the skeleton |

improve the efficiency of the skeleton network extraction algorithm. Moreover, the study can allow counterterrorism to choose a suitable technique that highlights the critical nodes and possible incidents if and when the need arises. Table 10 summarizes strength and weakness of different methods regarding their ability to evaluating the influence value of the nodes and uncovering a set of nodes that significant for the core network structure.

From Table 10, the five main conclusions are drawn as follows:

(1) In terms of theoretical analysis, TPKS and TPPR with the distinctive mass parameter are more reasonable to evaluate the influence value of the nodes in the real-world network than the classical TP. From the experimental analysis, however, the TP method provides high accuracy in identifying the influential nodes of the global terrorism network, and the results are more consistent with the ground-truth from the real spreading process. Therefore, the simple measure by TP is useful for very large-scale networks that are complex or contain spurious connections among nodes.

(2) TPKS is good at the theoretical concept, and experimental performance since the method can provide a relatively high node's ranking correlations with the real node's spreading capability. TPKS highlight highly influential nodes on broad ranges of medium and large-sized networks, as well as networks with densely interconnected parts. Hence the method is suitable for extracting the best-connected core network structure, then can give a small size of the skeleton network with reasonable set of nodes and edges.

(3) As compared to other methods, TPPR reveals the lowest node's ranking correlation. On the other hand, the precision of TPPR could be seen by its ability to identify the highest influential nodes at the top of the ranking list. The skeleton network that is generated from TPPR nodes conserves strongly the most critical nodes that produce a perfect structural similarity to the original network. Thus, TPPR is suitable for extracting a skeleton network that focuses on the imitation model or patterning optimization, where the ranking correlation of overall nodes does not have a significant role.

(4) The GCI method does not fit the study of node identification concerning the nodes that must be strictly classified into different ranks. The size of the GCI-skeleton network is also small with a sufficient number of nodes and edges. However, the extracted network from GCI generates node ranking, which diverges from the original networks and other skeleton networks, as could be observed from Table 8. GCI mostly relies on the k-shell decomposition that focuses on the globally connected part of the network. Therefore, this method does not appropriate when apply for extracting the skeleton over a fully connected network [60].

(5) The IC and MAC fail to achieve high performance in the node's ranking correlation and may be less significant for evaluating the influence ability of the individual nodes in this global terrorism network. The methods provide a deviated set of nodes, most of which differ from SIR and other methods. With very few terrorist nodes and edges in IC-skeleton and MAC-skeleton, the methods are not suitable for observing terrorist-to-terrorist relationships. Nevertheless, both of them function well in skeleton network extraction as far as retaining the network structural-equivalence is concerned.

We can notice that a small number of key nodes encircle with the important edges can give a precise solution to understand the real characteristics of the global terrorism network. Six comparative methods capture different aspects

**Table 11**
The potential applications of different node identification methods concerning the objectives of skeleton network extraction.

| | Network extraction objective | Suggested method |
|---|---|---|
| Objective 1: | Construct a skeleton network with high network structural correlation or the network that imitate greatly the complex behavior and pattern of the original network. | TP, TPPR |
| Objective 2: | Construct a skeleton network by considering cost issues and accelerating other network analysis algorithms; the network that is reduced in size and complexity through preserving a small number of nodes and edges. | TPKS, GCI |
| Objective 3: | Construct a skeleton network that useful for observing some previously unknown property or estimating hidden behavior among nodes e.g., cooperation between terrorist organizations. | TP, TPKS, TPPR |
| Objective 4: | Construct a skeleton network that benefits studying community structure, the network that provides ground-truth communities. | TP, TPPR, GCI |
| Objective 5: | Construct a skeleton that capturing at the most affected areas or vulnerable entities attacked by the deadliest terrorist groups. | IC, MAC |

of network topology, and thus discover a different set of highly influential nodes, accordingly, different skeleton network structures.

As far as concerning the objectives of skeleton network extraction presented in Table 11, the skeleton networks constructed by TP and TPPR work effectively in preserving network structural-equivalence. Both of them are suggested as the best methods to imitate the complex behavior with acceptable network sizes. The TPKS and GCI capture significant nodes and edges, which make a skeleton network small enough and less complex. All TP-based methods explicitly reveal good node ranking correlation and maintain sufficiently terrorist nodes in the extracted network. These set of nodes can be further explored to find some unknown relationships among terrorists. The skeleton network of TP, TPPR, and GCI contains community structure. These methods thus benefit analyzing network with densely connected subgroups and useful for explaining similarity among community members. The IC and MAC neglect a large number of highly influential terrorists. Instead, the methods generate skeleton networks which mostly contain target nodes around key terrorist in star-like and disjointed communities. Therefore, IC and MAC are of great use to identify all vulnerable entities that were attacked by the deadliest terrorist organizations.

In this paper, we have focused on the comparison across different influential node identification methods for the sake of skeleton network extraction. However, there are some limitations to our study. Our study does not address the computational complexity of the comparative methods since they are implemented in different software environments. Moreover, the validation of a set of core nodes has relied on only one simple skeleton network extraction algorithm. There have plenty of factors that might affect the extraction results and might be required for future research. To this end, we anticipate that our experiments could explore the minimal requirement to explain the significance of influential nodes to the core network structure. Last of all, our findings could be useful for further research on skeleton network extraction.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## CRediT authorship contribution statement

**Kanokwan Malang:** Conceptualization, Methodology, Investigation, Writing - original draft, Writing - review & editing, Visualization. **Shuliang Wang:** Supervision. **Aniwat Phaphuangwittayakul:** Software, Validation, Data curation. **Yuanyuan Lv:** Software. **Hanning Yuan:** Writing - review & editing. **Xiuzhen Zhang:** Writing - review & editing.

## Acknowledgments

The authors would like to thank anonymous reviewers who made valuable suggestions to improve the quality of this research. This work is supported by the National Key Research and Development Program of China [grant number 2016YFC080300], The National Natural Science Fund of China [grant number 61472039], and Beijing Significant Fund of Science and Technology [grant number Z171100005117002].

## References


[1] J. Coaffee, Terrorism, Risk and the Global City: Towards Urban Resilience, Routledge, 2016.
[2] S.S. Husain, K. Sharma, V. Kukreti, A. Chakraborti, Identifying the global terror hubs and vulnerable motifs using complex network dynamics, Physics (2018) arXiv:1802.01147.
[3] S.T. Zech, M. Gabbay, Social network analysis in the study of terrorism and insurgency: From organization to politics, Int. Stud. Rev. 18 (2) (2016) 214–243.
[4] E.D. Mainas, The analysis of criminal and terrorist organisations as social network structures: a quasi-experimental study, Int. J. Police Sci. Manag. 14 (3) (2012) 264–282.
[5] P.A.C. Duijn, V. Kashirin, P.M.A. Sloot, The relative ineffectiveness of criminal network disruption, Sci. Rep. 4 (1) (2015).
[6] P. Choudhary, U. Singh, A Survey on Social Network Analysis for Counter- Terrorism, 2015.
[7] K.M. Carley, J.-S. Lee, D. Krackhardt, Destabilizing networks, Connections 24 (3) (2001) 31–34.
[8] F. Xu, D. Sun, Z. Li, B. Li, Exploring structural features of terrorist organization's online supporting community via social network modeling, in: 2017 3rd IEEE International Conference on Computer and Communications (ICCC), Chengdu, 2017, pp. 274–278.
[9] D. Xuan, H. Yu, J. Wang, A Novel Method of Centrality in Terrorist Network, in: 2014 Seventh International Symposium on Computational Intelligence and Design, Hangzhou, China, 2014, pp. 144–149.
[10] V. Latora, M. Marchiori, How the science of complex networks can help developing strategies against terrorism, Chaos Solitons Fractals 20 (1) (2004) 69–75.
[11] J. Alison, L. Deng, CS224W final project report: Uncovering the global terrorism network, 2017, p. 10.
[12] M. Lauchs, R.L. Keast, Social network analysis of terrorist networks: can it add value?, p. 17.
[13] K. Sharma, et al., A complex network analysis of ethnic conflicts and human rights violations, Sci. Rep. 7 (1) (2017) 8283.
[14] P.V. Fellman, R. Wright, Modeling Terrorist Networks - Complex Systems at the Mid-Rangei, p. 14.
[15] Y.-H. Fu, C.-Y. Huang, C.-T. Sun, Using global diversity and local features to identify influential social network spreaders, 2014, pp. 948–953.
[16] Q. Ma, J. Ma, Identifying and ranking influential spreaders in complex networks with consideration of spreading probability, Phys. Stat. Mech. Appl. 465 (2017) 312–330.
[17] J. Bae, S. Kim, Identifying and ranking influential spreaders in complex networks by neighborhood coreness, Phys. Stat. Mech. Appl. 395 (2014) 549–559.
[18] Jianxin Wang, Min Li, Huan Wang, Yi Pan, Identification of essential proteins based on edge clustering coefficient, IEEE/ACM Trans. Comput. Biol. Bioinform. 9 (4) (2012) 1070–1080.
[19] Z. Sun, B. Wang, J. Sheng, Y. Hu, Y. Wang, J. Shao, Identifying influential nodes in complex networks based on weighted formal concept analysis, IEEE Access 5 (2017) 3777–3789.
[20] T. Kumari, A. Gupta, A. Dixit, Comparative study of page rank and weighted page rank algorithm, Int. J. Innov. Res. Comput. Commun. Eng. 2 (2) (2014) 2929–2937.
[21] P. Choudhary, U. Singh, Ranking terrorist nodes of 26/11 mumbai attack using analytical hierarchy process with social network analysis, in: 11th Annu. Symp. Inf. Assur., 2016, p. 6.
[22] Y. Han, D. Li, T. Wang, Identifying different community members in complex networks based on topology potential, Front. Comput. Sci. China 5 (1) (2011) 87–99.
[23] S. Wang, W. Gan, D. Li, D. Li, Data field for hierarchical clustering:, Int. J. Data Warehous. Min. 7 (4) (2011) 43–63.
[24] T. Wang, Y. Han, J. Wu, Evaluate nodes importance in directed network using topological potential, in: Information Engineering and Computer Science (ICIECS), 2010 2nd International Conference on, 2010, pp. 1–4.
[25] D. Zhang, L. Gao, Virtual network mapping through locality-aware topological potential and influence node ranking, Chin. J. Electron. 23 (1) (2014) 61–64.
[26] D. Li, S. Wang, L. Deyi, Spatial Data Mining Theory and Application, 2015.
[27] Y. Wang, J. Yang, J. Zhang, J. Zhang, H. Song, Z. Li, A Method of Social Network Node Preference Evaluation Based on the Topology Potential, 2015, pp. 223–230.
[28] R. Sun, W. Luo, Using topological potential method to evaluate node importance in public opinion, in: Presented at the 2017 International Conference on Electronic Industry and Automation (EIA 2017), 2017.
[29] X. Lei, Y. Zhang, S. Cheng, F.-X. Wu, W. Pedrycz, Topology potential based seed-growth method to identify protein complexes on dynamic PPI data, Inform. Sci. 425 (2018) 140–153.
[30] M. Li, Y. Lu, J. Wang, F.-X. Wu, Y. Pan, A topology potential-based method for identifying essential proteins from PPI networks, IEEE/ACM Trans. Comput. Biol. Bioinform. 12 (2) (2015) 372–383.
[31] Q. Han, H. Wen, M. Ren, B. Wu, S. Li, A topological potential weighted community-based recommendation trust model for P2P networks, Peer–Peer Netw. Appl. 8 (6) (2015) 1048–1058.
[32] K. Malang, S. Wang, T. Dai, Analyzing community structure based on topology potential over complex network system, in: International Conference on Geo-Spatial Knowledge and Intelligence, 2017, pp. 56–68.
[33] Z. Wang, Z. Chen, Y. Zhao, S. Chen, A community detection algorithm based on topology potential and spectral clustering, Sci. World J. 2014 (2014) 1–9.
[34] X. Ding, Z. Wang, S. Chen, Y. Huang, Community-based collaborative filtering recommendation algorithm, Int. J. Hybrid Inf. Technol. 8 (2) (2015) 149–158.
[35] Z. Wang, Z. Li, G. Yuan, Y. Sun, X. Rui, X. Xiang, Tracking the evolution of overlapping communities in dynamic social networks, Knowl.-Based Syst. 157 (2018) 81–97.
[36] Q. Han, et al., A P2P recommended trust nodes selection algorithm based on topological potential, in: Communications and Network Security (CNS), 2013 IEEE Conference on, 2013, pp. 395–396.
[37] H. Yuan, K. Malang, Y. Lv, A. Phaphuangwittayakul, From complex network to skeleton: mj-Modified topology potential for node importance identification, in: International Conference on Advanced Data Mining and Applications, 2018, pp. 413–427.
[38] M. Kitsak, et al., Identification of influential spreaders in complex networks, Nat. Phys. 6 (11) (2010) 888–893.
[39] Z. Wang, Y. Zhao, Z. Chen, Q. Niu, An improved topology-potential-based community detection algorithm for complex network, Sci. World J. 2014 (2014) 1–7.
[40] S. Brin, L. Page, The anatomy of a large-scale hypertextual Web search engine, Comput. Netw. ISDN Syst. 30 (1) (1998) 107–117.
[41] W. Zhi-Xiao, L. Ze-chao, D. Xiao-fang, T. Jin-hui, Overlapping community detection based on node location analysis, Knowl.-Based Syst. 105 (2016) 225–235.
[42] L.-L. Ma, C. Ma, H.-F. Zhang, B.-H. Wang, Identifying influential spreaders in complex networks based on gravity formula, Phys. Stat. Mech. Appl. 451 (2016) 205–212.
[43] Z. Wang, C. Du, J. Fan, Y. Xing, Ranking influential nodes in social networks based on node position and neighborhood, Neurocomputing 260 (2017) 466–477.

[44] D. Liu, H. Nie, B. Zhang, A novel method for identifying influential nodes in complex networks based on multiple attributes, Internat. J. Modern Phys. B 32 (28) (2018) 1850307.
[45] Start.umd.edu |, 2018, [Online]. Available: https://www.start.umd.edu/ [Accessed: 21-Dec-2018].
[46] G. Codebook, Codebook: Inclusion criteria and variables, 2012.
[47] T. Opsahl, Triadic closure in two-mode networks: Redefining the global and local clustering coefficients, Soc. Netw. 35 (2) (2013) 159–167.
[48] R. Medina, G. Hepner, Geospatial analysis of dynamic terrorist networks, in: I.A. Karawan, W. McCormack, S.E. Reynolds (Eds.), Values and Violence, Vol. 4, Springer Netherlands, Dordrecht, 2009, pp. 151–167.
[49] X. Zhao, F. Liu, J. Wang, T. Li, Evaluating influential nodes in social networks by local centrality with a coefficient, ISPRS Int. J. Geo-Inf. 6 (2) (2017) 35.
[50] G. Lawyer, Understanding the influence of all nodes in a network, Sci. Rep. 5 (1) (2015).
[51] M.G. Kendall, The treatment of ties in ranking problems, Biometrika 33 (3) (1945) 239–251.
[52] W.R. Knight, A computer method for calculating Kendall's tau with ungrouped data, J. Amer. Statist. Assoc. 61 (314) (1966) 436–439.
[53] Institute for economics & peace, Global Terrorism Index 2017, 2017, Institute for economics & peace.
[54] Institute for economics & peace, Global Terrorism Index 2018-Measuring the impact of terrorism, 2018.
[55] D. Grady, C. Thiemann, D. Brockmann, Robust classification of salient links in complex networks, Nature Commun. 3 (1) (2012).
[56] H. Yuan, Y. Han, N. Cai, W. An, A multi-granularity backbone network extraction method based on the topology potential, Complexity 2018 (2018) 1–8.
[57] D. Krackhardt, QAP partialling as a test of spuriousness, Soc. Netw. 9 (1987) 171–186.
[58] James Lani, Moran Glase, Ed Isaidys, Rachel Justine, David Tom, Correlation (Pearson, Kendall, Spearman) - Statistics Solutions, 2019, [Online]. Available: https://www.statisticssolutions.com/correlation-pearson-kendall-spearman/, [Accessed: 21-Sep-2019].
[59] Timothy A. Ebert, How to do correlation analysis with two variable in different sample size, ResearchGate (2016) [Online]. Available: https://www.researchgate.net/post/How_to_do_correlation_analysis_with_two_variables_in_different_sample_size, [Accessed: 21-Sep-2019].
[60] L. Dai, B. Derudder, X. Liu, Transport network backbone extraction: A comparison of techniques, J. Transp. Geogr. 69 (2018) 271–281.