2017-07

# Three is The Answer: Combining Relationships to Analyze Multilayered Terrorist Networks

Gera, Ralucca; Miller, Ryan; Saxena, Akrati; MirandaLopez, Miguel; Warnke, Scott

ACM

Gera, Ralucca, Ryan Miller, Akrati Saxena, Miguel MirandaLopez, and Scott Warnke. "Three is the answer: Combining relationships to analyze multilayered terrorist networks." In Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017, pp. 868-875. ACM, 2017.

http://hdl.handle.net/10945/57262

2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining

# Three is The Answer: Combining Relationships to Analyze Multilayered Terrorist Networks

Ralucca Gera, Ryan Miller
Department of Applied Mathematics,
Naval Postgraduate School, Monterey, CA
Email: {rgera, remiller1}@nps.edu

Akrati Saxena
Department of Computer Science and Engineering,
Indian Institute of Technology Ropar, India
Email: akrati.saxena@iitrpr.ac.in

Miguel MirandaLopez,
Department of Computer Science,
Naval Postgraduate School, Monterey, CA
Email: mamirand@nps.edu

Scott Warnke
Department of Applied Mathematics,
Naval Postgraduate School, Monterey, CA
Email: sdwarnke@nps.edu

***Abstract*—In this paper we introduce a methodology to create multilayered terrorist networks, taking into account that the main challenges of the data behind the networks are incompleteness, fuzzy boundaries, and dynamic behavior. To account for these dark networks' characteristics, we use knowledge sharing communities in determining the methodology to create 3-layered networks from each of our datasets. We analyze the resulting layers of three terrorist datasets and present explanations of why three layers should be used for these models. We also use the information of just one layer, to identify the Bali 2005 attack community.**

## I. Introduction and Motivation

The current research considers how attribute enriched data about terrorist organizations can be aggregated in order to (1) be effectively used as a network, and (2) gain an understanding of the groups of people in it. Since most terrorist networks work hard to hide their interactions, the data must be modeled in an appropriate type of the network in order for it to be useful.

There are many difficulties associated with mapping and analyzing dark networks. Krebs [1] uses the September $11^{th}$ 2001 terrorist attack in the United States as a case study in dark network mapping and analysis. He describes three challenges previously identified by Sparrow [2] that are specifically associated with mapping and analyzing criminal social networks: 1. incompleteness, 2. fuzzy boundaries, and 3. dynamic behavior.

The resiliency of a dark network is qualitatively described by Krebs [1] as strong due to the high redundancy of trust relationships which includes classmates, kinship, or participating in terrorist related training and operations. This highlights the differences between social network and covert network. The classification of relationships as strong or weak ties is entirely dependent on the type of network being analyzed. For dark networks, trusted prior contacts is typically considered a strong tie between two vertices whereas the two people connected by the same nationality could be viewed as a weak tie. The strong tie clearly emphasizes a close relationship, whereas a weak tie viewed by itself may offer only ambiguity on the relationship status. Analysis of strong ties in social networks usually produces the "cluster of network players" [1]. However, network players in dark networks may visibly appear to only have weak ties [1]. Everton [3] supports the claim that an optimal combination of both weak and strong ties is ideal for dark network analysis. This claim highlights the notion that multiple relationships of data must be included when analyzing the network. The incomplete and secret nature of dark networks requires weak ties to help illuminate potentially hidden strong ties.

Krebs offers a strategy for disrupting terrorist networks through information aggregation and knowledge sharing. Under this strategy, the key vertices to target in a network are vertices with unique skills and vertices that have deep rooted trust relationships with other groups. For more information on understanding dark networks and using topological characteristics to disrupt them see [4].

Krebs and Sparrow emphasize the sparse nature of dark networks and in turn, Taylor et al. [5] suggest aggregating similar relationships into categories selective on relationship choices. They bring the awareness against the dangers of too many relationships and redundancy. Didier et al. [6] present problems associated with aggregating relationships.

The research highlighted here serve as the foundational understanding and inspiration that enabled us to develop our methodology for obtaining information from data, by creating multilayered terrorist networks. We propose methodology to create networks from three existing terrorist data with the goal of making these knowledge sharing communities (KSC) evident, while addressing the sparsity of the terrorist networks.

*ASONAM '17*, July 31 - August 03, 2017, Sydney, Australia
© 2017 Association for Computing Machinery.
ACM ISBN 978-1-4503-4993-2/17/07?/$15.00
http://dx.doi.org/10.1145/3110025.3110154

We present a methodology that creates meaningful multilayered terrorist networks based on the data tagged with attributes between the individuals. We show how too many layers may not be the best option as they are incomplete and too sparse. Similarly, collapsing all the layers into a single network without differentiating between the edge type is also not desired, as the information between strong and weak ties is lost, and it makes the communities boundaries blurred. Our methodology produces synthetic layers of multilayered networks with the goal of identifying KSC.

We present an analysis of these networks at different granularity levels, and we further emphasize the communities obtained in each case. We validate the proposed methodology by analyzing the KSC communities obtained and comparing them to the ones obtained by randomly combining 3, 4 and 5 layers. We also present an application that focuses on identifying targeted communities based on the Bali attack of 2005.

## II. Community Metrics

We first present methods used to quantify community detection. We use these metrics to compare different community structures obtained using Louvain method [7], by varying the combinations of attributions.

### A. *Community Quality Metrics*

We use two metrics to measure the quality of the community detection:

1) **Modularity** (*Q*) of a community partition is defined as

$$Q = \frac{1}{4m}\sum_{ij}\big(A_{ij} - \frac{k_i k_j}{2m}\big)\delta_{ij}, \tag{1}$$

where $m$ is the number of edges in the network, $A_{ij} = 1$ if $(i, j)$ is an edge and 0 otherwise, $k_i$ (and $k_j$) is the degree of node $i$ (and $j$), $\delta_{ij} = 1$ if both $i$ and $j$ are in the same community and 0 otherwise [8].

2) **Cluster adequacy** ($Q'$), normalizes graph modularity, $Q$, by dividing the measured $Q$ by the best possible $Q$ for a given number of communities, $m$. The best possible $Q$ is determined as a function of the number of communities [4], [3]. Bogartti et. al [9] defined the cluster adequacy as,

$$Q' = \frac{Q}{1 - \frac{1}{comm\_count}}, \tag{2}$$

where $Q$ is the best possible modularity, and *comm_count* is the number of communities for the measured $Q$.

Looking purely at the measured $Q$, it is possible to mistakenly conclude that the communities are mediocre quality. However, by comparing the measured value of $Q$ to the best possible modularity for a given number of communities, cluster adequacy reveals that the community quality is much higher.

Cluster adequacy favors a uniform distribution of vertices into equal sized communities, which is rarely possible in real networks. Orman et al. [10] argue that similar to a degree distribution, community size tends to follow a power-law distribution as well.

### B. *Metrics for Comparing Two Community Partitions*

There are several metrics used to compare two networks' communities, each having its own benefits, none of them being the standard. They all build on the confusion matrix of two partitions $P^a$ and $P^b$ each of the two networks/subnetworks. The confusion matrix displays counts of the number of nodes in common to each set in the partition $P^a$ to each set in the partition $P^b$. The four metrics, Normalized Mutual Information (NMI) [11], [12], Purity (or fraction of correctly classified vertices) [13], [14], Rand Index [15] and Adjusted Rand Index (ARI) [16], [17] combine the elements of the confusion matrix differently, producing a value in the interval $[0, 1]$, with the understanding that closer the value is to 1 more similar the two partitions $P^a$ and $P^b$ are. While we present the values for each of these metrics, we also use the sum of four metrics to capture the contribution of each of them.

We used the four metrics to compare the quality of identified communities and their formal definitions are:

1) **Normalized Mutual Index (NMI)**:

$$NMI(P^a, P^b) = \frac{-2\sum_{i=1}^{k_a}\sum_{j=1}^{k_b} n_{i,j}^{ab}\log(\frac{n_{i,j}^{ab}\cdot n}{n_i^a \cdot n_j^b})}{\sum_{i=1}^{k_a} n_i^a \cdot \log(\frac{n_i^a}{n}) + \sum_{j=1}^{k_b} n_j^b \cdot \log(\frac{n_j^b}{n})}, \tag{3}$$

where $n_{i,j}^{ab}$ the vertices identified by the algorithm to be in community $i$ in $P^a$ while they are in a different community $j$ in $P^b$.

2) **Purity** identifies the likely community counterparts in two separate network partitions based on the idea that a community in $P^a$ corresponds to the community in $P^b$ with the highest number of mutual nodes [13]. Equation 4 defines the purity measure.

$$Purity(P^a, P^b) = \frac{1}{n}\sum_{i=1}^{k} max_j |n_i^a \cap n_j^b|, \tag{4}$$

where the purity is found by summing the maximum values of the for each row (or column) of the confusion matrix.

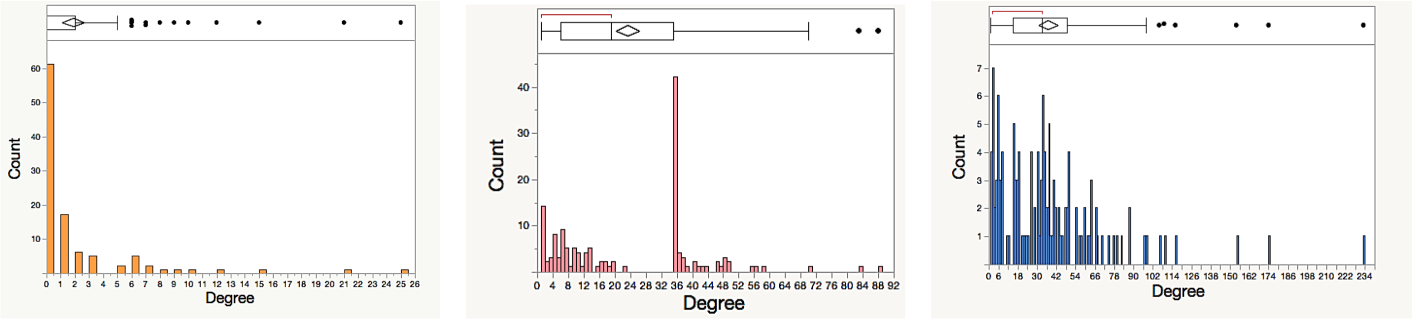
3) **Rand Index** is given by

$$Rand\,Index = \frac{a+b}{\binom{n_{samples}}{2}}, \tag{5}$$

where $a$ is the number of node pairs that are int he same community in $P^a$ and $P^b$, while $b$ is the number of pairs that are in different communities in $P^a$ and $P^b$, $n$ is the number of nodes in the network.

4) **Adjusted Rand Index** is given by

$$ARI = \frac{RI - E[RI]}{max(RI) - E[RI]}, \tag{6}$$



(a) Boko Haram (b) FARC (c) Noordin

Fig. 1: The three networks' weighted degree distribution.

where $RI$ is the Rand Index and $E[RI]$ is the expected value of $RI$.

Since a combination of these metrics captures the comparison of two community partitions, we combine them into a single value to guide the conclusions of our analysis:

$$\mathbf{ARPN} = \mathbf{ARI} + \mathbf{Rand\,Index} + \mathbf{Purity} + \mathbf{NMI}. \qquad (7)$$

## III. Data Description

Our methodology is general for multiple attribute terrorist data, but for the purposes of this research, three small, real, and dark multiplex network case studies were examined:

- Boko Haram Terrorist Network data set contains 44 vertices,
- FARC Terrorist Network data set contains 142 vertices,
- Noordin Top Network data set containing 133 vertices,

with the respective edges presented in the Tables I-III.

The degree distributions of the three networks are presented in Figure 1. Each edge is a particular type of correlation/interaction relationship between terrorists. Each type of relationship gives rise to a relationship subnetwork summarized in the rows of Tables I, II and III. We use these tables as baseline in our comparison, as each of them would be an easy default for a layer in multilayered networks.

We only use some of the attributes as relationships, as the other ones either were classified as weak and redundant or irrelevant attributes. We will propose three consistent categories/layers for each network in SectionV. The correlated relationships are colored based on the one of the three categories for an easy of correlating the relationships that form different layers.

We follow them with a global overview of the networks presented at the relationship level. The attributes were tagged in the original data, and for each used attribute we create a relationship subnetwork: capturing the Total Number of Vertices of degree greater than 0 (V), Total Number of Edges (E), Average Degree (AD), Average Weighted Degree (AWD), Network Diameter (Di), Graph Density (De), Modularity (M), Average Clustering Coefficient (ACC), Average Path Length (APL), and Number of Partitioned Components (P) for each relationship.

### A. *Boko Haram Terrorist Network*

The Boko Haram Terrorist Network contains the relationship information of 44 terrorists that belong to an Islamic sect that primarily operates in northern Nigeria since 2002. According to Walker [18], the group believes the current government in Nigeria is corrupted by false Muslims. The network is extremely sparse due to its relatively young cell-like structure, and lack of collective leadership. This network dataset was created by Cunningham [19] using a variety of open source documents. We re-organized the available relationship data into edge lists to build 9 separate layers for the case study on the Boko Haram Terrorist Network.

TABLE I: Boko Haram Network topological characteristics by relationship subnetwork.

| Relationship Name | V | E | AD | AWD | Di | De | M | ACC | APL | P |
|---|---|---|---|---|---|---|---|---|---|---|
| Colleagues | 9 | 8 | 1.78 | 1.78 | 4 | 0.22 | 0.41 | 0.00 | 2.33 | 1 |
| Kinship | 6 | 3 | 1.00 | 1.00 | 1 | 0.20 | 0.67 | NA | 1.00 | 3 |
| Superior | 18 | 17 | 1.89 | 1.89 | 3 | 0.11 | 0.54 | 0.18 | 1.93 | 4 |
| Supporter | 5 | 3 | 1.20 | 1.20 | 2 | 0.30 | 0.44 | 0.00 | 1.25 | 2 |
| Financial Ties | 2 | 1 | 1.00 | 1.00 | 1 | 1.00 | 0.00 | NA | 1.00 | 1 |
| Communication | 2 | 1 | 1.00 | 1.00 | 1 | 1.00 | 0.00 | NA | 1.00 | 1 |
| Membership | 14 | 32 | 2.71 | 4.57 | 2 | 0.21 | 0.30 | 0.93 | 1.34 | 4 |
| Shared Events | 16 | 21 | 2.63 | 2.63 | 2 | 0.18 | 0.40 | 0.81 | 1.22 | 5 |
| Collaboration | 13 | 13 | 2.00 | 2.00 | 7 | 0.17 | 0.47 | 0.35 | 2.84 | 2 |
| Average | 9 | 11 | 1.69 | 1.90 | 3 | 0.38 | 0.36 | 0.38 | 1.55 | 3 |

### B. *FARC Terrorist Network*

The FARC Terrorist Network data set includes the relationship information of 142 terrorists known as the Revolutionary Armed Forces of Colombia that primarily operates in Columbia and Venezuela since 1964. According to Weimann [20], the organization believes in Marxist ideology and seeks to overthrow the Colombian government. The network is sparse for most layers, but has a well-documented hierarchical structural layer due to social media [20]. This network data set was created by Cunningham et al. [21] using a variety of open source documents. We re-organized the available relationship data into edge lists to build 10 separate layers for the case study on the FARC Terrorist Network.

TABLE II: FARC Network topological characteristics by relationship subnetwork.

| Relationship Name | V | E | AD | AWD | Di | De | M | ACC | APL | P |
|---|---|---|---|---|---|---|---|---|---|---|
| Friendship | 2 | 1 | 1.00 | 1.00 | 1 | 1.00 | 0.00 | NA | 1.00 | 1 |
| Kinship | 8 | 8 | 2.00 | 2.00 | 1 | 0.29 | 0.41 | 1.00 | 1.00 | 3 |
| Superior | 17 | 12 | 1.41 | 1.41 | 2 | 0.09 | 0.74 | 0.00 | 1.52 | 5 |
| Supporter | 3 | 2 | 1.33 | 1.33 | 2 | 0.67 | 0.00 | 0.00 | 1.33 | 1 |
| Lovers | 8 | 4 | 1.00 | 1.00 | 1 | 0.14 | 0.75 | NA | 1.00 | 4 |
| Radicalizer | 2 | 1 | 1.00 | 1.00 | 1 | 1.00 | 0.00 | NA | 1.00 | 1 |
| Communication | 9 | 7 | 1.56 | 1.56 | 4 | 0.19 | 0.46 | 0.00 | 2.23 | 2 |
| Meetings | 17 | 30 | 3.53 | 3.53 | 3 | 0.22 | 0.43 | 0.91 | 1.44 | 4 |
| Shared Orgs | 120 | 1577 | 24.6 | 26.3 | 4 | 0.21 | 0.50 | 0.95 | 1.87 | 5 |
| Collaboration | 13 | 8 | 1.23 | 1.23 | 2 | 0.10 | 0.78 | 0.00 | 1.27 | 5 |
| Average | 20 | 165 | 3.87 | 4.04 | 2 | .39 | .41 | .41 | 1.37 | 3 |

### C. *Noordin Top Terrorist Network*

The Noordin Top network data set contains the relationship information of 139 terrorists that belong to five major parent terrorist organizations operating in Indonesia [22]. The network is named after the key broker, Noordin Top, who was known for coordinating between terrorist organizations for training and operations. This network was primarily developed from the information provided by an article published by the International Crisis Group in 2006, *Terrorism in Indonesia: Noordin's Networks* [4]. Roberts et al. [22] used this information to construct a possible total of 36 attributes. We re-organized the attribute data into relationships captured by the edge lists to build the layers.

TABLE III: Noordin Network topological characteristics by relationship subnetwork.

| Relationship Name | V | E | AD | AWD | Di | De | M | ACC | APL | P |
|---|---|---|---|---|---|---|---|---|---|---|
| Classmates | 44 | 217 | 9.32 | 9.86 | 7 | 0.22 | 0.35 | 0.76 | 2.48 | 1 |
| Kinship | 44 | 49 | 2.23 | 2.23 | 2 | 0.05 | 0.87 | 0.95 | 1.09 | 15 |
| Soulmates | 13 | 17 | 2.62 | 2.62 | 2 | 0.22 | 0.65 | 0.89 | 1.23 | 3 |
| Friends | 83 | 158 | 3.71 | 3.81 | 9 | 0.05 | 0.71 | 0.55 | 4.01 | 3 |
| Mentor Ideological | 21 | 15 | 1.43 | 1.43 | 5 | 0.07 | 0.68 | 0.00 | 2.03 | 7 |
| Mentor Supervisory | 46 | 51 | 2.22 | 2.22 | 6 | 0.05 | 0.57 | 0.40 | 2.50 | 6 |
| Mentor Technological | 13 | 13 | 2.00 | 2.00 | 5 | 0.17 | 0.34 | 0.00 | 2.20 | 2 |
| Recruiting | 27 | 24 | 1.78 | 1.78 | 3 | 0.07 | 0.75 | 0.37 | 1.78 | 5 |
| Meetings | 33 | 110 | 5.33 | 6.67 | 4 | 0.17 | 0.33 | 0.84 | 2.16 | 1 |
| Communication | 120 | 318 | 5.30 | 5.30 | 8 | 0.05 | 0.54 | 0.53 | 3.10 | 1 |
| Logistical Place | 34 | 106 | 5.71 | 6.24 | 3 | 0.17 | 0.28 | 0.83 | 1.73 | 5 |
| Operations | 60 | 490 | 15.63 | 16.33 | 2 | 0.27 | 0.51 | 0.94 | 1.67 | 4 |
| Training | 54 | 291 | 9.74 | 10.78 | 4 | 0.18 | 0.58 | 0.89 | 2.33 | 2 |
| Logistical Function | 49 | 592 | 22.61 | 24.16 | 2 | 0.47 | 0.28 | 0.89 | 1.53 | 1 |
| Average | 46 | 175 | 6.40 | 6.82 | 4 | 0.16 | 0.53 | 0.63 | 2.13 | 4 |

## IV. Example applications of the networks

In this section, we use the information of just one layer partial information, and compare it to the whole information of the multilayered network, using community detection. First, we focus on identifying a community of interest in Subsection IV-A, and then identify all the communities IV-B.

### A. *Community of Interest Identification*

Three years following the worst attacks in the history of Indonesia, notorious terrorist group Jemaah Islamiyah (JI) was once again responsible for the bombings in the beach towns of Kuta and Jimbaran. Three suicide bombers attacked three different cafes, which claimed the lives of 20 people, including Australian and Japanese nationals. The prime suspect, who became Indonesia's most wanted Islamist militant, was JI's financier, expert bomb-maker and, most importantly, the mastermind behind the Bali attacks: Noordin Mohammad Top [23].

We ran community detection on the Noordin monoplex, and identified one of the communities to have the individuals associated with this Bali attack. We call this Bali Community the Community of Interest (COI) Identification. We wish to identify this COI, from having access to partial information obtained from just the communication layer. The communication subnetwork contains 20% of the edges of the monoplex, and 120 of the 139 terrorists. Is the information captured in this communication subnetwork enough to identify the COI in the monoplex?

To pursue this, we run Louvain community detection the communication subnetwork. We identify the potential COI in the communication subnetwork and compare it to the real COI. While Louvain is not a deterministic algorithm, on our monoplex we get only two very similar partitions, and we base our analysis on the prevalent one. The only significant change we found between the two versions is that Noordin Top was placed different communities in the two versions. This is because he is connected at such a high degree, and as reflected in the supporting literature [23] he is a member of JI (one of the network's groups), yet coordinates all five groups.

The COI has comparatively higher clustering coefficient and is denser than the rest of the communities in the monoplex. Thus, we chose the community with the highest density among all the communities of the communication subnetwork. This successfully identified a community, whose nodes form a proper subset of the COI's nodes. The missing five nodes include three suicide bombers, Salik Firdaus, Misno, Aip Hidayat, and the recruiter, Jabir. Because of the process of recruiting suicide bombers [23], these individuals are not as closely associated with the rest of the actors in the communication layer, making it hard to detect them. This is an indication that not only are we able to identify a specific terrorist cell, but when we compare the results of a single relationship subnetwork to the results of the monoplex, we can identify members of the terrorist cell that have different roles than the majority of the members of the community.

Recall that community detection partitions the node set of the monoplex into communities based on the edges present. While the communication subnetwork contains about 20% of the monoplex's edge count, we are able to use its structure to identify the main actors of the COI.

Our COI only appears in three other single relationship subnetworks. Using the friendship attribute, the relationships subnetwork has a ten-node community with the same ten members as the communication subnetwork, the suicide bombers being excluded from the target community again. In the operations relationship subnetwork, again we see the target community identified, and it is larger than the COI. This community in the operations relationship subnetwork includes the suicide bombers. Perhaps the most intriguing relationship is the recruiting attribute: only nine of the ten individuals n the COI appear in this community missing Ardi Wibowo and the suicide bombers. This highlights some of the inherent qualities of our dark network. We are confident that Ardi Wibowo is a member of this terrorist organization, however when we analyze the community structure based on recruiting, he is absent. The three bombers and their recruited appear in a separate community of four nodes. Jabir is a member of the COI, yet he was never identified as a member of the potential COI in any single relationship subnetwork.

### B. Identifying All Communities from Partial Information

We now expand our process by trying to identify all the communities in the monoplex just from the communities in the subnetwork of each attribute. We evaluate each of the relationships subnetworks, to see how well each relationship individually can be used to determine the community structure of the full network. We will use the cluster adequacy, to evaluate how strong the community structure is in each relationship subnetwork. The higher the value, the better community structure we have.

We also use NMI/Purity/ARI/Rand Index to see how similar the community structure in a particular relationship subnetwork is to the one in the monoplex. We desire values close to 1, showing that a layer captures significant information of the monoplex. These results are presented in Tables IV, V and VI, for each of the three networks.

| Relationship | Comty. | ARI | Rand | Purity | Assty. | NMI | Clr. Adq. |
|---|---|---|---|---|---|---|---|
| Colleagues | 99 | 0.1327 | 0.9789 | 1 | -0.5000 | 0.9300 | 0.418 |
| Kinship | 102 | 0.0462 | 0.9778 | 1 | 0 | 0.9241 | 0.673 |
| Superior | 92 | 0.3100 | 0.9810 | 0.9905 | -0.1244 | 0.9396 | 0.540 |
| Supporter | 102 | 0.0299 | 0.9773 | 0.9905 | -0.5000 | 0.9203 | 0.449 |
| Financial Ties | 104 | -3.6E-4 | 0.9771 | 0.9905 | 0 | 0.9182 | 0 |
| Communication | 104 | 0.0156 | 0.9775 | 1 | 0 | 0.9212 | 0 |
| Membership | 96 | 0.0196 | 0.9795 | 0.9905 | 0.4711 | 0.9316 | 0.300 |
| Shared Events | 94 | 0.3390 | 0.9819 | 0.9905 | 0.7667 | 0.9386 | .408 |
| Collaboration | 96 | 0.1799 | 0.9789 | 0.9810 | 0.0758 | 0.9278 | 0.448 |
| Monoplex | 72 | 1 | 1 | 1 | 0.1488 | 1 | 0.508 |

TABLE IV: The number of communities and their metrics in each of the 9 relationships of the Boko Haram Network.

| Relationship | Comty. | ARI | Rand | Purity | Assty. | NMI | Clr. Adeq. |
|---|---|---|---|---|---|---|---|
| Friendship | 141 | 8.47E-4 | 0.8092 | 1 | 0 | 0.5494 | 0.557 |
| Kinship | 137 | 0.0068 | 0.8100 | 1 | 1 | 0.5534 | 0.334 |
| Superior | 130 | 0.0168 | 0.8108 | 0.9930 | -0.6145 | 0.5562 | 0.519 |
| Supporter | 140 | .0025 | 0.8094 1 | 1 | -1 | 0.5505 | 0.720 |
| Lovers | 138 | 0.0023 | 0.8093 | 0.9930 | 0 | 0.5489 | 0.537 |
| Radicalizer | 141 | 8.47E-4 | 0.8092 | 1 | 0 | 0.5494 | 0.656 |
| Communication | 136 | 0.0011 | 0.8087 | 0.9789 | -0.5 | 0.5418 | 0.343 |
| Meetings | 130 | 0.0212 | 0.8114 | 0.9859 | 0.8654 | 0.5550 | 0.759 |
| Shared Orgs | 30 | 0.8876 | 0.9675 | 1 | 0.2807 | 0.8881 | 0.875 |
| Collaboration | 134 | 0.0030 | 0.8090 | 0.9718 | -0.6 | 0.5421 | 0.280 |
| Monoplex | 9 | 1 | 1 | 1 | 0.2987 | 1 | 0.584 |

TABLE V: The number of communities and their metrics in each of the 10 relationships of the FARC Network.

Thus, the existing community structure in the multiplex cannot be discovered from any single relationship subnetwork. However, aggregating some of these relationships into layers based on their meaning will enable us to get a better detection of our communities, as we show next.

| Relationship | Comty. | ARI | Rand | Purity | Assty. | NMI | Clr. Adq. |
|---|---|---|---|---|---|---|---|
| Classmates | 101 | 0.0187 | 0.8136 | 0.8058 | -0.0162 | 0.4961 | 0.348 |
| Kinship | 110 | 0.0129 | 0.8279 | 0.8777 | -0.5678 | 0.5218 | 0.875 |
| Soulmates | 129 | -0.0021 | 0.8274 | 0.9425 | -0.3258 | 0.5331 | 0.656 |
| Friends | 65 | 0.0192 | 0.7960 | 0.6331 | -0.2198 | 0.4086 | 0.720 |
| Mentor-Ideo. | 126 | 0.0011 | 0.8279 | 0.9281 | 0.0531 | 0.5352 | 0.702 |
| Mentor-Supvr. | 103 | -0.0030 | 0.8189 | 0.8058 | -0.3258 | 0.4773 | 0.537 |
| Mentor-Tech. | 130 | -8.76E-5 | 0.8281 | 0.9568 | -0.1591 | 0.5405 | 0.343 |
| Recruiting | 117 | 0.0290 | 0.8293 | 0.9281 | -0.5524 | 0.5402 | 0.759 |
| Meetings | 110 | 0.0596 | 0.8316 | 0.9425 | -0.1519 | 0.5578 | 0.334 |
| Communication | 25 | 0.1307 | 0.7916 | 0.5899 | -0.1182 | 0.4055 | 0.557 |
| Logistical Place | 112 | 0.0334 | 0.8286 | 0.9209 | 0.1648 | 0.5383 | 0.210 |
| Operations | 85 | 0.3189 | 0.8615 | 0.9425 | -0.0466 | 0.6270 | 0.519 |
| Logistical Function | 93 | 0.0532 | 0.8165 | 0.8130 | 0.8653 | 0.4922 | 0.280 |
| Training | 88 | 0.2766 | 0.8531 | 0.9353 | 0.3860 | 0.6165 | 0.589 |
| Monoplex | 13 | 1 | 1 | 1 | 0.0531 | 1 | 0.388 |

TABLE VI: The number of communities and their metrics in each of the 14 relationships of the Noordin Top Network.

## V. Methodology For Multilayered Terrorist Networks

We organized the relationships of each of the three networks into a multilayered network with three layers, each layer being the union of related relationships. Layers are based on Kreb's observation that dark networks are sparse. The aggregation of similar attributes into layers reduces sparseness and increases network density for more accurate community detection [1]. As an example, Noordin network can be seen in Figure 2, and similar figures can be created for the other two.

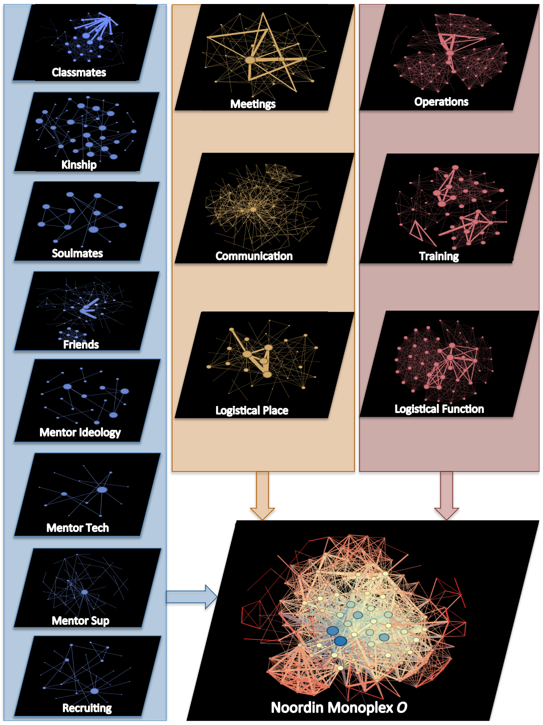


Fig. 2: The Noordin multilayered network: the 14 relationships organized by layer color with the monoplex $O$ representing the aggregation of all the relationships.

This allows us to summarize the data to achieve analytical depth with a user's goal in mind. In our case, the goal is to find a structure for the data, that does not have too many layers (such as each relationship being its own layer) and at the same time not just a monoplex network in which we lose the richness of the data. A secondary goal, is to group the layers based on the relationships' correlation, so that community detection identifies meaningful communities. Meaningful communities are user-inspired categorical communities based upon the analytical needs, whose structural properties enhance the customer's understanding of the network in order to achieve the customer's objective.

In the absence of a customer to drive the objectives, we use the Joint Improvised Explosive Device Defeat Organization (JIEDDO). JIEDDO's Attack-the-Network philosophy to infer the customer objectives. Martin et al. [24] describe that the mission of JIEDDO is 'to focus, lead, advocate, and coordinate all Department of Defense actions in support of the Combatant Commanders' and their respective Joint Task Forces' efforts to defeat Improvised Explosive Devices as weapons of strategic influence." Understanding of these customer objectives provides context and focus in finding communities for the purposes of disrupting its ability to function. For each of the three networks, different attributes form each of the three layers, working towards identifying meaningful communities as shown in Table VII.

TABLE VII: The attributes per layer in each network

| **Layer** | Boko | FARC | Noordin |
|---|---|---|---|
| Trust | colleagues, kinship, superior, supporter | friendship, kinship, superior, supporter, lovers, radicalized | classmates, kinship, soulmates, friends, mentor-ideo, mentor-tech, recruiting |
| LOC | financial ties, communication, membership | communication, meetings, shared orgs | meetings, communication, logistical places |
| Knowl. | shared events, collaboration | collaboration | operations, logistical function, training |

*Definition 1:* **Knowledge Sharing Communities (KSC):** Given the Noordin Network and JIEDDO, the **Knowledge Sharing Communities (KSC)** are the intersection of Trust, Lines of Communication (LOC), and Knowledge communities based on the need to disrupt intra-organizational coordination in the Noordin Network.

For each of the three layers, we now present on overview of the obtained data, similar to Tables I-III. The value of $V$ counts the number of nodes that have degree at least 1.

TABLE VIII: Noordin Network topological characteristics by layer, $V$ counts nodes of degree greater than 0 in each layer.

| **Layer** | **V** | **E** | **AD** | **AWD** | **Di** | **De** | **M** | **ACC** | **APL** | **P** |
|---|---|---|---|---|---|---|---|---|---|---|
| Trust | 111 | 544 | 7.53 | 9.80 | 7 | 0.07 | 0.51 | 0.66 | 3.10 | 3 |
| LOC | 121 | 534 | 6.33 | 8.83 | 7 | 0.05 | 0.38 | 0.57 | 2.92 | 1 |
| Knowl. | 106 | 1373 | 22.4 | 25.9 | 5 | 0.21 | 0.41 | 0.79 | 1.93 | 3 |
| Average | 113 | 817 | 12.0 | 14.8 | 6 | 0.11 | 0.43 | 0.89 | 2.65 | 2 |
| Mnplx. | 133 | 2451 | 22.5 | 36.9 | 5 | 0.17 | 0.35 | 0.71 | 2.13 | 1 |

TABLE IX: Boko Haram Network topological characteristics by layer, $V$ counts nodes of degree greater than 0.

| **Layer** | **V** | **E** | **AD** | **AWD** | **Di** | **De** | **M** | **ACC** | **APL** | **P** |
|---|---|---|---|---|---|---|---|---|---|---|
| Trust | 29 | 31 | 2.07 | 2.14 | 4 | 0.07 | 0.56 | 0.26 | 2.28 | 5 |
| LOC | 17 | 34 | 2.47 | 4.00 | 2 | 0.15 | 0.33 | 0.92 | 1.48 | 5 |
| Knowl. | 21 | 34 | 2.95 | 3.24 | 7 | 0.15 | 0.46 | 0.56 | 2.64 | 4 |
| Average | 22 | 33 | 2.50 | 3.13 | 4 | 0.12 | 0.45 | 0.58 | 2.13 | 5 |
| Mnplx. | 105 | 99 | 3.32 | 4.50 | 5 | 0.08 | 0.50 | 0.50 | 2.42 | 6 |

TABLE X: FARC Network topological characteristics by layer, $V$ counts nodes of degree greater than 0 in each layer.

| **Layer** | **V** | **E** | **AD** | **AWD** | **Di** | **De** | **M** | **ACC** | **APL** | **P** |
|---|---|---|---|---|---|---|---|---|---|---|
| Trust | 32 | 28 | 1.68 | 1.75 | 3 | 0.05 | 0.81 | 0.39 | 1.61 | 8 |
| LOC | 130 | 1614 | 23.2 | 24.8 | 6 | 0.18 | 0.51 | 0.93 | 2.28 | 3 |
| Knowl. | 13 | 8 | 1.23 | 1.23 | 2 | 0.10 | 0.78 | 0.00 | 1.27 | 5 |
| Average | 58 | 550 | 8.70 | 9.26 | 4 | 0.11 | 0.70 | 0.44 | 1.72 | 5 |
| Mnplx. | 142 | 1650 | 21.5 | 23.2 | 8 | 0.15 | 0.52 | 0.91 | 2.90 | 1 |

## VI. Results

In Subsection VI-A, we present the metrics of comparing the community detection in each of the three KSC-driven layers with the communities of the monoplex. Subsection VI-B presents a comparison to other possible combinations of the relationships.

### *A. Results for the Knowledge Sharing Communities*

The comparison of the layer community detection to the one of each relationship subnetwork shows fewer communities per layer, closer to the number of communities in the monoplex. Also, the communities in the KSC layers have higher values for ARI, Rand index, Purity and NMI. These values show that the communities identified in each layer better represent the community structure of the monoplex, compared to the communities obtained from each relationship subnetwork. A practical application better validate these communities, and that has been considered in Miller's thesis [25]

TABLE XI: Boko Haram Networkcomparison: layer against monoplex.

| **Layer** | **Comty.** | **ARI** | **Rand** | **Assty.** | **Purity** | **NMI** | **Clr. Adq.** |
|---|---|---|---|---|---|---|---|
| Trust | 83 | 0.5628 | 0.9848 | -0.2709 | 0.784 | 0.9555 | 0.564 |
| LOC | 94 | 0.2000 | 0.9788 | 0.3549 | 0.719 | 0.9297 | 0.332 |
| Knowl. | 89 | 0.3994 | 0.9813 | 0.6270 | 0.806 | 0.9368 | 0.467 |
| Mnplx. | 72 | 1 | 1 | 0.1488 | 1 | 1 | 0.508 |

TABLE XII: FARC Network comparison: layer against monoplex.

| Layer | Comty. | ARI | Rand | Assty. | Purity | NMI | Clr. Adq. |
|---|---|---|---|---|---|---|---|
| Trust | 118 | 0.0453 | 0.8141 | -0.5188 | 0.9860 | 0.5676 | 0.818 |
| LOC | 21 | 0.8812 | 0.9646 | 0.2935 | 0.9860 | 0.8972 | 0.537 |
| Knowl. | 134 | 3.02E-3 | 0.8090 | -0.6000 | 0.9718 | 0.5421 | 0.787 |
| Mnplx. | 9 | 1 | 1 | 0.2987 | 1 | 1 | 0.584 |

TABLE XIII: Noordin Network comparison: layer against monoplex.

| Layer | Comty. | ARI | Rand | Assty. | Purity | NMI | Clr. Adq. |
|---|---|---|---|---|---|---|---|
| Trust | 37 | 0.2817 | 0.8316 | 0.1130 | 0.7842 | 0.5865 | 0.44 |
| LoC | 26 | 0.2990 | 0.8248 | -0.0508 | 0.7194 | 0.5718 | 0.41 |
| Knowl. | 40 | 0.3554 | 0.8432 | 0.0892 | 0.8058 | 0.6007 | 0.58 |
| Mnplx. | 13 | 1 | 1 | 1 | 1 | 1 | 0.388 |

There is still a large number of communities in some layers, as many nodes are still of degree 0 (the count of nonzero degree nodes is captured in Tables VIII - X). Running the community detection on just the connected part of the network would decrease these values, but for consistency with the rest of our analysis, we kept all the nodes in each layer.

We investigated the possible correlation between the performance of the community metrics and cluster adequacy. We found that there is no general pattern identifying that an increase in cluster adequacy influences the ARPN values in the three networks.

### B. Comparing to possible multilayered networks

To demonstrate the benefit of the KSC communities as a goal for optimally aggregating network relationships into layers, we conducted a series of experiments on randomly configured multilayered graphs from the existing attributes.

For Fig. 3 - Fig 5, we create several choices of multilayered networks: a monoplex with 3 layers was created by partitioning at random the relationships into 3 layers, and the average of ARPN for 50 randomly sampled networks is presented as the $y$-value for the $3L-Avg$, which stands for “the average of ARPM of all 50 3-layers networks”. The process is then repeated for partitioning the relationships into four layers ($4L-Avg$) and five layers ($5L-Avg$). The first set of measures are for the KSC layers.

Figure 3 shows that the KSC multilayered network has the highest ARPN values compared to random partition of relationships into three, four or five layers, to emphasis that the grouping of the relationships as proposed in this work is meaningful.

Figure 4 presents the ARPN values (the sum of ARI, Rand Index, Purity, and NMI) by layer rather than multilayered network. The first 9 bars in the bar chart represent the values for the KSC multilayered network. They are compared against the average of the ARPN values of 50 randomly chosen three, four and five-layered networks. The three layers of each 3-layer network are labeled $3L-L1, 3L-L2$ and $3L-L3$, while the four layers of the 4-layered network are labeled $4L-L1, 4L-L2, 4L-L3$ and $4L-L4$, and similarly for the 5-layered network. The values of ARPN over the first layer of each of the 50 networks is averaged and plotted as the $y$-value for $3L-L1$, and them similarly for the 2nd layer of each of the 50 networks being plotted as $3L-L2$, and so on.

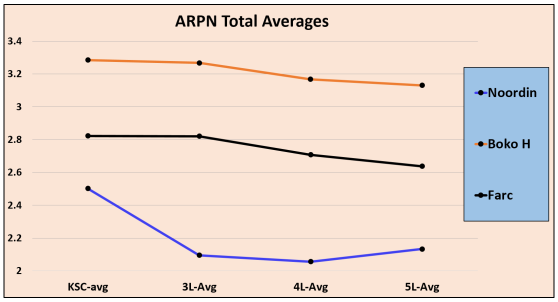


Fig. 3: The network average of the ARPN (the sum of ARI, Rand Index, Purity, and NMI)for each dark network, against the average of 50 randomly selected relationships to created dark multilayered networks with 3, 4 or 5 layers, respectively

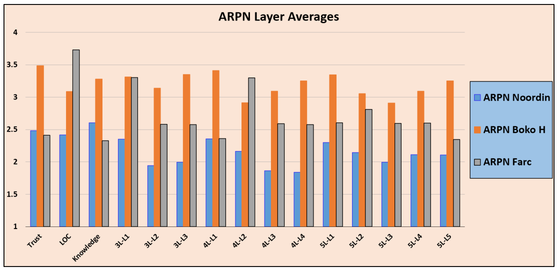


Fig. 4: The layer ARPN (the sum of ARI, Rand Index, Purity, and NMI) for each dark network, against the average of 50 randomly selected relationships to created dark multilayered networks with 3, 4 or 5 layers, respectively

While the KSC layers, namely Trust, LOC and Knowledge, may not be the absolute best compared against all cases, nothing is gained by modeling the network with four or five layers.

This can be seen better in Figure 5 where we present the average for all layers in a 3, 4 or 5-layer network.

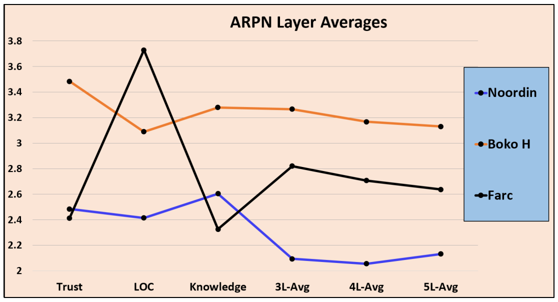


Fig. 5: The layer average of the ARPN (the sum of ARI, Rand Index, Purity, and NMI) for each dark network, against the average of 50 randomly selected relationships to created dark multilayered networks with 3, 4 or 5 layers, respectively

We particularly guide our analysis by Noordin's results, as it is the most established network of the three, for which we have the most information about. We consider Boko and FARC to be less informative, rather presenting an evolving network.

## VII. Conclusion

The current research introduces a methodology that creates multilayered networks from attributed enriched existing terrorist datasets. We based our work on the analysis of three datasets that were publicly available: Noordin Top, Boko Haram, and FARC. Our methodology thus depends on the observed datasets and literature review of the terrorist networks. It may be refined based on analysis of the topology of more terrorist networks or an augmentation of understanding of terrorist behavior.

We first use the subnetwork given by each one of the relationships to create a relationship subnetwork. We compared each relationship subnetwork to the monoplex with respect to standard network metrics and network's community identification. Some relationship subnetwork proved to be useful in identifying a community of interest, such as the community for the Bali 2005 attack. This was identified using the denseness of the subnetwork's community, that captured enough information as described in Section IV. However, identifying all the communities was not possible due to the general lack of information that a single attribute provides.

Based on the existing literature and the reason above, to enhance the information captured, we then group relationships into categories that we call KSC layers. We thus propose a methodology to create 3-layered networks, and created them for each of the datasets in our possession. We present an analysis of these KSC layers, and compare them against other possible multilayered networks that could be created using the relationships that the data was tagged with. This analysis shows that no extra information is gained from building four or five layers.

As future direction, it would be useful to perform a deeper theoretical analysis of each of the three layers, of each of the three networks. The goal would be to see if there are common characteristics, based on which synthetic generation of dark networks could be possible. Also, more validation of the communities with a scenario or on data that is tagged with real communities for this purpose would be desired.

## VIII. Acknowledgements

R. Gera, R. Miller, M. Miranda Lopez and S. Warnke thank the DoD for partially sponsoring the current research. In particular, R. Gera also thanks the Asymmetric Warfare Group and the West point Network Science Center for partially sponsoring this work. This article contains parts of authors' theses [25], [26].

## References

[1] Valdis E Krebs. Mapping networks of terrorist cells. *Connections*, 24(3):43–52, 2002.

[2] Malcolm K Sparrow. The application of network analysis to criminal intelligence: An assessment of the prospects. *Social networks*, 13(3):251–274, 1991.

[3] Sean F Everton. Network topography, key players and terrorist networks. In *annual conference of the Association for the Study of Economics, Religion and Culture in Washington, DC*, 2009.

[4] Sean F Everton. *Disrupting dark networks*. Number 34. Cambridge University Press, 2012.

[5] Michael AP Taylor, Somenahalli VC Sekhar, and Glen M D'Este. Application of accessibility based methods for vulnerability analysis of strategic road networks. *Networks and Spatial Economics*, 6(3-4):267–291, 2006.

[6] Gilles Didier, Christine Brun, and Anaïs Baudot. Identifying communities from multiplex biological networks. *PeerJ*, 3:1–20, 2015.

[7] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):1–13, 2008.

[8] Mark EJ Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582, 2006.

[9] Stephen P Borgatti, Martin G Everett, and Linton C Freeman. *Ucinet for Windows: Software for social network analysis*. Analytic Technologies, 2002.

[10] Günce Keziban Orman and Vincent Labatut. A comparison of community detection algorithms on artificial networks. In *Discovery science*, pages 242–256. Springer, 2009.

[11] Günce Keziban Orman, Vincent Labatut, and Hocine Cherifi. On accuracy of community structure discovery algorithms. *arXiv preprint arXiv:1112.4134*, 2011.

[12] LNF Ana and Anil K Jain. Robust data clustering. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 2, pages II–128. IEEE, 2003.

[13] Alexander Strehl, Joydeep Ghosh, and Raymond Mooney. Impact of similarity measures on web-page clustering. In *Workshop on Artificial Intelligence for Web Search (AAAI 2000)*, pages 58–64, 2000.

[14] A Dean Forbes. Classification-algorithm evaluation: Five performance measures based onconfusion matrices. *Journal of Clinical Monitoring*, 11(3):189–206, 1995.

[15] William M Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association*, 66(336):846–850, 1971.

[16] Lawrence Hubert and Phipps Arabie. Comparing partitions. *Journal of classification*, 2(1):193–218, 1985.

[17] Francesco Ricci, Lior Rokach, and Bracha Shapira. *Introduction to recommender systems handbook*. Springer, 2011.

[18] Andrew Walker. *What is Boko Haram?* US Institute of Peace, Washington, DC, 2012.

[19] Cunningham, Daniel. The Boko Haram Network. [Machine-readable data file]. https://sites.google.com/site/sfeverton18/research/appendix-1, June 2014. Online; Accessed 30 January 2017.

[20] Gabriel Weimann. *Terror on the Internet: The new arena, the new challenges*. US Institute of Peace Press, Washington, DC, 2006.

[21] Dan Cunningham, Sean Everton, Greg Wilson, Carlos Padilla, and Doug Zimmerman. Brokers and key players in the internationalization of the FARC. *Studies in Conflict & Terrorism*, 36(6):477–502, 2013.

[22] Roberts, Nancy and Sean F. Everton. Terrorist Data: Noordin Top Terrorist Network. https://sites.google.com/site/sfeverton18/research/appendix-1, June 2011. Online; Accessed 30 Jan 2017.

[23] S Jones. Terrorism in indonesia: Noordin's networks. Technical Report 114, Asia Report, 2006.

[24] Brad Martin, Thomas Manacapilli, James C Crowley, Joseph Adams, Michael G Shanley, Paul Steinberg, and Dave Stebbins. Assessment of joint improvised explosive device defeat organization (jieddo) training activity. Technical report, DTIC Document, 2013.

[25] Ryan Miller. Purpose-driven communities in multiplex networks: Thresholding user-engaged layer aggregation. Master's Thesis, 2016.

[26] Scott Warnke. Partial information community detection in a multilayered network. Master's Thesis, 2016.