Covert Network Forensics Using Graphical Database Sensitivity 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 sensitivity 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 have lower mean degree and secrecy (resilience) than Al-Qaida, making it a qualitatively different network from the ideologically-related Al-Qaida.

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. 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[1], 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 [2] and the 3/11 Madrid bombing network[3]. Studies have also mapped the larger-scale structure of entire covert organizations (such as

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

Al-Qaida[4], Jama'a Islamiya[5] and FTP[6]), 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[7]. 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[8], 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.

Unfortunately, both methods suffer a certain loss in translation. The use of limited information creates a logical leap when non-uniform data is transferred into a multi-relational or multi-modal network (although both approaches are superior to using a simple graph). 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, but we usually don't know if they interacted in the past, in what ways, and for how long.

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 has 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 sensitivity analysis (GDSA). At the basis of GDSA is the application of the recently-introduced

---

[4] 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

[5] Magouirk et al. http://www.artisresearch.com/articles/Atran_Connecting_Terrorist_Networks.pdf
N Roberts, SF Everton - 2011 Strategies for Combating Dark Networks

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

[7] 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

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

technology of graphical databases, to represent the information about a covert network. A graphical 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 (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 called CYPHER for extracting information, analogously to how table databases are queried using the SQL language. This query language is used to reconstruct the structure of the covert network by listing only relationships that satisfy a particular condition. For example, a query can list all pairs of persons who were present in the same location within 7 days of a terrorist attack, or were involved in the same covert act or terrorist attack.

The query language is a key advance of the proposed GDSA for covert networks: while some network software tools have the ability to store multiple types of nodes and edges, it is not generally possible in these systems to perform automated extraction of network information based on complex criteria. In contrast, the graphical database is well-suited to store and query detailed temporal, spatial and other attributes about the nodes and relationships. Furthermore, as new data is added to the database, the query can instantly report the updated structure of the covert network based on the 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[9].

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 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?].

[9] http://www.sciencedirect.com/science/article/pii/S0378873305000511

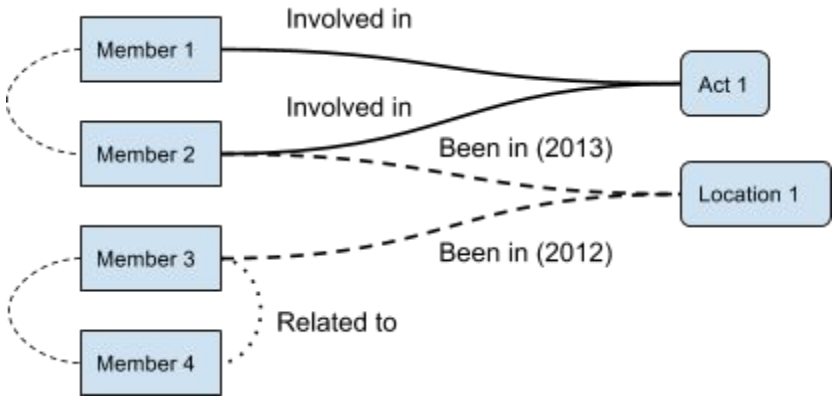

Fig.1. Example of a graphical database representation of a four-node covert network (boxes and solid lines) with inferred ties (arcs on the left). Tie from member 1 to member 2 is inferred because of their involvement in the same act. Members 2 and 3 were in the same location, but at a different year, so they are not tied. Ties 3-4 are 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 may be important for covert network forensics.

We apply the GDSA 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[10]. 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. [11] 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 Museum of Belgium (May 2014), the interdicted plot in Verviers (January 2015)[12], and the high-speed train attack on the Brussels-Paris line (August 2015)[13]. However, IS-E does not include the Charlie Hebdo attacks, which are attributed to an unrelated group affiliated with Al-Qaida[14] nor does it include 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

[10] Global Terrorism Database: http://www.start.umd.edu/gtd/
[11] http://atimes.com/2015/11/paris-made-in-libya-not-syria/
[12] http://www.telegraph.co.uk/news/worldnews/europe/greece/11353214/Greek-police-detain-suspected-ringleader-of-Belgian-terror-cell-says-source.html
[13] http://www.nytimes.com/interactive/2015/11/15/world/europe/manhunt-for-paris-attackers.html?_r=0
[14] 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

human agents and analyzed the IS-E network. Specifically, we aim to provide a visual summary of the overall plot, and 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 GDSA, we then perform a sensitivity analysis to evaluate how the measurements depend on how the ties are inputted. Finally, we compared the IS-E network to past operations by radical Islamist groups against OECD countries (cf. Krebs 2002, Rodriguez 2004, Gutfraind 2010, Sageman 2011). We found that IS-E has lower structural secrecy (defined in[15]) as compared to Al-Qaida, that contributed to some of its failures. If the group learns to employ more resilient network structures, the danger from IS would grow.

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. Lindelauf et al.[16] argued that terrorist networks could be represented as a bimodal network linking individuals based on their affiliation to missions or targets. Xu and Chen[17], Everton et al.[18] 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). In effect, each activity in our framework corresponds to a distinct layer of the standard framework. Moreover, our framework allows the relationship to an activity to contain 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 Sensitivity Analysis can be performed using any modern graphical database software. In our analysis we used the free and open source Neo4j database[19].

---

[15] Lindelauf: http://www.sciencedirect.com/science/article/pii/S0378873309000021
[16] Lindelauf: http://www.sciencedirect.com/science/article/pii/S0378873312000433
[17] Xu, Chen: http://dl.acm.org/citation.cfm?id=1400198
[18] Roberts, Everton: http://calhoun.nps.edu/handle/10945/41260
[19] Version: 2.2.5 © 2015 Neo Technology, Inc. San Mateo, CA

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 only partially-successful and interdicted plots)
3. Activities related to terrorism (in/exfiltrations, weapon preparation, etc.)
4. Activities not related to terrorism (e.g. INSERT)
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

Because public media sources name multiple individuals, not all of them are involved in a covert plot. In general, most of the named in the media should be added to the database, and subsequently it should be decide whether they should or should not be listed as members of the covert network based on the weight of evidence, assisted by the graphical database.

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 in a terrorism-related task (as organizers or in logistical roles)
3. Wanted or arrested in relation to the attacks for unspecified reasons

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).

Relationship types

Relationships in the database were drawn based on information directly reported in media sources, based on the following types:

1. Attacked: An individual that attacked a particular site
2. Involved-In: An individual who participated in a terror mission or task
3. Linked-To: A pair of socially linked individuals, whenever possible noting whether the tie is family, friendship or acquaintance
4. Present-In: An individual connected to a location for terror-related activity of unspecified nature
5. Affiliated-With: An individual connected with a non-terror 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 terror mission).

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, involved in a common terror-related activity, were present in the same terror-related location. A pre-existing strong tie (familiar, friendship or other) was also grounds for a tie, as long as both members were members of IS-E (see above criteria).

Sensitivtiy Analsyis

Using sensitivity analysis, we also considered two alternatives:

1. IS-E limited: this network only has ties if persons were involved in attacks or terror-related activities. Sharing of space is not considered, thus it excludes ties of members who e.g. briefly shared a safe house together.
2. 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-terror 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[20].

The network was compared structurally based on several metrics that were previously used to characterize covert networks[21]. 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, based on the average distance[22] and secrecy S1[23], a measure of its ability to survive capture of individual operatives, which often conflicts with efficiency[24]. 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)[25], which enables them to operate covertly and avoid accidental

---

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

[21] 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

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

[23]
Lindelauf: http://www.sciencedirect.com/science/article/pii/S0378873309000021

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

[25]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

exposure. To measure the modularity of the metrics, we used a modularity score originally proposed by Newman and Girvan[26] and calculated using the approach of Blondel et al.[27] 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).

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[28]. 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 NETS). 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: The 9/11 attacks in the US and the March 11 attack in Madrid.

Table NETS: Network datasets which were compared to IS-E

Network	**Dates Active**	**Type**	**Movement**	**Source**
9/11 attackers and assistants	2001	Terrorist	Islamist radicals	Krebs[29]
Strasbourg Cathedral plot	2000	Terrorist	Islamist radicals	Xu & Chen[30]
Madrid bombers (M11)	2003	Terrorist	Islamist radicals	Rodrigues[31]

[26] Finding and evaluating community structure in networks MEJ Newman, M Girvan - Physical review E, 2004
[27] Fast unfolding of communities in large networks
VD Blondel, JL Guillaume, R Lambiotte… - Journal of Statistical …, 2008 - iopscience.iop.org
[28] https://github.com/sashagutfraind/Nov13
[29] https://www.aclu.org/sites/default/files/field_document/ACLURM002810.pdf
[30] Xu, Chen: http://dl.acm.org/citation.cfm?id=1400198
[31] Rodriguez J. The march 11th terrorist network: In its weakness lies its strength. Working Papers EPP-LEA. Barcelona: University of Barcelona

Indonesian Jama'a Islamiya (JI)	2000–2005	Terrorist	Islamist radicals	Magouirk et al.,Roberts/Everton[32]
Al-Qaida global organization	2001	Terrorist	Islamist radicals	Xu & Chen[33]
Francs-tireurs et Partisans (FTP)	1941-1945	Urban insurgency	French resistance	Gutfraind[34]

Results

The code of the IS-E network was visualized by retrieving from the graphical database a subset containing just the attackers and sites is in Fig.2.

[32] Magouirk et al. http://www.artisresearch.com/articles/Atran_Connecting_Terrorist_Networks.pdf Roberts, Everton: http://calhoun.nps.edu/handle/10945/41260
[33] Xu, Chen: http://dl.acm.org/citation.cfm?id=1400198
[34] Gutfraind: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0013448

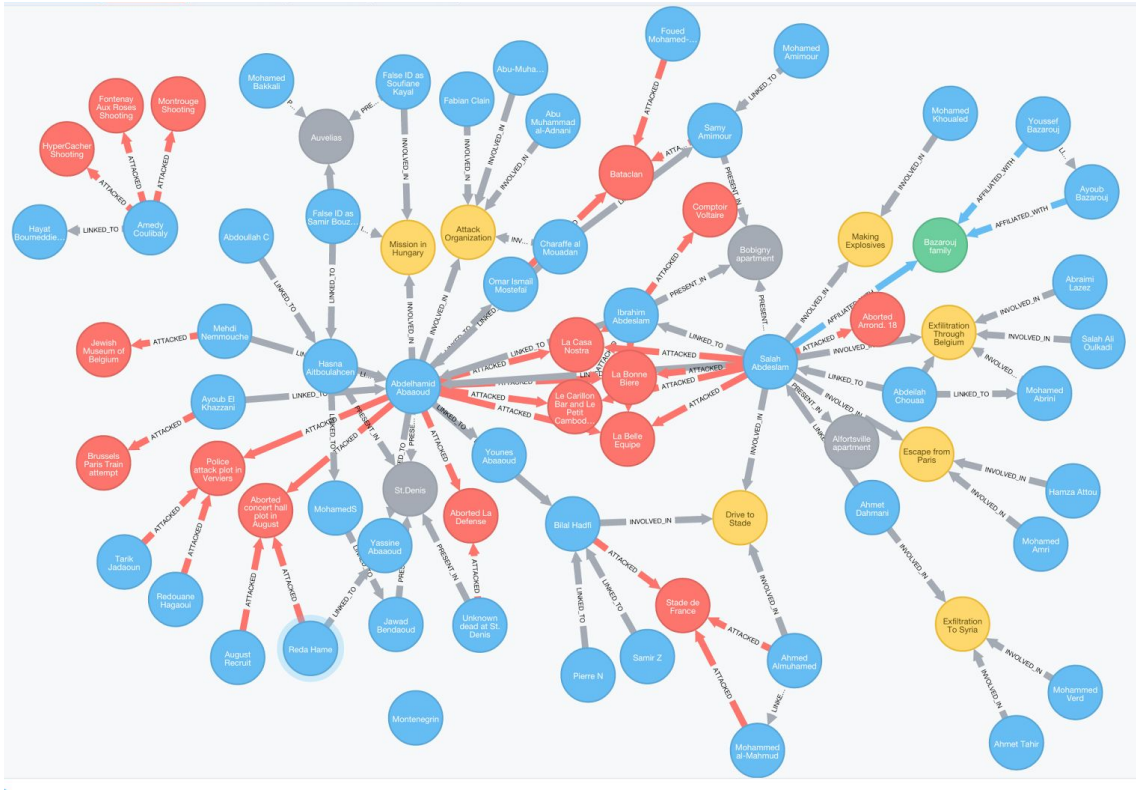

Fig.2: The nodes and relationship in the IS-E network in the graphical database representation (excluding countries): targets (red), terror activities (yellow), agents (blue), sites (gray) and social foci (green).

A. Abaaoud was involved in the Nov. 13 attacks and has known links to perpetrators of the high-speed train attack, the attack on the Jewish Museum of Belgium and the attacks in Verviers.

The full database has 81 nodes and 141 relationships (Fig.3). Many of the actors have multiple paths to each other.

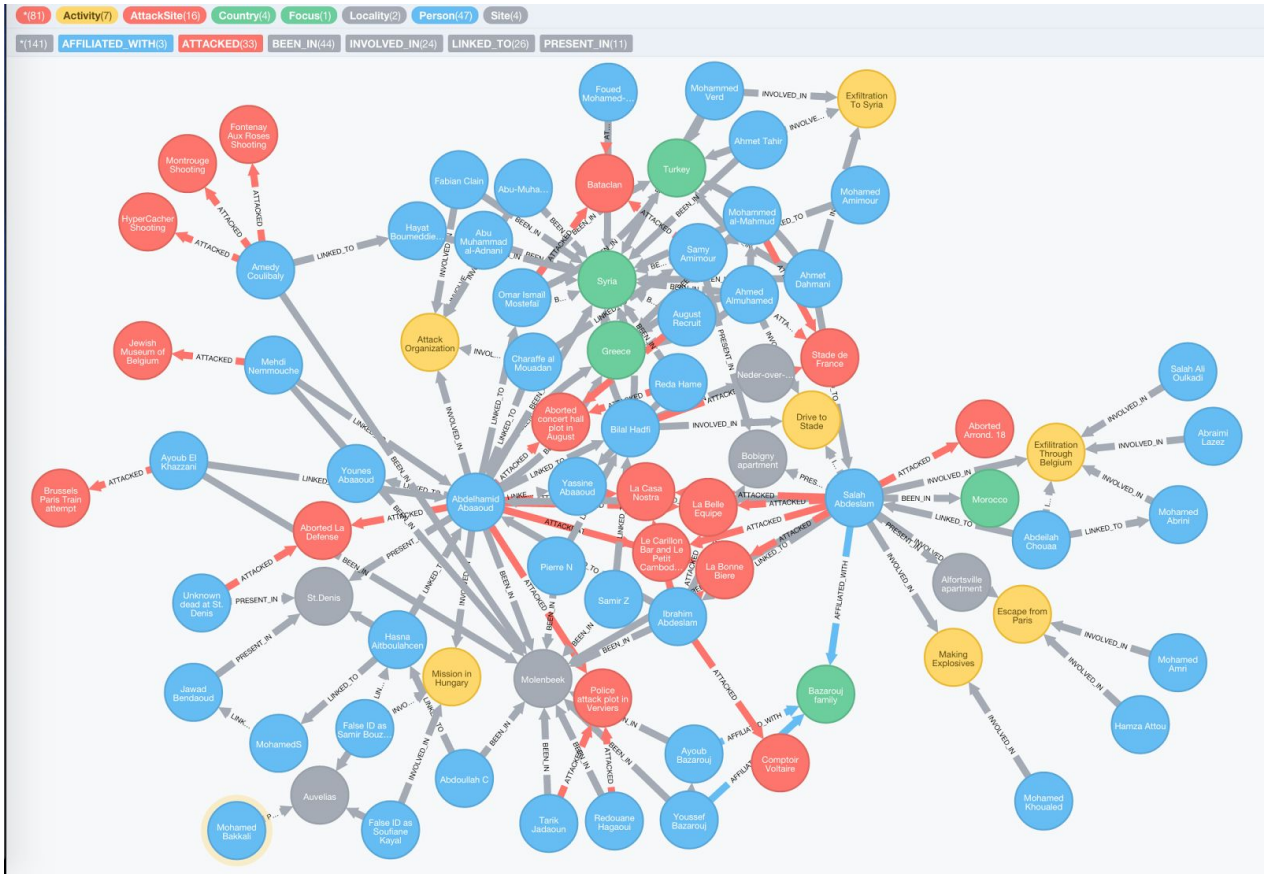

Fig.3: The nodes and relationship in the IS-E network in the graphical database representation: targets (red), terror activities (yellow), agents (blue), sites/localities/countries (gray) and social foci (green).

Applying the link inference procedure (see Methods), we obtained a social network with 47 individuals and 71 undirected relationships (Fig.4). The plot reveals the cells involved in the attacks. 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 [35]. 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[36]. 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.

[35] Krebs: https://www.aclu.org/sites/default/files/field_document/ACLURM002810.pdf
[36] 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

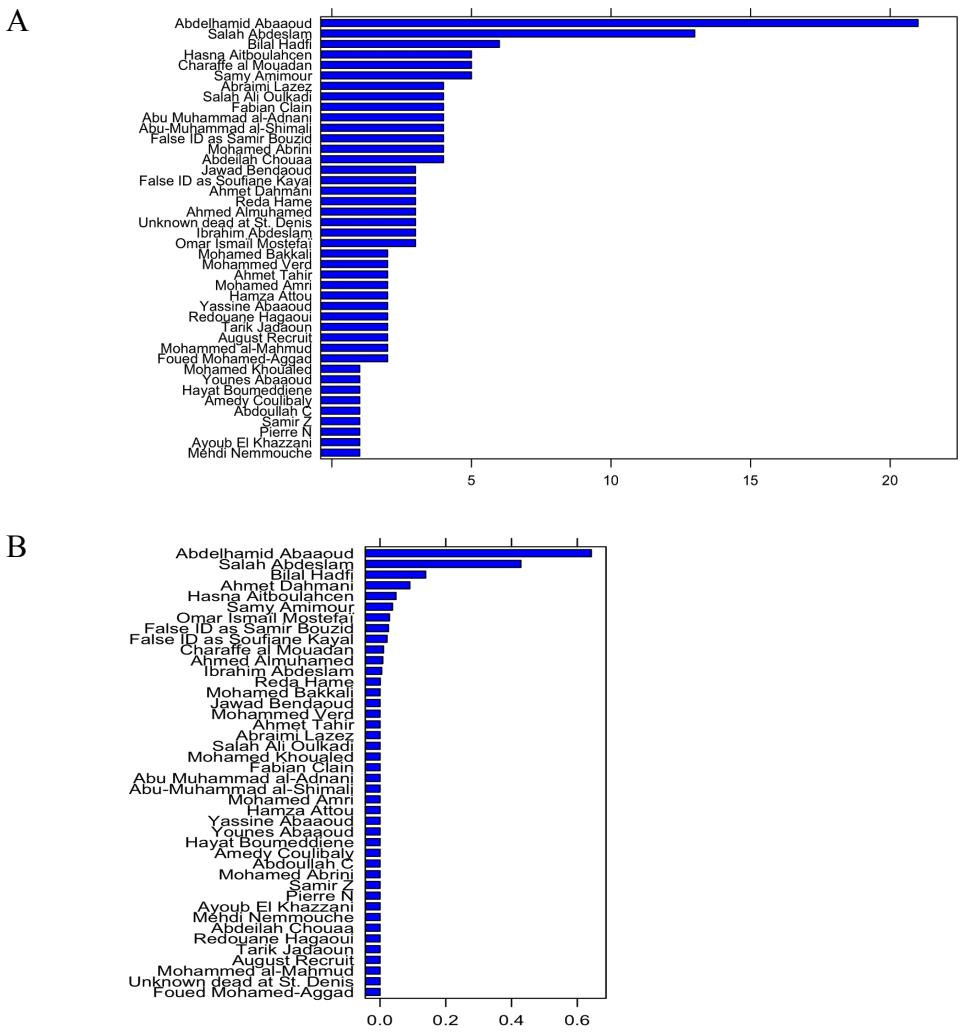

Fig.5: The degree distribution (A) and betweenness centrality (B) of the social network.

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
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. Taken together, IS-E is a relatively loosely-knit grouping.

Although IS-E is the best estimate of the true network, the GDSA 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 contained 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, low efficiency and high modularity. The secrecy and transitivity is comparable to Al-Qaida operational networks. We hypothesize that IS-E might use smaller cells as compared to Al-Qaida.

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
Strasbourg	18	26	2	5	0.17	2.89	0.13	0.45	0.78	0.38
9/11	62	152	1	5	0.08	4.90	0.36	0.40	0.90	0.53
Madrid M11	70	240	7	6	0.10	6.86	0.57	0.37	0.89	0.46

Discussion

Our study proposed a method termed GDSA to document and infer clandestine networks based on data drawn from large-circulation open media sources accessible through the Internet. The GDSA 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[37]. 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.

Even with more time and more data, there are basic limitations to the process of reconstructing network data about terror plots 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, but 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 it is based in Syria, and hence many of its members are not publically known. Similarly, because IS-E is a unit of IS, members of IS might be involved in IS-E operations. Nevertheless, because information about 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. Ultimately, the only reliable approach would be to introduce a consistent coding methodology, perhaps along the lines we propose in this study.

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

Conclusion

This study proposes to use graphical database sensitivity analysis to obtain a high-resolution map of covert or terrorist networks based on public media sources, and perform link 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 attempt to a comparative study of terrorist networks. We found that IS-E, 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 sensitivity analysis offers an efficient way to do so.

Acknowledgements

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