Covert Network Forensics Using Graphical Database Analysis: An Application to the November 13 Paris Attacks

Abstract

Covert network forensics is the problem of inferring the structure of a covert network from heterogeneous and incomplete sources. This paper proposes a system for recording information about covert networks, and then systematically inferring the network structure using graphical database analysis. It applies the method to infer the terrorist network of the recent November 13 attacks in Paris and related plots in Europe by the Islamic State group (IS). The resulting network was found to be a qualitatively different network from the ideologically-related Al-Qaida, having a lower resilience and lower mean degree.

Keywords (6 max): covert network, graphical database, terrorism, Islamic State, ISIS

Introduction

Covert network forensics studies the problem of reconstructing the relationships in a covert network, much like a detective trying to piece together the details of a crime while sifting through a multitude of clues. Since data about a covert network is usually obtained after the network has already carried out its mission, there is a great deal of uncertainty as to which ties and nodes to include[1]. Unlike with non-covert networks, it is not usually possible to determine the structure of a covert network using methods such as analysis of transactional data, name generators and name interpreters. To compound this challenge, open source information about covert networks (including terrorist networks) is often quite unreliable and restricted[2], and to infer their structure we must rely on data from police reports, news stories, and press releases.

Undaunted by these challenges, researchers have mapped the structure of about half-dozen terrorist networks, and this includes the 9/11 network [3] and the 3/11 Madrid bombing network [4]. Studies have also mapped the larger-scale structure of entire covert organizations (such as

[1] Breiger et al.
[2] MAGOUIRK et al. http://www.artisresearch.com/articles/Atran_Connecting_Terrorist_Networks.pdf
[3] Krebs 2002 - Mapping Networks of Terrorist Cells - Connections
[4] "March 11th terrorist network in its weakness lies its strength José A. Rodríguez**

Al-Qaida[5], Jama'a Islamiya[6] and FTP[7]), that included, as subnetworks the operational networks of multiple attacks.

The most common approach to representing covert networks is probably the multi-relational (or multiplex) framework[8]. In this approach, the investigators first identify all members of the organizations and their contacts and then separate relationships by type (e.g. family, friendship, having the same religious teacher, living in the same place, and so on); each relationship type in stored in its own matrix. The covert network is then usually the combination of all relationship types. Another approach to representing covert networks is via a multi-modal graph[9], where the members are identified and then linked to units (a link referring to a participation in an unit, such as an attack cell). To compute the covert network, the multi-modal network is projected to a one-mode network of persons.

Although both approaches are richer than using a simple graph, a lot of information about the network is lost in translation. Because of the clandestine nature of covert networks, we usually do not know the true nature of the interaction between any given two operatives. Instead, most of the available information is indirect and connects the two operations to an activity or site. For instance, we might know that two members of a terrorist organization were involved in the same attack, or were co-present at a location at the same time. The analysis, particularly in the multiplex framework, then takes this information into a social network, resulting in loss of information.

Furthermore, it is very difficult to measure ties in a uniform manner, such as to distinguish friends from mere acquaintances or decide what a certain link actually means for a personal tie. Instead researchers are forced to infer ties from incomplete, uncertain, and sometimes changing information. The covert network that is graphed and characterized masks the multiple coding decisions the researchers have made regarding how the ties between nodes were measured in addition to which nodes were included. To some extent this coding is unavoidable in all network studies, but covert networks pose a special challenge given the indirect nature of the data. Thus, when performing forensics on covert networks, it would be highly desirable to store any information as it is available, i.e. at the highest possible resolution, and to state any coding decisions more explicitly.

To respond to these challenges, this paper proposes a method termed graphical database analysis (GDA). At the basis of GDA is the application of the recently-introduced technology of graphical databases, to represent the information about a covert network. A graphical

---

[5] Analyzing Terrorist Networks: A Case Study of the Global Salafi Jihad Network
Jialun Qin1, Jennifer J. Xu1, Daning Hu1, Marc Sageman 2, and Hsinchun Chen1
http://link.springer.com/chapter/10.1007/11427995_24
[6] Magouirk et al. http://www.artisresearch.com/articles/Atran_Connecting_Terrorist_Networks.pdf
N Roberts, SF Everton - 2011 Strategies for Combating Dark Networks
[7] Gutfraind: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0013448
[8] Alhajj & Rodke: https://books.google.com/books?id=abKXmQEACAAJ
Morselli: http://www.sciencedirect.com/science/article/pii/S0378873306000268
Qin et al. http://link.springer.com/chapter/10.1007/11427995_24
[9] Lindelauf: http://www.sciencedirect.com/science/article/pii/S0378873312000433

database is a knowledge representation system which stores entities and relationships between them, rather than storing tables of rows and columns as found in conventional databases. Thus, a graphical database is very well-suited to store the information about covert networks, including members, activities, events and the relationships among them, as well as the attributes of the entities and relationships (Fig.1).

A key advantage of graphical databases is their efficiency in storing multi-modal and multi-plex information and then rapidly updating or correcting it as new information comes in. While this is in principle possible to do with conventional methods using matrices, doing so is extremely tedious, time-consuming, and prone to error. This is partly the reason why most analysts of covert networks choose to simplify the way ties are represented.

A graphical database is also equipped with a fast and versatile query language for extracting information, analogously to how relational databases are queried using the SQL language. The query language is a key advance of the proposed GDA for covert networks because it allows systematic extraction of information from the available raw data. Using a query, it is possible to determine the members of a covert networks, and the relationships between them by listing only members and relationships that satisfy a particular condition. Furthermore, as new data is added to the database, the query can instantly (i.e., with one command) report the updated structure of the covert network based on any new information. The query also facilitates easy sensitivity analysis to determine whether the covert network is dependent upon the way the data is being coded[10].

While the most advanced network software tools have the ability to store multiple types of nodes and edges as well as their attributes, it is not generally possible in these systems to perform automated extraction of network information based on complex logical criteria. In contrast, the graphical database is well-suited to store and query detailed temporal, spatial and other attributes about the nodes and relationships. For example, a query can list all pairs of persons who were present in the same location within 7 days of a terrorist attack and who also visited a specific country in the past 5 years; another query could list all persons who were involved in a covert act as a known attacker. The distinction is mathematical: a query returns any logical path (a set of one or more linked entities) which satisfies specified logical conditions[11], while existing frameworks are designed for filtering of single nodes and relationships (e.g., all nodes of a particular kind).

Another advantage of the graphical databases, which it shares with the multi-modal framework, is that it assists in a more informative visualization of the network. To visualize the graphical database, we draw all the nodes stored in it (including people, activities, tasks and locations), see Fig.1, and show the relationships of the nodes (distinguishing type). When individuals in a graphical database have a common activity or task, they form a star around it, while in conventional multi-relational social network approaches these individuals are visualized as cliques. As a result, the graphical database tends to have more nodes but a lower

[10] http://www.sciencedirect.com/science/article/pii/S0378873305000511
[11] See e.g. Zouzias: http://alumni.cs.ucr.edu/~mvlachos/pubs/templated.pdf

density of edges, and is thus easier to study visually. Indeed, as we show in the results, the graphical database representation of the IS-E network is more legible and provides more information (activities) than the multi-relational network visualization [citation needed to visual. of multi-modal nets?].

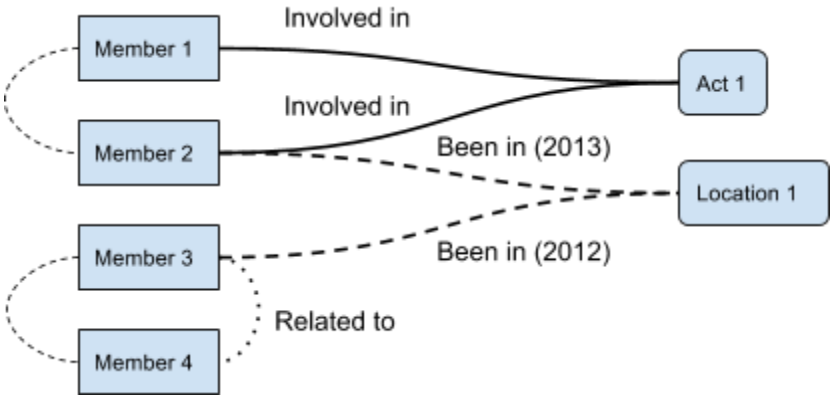

Fig.1. Example of a graphical database representation of a four-node covert network, with inferred ties (arcs on the left side). Tie 1-2 (from member 1 to member 2) is inferred because of their involvement in the same act. Members 2 and 3 were in the same location, but at a different year, so they are not tied. Tie 3-4 is inferred because members 3 and 4 have a familial tie to each other. The graphical database often stores multiple attributes which cannot be visualized here, but which may be important for covert network forensics.

To summarize, GDA provides the following advantages over existing frameworks. First, the data is represented in a raw form in a graphical database, obviating the need to make coding decisions. Second, the query language provides a way of systematically extracting structural information, whereas existing methods are limited to simple filtering and projecting. Third, the combination of more data and better queries also enables a more comprehensive set of sensitivity analyses. Finally, as compared to the multiplex framework, GDA enables better visualization of the network by using a more sparse representation.

We apply the GDA method to a major recent terrorist event. On Nov. 13, 2015 a set of coordinated attacks in Paris, France took the lives of 130 people and caused more than 350 injuries. The attack is considered one of the worst terrorist atrocities in France, the deadliest terrorist attack in Europe since the Madrid bombings and one of the deadliest attacks in an OECD country since the 9/11/2001 attacks in the US[12]. The Paris attacks have been carried out by the Islamic State (IS), or more specifically by a group of French-speaking radicals that might be organized as a unit of IS known as Katibat al-Battar al-Libi. [13] The network, which we called IS-Europe (or IS-E) has been operating in Western Europe for several years, and could be linked to three additional attacks prior to the Paris attacks: the attack on the Jewish

[12] Global Terrorism Database: http://www.start.umd.edu/gtd/
[13] http://atimes.com/2015/11/paris-made-in-libya-not-syria/

Museum of Belgium (May 2014), the interdicted plot in Verviers (January 2015)[14], and the high-speed train attack on the Brussels-Paris line (August 2015)[15]. However, IS-E does not include the Charlie Hebdo attacks, which are attributed to an unrelated group affiliated with Al-Qaida[16] nor does it include acts by self-starting Islamists.

Because of the extensive interest in the Paris attacks and related plots, considerable information is available about the IS-E network in the open-source media. Based on newspaper reports following the attacks, a graphical database was used to report on IS-E members, attacks, terrorist activities, sites, locations, and countries (see Methods). Following the creation of the database, we inferentially built a social network between the human agents and analyzed the IS-E network. Specifically, we aimed to provide a visual summary of the overall plot, and to understand the structure of the social network that made the attack possible.

We evaluated the IS-E network using measures like edge density, transitivity, diameter, and several specialized measures. Using GDA, we then perform a sensitivity analysis to evaluate how the measurements depend on how the ties are inferred. Finally, we compared the IS-E network to other organizational networks. We found that IS-E has lower mean degreat as compared to Al-Qaida, as well as, intriguingly lower secrecy (defined in[17]).

Our survey of the social network and covert network literature has found that the graphical database method is substantially novel, with only a handful of papers using the graphical database technology for network analysis, largely by engineers and computer scientists, and no published studies applied it to covert networks and forensics. The related method of multi-modal network analysis has been proposed by Breiger et al.[18] and Lindelauf et al.[19], who argued that terrorist networks could be represented as a multi-modal networks linking individuals based on their affiliation to missions or targets. Xu and Chen[20], Everton et al.[21] and other studies used a closely related method of multi-relational networks, common in social network analysis. In the multi-relational network, the network allows for different types of relationships (family, friendship, recruiter, etc.) but all the nodes in the network are of the same type. This is distinguished from a graphical database framework proposed here in which the nodes can be of several classes, both human agents and non-human nodes (locations, events, attacks). Moreover, our framework allows the relationship to an activity to contain

---

[14] http://www.telegraph.co.uk/news/worldnews/europe/greece/11353214/Greek-police-detain-suspected-ringleader-of-Belgian-terror-cell-says-source.html

[15] http://www.nytimes.com/interactive/2015/11/15/world/europe/manhunt-for-paris-attackers.html?_r=0

[16] NYT: "Al Qaeda Trained Suspect in Paris Terror Attack, Official Says" http://www.nytimes.com/2015/01/09/world/europe/paris-terror-attack-suspects.html

[17] Lindelauf: http://www.sciencedirect.com/science/article/pii/S0378873309000021

[18] Breiger et al. http://www.sciencedirect.com/science/article/pii/S0378873313000300

[19] Lindelauf: http://www.sciencedirect.com/science/article/pii/S0378873312000433

[20] Xu, Chen: http://dl.acm.org/citation.cfm?id=1400198

[21] Roberts, Everton: http://calhoun.nps.edu/handle/10945/41260

information such as time (e.g. years of attending a training camp), level of involvement etc, which is not possible in the multi-relational network framework.

Thus, our main contribution is to propose a new methodology for systematic representation and analysis of covert networks, and secondarily, to map the IS-E/IS network in Europe. A practical outcome of this work for counter-terrorism is to better understand the structure and any vulnerabilities of terrorist networks, assisting future identification of leaders and other counter-terrorism interdiction activities.

Methods

Representation of covert network data in a graphical database

Graphical Database Analysis can be performed using any modern graphical database software. In our analysis we used the free and open source Neo4j database and its CYPHER query language[22].
The following classes of nodes were used:

1. Human agents with the attributes age, gender, citizenship and status (free, wanted or dead)
2. Attack sites (including successful, partially-successful and interdicted plots)
3. Activities related to covert missions or terrorism (in/exfiltrations, weapon preparation, etc.)
4. Activities not related to covert missions or terrorism (e.g. sports clubs[23])
5. Sites (residence, safe houses and staging sites)
6. Localities (larger areas such as communities commonly reported as places of residence)
7. Countries (reported as bases of operation and training)

Attack sites were only included if they were carried out by IS (or IS-inspired operatives) and were also orchestrated by IS-E leadership (A. Abaaoud). IS-E is responsible for several attacks before Nov.13, but not the attacks on Charlie Hebdo and others. Although IS inspired and orchestrated other attacks in Europe, it would be misleading to study them together with the IS-E attacks because there is little evidence of any operational connections, and indeed many IS attacks were by self-radicalized individuals rather than formal IS members.

Network Boundary

For the IS-E network, we used the following criteria for inclusion:

1. Attackers (i.e. used/planned to use a weapon on/at a particular target)
2. Known to have been involved with an attacker/operative in a covert/terrorism-related task (as organizers or in logistical roles)
3. Wanted or arrested in relation to the attacks for unspecified reasons

[22] Version: 2.2.5 © 2015 Neo Technology, Inc. San Mateo, CA
[23] What Genkin and Gutfraind called "neutral magnets" in mobilization

Thus, for IS-E, any individual that assisted a wanted or dead member of IS-E is included, but not individuals who played no known role (such as innocent family members).

In general, a graphical database provide the facilities for a robust determination of measurement in a covert operation. First, all broadly relevant persons might be entered and then logical inclusion criteria would be formulated based on any information that might be available. The criteria would be formulated as a logical query on the database yielding a consistent and automatic determination of network membership.

Relationships

The database contains person-entity relationships, as well as a minority of person-person relations. Covert network forensics is often reliant on information, which belongs to several fairly well-characterized types:

1. Attacked: An individual that attacked a particular site
2. Involved-In: An individual who participated in a covert mission or task
3. Linked-To: A pair of linked individuals, whenever possible noting whether the tie is family, friendship or, if sources exist, a relationship formed after recruitment to the covert network[24].
4. Present-In: An individual connected to a location for covert activity of unspecified nature
5. Affiliated-With: An individual connected with a non-covert group or activity
6. Been-In: An individual connected to any places where he or she is known to have resided in, stayed in (temporarily) or countries visited (if relevant to the covert mission).

When recording the relationship in the graphical database, we recoded the specific source that reported it. In the case of IS-E, most of the sources were large circulation media sources in English or French.

For the IS-E network, two individuals were tied to each other in the social network on one of several grounds: if they were involved in an attack on the same target, jointly involved in a covert-related activity, were present in the same location for a covert activity. A pre-existing strong tie (familiar, friendship or other) was also grounds for a tie, as long as both members were included as members of IS-E (see above criteria).

Sensitivity Analysis

We also considered several variants of the network:

1. IS-E: the most probable network

[24] While this methodology avoids inference of person-to-person in a covert network ties until a later stage, such relationships are sometimes reported in credible sources. In this situation the tie was coded as Linked-To relationship.

2. IS-E limited: this network only has ties if persons were involved in attacks or covert activities. Sharing of space is not considered, thus it excludes ties of members who e.g. briefly shared a safe house together.
3. IS-E expanded: expanded the IS-E network by tying two members whenever they were at the same locality (for IS-E members, it was often Molenbeck) or affiliated with a non-covert activity.

The graphical database was used to store the sources of knowledge. Specifically, each node and edge has a "ref1" attribute (sometimes ref2 and so on) which indicates the source (news article, press release, and so forth). From the graphical database, we generated projections which were used for structural network analysis. All networks were studied using the open source R/igraph library[25].

The network was compared structurally based on several metrics that were previously used to characterize covert networks, particularly terrorist networks[26]. Our basic metrics are: density (number of edges divided by the number of edges in a complete network), transitivity (number of actual triads divided by the possible) and the diameter - a measure of the communication time across the network. We also considered efficiency, a measure of its ability to perform terrorist operations as computed from the harmonic mean of person-to-person distances[27] and secrecy S1[28], a measure of its ability to survive capture of individual operatives[29]. Both efficiency and secrecy range in [0,1] with 1 considered the maximal possible for a network. Terrorist networks tend to be highly modular (organized into cells)[30], which enables them to operate covertly and avoid accidental exposure. To measure the modularity of the metrics, we used a modularity score originally proposed by Newman and Girvan[31] and calculated using the approach of Blondel et al.[32] This modularity measure identifies the optimal separation of the network into communities (including their number) and computes a modularity score in [0,1] which reflects how strongly is the network separable (1 being the highest).

---

[25](Version 1.0.1 © igraph team http://igraph.org/r/

[26] Krebs: https://www.aclu.org/sites/default/files/field_document/ACLURM002810.pdf
Memon [why are using a guy who was retracted]:
https://books.google.com/books?hl=en&lr=&id=swm1tfH_vykC&oi=fnd&pg=PA1&dq=Nasrullah+Memon&ots=JmhbI0FOW7&sig=s3t7YkHwT9XAbcP_gjLxsi7fbQA#v=onepage&q=Nasrullah%20Memon&f=false
Lindelauf: http://www.sciencedirect.com/science/article/pii/S0378873309000021
Helfstein: http://jcr.sagepub.com/content/55/5/785

[27] Gutfraind: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0013448

[28]
Lindelauf: http://www.sciencedirect.com/science/article/pii/S0378873309000021

[29] Morselli: http://www.sciencedirect.com/science/article/pii/S0378873306000268

[30]Sageman:
https://books.google.com/books?hl=en&lr=&id=SAQ8Oa6zWF4C&oi=fnd&pg=PR7&dq=terrorist+cells+social+networks&ots=qUA0yP9iTk&sig=pFjo6-EeZOzHlH7ZdT9erRcuLcs#v=onepage&q=terrorist%20cells%20social%20networks&f=false

[31] Finding and evaluating community structure in networks MEJ Newman, M Girvan - Physical review E, 2004

[32] Fast unfolding of communities in large networks
VD Blondel, JL Guillaume, R Lambiotte... - Journal of Statistical ..., 2008 - iopscience.iop.org

Sources of data

Data from media sources was collected from Nov. 17, 2015 to Jan. 3, 2016 by reading large-circulation media sources, particularly French and English-language daily newspapers and news channels. Full list of source of the articles are included in an online website complementary to this article[33]. In a few cases, some of the details published in the first week after the attack were contradicted by later reports, and in these cases the later reports were viewed as more credible.

Comparison of terrorist networks

We compared the IS-E network to several datasets from past covert networks, including terrorist and underground resistance networks (Table 1). A particularly significant empirical reference are networks affiliated by Al-Qaida, which has ideological similarity to IS but is also in competition with IS. Included are several plots conducted by Al-Qaida in the previous decade including the 9/11 attacks in the US and the March 11 attack in Madrid.

Table 1: Network datasets which were compared to IS-E and the Nov.13 operation network

Network	**Dates Active**	**Type**	**Movement**	**Source**
9/11 attackers and assistants	2001	Terrorist, Operational	Islamist radicals	Krebs[34]
Strasbourg Cathedral plot	2000	Terrorist, Operational	Islamist radicals	Xu & Chen[35]
Bali Bombings network	2002	Terrorist, Operational	Islamist radicals	Atran[36]
Madrid bombers (M11)	2003	Terrorist, Operational	Islamist radicals	Rodrigues[37]
Indonesian Jama'a Islamiya (JI)	2000–2005	Terrorist, Operational	Islamist radicals	Magouirk et al.,Roberts/Everton[38]

[33] https://github.com/sashagutfraind/Nov13
[34] https://www.aclu.org/sites/default/files/field_document/ACLURM002810.pdf
[35] Xu, Chen: http://dl.acm.org/citation.cfm?id=1400198
[36] Scott Atran http://doitapps.jjay.cuny.edu/jjatt/data.php
[37] Rodriguez J. The march 11th terrorist network: In its weakness lies its strength. Working Papers EPP-LEA. Barcelona: University of Barcelona
[38] Magouirk et al. http://www.artisresearch.com/articles/Atran_Connecting_Terrorist_Networks.pdf
Roberts, Everton: http://calhoun.nps.edu/handle/10945/41260

Jakarta hotel bombings	2009	Terrorist, Operational	Islamist radicals	Atran[39]
Al-Qaida global	2001	Terrorist, Organizational	Islamist radicals	Xu & Chen[40]
Francs-tireurs et Partisans (FTP)	1941-1945	Urban insurgency, Organizational	World-war II Resistance	Gutfraind[41]
Revolutionary Org. November 17 (17N)	1975-2002	Terrorist, Organizational	Marxist	Atran[42]

Results

Applying the GDA method to the IS-E network resulted in a database of 81 nodes and 141 relationships.

Our first analysis visualizes the core of the IS-E network, by showing only the most important nodes and the relationships between them. For this visualization, we remove sites, localities and countries, since they weakly connect a large number of nodes, and show just the members, attack sites and covert/non-covert activities (Fig.2).

[39] Scott Atran http://doitapps.jjay.cuny.edu/jjatt/data.php
[40] Xu, Chen: http://dl.acm.org/citation.cfm?id=1400198
[41] Gutfraind: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0013448
[42] Atran: http://doitapps.jjay.cuny.edu/jjatt/history.php

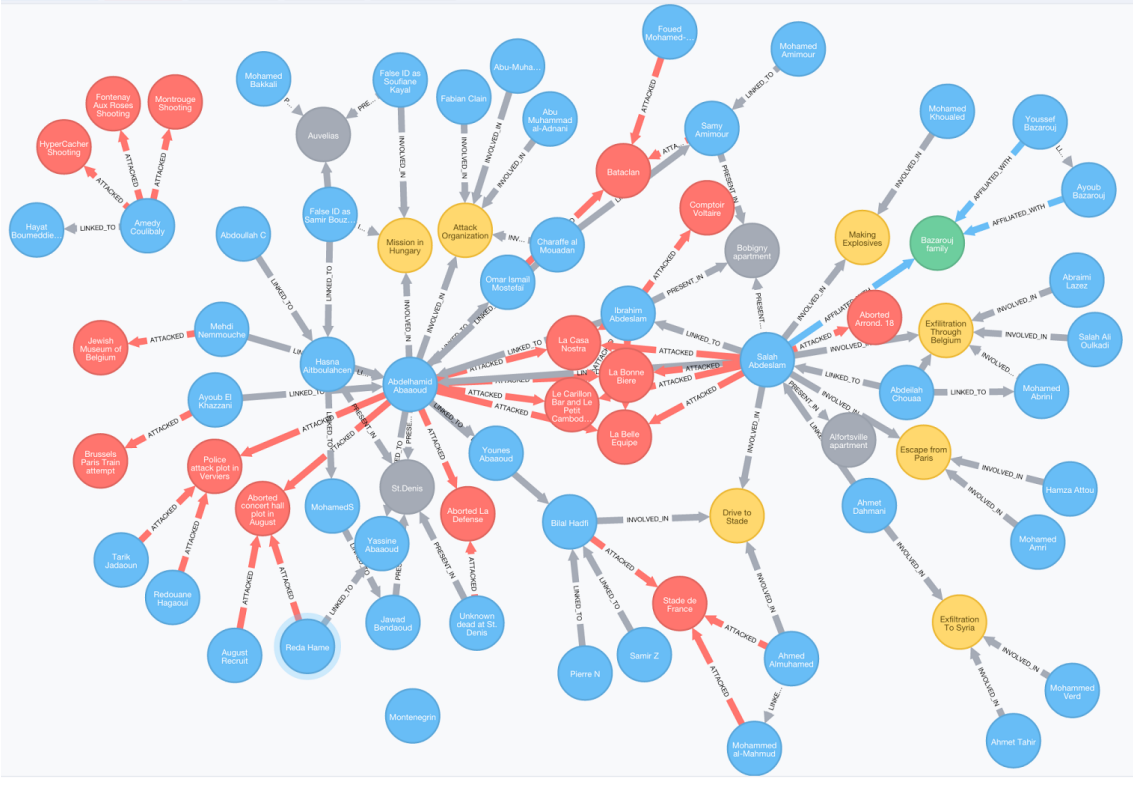

Fig.2: The nodes and relationship in the IS-E network in the graphical database representation (excluding countries): targets (red), terror activities (yellow), agents (blue), sites (gray) and social foci (green).

A striking feature of the network is the high centrality of the A. Abaaoud node, who was one of the perpetrators of the Nov. 13 attacks, as well as having links to other attacks in 2014 and 2015. The node of S. Abdeslam also has an important role and has a high degree.

All the relationships in the database are shown in Fig.3. Notice that now many of the actors have multiple paths to each other.

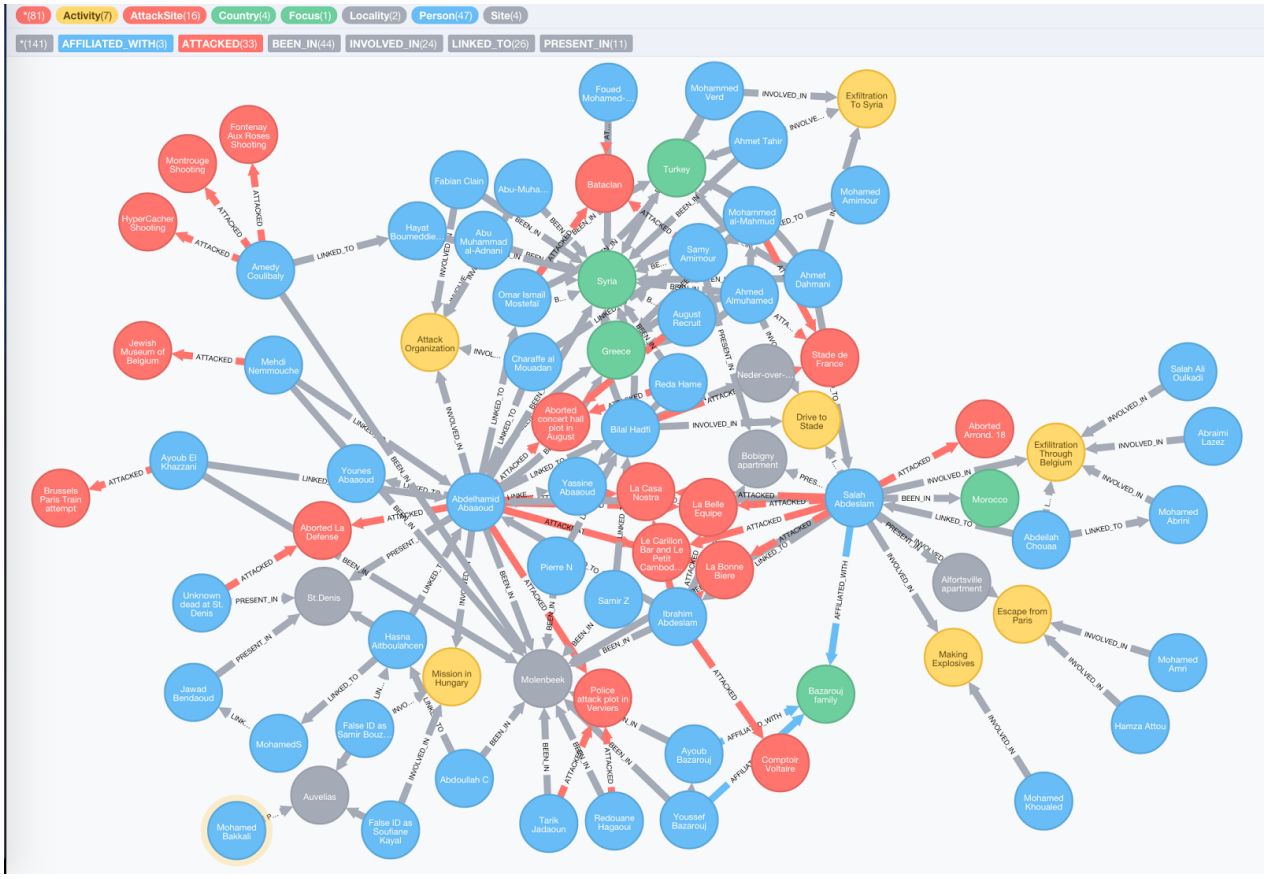

Fig.3: The nodes and relationship in the IS-E network in the graphical database representation: targets (red), terror activities (yellow), agents (blue), sites/localities/countries (gray) and social foci (green).

Applying the link inference procedure (see Methods), we obtained a covert social network with 47 individuals and 71 undirected relationships. The network plot reveals the cells involved in the attacks, which appear as cliques (Fig.4). Each of the attacks cells (Stade de France, Café and Bataclan) are clearly distinguished. Other cells are smuggling cell of S. Abdeslam, the cell responsible for exfiltration of Ahmt Dahmani, the group around St. Denis and finally the IS leadership grouping, which is connected to A. Abaaoud. Abaaoud and S. Abdeslam are clearly vital for connecting the network.

Fig.4: Plot of the IS-E actor network. The network nodes represents all the individuals involved in the covert organization, and the edges link members that interact with each other in the course of performing their missions. Additional nodes (7), not shown, consisting of one dyad and 5 singleton are not connected to the main component.

The degree of a node, i.e. the number of its connection in the network, can highlight its relative importance to the network [43]. Examining the degrees of the nodes in IS-E (Fig.5A) it could be seen that A. Abaaoud, the has the highest degree followed by S. Abdeslam. The same nodes also have the highest betweenness centrality (Fig.5B).
Indeed, Abdelhamid Abaaoud is believed to be a leader of IS-E and he was responsible for multiple terror attacks in Europe while S. Abdeslam is known as one of the co-leaders, possibly with an additional person[44]. People loosely tied to the plot, have the lowest degree. Thus, in IS-E, the degree of the node appears to be a reliable indicator of leadership role - a finding that may assist future counter-terrorism investigations.

[43] Krebs: https://www.aclu.org/sites/default/files/field_document/ACLURM002810.pdf
[44] NYT7: "Cellphone Contacts in Paris Attacks Suggest Foreign Coordination" http://www.nytimes.com/2015/12/31/world/europe/cellphone-contacts-in-paris-attacks-suggest-foreign-coordination.html?_r=0

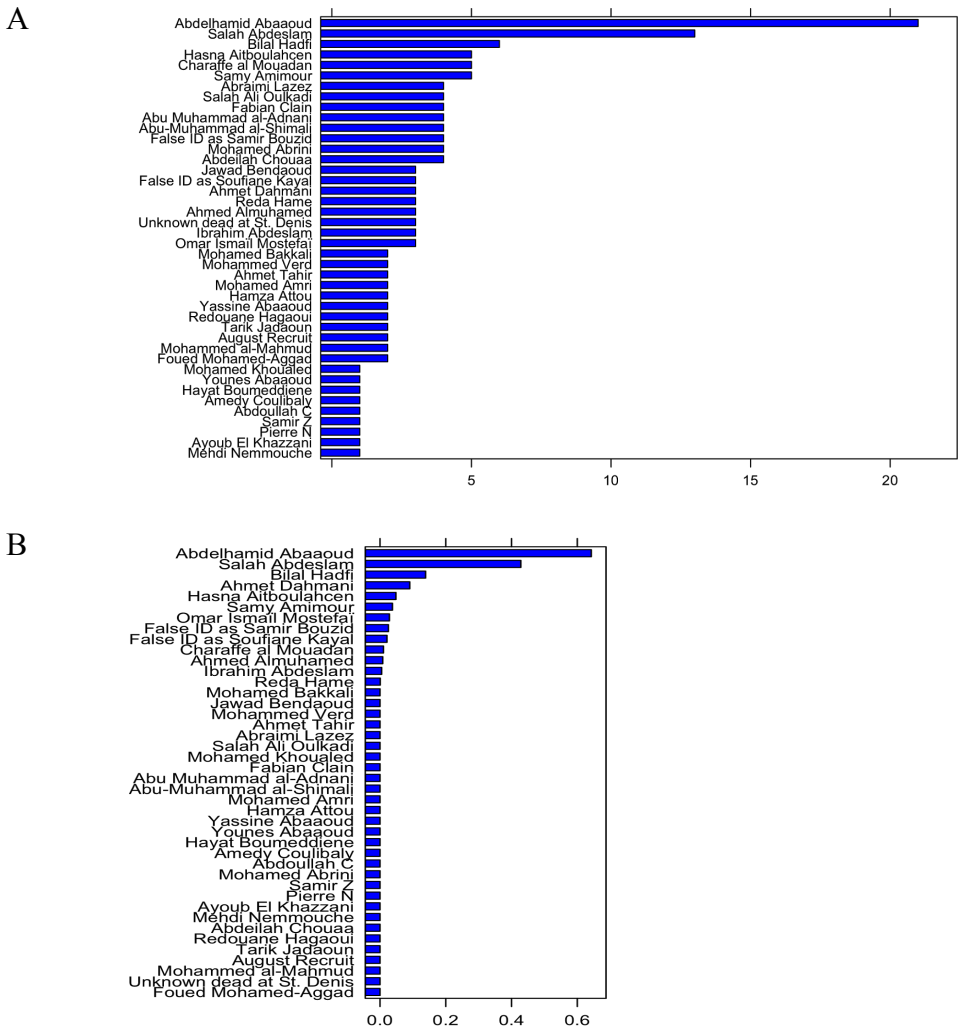

Fig.5: Node centrality of the IS-E agent network. (A) Node degrees and (B) Betweenness centrality.

The agents have a mean age of 27 [13, 39]. The gender distribution is 95% male (40 male and 2 female).

Structural network analysis and Comparison to other covert networks

Table 1 shows the structural properties of the IS-E network and compares IS-E to other networks in our dataset.

Table 1. Terrorist Networks Properties of reconstructed covert networks

Organization	Nodes	Edges	Comp't s	Diameter	Edge density	Mean degree	Transitivity	Efficiency	Secrecy	Modula-rity
IS-E	47	71	7	5	0.07	3.02	0.27	0.32	0.91	0.56
17N	18	46	1	2	0.30	4.11	0.50	0.65	0.66	0.21
Al-Qaida	368	1481	5	7	0.02	8.05	0.16	0.33	0.98	0.56
FTP	174	264	1	9	0.02	3.03	0.41	0.20	0.98	0.81
JI	79	623	2	14	0.20	15.77	0.56	0.56	0.79	0.33
IS-E extended*	47	141	4	5	0.13	6.00	0.65	0.45	0.85	0.43
IS-E restricted**	47	63	9	5	0.06	2.68	0.27	0.29	0.92	0.57

*IS-E limited: includes only ties based on known terrorist activity, attacks or known links.
**IS-E extended: includes ties also based on sites and locality of residence or non-terrorist affiliation.

From Table 1 observe that IS-E is the smallest network, and it has a relatively low mean degree as compared to all other organizations. It also have a low secrecy, as compared to Al-Qaida and FTP, but higher than the secrecy in JI. The transitivity of IS-E is lower than FTP and JI and modularity is intermediate.

Although IS-E is the best estimate of the true network, the GDA allows us to evaluate the sensitivity of the results by considering alternative tie-inferring assumptions. We considered two alternative assumptions with corresponding variant networks: (1) the only real ties are those around terrorist activities or known links (IS-E restricted), and (2) many ties were unreported and individuals should be tied when they shared site, location or a non-terrorist affiliation (IS-E extended). Under this evaluation, the relative low degree and low secrecy of the IS-E network as compared to Al-Qaida is robust (Table 1). However, the transitivity increases substantially, which implies that this measure might be dependent on the tie inference assumptions.

Focusing the Nov. 13 attacks , we found that subnetwork of the IS-E network responsible directly for these attacks contains ¾ of the original network. We also compared the Nov. 13 network to operational networks from past attacks by Al-Qaida (Table 2). In this comparison, the Nov. 13 network was found to have a relatively low mean degree (but not the lowest), low efficiency and high modularity. The secrecy and transitivity is comparable to Al-Qaida operational networks. From the lower mean degree figure, we hypothesize that IS-E might

use smaller cells as compared to Al-Qaida. The Nov.13 has the highest modularity except for the Jakarta 2009 attacks by Jama'a Islamiya, which also has a lower mean degree.

Table 2: Comparison of operational networks

Operations	Nodes	Edges	Comp'ts	Diameter	Edge density	Mean degree	Transitivity	Efficiency	Secrecy	Modularity
Nov13	37	59	6	5	0.09	3.19	0.38	0.34	0.89	0.54
9/11	62	152	1	5	0.08	4.90	0.36	0.40	0.90	0.53
Bali 2002	17	158	1	3	0.45	11.7	0.72	0.72	0.53	0.22
Jakarta 2009	28	38	3	4	0.10	2.71	0.36	0.22	0.87	0.61
Madrid M11	70	240	7	6	0.10	6.86	0.57	0.37	0.89	0.46
Strasbourg	18	26	2	5	0.17	2.89	0.13	0.45	0.78	0.38

Discussion

Our study proposed a method termed GDA to document and infer clandestine networks based on data drawn from large-circulation open media sources accessible through the Internet. The GDA method was applied to represent the network of the Islamic State group in Europe (IS-E).

For IS-E, it was found that IS-E leaders at the operational level could be identified by their much higher degree and betweenness centrality, which is consistent with past studies on Al-Qaida[45]. The entire IS-E network was then compared to previously published networks on the 2001 attacks in the US and the 2005 attacks in Spain, and others. Examination of the IS-E network suggested that its network could be characterized by relatively low mean degree and low secrecy, as compared to networks such as Al-Qaida. Taken together, IS-E could be characterized by relatively small cells of about three persons, which are relatively loosely, even opportunistically, knit together.

Even with more time and more data, there are basic limitations to the process of reconstructing covert network structure when relying on exclusively public media sources. It is likely that some ties between the individuals might not be revealed for security reasons until many years after the attack. However, those ties that were reported in the media are likely to be real. There are also particular difficulties in the case of the IS-E because its commanders are based in Syria, and hence some of its higher echelons are not characterized in detail

[45] Krebs: https://www.aclu.org/sites/default/files/field_document/ACLURM002810.pdf

(although they were reported and included in our database). Similarly, because IS-E is a unit of IS, members of IS might be involved in IS-E operations. Nevertheless, because information about arrests and arrest warrants is public, the public record has good coverage of IS-E operations surrounding the recent attacks.

The comparison of terrorist networks requires that they are coded using the same framework but that is not the case today. Indeed, there are differences in how various authors define membership and relationships in these networks. To some extent, we overcome this problem using the sensitivity analysis procedure.

Conclusion

This study proposes to use graphical database analysis to obtain a high-resolution map of covert or terrorist networks based on public media sources, and perform structural inference. The analysis method was applied to study the Islamic State network responsible for the Nov.13 2015 attacks in Paris and related plots. Our study is one of the first to study the structure of a network operated by the Islamic Stage, and to compare it to covert networks from the past. We found that the network, as compared to Al-Qaida operations appears to have lower secrecy, which may make it relatively more vulnerable to interdiction.

An important direction for future research is to standardize the coding of clandestine and terrorist networks, possibly using a graphical database, to enable more consistent comparison of their structure. This promises to highlight universal patterns in the organization of these networks. Graphical database analysis offers an efficient way to do so.

Acknowledgements

AG thanks Uptake Technologies, Inc. for supporting this research.