

Research Article

Extracting Skeleton of the Global Terrorism Network Based on m-Modified Topology Potential

Shuliang Wang ^{1,2}, Kanokwan Malang ¹, Hanning Yuan ¹,
Aniwat Phaphuangwittayakul,³ Yuanyuan Lv,¹ Matthew David Lowdermilk,¹
and Jing Geng ^{1,2}

¹School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China

²Institute of E-Government, Beijing Institute of Technology, Beijing 100081, China

³International College of Digital Innovation, Chiang Mai University, Chiang Mai 50200, Thailand

Correspondence should be addressed to Kanokwan Malang; kanokwan.malang@yahoo.com, Hanning Yuan; yhn6@bit.edu.cn, and Jing Geng; janegeng@bit.edu.cn

Received 14 February 2020; Accepted 27 April 2020; Published 16 June 2020

Academic Editor: Qingling Wang

Copyright © 2020 Shuliang Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Skeleton network extraction is a crucial context in studying the core structure and essential information on complex networks. The objective of this paper is to introduce the novel network extraction method, namely, TPKS-skeleton, for investigating the global terrorism network. Our method aims to reduce the network's size while preserving key topology and spatial features. A TPKS-skeleton comprises three steps: node evaluation, similarity-based clustering, and skeleton network reconstruction. The importance of skeleton nodes is quantified by the improved topology potential algorithm. Similarity-based clustering is then integrated to allow detecting high incident concentrations and allocating the important nodes according to the event features and spatial distribution. Finally, the skeleton network can be reconstructed by aggregating high-influential nodes from each cluster and their simplified edges. To verify the efficiency of the proposed method, we carry out three classes of a network assessment framework: node-equivalence assessment, network-equivalence assessment, and spatial information assessment. For each class, various assessment indexes were performed using the original network as a benchmark. The results verify that our proposed TPKS-skeleton outperforms other competitive methods in particular node-equivalence by Spearman rank correlation and high network structural-equivalence defined by quadratic assignment procedure. In the spatial perspective, the TPKS-skeleton network preserves reasonably all kinds of spatial information. Our study paves the way to extract the optimal skeleton of the global terrorism network, which might be beneficial for counterterrorism and network analysis in wider areas.

1. Introduction

For over the past decade, complex network analysis has become a vital component to understand behavioral patterns of the global terrorism network. Fertile knowledge discovered from investigating various complex network properties helps to reduce the risk or effect of any attack that would happen in the future [1, 2]. More than that, it can deliberately carry out feasible disruption plans or demolition of terrorist organizations [3, 4].

In recent years, mining and analyzing the entire global terrorism network is, however, a nontrivial task. There remain various issues to complex network analysis. First, an

excessive amount of data has far exceeded the capability of most available technologies to interpret and analyze the large-scale network appropriately. The statistical evidence from START [5] shows that nearly 182,000 terrorism incidents occurred globally from 1970 to 2017. Since the world experienced on September 11th, the total number of terrorism incidents have rapidly increased and reached its peak in 2014. Despite the fall in deaths for the later three consecutive years, the terrorist incidents in Europe, nevertheless, increased [6]. Second, terrorism is growing not only in the network's size but also in the complexity of the social relationships involved. This fact could be confirmed by the suicide bombing attacks of Sri Lanka on Easter Sunday, 2019

[7]. The information implies that the collaboration among terrorist groups is not a straightforward matter. Local terrorists, alone, could have attempted but may not succeed in executing the attack on their own unless they have collaborators. Third, it is undeniable that the large-scale network always comes along with noisy information. The terrorism data are collected from multiple sources containing a number of unimportant elements and are also inconsistent. Directly modeling a network may cause misleading relationships and makes the network far from its actual structure. The extent of these problems affects network analysis and the discovery process to be computationally very expensive and constantly faced with high difficulties.

In this paper, the advantages of analyzing the global terrorism network are under the discrimination of the network's core structure. Generally speaking, instead of investigating the significant information from the entire network, we pay attention to the skeleton network [8]. The skeleton network relies on the concept of minimizing the network's scale and complexity [9]. It is the simplified form of any original network [10] that would allow network analysis methods to perform faster and more efficient [11]. The skeleton network helps to easily visualize, identify, and understand relevant information over the network's natural abstraction. At the same time, it still preserves the most useful insights and the major functions of the original network. Therefore, research on skeleton network extraction may help to discover the ground-truth relationship of the global terrorism network much easier.

The main objectives of this paper are as follows: first, we propose an idea of extracting the skeleton network using the influential nodes identified by an improved topology potential algorithm (TPKS-skeleton). The extraction strategy is conducted under the assumption that a skeleton network could retain its functions and behavior patterns with a small portion of the important nodes. Second, we apply the TPKS-skeleton method to obtain intrinsic insight of the global terrorism network. The influence value of all nodes in the network is evaluated to identify the optimal skeleton nodes with high significance. The integrated process of the clustering algorithm over the similarity-based network is applied to detect the group of nodes regarding their region in space and characteristics of terrorism events. Consequently, the skeleton network is reconstructed by a set of highly influential nodes from each cluster and the simplified edges. Through evaluation with the quadratic assignment procedure (QAP) and Voronoi region, our skeleton network extraction method is efficient in preserving the topological structure and spatial information as consistent with the original network. Results from our study can also provide valuable contributions to the field of social network analysis and related areas.

The remainder of this paper is organized as follows: in Section 2, we describe the current situation on skeleton network extraction and its lack of contribution to the terrorism domain. In Section 3, the concept of TPKS-skeleton for network extraction is presented. In Section 4, we briefly describe the details of the global terrorism dataset, the construction process of the original network,

and the network assessment framework. Furthermore, we discuss the deployment of TPKS-skeleton for extracting the core network structure and its results in Section 5. Finally, Section 6 gives conclusions to this paper.

2. Current Situation on Skeleton Network Extraction

The literature on skeleton network extraction takes a number of different approaches. Most of them have introduced techniques to deal with the edge reduction problem, for instance, edge weighting [12], edge sampling [13], statistical importance of links [14], and unexpected noise connections [15]. These methods try to remove edges as many as possible but retain countless of vertices. We feel that only a small number of important nodes and their connections might be satisfactory for network functionalities. The basic thought of our approach is that a set of the most influential nodes can sufficiently describe the ground-truth information of the global terrorism network.

Another argument shows that recent network extraction methods have a successful application in the social network domain. They are developed and applied to different networks, ranging from coauthorship networks [16], online social networks (e.g., Youtube, Live Journal, and Goodreads blogging community) [11], and communication networks [17, 18]. Unfortunately, it is rare to find empirical investigations of the global terrorism network, notably via the shape of a skeleton. In network science, recent studies focus on resilience inside the terrorism network as a consequence of network disruption strategies [19, 20]. Rostami and Mondani [21] work on comparing different criminal networks to illustrate the complexity, similarity, and reliability of measurement in view of social network analysis (SNA). Research from [22] highlights the evolution of hub nodes via the backbone network structure. Their work is the only one that closely related to our research. The existing studies give evidence that skeletonizing discipline has not been confirmed sufficiently to its performance, especially in extracting the hidden behaviors and meaningful insights into the global terrorism network.

In spatial context, efficient network extraction methods are not only used to reduce network size but are also supposed to preserve the key topology and spatial features [23]. It is worth to remind that impact of terrorism remains widespread over the targeted geographical area. The global terrorism network contains attributes associated with geolocation where the incidents occurred. Although some of the existing methods can be applied to spatial networks, they have not taken into account the spaces occupied by vertices and spatial distribution. Existing methods are purely statistical, which has negligible implications in a spatial notion.

Besides, the efficiency of empirical methods to retain spatial features has not yet been clarified. It is still unclear which methods are dominant in the network that contains spatial attributes. Disregarding spatial features may lead to frustration in understanding terrorism behavior. The usage of existing applications for spatial networks might be doubtful. For this reason, it requires specific skeleton

network extraction algorithms that are sensitive to the network topology and geography. This situation calls for a novel approach that could potentially be used to improve a comprehensive understanding of the global terrorism network and other network mining processes. The question arises whether the skeleton network extraction based on topology potential can sufficiently preserve both topological information and spatial information of the original network.

3. Method

The proposed network extraction method consists of three consecutive processes: (1) evaluation of node influence value by an improved topology potential algorithm, (2) identification of important nodes associated with the geographical area and event features, and finally, (3) reconstruction of the skeleton network.

3.1. Influential Nodes by an Improved Topology Potential. Topology potential (TP) is brought as a key technique to measure the influence of the nodes over recent years. The efficiency of TP can be proved by evidence that covers the analysis methods of social networks [24, 25], medical science [26, 27], and information search area [28, 29]. TP was introduced in light of the data field [30, 31]. It is based on the concept that “the importance of nodes which are adjacent to the important nodes” [32]. Such that, TP is an indicator to reflect the node’s ability, which is influenced by the differential position of each node and its neighbors [33].

In this research, the subject of a global terrorism network is actors (terrorists and targets) that can be represented by nodes. Edges illustrate the interactive relationships between different types of nodes. Given a directed complex network $G = (V, E)$, G consists of the set of nodes V and the set of edges E . $V = \{v_1, v_2, \dots, v_n\}$ represents each unique node, while $E = \{e_1, e_2, \dots, e_m\}$ denotes the interactive relation from the node v_i to the node v_j , i.e., attackers’ action. TP is defined in the form of Gaussian function [34] and can be formalized as

$$\varphi(v_i) = \sum_{j \in N} \left(m_j * \exp^{-\left(d_{i \rightarrow j} / \sigma\right)^2} \right), \quad (1)$$

where $\varphi(v_i)$ is the TP of the node v_i , for $v_i \in V$. $d_{i \rightarrow j}$ expresses the shortest distance from the node v_i to the node v_j . The parameter σ is a factor used to control the influence region of each node. m_j represents the mass of the node, which is redefined by k -shell centrality, and $m_j = KS_v$ [8]. Accordingly, the influence value of the nodes is induced by a different value of m_j in an improved TP algorithm, namely, TPKS.

TPKS describes the influence of the nodes itself and its neighbors [35]. The source node which is terrorist can pass its impact to the target through the reachable path in network topology structure [36]. As previously tested and confirmed by [37], TPKS provides a great potential application in the context of skeleton network extraction. The reason is straightforward: TPKS is more likely to create a skeleton network from a small set of highly influential nodes.

This set of nodes is sufficient and significant for the network’s key functions. It is showing that the extracted network from TPKS is much reduced in size and complexity as concerning the objective of this paper. Therefore, the resulting network would solve some cost issues in the large-scale network and allows other network analytical algorithms to perform precisely.

3.2. Clustering over Similarity Network. The selection of skeleton nodes can be manipulated by the specific clustering algorithm, which relied on both geolocation and event features. To our knowledge, clustering is a fundamental qualitative concept that is described by assigning all observations into different clusters [38]. A resulting cluster is a group of densely connected nodes that are more similar than those in different clusters. Thus, the distribution of the nodes in the global terrorism network and their feature similarities can be used to determine the influences of the core node to other nodes. The significance of clustering in this study is to detect high incident concentrations and allocate the high influential nodes of each cluster accordingly. It is also mainly done to break up the dataset into subgroups and correctly select essential nodes as the core membership for a skeleton network.

To complete this task, we employ the similarity network model [39] to quantify any resemblance between terrorist organizations from the given terrorists-to-targets network. The algorithm creates a summary vector for each terrorist based on the associated targets. This algorithm requires extremely high computational demands and exceeds the capability of our available resources. To overcome this issue, we include only the terrorist events that started in 2001, the year of the tragic incident on September 11th [40]. After the summary vector is obtained, the algorithm continually computes the similarity scores according to the TPKS value and the following event features: (1) Geolocation: the geographical attribute could influence the accessibility of spatial nodes in the global terrorism network. It is, therefore, realistic to define the dissimilarity between nodes by their location in the physical space. For each terrorist, we calculate the mean value of geolocation coordinates (latitude and longitude) of all its targeted attacks. It is for estimating the active geographic area of the terrorism events caused by a certain terrorist organization. (2) Peak year: this is a year where the terrorist performs the most events. The peak year is to show how active the terrorist is associated with the time dimension. (3) Lethality: the summation of deaths and the number wounded from all attacks by a terrorist. (4) Attack type: the general method or tactics that the terrorist used the most in all of its events. (5) Weapon type: the type of weapon that the terrorist used the most in all of its events. (6) Target type: the victim types that terrorists targeted the most in all of its events.

Algorithm 1 explains the detailed description for constructing a similarity network based on geolocation and event features. In line with this, a network element is defined as a small subgraph containing a terrorist node with many attacked targets. Once the similarity score between any pair

of subgraphs reaches the threshold value, the algorithm connects two subgraphs by adding a tie between terrorist nodes. Subsequently, we apply a k-means clustering algorithm to partition similarity network into different k clusters [41]. To obtain a value of k , we use the `factoextra::fviz_nbclust()` integrated with Elbow analysis in the *R* platform. Consequently, $k = 4$ is suggested as the optimum number of clusters for the similarity network. Each node that belongs to the same cluster is detected from the nearest means by Euclidean distance and has the greatest possible distinction to other nodes in different clusters.

3.3. Reconstruction of Skeleton Network. The practical procedure of skeleton network reconstruction includes four steps. In the first step, the initial set of terrorist nodes from the similarity network is identified. For each cluster, the algorithm selects high-influential terrorists with a number of top nodes that satisfied α . From this step, we could obtain the terrorist network that covers some disconnected components. This is because the top terrorist nodes from different clusters are dissimilar, and they might have originally no connection with each other. In order to maintain the network’s key structure, we focus only on the connected network $netE$, which has the highest degree of connectivity.

In the second step, the relative targets are discovered with the identical parameter α . This means all the target nodes, which have at least one connection to any of the terrorist nodes from the first step, are ranked in decreasing order. Then, the algorithm picks only the top target nodes according to their TPKS value.

In the third step, the algorithm merges a set of top terrorists $netE$ and the top targets $netU$. Both terrorist and target nodes and their formerly connected edges are aggregated to form the whole skeleton network.

Finally, all edges between two nodes are mapped into a single one, and the summary statistics or attributes of the edges are combined. Such that the node set V' and edge set E' of the skeleton G' contain only the most important V and E , and the size of the new network is much smaller than the original network. The summary description of the TPKS-skeleton is given in Algorithm 2.

4. Experimental Settings

In this section, we introduce the network data and the network assessment framework used in this study. The TPKS algorithm is developed using Python programming language. The skeletonizing experiment and comparison were mainly conducted on an open package available in the *R* platform. Rstudio has been used to generate data visualization.

4.1. Data Source Description. We use a dataset from the Global Terrorism Database (GTD) to construct the network. The GTD is retrieved and maintained by the National Consortium for the Study of Terrorism and Responses to Terrorism (START) [5]. The terrorism incidents are recorded from several publicity sources that include media

articles, electronic news, books, journal, and legal documents. The observations span from 1970 to 2017. However, our study excludes the incidents from 1993, as there was only 15% of attacks could be reliably identified [42]. Moreover, geolocations of 1993’s incidents that could benefit our study were not available. The entire GTD dataset contains nearly 182,000 observations with 135 attributes, e.g., perpetrator information, victim information, geolocation, number of casualties and consequences, and so forth.

The GTD is a large-scale network dataset that is continually increasing with time. A cleaning process is necessary for preparation before the study. The incidents in the GTD are often ambiguous between terrorism and other forms of crime or political violence. In order to make a clear determination regarding this issue, we filtered out all incidents that the GTD suggested to be unclear. In cases that incidents were carried out by “Unknown Terrorist” and “Unknown Target”, they were removed from the dataset. Since our study relies on spatial information, any event with unavailable geolocation was filtered out. Our work does not include the attacks that were attempted but were ultimately unsuccessful. Moreover, we eliminate all incidents that were targeted in the multinational area or occurred in the international area. It is to maintain network modularity to a particular nation. The cleaning process results with 70,113 observations remaining in the global terrorism dataset.

4.2. Original Network Construction. The global terrorism network contains all actors and their relationships involved in each terrorism incident. Over the entire network, two types of actors (terrorist nodes and target nodes) are connected by directed links. A relationship is defined whenever a terrorist attacks a target in the event E_1 at the timeframe t . A directed edge then emits from the terrorist node a_1 to the target node a_2 , where t_0 is the initial timeframe. Each target node is associated with the geolocation L where the incident occurred. The severity of the attack obtained by the number of deaths and wounded can be used to define the weight of an edge. According to this structure, the generated network is a directed, weighted, and bipartite network (DNWB) whose relation’s structure is presented across a particular set of nodes. Thus, the connection between nodes within the same group (i.e., terrorist-to-terrorist, target-to-target) is not allowed.

4.3. Assessment Framework for Skeleton Network Extraction. To verify the accuracy and performance of the network extraction method, we proposed the network assessment framework, which is comprised of three classes: node-equivalence assessment, structural-equivalence assessment, and spatial information assessment. Several measures for the extracted skeleton network will be computed and compared with the original network.

Node-equivalence assessment is to find the similarity of all preserved nodes in the skeleton network that cover the important nodes in the original network. It is to discover the significant correlation in the degree of dominance from the individual actors. The evaluation indexes which are

Similarity Network based on Geolocation and Event Features
Input Original Network G , Terrorist Node T
Output $simG$

- (1) **m-Modified Topology Potential function (G)**
- (2) $TPKS \leftarrow 0$
- (3) **for** each node v_i **to** $v_n \in G$ **do**
- (4) calculate $TPKS$ of v_i within i hops
- (5) **end for**
- (6)
- (7) **similarityScore function (v_1, v_2)**
- (8) $score \leftarrow 0$
- (9) **for** $i = 1$: *Vector Dimension* **do**
- (10) **if** i^{th} feature is unbounded numeric value **then**
- (11) $curr = \min(v_1[i], v_2[i]) / \max(v_1[i], v_2[i])$
- (12) **if** i^{th} feature is bounded numeric value **then**
- (13) $curr = \exp(-\text{abs}(v_1[i] - v_2[i]))$
- (14) **if** i^{th} feature is categorical **then**
- (15) $curr = 1 [v_1[i] == v_2[i]]$
- (16) $curr = 0 [v_1[i] \neq v_2[i]]$
- (17) **if** i^{th} feature is geolocation **then**
- (18) $distance = \text{geographic distance by latitude \& longitude}$
- (19) $curr = \exp(-distance)$;
- (20) $score += curr$;
- (21) **return** $score$
- (22)
- (23) **similarityNetwork function (T, threshold)**
- (24) **for** $t_1, t_2 \in T$ **and** $t_1 \neq t_2$ **do**
- (25) $t1_subg \leftarrow \text{subGraph}(t_1)$
- (26) $t2_subg \leftarrow \text{subGraph}(t_2)$
- (27) $v_1 \leftarrow \text{getGroupVector}(t1_subg)$
- (28) $v_2 \leftarrow \text{getGroupVector}(t2_subg)$
- (29) $similarity \leftarrow \text{similarityScore}(v_1, v_2)$
- (30) **if** $similarity \geq \text{threshold}$ **then**
- (31) add tie between t_1 and t_2
- (32) **return** $simG$

ALGORITHM 1: Pseudocode explaining the similarity network of geolocation and event features (source, Alison and Deng, 2017 [39]).

Skeleton Network Extraction Model (TPKS-skeleton)
Input Original Network G , Similarity Network $simG$, nodes in different clusters $V_{clusters}$
Output Skeleton Network G'

- (1) **TPKS-skeleton ($G, \alpha, simG, V_{clusters}$)**
- (2) $V_{terrorist_\alpha} \leftarrow \text{getTopAlpha}(\text{sort}(TPKS(v_e))); v_e \in V_{clusters}$
- (3) $simG_\alpha \leftarrow \text{subGraph}(V_{terrorist_\alpha}); V_{terrorist_\alpha} \in \text{getV}(SimG)$
- (4) $netE = \max(\text{degree}(simG_\alpha))$
- (5) $V_{terrorist_{netE}} \leftarrow \text{getV}(netE)$
- (6) $V_{target_{netE}} \leftarrow \text{getV}(\text{subGraph}(\text{getV}(G) \neq V_{terrorist_{netE}}))$
- (7) $V_{target_\alpha} \leftarrow \text{getTopAlpha}(\text{sort}(TPKS(va))); va \in V_{target_{netE}}$
- (8) $netU \leftarrow \text{subGraph}(V_{target_\alpha})$
- (9) $G' = \text{merge}(netE, netU)$
- (10)
- (11) **for** each pair of nodes $v_i, v_j \in G'$
- (12) **if** no. of edge between $(v_i, v_j) > 1, \neq 0$ **then**
- (13) map all edges into a single tie
- (14) **else**
- (15) remain a tie between (v_i, v_j)
- (16) **end if**
- (17) until the network is connected
- (18) **end for**
- (19) **return** Skeleton Network G'

ALGORITHM 2: Pseudocode explaining the TPKS-skeleton network extraction.

integrated into this class are the measurement of Spearman's rank-order correlation [43], and the feature's correlation matrix [44].

Network-equivalence assessment is to find a significant equivalence of two networks in the overall topological structure. The validation strategies in this class are the basic structural features of the network and the network's correlation by QAP [45].

Spatial information assessment is to find how the skeleton network and the original network exhibit relationship among spatial nodes. In this class, different kinds of spatial information on the Voronoi diagram [46] are explored and described whether the information pattern of the two networks is consistent.

5. Results and Discussion

5.1. Reduction Effectiveness of TPKS-Skeleton. On extracting the skeleton network, the higher control parameter α retained many significant nodes that have high TPKS value, as well as noise nodes. Lower α , on the other hand, filters a larger number of nodes that were the key to the network information, leading to a risk of network destruction. Setting the control parameter too low or too high may result in extraction inefficiency. Therefore, it is necessary to examine the suitable extraction level beforehand. For this purpose, we refer to the empirical research from [47]. Their proposed metric of reduction effectiveness guides us to specify the extraction capacity, which is appropriate for a certain network. The reduction effectiveness could be found by quantifying the size of isolated subnetworks. Owing to the fact that the skeleton network contains the most significant connected nodes that hold key information. Removing the skeleton will cause the entire network structure to break into small pieces. Without a skeleton structure, the network seems to malfunction as the remaining subgraphs cannot operate further. For this experiment, the networks were generated at different compression ratios. We set the control parameter α ranging from 0.01 to 1. Then, the number of isolated subnetworks could be derived by removing the skeleton network from the original network at different scales.

Figure 1 shows the compression ratio and the number of subnetworks. The fitted curve is continually decreasing after reaching the peak at the compression ratio of 0.2 (skeleton of top 20% nodes). At the compression ratio of 0.5 to 0.8, the extracted network reveals a relatively high number of subnetworks. It implies that when the backbone structure of a particular compression ratio is destroyed, the global terrorism network could collapse easily. Although the extracted network from the compression ratio 0.5 to 0.8 provides voluminous subnetworks, the network remained bulky in size. Thus, continually compressing the network is no longer effective. It can be concluded that whenever the number of an isolated subnetwork is comparatively large, the probability that skeleton nodes are sufficiently small, and the entire network structure G could be collapsed after eliminating the skeleton G' . This experiment gives direct evidence of measuring the performance of TPKS-skeleton should exclude the situation of inefficient compression ratio.

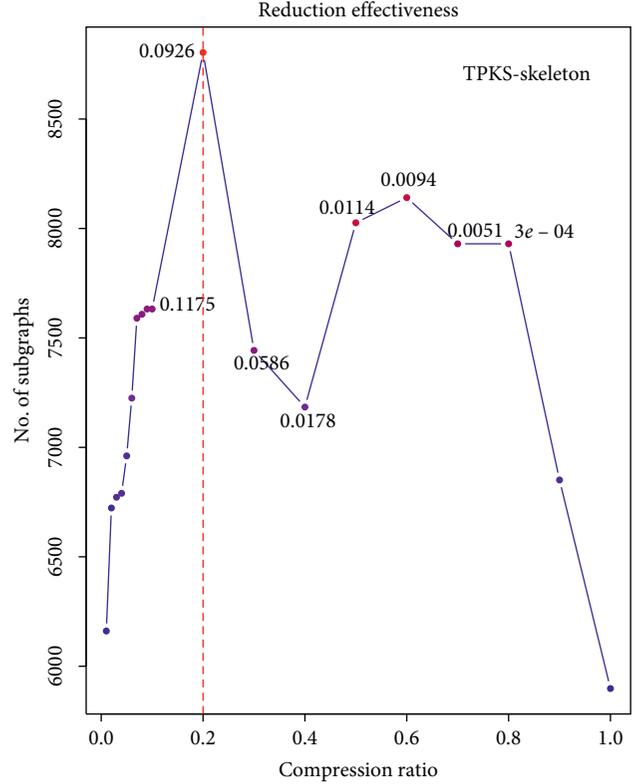


FIGURE 1: The range of reduction effectiveness denotes the number of generated isolated subnetworks at different compression ratios. The values of the coordinate points indicate the coverage ratio of an extracted network.

Together with the reduction effectiveness, the coverage ratio is applied. The coverage ratio is empirically used to measure the overall skeleton nodes from different granularities that cover the important nodes of the original network [47]. Based on the TPKS value, all the nodes are ranked in descending order. Then, the coverage of each node can be calculated:

$$\text{coverage}(v) = \frac{\text{rank}(v_i)}{\text{Rank}(v_i)}, \quad (2)$$

$$\text{coverage ratio} = \frac{\sum_{v \in V(\text{skeleton})} \text{coverage}(v)}{|V(\text{skeleton})|}. \quad (3)$$

For the arbitrary node v_i , $\text{rank}(v_i)$ indicates its ranking value in the extracted skeleton network, while $\text{Rank}(v_i)$ is its sorting value in the original network. A node in a skeleton network which has the larger coverage value implies the better quality. Therefore, the prominence of the skeleton network at a particular scale depends on the distribution of all nodes' coverage. The coverage ratio of a network is defined in equation (3). The larger the coverage ratio, the higher the accuracy of the extracted skeleton.

The significant coverage ratio of the skeleton network is declared in Figure 1. The ratio was normalized, indicating 1 as the maximum score, while 0 is the minimum. It is depicted that the network in the range of compression ratio 0.1 and 0.2 can provide a higher coverage ratio beyond

others: 0.1175 and 0.0926, respectively. When the compression ratio increases up to 0.3, the coverage ratio gradually changes downwards, and the network preserves a much higher number of nodes. Therefore, the compression ratio 0.2 was chosen as the optimal ratio for extracting the network by the TPKS-skeleton. This ratio is more efficient than others both in terms of breaking connections over the entire network, while the coverage of important nodes is well preserved.

Figure 2 illustrates the graphical representation of the original network and the extracted network with the compression ratio 0.2. The original network in Figure 2(a) exhibits the hairball structure where its real topological structure is hidden in the noise. Some of the periphery nodes appeared in the network; those can be considered as isolated nodes with few or weak connections. Compared to the original network, the TPKS-skeleton network in Figure 2(b) contains fewer high-influential terrorists and their corresponding targets (top 20% nodes from each cluster). There is only a single connection between a node pair. Thus, the network is comparatively smaller and less complex than the original network. The TPKS-skeleton network at this scale intends to sufficiently explain the behavioral patterns of the original global terrorism network hereafter.

5.2. Node-Equivalence and Spearman Ranking Correlation.

To evaluate the node importance ranking, we investigate the relationship between the top nodes from two networks. We list the top 20 terrorists and target nodes from the skeleton network to compare with the top 20 terrorists and targets from the original network.

As presented in Table 1, the association between the highly influential nodes is measured and ranked by TPKS values, which are quantified by equation (1). With the removal of some peripheral nodes at a specific compression ratio, the TPKS-skeleton network contains nearly all key nodes, most of which are positioned at the top of the ranking. Taliban ranked as the deadliest terrorist. It was followed by ISIL, Muslim Extremists, Al-Qaida in Iraq, IRA, ETA, and ISI. However, Boko Haram, one of the most dangerous terrorists in 2019 [48], is out of the list from our TPKS-skeleton. Excluding Boko Haram in the skeleton network implies that this terrorist group has a weak relationship to other key terrorists.

In this exploration, different features of all terrorist nodes are expanded to plot the correlation matrix [44]. The plots illustrated in Figure 3 enable understanding positive, negative, weak, or strong relationships between one feature and other features. The correlation coefficient indicated by the color of the cell accounts for the strength of correlation. In examining attack type for both the original network and skeleton network, we can see a strong, positive correlation with weapon type. This information signifies that terrorists tend to perform different types of attacks considering the type of accessible weapons. However, geolocations (latitude and longitude) for both networks show no sign of correlations. Focusing on the original network, most of the event features are lacking in relationships with other features. The

correlation matrix manifests the global terrorism network is extremely unpredictable, as the cooccurrence of the incidents is varying. However, the correlation among terrorist features in the skeleton network more correlates than with those in the original network. The skeleton network exhibits a weak similarity and positive correlation between year, lethality, attack type, target type, and weapon type. The TPKS-skeleton not only maintains a similar ranking correlation to the original network but also makes the relationship between two nodes' properties more pronounced.

When only the target nodes in Table 2 were considered, Private Citizens & Property, Business, and Government in Belfast are positioned at the top 3 nodes in both the skeleton network and the original network. The list of top targets showed that Belfast held the most vulnerable area for global attacks. Cities in European Union countries, including Paris, Lurgan, Londonderry, Newtownabbey, London, and Manchester, have dominated the rest of top 20 targets in the TPKS-skeleton. Likewise, Baghdad, Istanbul, and Basra from the middle east region countries are also included. However, there was less dominance in the original network by target nodes from South America.

To confirm the ability of the TPKS-skeleton method in terms of node-equivalence, the correlation analysis is employed. We adopt Spearman rank-order correlation [49] to find a statistically significant relationship between the influential values of the nodes from the original network and ones from the skeleton network. Spearman's rho $\rho = 1 - (6(\sum d_i^2)/n(n^2 - 1))$ is calculated based on the ranked value and does not carry any assumption about the distribution of the data. Where ρ represents the Spearman rank correlation. d_i is the difference between ranks of the corresponding variables, while n denotes the number of observations. The value of correlation coefficient can be interpreted between +1 and -1. A perfect degree of association between the two variables is indicated by 1. As the correlation coefficient goes towards 0, the relationship will be weaker. The negative sign indicates antirelationship among two variables.

The performance of TPKS-skeleton is compared with five competitive methods, including disparity filter (DF) [12], global weight thresholding (GWT) [50], k-core decomposition (k-core) [51], minimum spanning tree (MST) [52], and w-skeleton [53]. Inevitably, the network extraction methods rely on different disciplines. Choosing a suitable parameter for each method is difficult for an impartial judgment. For this reason, we study the reduction effectiveness of each method and select one best skeleton at a certain granularity. The optimal skeleton network, which can provide the highest number of subgraphs, is chosen as a competitor. However, the MST is an exception to this criterion because the method does not require any control parameter as an input. The best-sized tree is directly selected and constructed by applying *mst()* function available in the R platform.

The traditional correlation coefficient is defined as a paired observation. The calculation of Spearman correlation is limited for networks that have different sizes and contain unpaired nodes. To find the correlation between two sets of

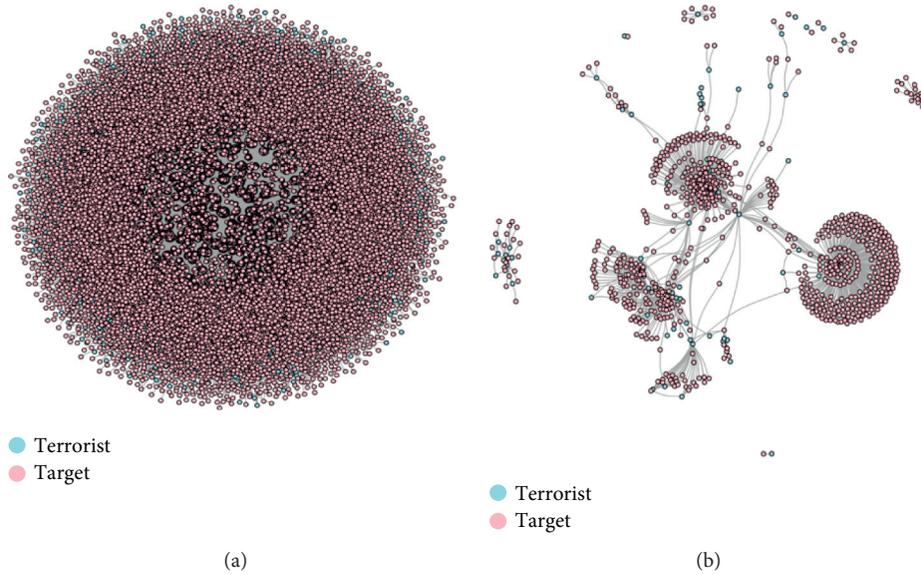


FIGURE 2: The graphical representation of (a) the original network and (b) the extracted TPKS-skeleton network both exhibit relationships between two classes of nodes: terrorists and targets.

TABLE 1: Top 20 terrorists of the skeleton network comparing with the original global terrorism network.

Rank	Original network		Rank	Skeleton network	
	Terrorist nodes	TPKS		Terrorist nodes	TPKS
1	Taliban	152.78	1	Taliban	152.78
2	Islamic State of Iraq and the Levant (ISIL)	145.25	2	Islamic State of Iraq and the Levant (ISIL)	145.25
3	Shining Path (SL)	114.27	3	Muslim Extremists	69.18
4	Revolutionary Armed Forces of Colombia (FARC)	85.76	4	Al-Qaida in Iraq	61.81
5	Muslim Extremists	69.18	5	Irish Republican Army (IRA)	57.54
6	National Liberation Army of Colombia (ELN)	62.48	6	Basque Fatherland and Freedom (ETA)	55.92
7	Al-Qaida in Iraq	61.81	7	Islamic State of Iraq (ISI)	43.31
8	Irish Republican Army (IRA)	57.54	8	Gunmen	42.03
9	New People's Army (NPA)	56.03	9	Tehrik-i-Taliban Pakistan (TTP)	42.02
10	Basque Fatherland and Freedom (ETA)	55.92	10	Palestinians	36.92
11	Boko Haram	55.17	11	Liberation Tigers of Tamil Eelam (LTTE)	33.71
12	Al-Shabaab	54.86	12	Death Squad	32.06
13	Communist Party of India-Maoist (CPI-Maoist)	54.78	13	Tawhid and Jihad	28.83
14	Tupac Amaru Revolutionary Movement (MRTA)	52.41	14	Protestant Extremists	27.40
15	Farabundo Marti National Liberation Front (FMLN)	51.63	15	Popular Liberation Army (EPL)	26.54
16	Kurdistan Workers' Party (PKK)	51.35	16	Hamas (Islamic Resistance Movement)	25.67
17	Maoists	46.46	17	Separatists	24.81
18	Islamic State of Iraq (ISI)	43.31	18	Continuity Irish Republican Army (CIRA)	24.53
19	M-19 (Movement of April 19)	42.47	19	Houthi Extremists (Ansar Allah)	23.78
20	Gunmen	42.03	20	Dissident Republicans	23.66

variables (i.e., nodes in the extracted network and nodes in the original network), we design the correlation measure using two techniques. First, ρ could be found by directly removing unpaired observations from the original network. Such that, nodes of the original network that does not appear in the skeleton network are filtered out before computing ρ . In the second technique, we adopted an idea of samples at random and representative samples, as was suggested in [54]. We run the randomization test 1000 times for each method. Then, the average result from a random set is used to reflect the real node's importance of equivalence between the skeleton network and the original network.

The information in Table 3 demonstrates that ρ of the skeleton nodes from all methods is all positively and profoundly correlated with the influence capability of the nodes in the original network. As a result, it is no surprise that the extracted skeleton from the MST always held a high correlation value in both $\text{corr}_{\text{rand}}$ and corr_{cut} . The reason is that the MST preserves all the nodes as equal to the original network. We argue that although the MST can give the highest correlation, the method fails to reduce the network size. In this case, it is meaningless in the application level since the key nodes could not be recognized, and the resulting network is still complex and remains trouble for

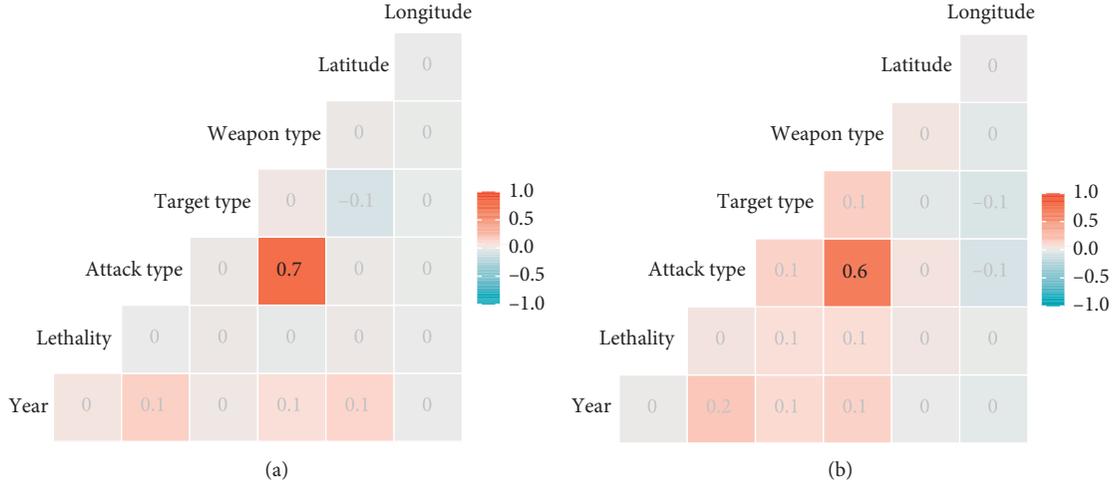


FIGURE 3: The correlation matrix of the terrorist’s features comparing between (a) the original network and (b) the skeleton network.

TABLE 2: Top 20 targets of the skeleton network comparing with the original global terrorism network.

Original network			Skeleton network		
Rank	Target nodes	TPKS	Rank	Target nodes	TPKS
1	Private Citizens & Property Belfast	25.34	1	Private Citizens & Property Belfast	25.34
2	Business Belfast	23.69	2	Business Belfast	23.69
3	Government (General) Belfast	23.25	3	Government (General) Belfast	23.25
4	Business Paris	21.72	4	Business Paris	21.72
5	Business San Miguel	21.19	5	Private Citizens & Property Lurgan	21.09
6	Private Citizens & Property Lurgan	21.09	6	Police Belfast	21.01
7	Police Belfast	21.01	7	Terrorists/Nonstate Militia Belfast	20.57
8	Private Citizens & Property Portadown	20.82	8	Private Citizens & Property Londonderry	20.29
9	Private Citizens & Property Lima	20.63	9	Private Citizens & Property Newtownabbey	19.17
10	Government (General) Lima	20.59	10	Private Citizens & Property Baghdad	18.87
11	Terrorists/Nonstate Militia Belfast	20.57	11	Transportation London	18.26
12	Business Lima	20.49	12	Business Baghdad	17.56
13	Private Citizens & Property Londonderry	20.29	13	Government (General) Baghdad	17.25
14	Private Citizens & Property Concepcion	19.65	14	Private Citizens & Property Istanbul	16.97
15	Airports & Aircraft Trujillo	19.54	15	Business Santiago	16.94
16	Private Citizens & Property San Miguel	19.34	16	Private Citizens & Property Armagh	16.77
17	Private Citizens & Property Newtownabbey	19.17	17	Business Manchester	16.55
18	Private Citizens & Property Baghdad	18.86	18	Government (General) Tuulo Barwaqo	16.52
19	Private Citizens & Property Ballymena	18.64	19	Military Basra	16.52
20	Private Citizens & Property Craigavon	18.61	20	Journalists & Media Bogota	16.52

actual uses. By measuring corr_{cut} , the results from GWT and k-core are highly related to the original network. Nevertheless, they gave a low value in a random method $\text{corr}_{\text{rand}}$. It can be clearly seen that the skeleton network from DF contains the lowest correlation value, whereas w-skeleton gives the best results in corr_{cut} , but its $\text{corr}_{\text{rand}}$ is comparatively low. By emphasizing both sides and the average value of correlation, the nodes’ importance ranking from the TPKS-skeleton has the strongest correlation as compared to other traditional methods.

5.3. Network Features and Graphical Representation of Different Skeletons. At this point, we study the cumulative degree distribution of the global terrorism network, where $p(x)$ denotes probability density function and 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 allows us to discriminate against the mechanisms of a network. Seeing that the cumulative degree distribution provided in Figure 4(a) shows a wide variety and long-tailed distribution. The distribution reveals the heterogeneous topology in the node’s activities, which follows a linear downward sloping trend. From this information, the extracted skeleton network is probably a scale-free network. As similar to the original network, 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 few links.

The same applies when investigating overall weight distribution drawn in Figure 4(b). Edge weight is compared to the distribution of the remaining edge weight after the

TABLE 3: The Spearman rank correlation $\rho(\cdot)$ of the corresponding skeleton network extraction methods.

	$\rho(DF)$	$\rho(GWT)$	$\rho(kCore)$	$\rho(MST)$	$\rho(w\text{-skeleton})$	$\rho(TPKS\text{-skeleton})$
$corr_{rand}$	0.6796	0.7579	0.6883	0.9998	0.7603	0.9268
$corr_{cut}$	0.9895	0.9977	0.9974	0.9998	0.9998	0.9985
Average	0.8345	0.8429	0.8778	0.9998	0.8801	0.9626

$corr_{rand}$ represents the correlation value obtained by randomly selected nodes. $corr_{cut}$ is the correlation value by removing unpaired nodes.

skeleton network is extracted successfully. The cumulative weight distribution of the TPKS-skeleton deviates, and it is more linear than the one observed from the original network. The distribution implies that the heavyweight edges are conserved in the skeleton because many edges are merged into a single one, and their weight attribute is combined.

The strength distribution of both the skeleton network and the original network is demonstrated in Figure 4(c). We compute the strength of a vertex v_i by $s_i = \sum_{j \in N_{(i)}} w_{ij}$, where $N_{(i)}$ indicates the neighbor of node i and w_{ij} is the weight of an edge (i, j) . This metric is used to verify having a node whose strength is larger than s . The strength of all nodes extends from the majority of vertices whose links have a minor effect on lethality. Only a fraction of nodes placed with multiple severe attacks. By filtering unimportant nodes, the edges connected to a particular node were also removed. It is the cause of a decrease in the scale of node strength. At the same time, the downward pattern keeps on a related path to the original network.

Another essential characteristic is the clustering coefficient distribution. Note that some network measures cannot be applied directly to the bipartite network. Hence, we applied the network projection method from *tnet* [55] to reveal the clustering coefficient from the terrorist-to-target network. Figure 4(d) illustrates the distribution of the local clustering coefficient. The clustering coefficient of all nodes from both skeleton and the original network is plotted against the cumulative frequency. The node clustering increases as the cumulative frequency increases. Both networks have some more precise clustered parts and some less clustered parts. The plot entails some of the nodes belong to very dense subgroups, and these subgroups are connected through hubs.

The local clustering ensures that the hub nodes are proportionally discovered across different dense subgroups. Since the global terrorism network is scale-free, the characteristic of hub nodes from different groups makes the network resilient to a random attack. As emphasized in [3], there are many more less-degree nodes than high-degree nodes in the network. Thus, randomly removing a terrorist node has a lower chance of selecting a high priority one. When the boundary nodes are removed, it shows little to no damage to the entire network. Nonetheless, directly targeting known hubs or relatively high-influential nodes, even only a small portion of them, could destroy the whole network structure. Thus, identifying the skeleton network structure by TPKS-skeleton allows preserving native characteristics of the network and significantly offers a great opportunity to destabilize the global terrorism network.

To confirm the power of TPKS-skeleton, we demonstrate the graphical representation of different skeleton networks in Figure 5. The resulting networks reveal their distinguishing features regarding the mechanism which the method relies on. Table 4 also provides several topological features of the original network and different skeleton networks. When the basic network characteristics are carefully investigated, more detailed distinctions from the network structural-level emerge. Different skeleton networks hint at uneven matches with the original network.

In Table 4, the DF method prunes the connections of the original network and produces a skeleton network which includes more than 11% of crucial terrorists and targets and over 7% of involving crucial incidents. DF and GWT are similar both in terms of network visualization (Figures 5(a)–5(b)), and topological features, e.g., number of nodes, edges, and average degree. DF holds a higher number of communities (by ignoring the case of the MST) but receives a lower global clustering than GWT. GWT produces the clustering coefficient value (0.3819), nearest to that of the original network (0.3302). The skeleton network from k-core method remains core behavior by neglecting peripheral nodes, as presented in Figure 5(c). k-core uncovers densely connected nodes with the highest maximum degree of connectivity ($k_{max} = 2864$). Compared to other methods, the k-core preserves a large number of edges. The network thus highlights a small set of nodes with the most massive connectivity and is significant for the network core structure.

In Figure 5(d), the MST has roughly the same structure as the original network. The skeleton conserves all the nodes, and only half of the edges were reduced. There are sparse connections among a large number of vertices that diminish the characteristic of cliquishness in the MST and lead the clustering coefficient to drop to zero. Since the number of nodes in the MST is fixed as equal to the original network, it might fail to achieve the objective of the skeletonizing process and network simplification defined in this paper. For w-skeleton, the method only focuses on the largest connected component of the weight thresholding network. Thus, the extracted network in Figure 5(e) merely contains one large connected network and does not include any isolated vertices or other subgroups. w-skeleton conserves more nodes than other methods as a second place following by the MST. Simultaneously, its clustering coefficient is also close to the results of the original network.

In Table 4, TPKS-skeleton is fundamentally allowing a network extraction with the smallest number of nodes and edges (904 nodes and 1349 edges). By focusing on the graphic manifested in Figure 5(f), the TPKS-skeleton network exhibits the densely connected subgroups. The

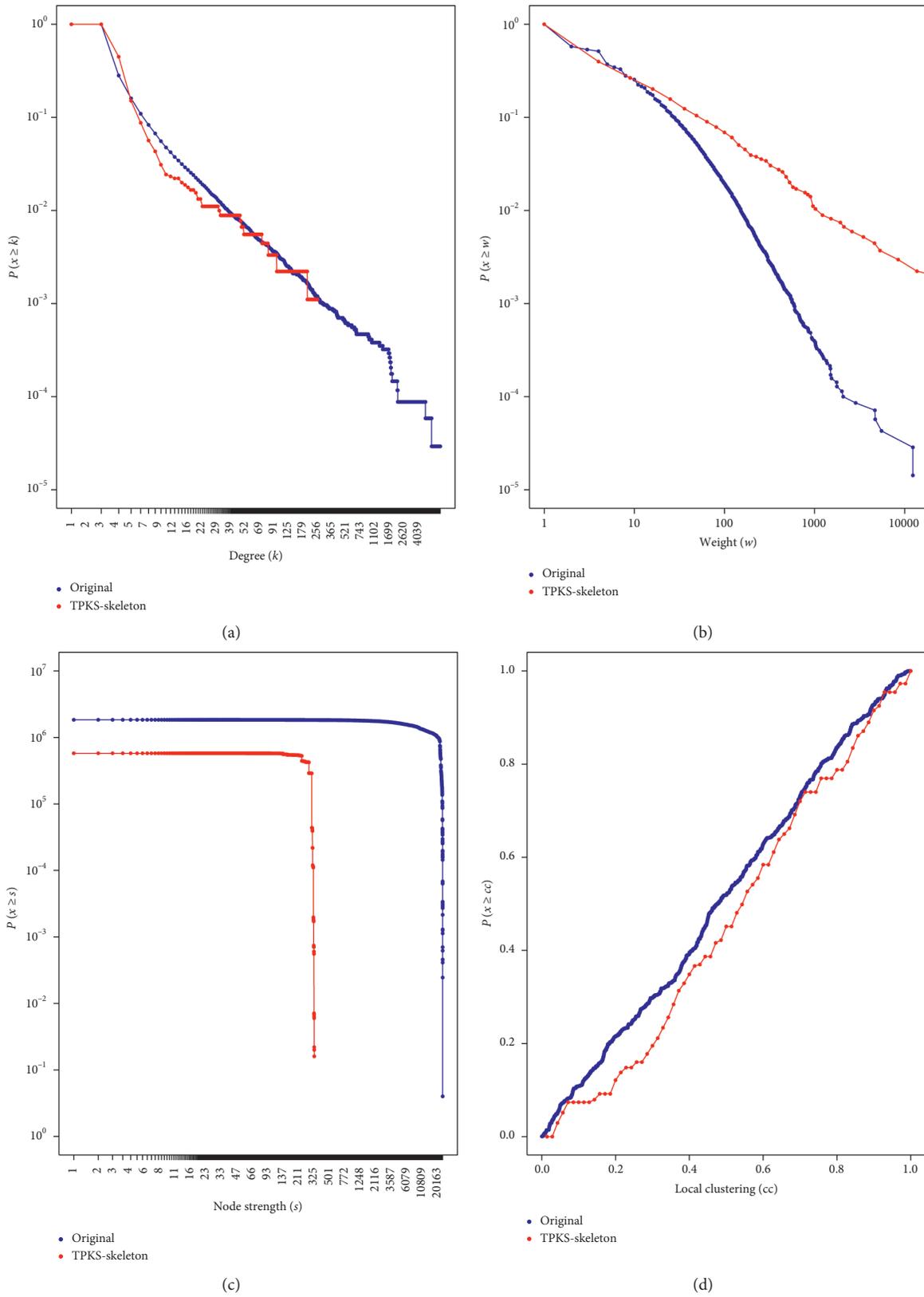


FIGURE 4: Basic network structural features of the original network and the extracted TPKS-skeleton: (a) the cumulative degree distribution, (b) the cumulative weight distribution, (c) the node strength distribution, and (d) the clustering coefficient.

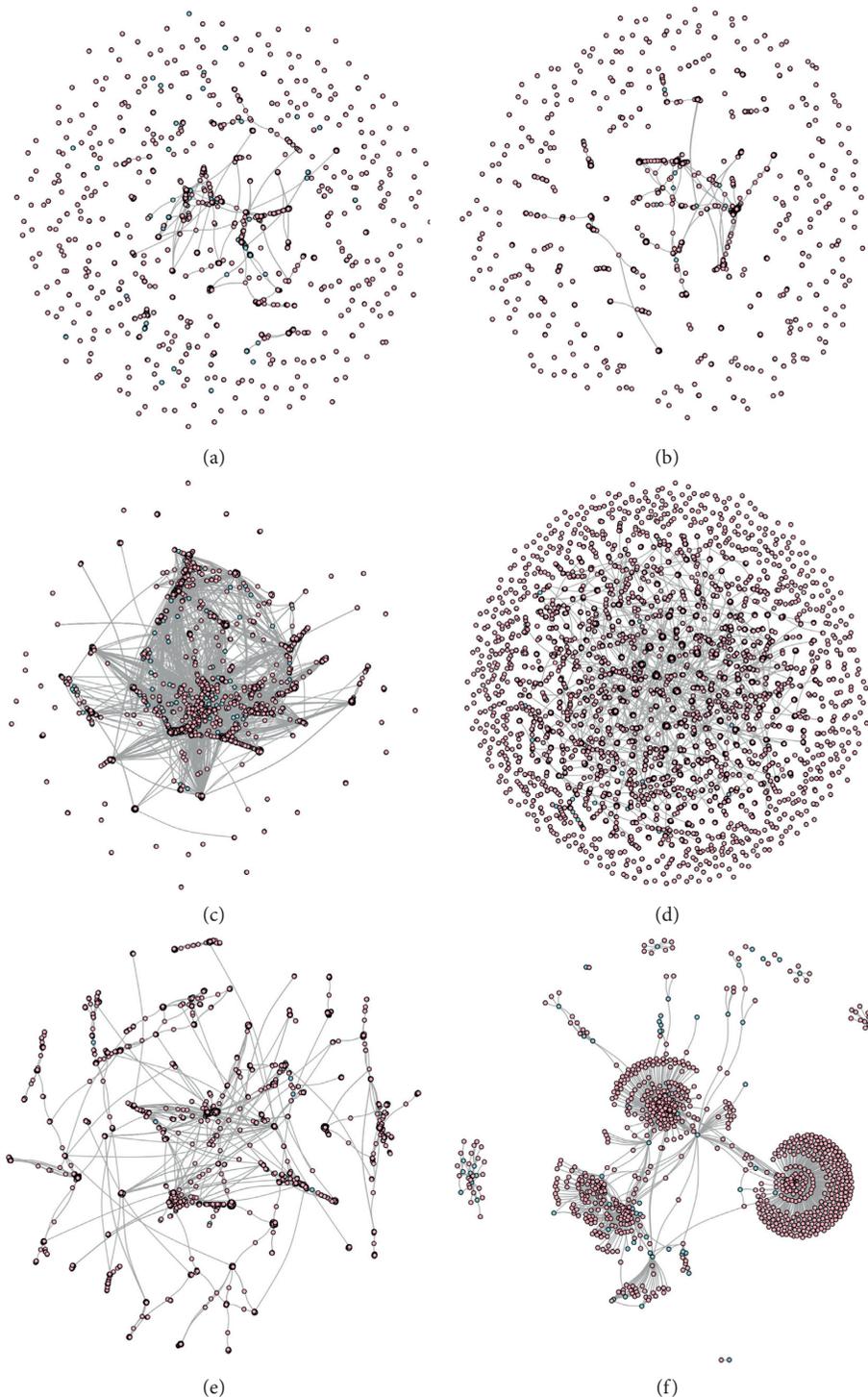


FIGURE 5: The graphical representation of the extracted skeleton network from different methods: (a) disparity filter, (b) global weight thresholding, (c) k-core decomposition, (d) minimum spanning tree, (e) w-skeleton, and (f) TPKS-skeleton.

network keeps the highest global clustering coefficient, which far exceeds that of the original network. The reason for the high clustering coefficient is that the skeletonizing process is combined with the similarity network model. Some ties among terrorist groups are added according to the similarity features and geolocations. TPKS-skeleton network contains connections across different classes of nodes, thus

leading to a high number of local cycles or cliques. The network displays some isolated subnetworks which are not tied to other subgroups. The network encompasses with nodes which are extremely significant regarding their TPKS value, and their similarity group. TPKS-skeleton reveals the more regionally oriented group of highly influential terrorists and targets with their geolocations over the globe. It

TABLE 4: Topological features of the original network compared with the skeleton network.

Network	n	m	k	k_{\max}	ϕ	d	dense	c	g
Original network	34199	70113	4.100	5246	1	1	$5.99E-05$	0.330	900
DF	3875	5206	2.686	369	12341	1	$3.47E-04$	0.193	417
GWT	3609	4514	2.501	696	12341	1	$3.47E-04$	0.382	328
k-core	3193	32792	20.54	2864	0.9998	1	$3.22E-03$	0.493	35
MST	34199	33299	1.947	2699	0.3785	1	$2.85E-05$	0	900
w-skeleton	6646	10651	3.205	1679	12341	1	$2.41E-04$	0.403	1
TPKS-skeleton	904	1349	2.985	302	126025	1	$1.65E-03$	0.836	10

n and m are the numbers of nodes and the number of edges. k and k_{\max} denote the average degree and the maximum degree. ϕ and d are the diameter and the average shortest path length. dense and c represent the network density and global clustering coefficient, while g stands for the number of communities that existed in the network.

means high TPKS nodes will not always be preserved in the skeleton network, but it depends on their role in a specific community as well. Specifically, the relationship among nodes in the TPKS-skeleton network covers not only terrorist-to-target but also terrorist-to-terrorist.

5.4. Network Structural Similarity by QAP Correlation. In this section, we employ a quadratic assignment procedure (QAP) to find the correlations between the two network metrics that retain the integrity of the observed structure [56]. The QAP is a nonparametric, permutation-based test commonly used in SNA. It is applied for identifying the systematic connections between different relations in the dyadic datasets. Based on the concept of the QAP, all pairs of nodes in the global terrorism network are analyzed. The QAP will generate a similar network’s structure by taking the matrix for the observed network and reshuffling (permutating) its rows and columns. The QAP controls permutating the network over a large number of iterations. We set 1000 iterations for this experiment. In doing so, the QAP should efficiently reflect the network’s structural correlation between the original network and different skeleton networks.

Table 5 accounts for the percentage of node reduction. The information is also including a fraction of edge reduction, weight reduction, and QAP correlation, respectively. Recalling from the application of the QAP defined above, the results show that QAP correlations between different skeleton networks and the original network are all statistically significant, except in the case of DF. The network structure of DF exhibits anticorrelation with the original network. It is followed by GWT, which bears the least structural similarity. The network structural-equivalence of the GWT and MST is both moderate (QAP=0.396 and 0.521, respectively). Interestingly, TPKS-skeleton reduces sharply in network sizes for all aspects (up to 97% for nodes, 98% for edges, and 68% for weights). At the same time, the network maintains a relatively high similarity to the original network (QAP=0.906). Despite this, the k-core network displays an almost-perfect QAP. A fraction of removed nodes, edges, and weights are smaller than those in the TPKS-skeleton. Indeed, there is always a trade-off between holding the network sizes as small as possible, while the network can retain the topology information as similar to the original network. From this matter, TPKS-skeleton outperforms other skeleton network extraction methods as it

best captures the overall network structure. The method is efficient in preserving high QAP correlation with the smallest number of nodes and edges involved in the skeleton network.

5.5. Spatial Information on Voronoi Diagrams. Voronoi diagrams have been widely adopted to evaluate spatial information embedded in maps [57]. Inspired by this, our paper employs some spatial statistics and quantitative measurements using Voronoi diagrams to calculate spatial information of the global terrorism network. The Voronoi diagram divides the entire network into polygonal partitions. The diagram represents the region governed by a particular spatial object based on the closure degree of distance to the object [58]. In these cases, the spatial objects are nodes. The classification of the two spatial nodes is the vertical bisector of the line connecting them. Such that each Voronoi region covers the location of exactly one terrorist/target who is involved in the terrorism incident. Figure 6 shows the Voronoi diagrams occupied by the spatial position of the nodes. We investigate spatial information over these diagrams in three aspects. First, the geometric information is related to the position, size, and shape of the network. Second, the spatial relationship information from neighboring features and their distribution [46]. Third, geometric balance and concentration are based on the concept of spatial equilibrium [59].

Considering the quantitative measure of information known as Shannon’s entropy [60], geometric information over the network can be evaluated by the relative space of the spatial node. Therefore, the geometric information of the global terrorism network is calculated by the entropy of the Voronoi region. The amount of geometric information suggests how spatial nodes occupy network space. In Table 6, TPKS-skeleton provides smaller geometric information $H(VG)$ than the networks generated by other methods. This characteristic implies that the space on the Voronoi region governed by each node is varied. We can assume that the area of devastating consequences from the terrorist attack also diverges. Figure 6 shows that the TPKS-skeleton method highlights a more distinct area of certain geolocation. This is because the network containing fewer nodes and regions. It can also be observed that nodes mostly correspond to a wider relative space, i.e., the blank space that surrounds the node. The larger the empty space around a certain node, the

TABLE 5: The reduction in size and QAP correlation.

Network	$\%R_{N_{\text{terrorist}}}$	$\%R_{N_{\text{target}}}$	$\%R_E$	$\%R_W$	QAP
DF	69.08%	90.55%	92.57%	31.50%	-0.0018
GWT	82.91%	90.07%	93.56%	42.12%	0.3961
k-core	81.71%	91.52%	53.22%	53.69%	1
MST	0%	0%	52.50%	61.39%	0.5212
w-skeleton	88.29%	79.82%	84.80%	32.47%	0.6066
TPKS-skeleton	97.67%	97.32%	98.07%	68.63%	0.9060

$\%R_{N_{\text{terrorist}}}$, $R_{N_{\text{target}}}$, $\%R_E$, and $\%R_W$ denote the percentage of network reduction from terrorist nodes, target nodes, edges, and weight.

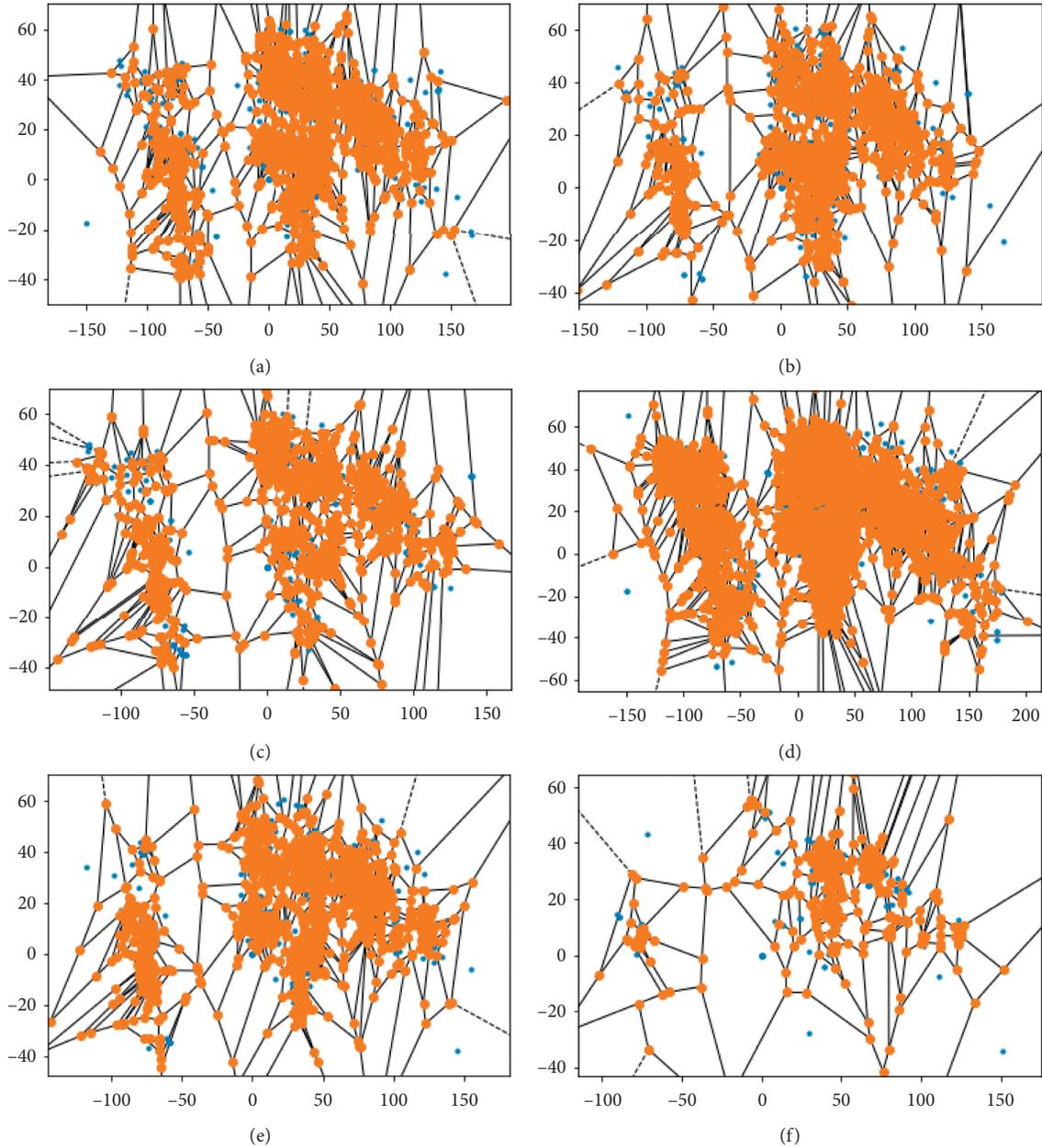


FIGURE 6: The Voronoi diagrams of extracted skeleton networks from different methods: (a) disparity filter, (b) global weight thresholding, (c) k-core decomposition, (d) minimum spanning tree, (e) w-skeleton, and (f) TPKS-skeleton.

TABLE 6: Spatial information of different skeleton networks over the Voronoi diagrams.

Network	$H(VG)$	$H_{\max}(VG)$	N_T	N_S	H_T	$E(VG)$	$H_e(VG)$
Original network	6.39362	14.38748	34199	-32354	-0.94605	0.44438	0.55561
DF	5.30150	11.27088	3875	17792	4.59148	0.47037	0.52962
GWT	5.04153	11.27961	3609	16796	4.65392	0.44696	0.55304
k-core	5.18420	10.69609	3193	12180	3.81459	0.48468	0.51531
MST	6.39359	14.38747	34199	124857	3.65089	0.44438	0.55561
w-skeleton	5.11077	12.25797	6646	29851	4.49157	0.41693	0.58306
TPKS-skeleton	3.61053	9.15228	904	3657	4.04535	0.39449	0.60551

$H(VG)$ and $H_{\max}(VG)$ denote the entropy and maximum entropy of geometric information quantified by the entropy of the Voronoi region. N_T is the total number of vertices in the dual graph, while N_S represents the sum of neighbors for all vertices. H_T is the average number of neighbors for each vertex, indicating spatial relationship information. $E(VG)$ is the geometric spatial equilibrium. $H_e(VG)$ stands for the geometric spatial concentration.

more effortlessly the node can be recognized and the more vulnerable the node will be. The remaining methods DF and k-core reflect a larger amount of geometric information, whereas the GWT and MST are also closely related to the information of the original network. Such substantial information means the relative space held by each node is similar, and the nodes in the network are more evenly distributed.

The Voronoi diagrams can reflect the adjacency relationships among nodes, which can be described by the spatial relationship information. This kind of information allows people to understand the relationship between spatial objects contained in the skeleton network. The types of spatial relationship information are typically based on three approaches: topological relationship, distance relationship, and directional relationship [57]. Our validation focuses on topological relationship information [46]. The topological relationship information can be quantified by the sum of Voronoi neighbors over the total number of nodes in the dual graph. Such that, the number of neighbors is somehow indicating the complexity of a dual graph. The index value of spatial relationships summarized in Table 6 explains that different methods have a variety of an average number of neighbors for each vertex H_T . The skeleton of GWT expresses the largest average number of neighbors, which also greatly differ from the original network. It can be stated that when filtering some meaningless edges and corresponding nodes, the relationship between the remaining spatial nodes in the skeleton becomes more dominant. By using the original network as a benchmark, the MST is the best as it can preserve the closest spatial relationship information to the original network. It is because the MST keeps all the nodes from the original network. The result from the MST could infer to the nodes have more resulting significance in spatial information and has a higher impact on the network structure beyond the edges. In another aspect, the MST provides the smallest H_T indicating less complexity as compared to other competitive methods. Meanwhile, the result from DF and w-skeleton holds an almost identical value of H_T . The spatial relationship from TPKS-skeleton is slightly more complex than the MST and k-core by having a larger H_T . Interestingly, the result from TPKS-skeleton does not deviate from other methods as it preserves relatively high spatial relationship information even though it contains only a few nodes and edges.

Next, the concept of geometric balance and concentration are applied to explain the spatial distribution. The measure of spatial equilibrium $E(VG)$ is commonly used to verify the balance of distribution within the sample space. The value of $E(VG)$ ranges between 0 and 1. The larger the value, the more the balance in the network's distribution. In contrast, the network's spatial concentration $H_e(VG)$ reveals the phenomenon of the spatial distribution that might concentrate in some particular areas. The higher the spatial equilibrium, the smaller the spatial concentration. In Table 6, all skeletons hold the characteristics of spatial distribution as the original network. The results show that all extracted skeletons have more tendency in the spatial concentration $H_e(VG)$. The highest $H_e(VG)$ is seen in the extracted network from TPKS-skeleton. However, at the same time, it provides the smallest $E(VG)$. The high $H_e(VG)$ of TPKS-skeleton is due to the method selects the major nodes from clusters and uses the intensity of incidents over the geographical area as a basis. It could be stating that the key incidents either did not happen equally or were not widespread over the globe but instead were denser in specific territories. To this end, we can conclude that our proposed TPKS-skeleton leads networks to be much reduced in size and retains reasonably all kinds of spatial information.

6. Conclusions

The global terrorism network is generally dense and contains many nodes and edges that are redundant and insignificant. The underlying structure of the real network is very likely hidden. Identifying the most influential nodes and edges would allow the real network's structure to become apparent. This process lies in the concept of skeleton network extraction. The study is based on the belief that a skeleton network would be less complex and contain only the most relevant information from a small portion of important vertices. As such, a well-extracted network would greatly simplify the analysis of complex network properties and the inherent behavior of the underlying complex system.

Our proposed skeleton network extraction method has broad prospects beyond the traditional methods. The influence value of all nodes is evaluated by the TPKS algorithm to obtain the optimal skeleton entities with high significance results. We used the integrated process of the clustering algorithm over the similarity-based network to detect the

group of nodes regarding their region in space and the event features. This is because nodes in the global terrorism network exhibit social relationships and are always associated with spatial units. Consequently, the skeleton network is simply reconstructed by a set of the high-influential nodes from each cluster and the simplified edges.

TPKS-skeleton allows one to measure how far the resulting node-equivalence, topological structure, and spatial information is from the original network. The extracted network covers the most important nodes from the original network and also provides the greatest node correlations compared to other competitive methods. Moreover, the TPKS-skeleton displays high structural-equivalence defined by QAP even though it preserves only a few nodes and edges. The validation experiments on spatial relationships using Voronoi diagrams provide a detailed distribution of information and the intensity of event features, which makes the results more understandable. Therefore, we are convinced that our skeleton network extraction method based on topology potential and similarity-based clustering is novel and unique. TPKS-skeleton network seems to provide a reasonable abstraction of the global terrorism network that helps us to understand the structure of the network easily. Its particular information could offer possible solutions for counterterrorism to monitor suspects and prevent terrorists from committing a terror attack together. More than that, we anticipate that our research findings could be useful for further academic research in complex network domains.

Data Availability

The dataset is publicly available in the Global Terrorism Database (GTD) which owned by the University of Maryland. The GTD is free for individual use and can be downloaded from <https://www.start.umd.edu/gtd/access/>. The source code used to support the findings of this study is available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by Science and Technology Innovation Research Project of the Ministry of Science and Technology of China (grant numbers ZLY201970 and ZLY201976-02), and National Natural Science Foundation of China (grant number 6147203).

References

- [1] M. Lauchs, R. L. Keast, and V. Le, "Social network analysis of terrorist networks: can it add value?" *Pakistan Journal of Criminology*, vol. 3, pp. 21–32, 2012.
- [2] P. V. Fellman and R. Wright, "Modeling terrorist networks—complex systems at the mid-range," 2008, <https://arxiv.org/abs/1405.6989>.
- [3] R. Medina and G. Hepner, "Geospatial analysis of dynamic terrorist networks," in *Values and Violence*, pp. 151–167, Springer Netherlands, Dordrecht, Netherlands, 2009.
- [4] H. T. Hawkins, "The geographical dimension of terrorism," *Eos, Transactions American Geophysical Union*, vol. 84, no. 40, p. 419, 2003.
- [5] University of Maryland, National Consortium for the Study of Terrorism and Responses to Terrorism, START, 2018 2010, 2019, <https://www.start.umd.edu/>.
- [6] Institute for Economics & Peace, *Global Terrorism Index 2018—Measuring the Impact of Terrorism*, Institute for Economics & Peace, Sydney, Australia, 2018, <http://visionofhumanity.org/app/uploads/2018/12/Global-Terrorism-Index-2018-1.pdf>.
- [7] Goldman R., Ramzy A., Garcia S. E., and Schmitt E., "Pointing a finger at a terrorist group in the aftermath of the Sri Lanka blasts—the New York times, 2019, <https://www.nytimes.com/2019/04/22/world/asia/sri-lanka-bombing-explosion.html>.
- [8] H. Yuan, K. Malang, Y. Lv, and A. Phaphuangwittayakul, "From complex network to skeleton: $\|\vec{m}_i - \vec{m}_j\|$ modified topology potential for node importance identification," in *Advanced Data Mining and Applications*, pp. 413–427, Springer, Berlin, Germany, 2018.
- [9] D. Grady, C. Thiemann, and D. Brockmann, "Robust classification of salient links in complex networks," *Nature Communications*, vol. 3, no. 1, 2012.
- [10] Z. Bu, Z. Wu, L. Qian, J. Cao, and G. Xu, "A backbone extraction method with local search for complex weighted networks," in *Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014)*, pp. 85–88, Beijing, China, August 2014.
- [11] G. Chowdhary and S. Bandyopadhyay, "Ties that matter," in *Proceedings of the 2015 IEEE International Conference on Big Data (Big Data)*, pp. 2398–2403, Santa Clara, CA, USA, October 2015.
- [12] M. A. Serrano, M. Boguna, and A. Vespignani, "Extracting the multiscale backbone of complex weighted networks," *Proceedings of the National Academy of Sciences*, vol. 106, no. 16, pp. 6483–6488, 2009.
- [13] N. Blagus, L. Šubelj, and M. Bajec, "Assessing the Effectiveness of Real-World Network simplification," *ArXiv150205156 Phys.*, 2015, <http://arxiv.org/abs/1502.05156>.
- [14] X. Zhang, Z. Zhang, H. Zhao, Q. Wang, and J. Zhu, "Extracting the globally and locally adaptive backbone of complex networks," *PLoS One*, vol. 9, no. 6, Article ID e100428, 2014.
- [15] M. Coscia and F. Neffke, "Network Backboning with Noisy Data," *ArXiv170107336 Phys.*, 2017, <http://arxiv.org/abs/1701.07336>.
- [16] R. J. Zhang, H. E. Stanley, and F. Y. Ye, "Extracting h-backbone as a core structure in weighted networks," *Scientific Reports*, vol. 8, no. 1, 2018.
- [17] G. Lu, J. Li, Z. Mo, F. Huang, and W. Liang, "Extract the backbone links and the weakness nodes via community detection," in *Proceedings of the 2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA)*, pp. 2490–2494, Hefei, China, June 2016.
- [18] N. Du, B. Wu, and B. Wang, "Backbone discovery in social networks," in *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence (WI07)*, pp. 100–103, Fremont, CA, USA, November 2007.
- [19] P. A. C. Duijn, V. Kashirin, and P. M. A. Sloot, "The relative ineffectiveness of criminal network disruption," *Scientific Reports*, vol. 4, no. 1, 2015.

- [20] K. Sharma, G. Sehgal, B. Gupta et al., “A complex network analysis of ethnic conflicts and human rights violations,” *Scientific Reports*, vol. 7, no. 1, p. 8283, 2017.
- [21] A. Rostami and H. Mondani, “The complexity of crime network data: a case study of its consequences for crime control and the study of networks,” *PLoS One*, vol. 10, no. 3, Article ID e0119309, 2015.
- [22] S. S. Husain, K. Sharma, V. Kukreti, and A. Chakraborti, “Identifying the global terror hubs and vulnerable motifs using complex network dynamics,” *Physica A: Statistical Mechanics and Its Applications*, vol. 540, Article ID 123113, 2020.
- [23] L. Dai, B. Derudder, and X. Liu, “Transport network backbone extraction: a comparison of techniques,” *Journal of Transport Geography*, vol. 69, pp. 271–281, 2018.
- [24] Y. Wang, J. Yang, J. Zhang, J. Zhang, H. Song, and Z. Li, “A method of social network node preference evaluation based on the topology potential,” in *Proceedings of the 2015 Eighth International Conference on Internet Computing for Science and Engineering (ICICSE)*, pp. 223–230, Harbin, China, November 2015.
- [25] R. Sun and W. Luo, “Using topological potential method to evaluate node importance in public opinion,” in *Proceedings of the 2017 International Conference on Electronic Industry and Automation (EIA 2017)*, Suzhou, China, July 2017.
- [26] X. Lei, Y. Zhang, S. Cheng, F.-X. Wu, and W. Pedrycz, “Topology potential based seed-growth method to identify protein complexes on dynamic PPI data,” *Information Sciences*, vol. 425, pp. 140–153, 2018.
- [27] M. Li, Y. Lu, J. Wang, F.-X. Wu, and Y. Pan, “A topology potential-based method for identifying essential proteins from PPI networks,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 12, no. 2, pp. 372–383, 2015.
- [28] Q. Han, H. Wen, M. Ren, B. Wu, and S. Li, “A topological potential weighted community-based recommendation trust model for P2P networks,” *Peer-to-Peer Networking and Applications*, vol. 8, no. 6, pp. 1048–1058, 2015.
- [29] X. Ding, Z. Wang, S. Chen, and Y. Huang, “Community-based collaborative filtering recommendation algorithm,” *International Journal of Hybrid Information Technology*, vol. 8, no. 2, pp. 149–158, 2015.
- [30] Y. Han, D. Li, and T. Wang, “Identifying different community members in complex networks based on topology potential,” *Frontiers of Computer Science in China*, vol. 5, no. 1, pp. 87–99, 2011.
- [31] S. Wang, W. Gan, D. Li, and D. Li, “Data field for hierarchical clustering,” *International Journal of Data Warehousing and Mining*, vol. 7, no. 4, pp. 43–63, 2011.
- [32] D. Zhang and L. Gao, “Virtual network mapping through locality-aware topological potential and influence node ranking,” *Chinese Journal of Electronics*, vol. 23, no. 1, pp. 61–64, 2014.
- [33] Y. Han, J. Hu, D. Li, and S. Zhang, “A novel measurement of structure properties in complex networks,” in *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, pp. 1292–1297, Springer, Berlin, Germany, 2009.
- [34] D. Li, S. Wang, and L. Deyi, *Spatial Data Mining Theory and Application*, Springer, Berlin, Germany, 2015.
- [35] K. Malang, S. Wang, and T. Dai, “Analyzing community structure based on topology potential over complex network system,” in *International Conference on Geo-Spatial Knowledge and Intelligence*, pp. 56–68, Springer, Berlin, Germany, 2017.
- [36] T. Wang, Y. Han, and J. Wu, “Evaluate nodes importance in directed network using topological potential,” in *Proceedings of the 2010 2nd International Conference on Information Engineering and Computer Science (ICIECS)*, pp. 1–4, Wuhan, China, December 2010.
- [37] K. Malang, S. Wang, A. Phaphuangwittayakul, Y. Lv, H. Yuan, and X. Zhang, “Identifying influential nodes of global terrorism network: a comparison for skeleton network extraction,” *Physica A: Statistical Mechanics and Its Applications*, vol. 545, Article ID 123769, 2020.
- [38] S. Wang, Q. Li, H. Yuan, J. Geng, T. Dai, and C. Deng, “Robust clustering with topological graph partition,” *Chinese Journal of Electronics*, vol. 28, no. 1, pp. 76–84, 2019.
- [39] J. Alison and L. Deng, *CS224W Final Project Report: Uncovering the Global Terrorism Network*, Stanford University, Stanford, CA, USA, 2017.
- [40] P. L. Bergen, September 11 attacks, facts & information, encyclopedia britannica, 2019, <https://www.britannica.com/event/September-11-attacks>.
- [41] S. Datta, C. Giannella, and H. Kargupta, “K-means clustering over a large, dynamic network,” in *Proceedings of the 2006 SIAM International Conference on Data Mining*, pp. 153–164, Bethesda, MD, USA, April 2006.
- [42] START, University of Maryland, GTD-Global Terrorism Database-CODEBOOK: INCLUSION CRITERIA AND VARIABLES, 2018.
- [43] J. H. McDonald, Spearman Rank Correlation—Handbook of Biological Statistics, 2015, <http://www.biostathandbook.com/spearman.html>.
- [44] D. Woods, R Correlation Tutorial, DataCamp Community, 2017, <https://www.datacamp.com/community/blog/r-correlation-tutorial>.
- [45] P. M. Pardalos and H. Wolkowicz, *Quadratic Assignment and Related Problems*, Vol. 16, American Mathematical Society, Providence, RI, USA, 1993.
- [46] Z. Li and P. Huang, “Quantitative measures for spatial information of maps,” *International Journal of Geographical Information Science*, vol. 16, no. 7, pp. 699–709, 2002.
- [47] H. Yuan, Y. Han, N. Cai, and W. An, “A multi-granularity backbone network extraction method based on the topology potential,” *Complexity*, vol. 2018, Article ID 8604132, 8 pages, 2018.
- [48] Institute for Economics & Peace, *Global Terrorism Index 2019-Measuring the Impact of Terrorism*, Institute for Economics & Peace, Sydney, Australia, 2019, <http://visionofhumanity.org/app/uploads/2019/11/GTI-2019-briefingweb.pdf>.
- [49] James L., Correlation (Pearson, Kendall, Spearman)—Statistics Solutions, 2019, <https://www.statisticssolutions.com/correlation-pearson-kendall-spearman/>.
- [50] Z. Neal, “The backbone of bipartite projections: inferring relationships from co-authorship, co-sponsorship, co-attendance and other co-behaviors,” *Social Networks*, vol. 39, pp. 84–97, 2014.
- [51] J. I. Alvarez-hamelin, L. D. asta, A. Barrat, and A. Vespignani, “Large scale networks fingerprinting and visualization using the k-core decomposition,” in *Advances in Neural Information Processing Systems 18*, Y. Weiss, B. Schölkopf, and J. C. Platt, Eds., pp. 41–50, MIT Press, Cambridge, MA, USA, 2006.
- [52] D.-H. Kim, J. D. Noh, and H. Jeong, “Scale-free trees: the skeletons of complex networks,” *Physical Review E*, vol. 70, no. 4, Article ID 046126, 2004.

- [53] X. Zhang and J. Zhu, "Skeleton of weighted social network," *Physica A: Statistical Mechanics and Its Applications*, vol. 392, no. 6, pp. 1547–1556, 2013.
- [54] T. A. Ebert, How to do the correlation analysis with two variable in different sample size, ResearchGate, 2016, https://www.researchgate.net/post/How_to_do_correlation_analysis_with_two_variables_in_different_sample_size.
- [55] T. Opsahl, "Triadic closure in two-mode networks: redefining the global and local clustering coefficients," *Social Networks*, vol. 35, no. 2, pp. 159–167, 2013.
- [56] D. Krackhardt, "Qap partialling as a test of spuriousness," *Social Networks*, vol. 9, no. 2, pp. 171–186, 1987.
- [57] P. Chen and W. Shi, "Measuring the spatial relationship information of multi-layered vector data," *ISPRS International Journal of Geo-Information*, vol. 7, no. 3, p. 88, 2018.
- [58] K. Xuan, G. Zhao, D. Taniar, W. Rahayu, M. Safar, and B. Srinivasan, "Voronoi-based range and continuous range query processing in mobile databases," *Journal of Computer and System Sciences*, vol. 77, no. 4, pp. 637–651, 2011.
- [59] H. Wang, *Research on quantity measurement of basic geographic information in topographic database*, Ph.D. thesis, Liaoning Technical University, Fuxin, China, 2010.
- [60] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.