1-1-2010

Bootstrapping a Terrorist Network

Robert D. Duval
West Virginia University, Bob.Duval@mail.wvu.edu

Kyle Christensen
Columbus State University, christensen_kyle@colstate.edu

Arian Spahiu
West Virginia University, aspahiu@mix.wvu.edu

Bootstrapping a Terrorist Network
A paper presented to the Political Networks Conference
May 2010

Robert D Duval, West Virginia University
Kyle Christensen, Columbus State University
and Arian Spahiu, West Virginia University

The advent of methods and tools for network analysis has brought about a paradigm shift in the way we examine data concerning terrorists, terrorist organizations and violent non-state actors in general. Relational data such as individual contacts, group membership, financial transactions, telephone calls, and web site links all provide information about social structure for the study of terrorism, political extremism, and violence. And indeed, the examination of terrorist networks has emerged as a prominent field of study in recent years (Sageman 2004; 2008; Duval and Christensen, 2010).

Recent work in Social Network Analysis (SNA) has shown that it provides the analyst with the means to describe and visualize communities of violent actors as networks. We can use SNA to infer both properties of the network as a whole, as well as internal characteristics that would not be apparent simply by examining the data with tallies, spreadsheets, or typical quantitative methods. A number of notable relationships and structures among the actors that have been studied have emerged from the complex ways in which they are linked. Most importantly, network analysis has produced significant knowledge while examining a limited class of data that simply identifies some form of contact, relationship or linkage between the subjects of interest.

Network analysis has emerged to prominence for several reasons. First, the increasing reliance on computer science for large scale data collection, processing, and mining leads readily to the adoption of the network model. The same terminology and methods that ascertain where bottlenecks exist in a local area computer network serve to identify and describe the central actors and vulnerabilities in complex human networks. Social network analysis, while originally finding occasional use in sociology, increasingly appears in a broader range of academic social science disciplines as a state-of-the-art methodology. The power of new visualization methods, data collection strategies, and innovative means to conjoin attribute data with the relational structures in the networks leads us to a powerful tool set for research, and ultimately tactical and policy decisions.

But perhaps network analysis has arrived in intelligence analysis primarily because of recent successes. Network analysis, or in its investigatory guise, link analysis, played a significant role in the capture of Saddam Hussein (Reed 2006), and the prosecution of the "Virginia Jihad" (Surface 2007). Network analysis was instrumental in focusing the hunt for Saddam Hussein by tracing the linkages of his ties to family members, Sunni tribal loyalists, and former protégés in Iraq. And in a notable direct success, the "Virginia Jihad Network" was investigated and prosecuted by link analysis based on the network of relationships among the 11 defendants. Ali al-Timimi's central role was revealed by the analysis of communications traffic. Recent congressional testimony highlighting the link between the "Virginia Jihad

Network" and Lashkar-e-Tayyiba training camps in Pakistan believed to have trained the Mumbai terrorists provides further indication of its current importance in counterterrorism (Van Duyn 2009).

The use of social network analysis (SNA) to examine the structure of terrorist organizations has gained significant credence in the intelligence community in recent years. At the same time, SNA has begun to emerge as a mainstream research method in political science and international relations (Maoz, Kuperman, et al. 2006; Maoz, Terris, et al. 2008; Hafner-Burton and Montgomery 2006). The collection of the relational data for such analyses can, however, be subject to significant constraints, especially in the case of terrorist groups. We may have data on contacts between known and/or suspected terrorists, but we know we likely do not have complete information. And given we are using tools that do not routinely let us assign confidence to our assertions, this makes our use of SNA largely descriptive rather than inferential. This paper is the initial examination of an innovative method for making inferences about networks even in the face of likely incomplete data.

Using Steven Koschade's (2006) data on Jemaah Islamiyah's attack on the Bali nightclub in 2002, and Valdis Krebs' (2002) data on the 9/11 hijackers, we investigate the robustness of their findings using resampling or bootstrapping methods. From this we seek to establish confidence intervals for a form of analysis that is routinely presented without statistical criteria. This research thus seeks to move intelligence analysis of networks from the largely descriptive to the inferential realm. Should our approach be found to have merit, it will be a valuable innovation for the use of SNA in intelligence work, as well as an important addition to our ability to test hypotheses on networks across many fields of research.

What We Have Learned from Network Analysis

Social Network Analysis provide a mechanism for the analysis and display of structural variables, as measured by contacts or connections between two or more actors, and composition variables, which are based on actor specific traits (Wasserman and Faust 1994, 29; Scott 2000, 2-3). These variables comprise the actors, and their attributes, and ties between the actors, which make up the social network. These are also referred to as nodes and edges in mathematical graph theory.

The primary application of SNA to the study of violent actors or terrorist groups has been to identify their social structure, ascertaining which actors are central, and who belongs to what "cliques" or subgroups (Reed 2007). These data are then used to generate a variety of actor and graph specific statistics. Similarly, these data may also be used to identify and examine sub-group characteristics as well. An analyst may choose to represent multiple aspects of social structure quickly and effectively using tools such as multi-dimensional scaling or clustering techniques to classify groups of actors, potentially identifying cells and less compliant groups within larger networks or movements (Sageman 2004). This affords the analyst a powerful means for visualization and discovery of new information. Such information is at the heart of battlefield tactical intelligence and Reed (2007) makes a strong case that the incorporation of network analysis into the Intelligence Preparation for Battle is a crucial step for modern warfare.

As the focus of our analysis is on terrorist or violent extremist networks, it is useful to note that there are many types (Williams 2008), and that network analysis has been used to reveal significant findings about a number of them: terrorist organizations (Sageman 2004), militias (Zhou et al. 2005), warlords and cartels (Milward and Raab 2006), insurgencies (Reed 2006), gangs (Xu and Chen 2003), and organized crime (Ferrara et al. 2008). And these actors relate to each other through contacts, meetings, activities, training, financial transactions, authority relationships, and many other structural relationships. All represent social systems that have close kinship, tribal, or regional proximity affiliations that foster the growth of network structures. In addition, their violent nature means that they are usually clandestine, and thus must trade organizational effectiveness for operational security (Clark 2005). This results in the reliance on pre-existing social networks, and constrains both operational detection on the part of state actors, and recruitment on the part of the organizations themselves.

Central to current analysis on terrorist groups is the ability to ascertain organizational structure and identify avenues of recruitment. Sageman's (2004) seminal work provided one of the first glimpses into the organizational structure of Al Qaeda identifying several substructures, or organizational cliques: the Central Staff, the Southeast Asians, the Maghreb Arabs, and the Core Arabs. Also from his limited network analysis, we came to understand recruitment to Al Qaeda as a more complex, and less demographically determined process. The realization that many members were middle class and married flew in the face of Western presuppositions. As we increase our ethnographic base for analyses such as these, our understanding of the components of recruitment and organizational structure will grow as well (Renzi 2006; Conway and McInerney 2008).

One of the more visible groups studying network analysis and a range of violent actors is the Dark Web Portal at the University of Arizona (Chen et al. 2004). This large-scale open source project has produced a large body of work on the use of the Internet by extremist groups and terrorist organizations, both domestic and international. The Dark Web project points to the utility of network analysis in ascertaining community structures or groupings. Analysis of linkages between sites has produced network clustering that matches substantive expertise with substantial understanding of the topic. When the organization structure of movements can be modeled on data that can be updated readily and frequently, we can move beyond from our previous sole reliance on post hoc analysis and move to real time monitoring. The Dark Web project has done extensive examination of extremists on the Web, with network analyses focusing on network extraction, group identification and clustering, subgroup detection and the discovery of network interaction patterns (Chen et al. 2003).

Network analysis of violent actors has been conducted over many diverse types of networks. Reed's analysis of trust networks involved immediate and extended family ties as well as friendships and bodyguards, while strategy and goals mapping utilized financial transactions and insurgency operations (Reed 2006). These mappings led to the inference that members of the Iraqi insurgency built upon their trust ties, leading ultimately to US operations that resulted in Saddam's capture.

Networks are also of interest because of their novel methods for engaging calls for action. Recent attention has been focused on swarming behavior, where networks of resisters or protesters are mobilized on very short time scales via mobile phones and other electronic

communications devices. The "Battle in Seattle" protest at the WTO in 1999 was a precursor to swarming as a tactic in social protests (Ronfeldt and Arquilla 2001). Such swarming behaviors and the rise of mini-blogs such as Twitter make this an important area of SNA that needs to expand and be monitored for extremist and terrorist utilization.

Increasingly, network analysis suggests that network reach is a key factor. The "Small World" or "Six Degrees of Separation" descriptions of the networked society may be too large to be of utility for network detection. Initial research suggests that two to three steps is all that is required to uncover the essential network structure (Travers and Milgram 1969; Surface 2007).

The ability to examine network structure and ascertain the characteristics of the nodes leads to tactical considerations. If we know what nodes are most central, we may be able to disrupt network operations through removal or interdiction of the node. Likewise when we see nodes that are closer to many other nodes, they represent targets for influence or misinformation approach. Quite simply, knowledge of network characteristics gives us information about tactical options.

Networks, as diffuse structures, grow unpredictably, and heal after damage (Cunningham 2003). Alternate communication routes are established through existing sets of actors, and new actors are recruited to provide additional pathways, not merely as replacements for nodes that have been removed. Networks are resilient due to the differentiation which emerges from confrontation and the need to adapt to operational conditions (Raab and Milward 2003). The robustness of network structures lies in the interconnectedness of the nodes. Rigidly hierarchical structures are vulnerable; diffuse networks are robust. Hierarchical structures have specialized paths that represent the only communications channels between the nodes, providing a formalized chain of command. In networks, additional connections provide redundancy through alternate paths, control is diffuse, and information channels are robust. The result is that terrorist groups have evolved networks rather than organized military organizations because their flexible command structures make them hardier than conventional force structures. These networks are organizationally resilient and adapt both organizational structure and operational tactics based upon their environment (Milward and Raab 2006).

As a result, SNA becomes invaluable because it identifies who is important, and where network vulnerabilities lie. Selected properties of networks become of great interest when we assemble the structures from the available relational data. We are interested in properties such as centrality, with particular interest in "closeness" and "betweenness." Nodes with greater closeness are connected to more individuals. They are rapid disseminators of information and other communications. Nodes with greater "betweenness" are the linkages between subgroups that otherwise may have little contact with each other. They are essential to communicate with different areas of the network and act as boundary spanners bringing ideas and innovation to both subgroups to whom they connect. Often these two assessments result in different tactics and different targeting strategies. In fact, research on disrupting networks suggests that while removal of emerging leaders is a desirable second-best strategy, random node removal is the best strategy for network destabilization (Carley et al. 2001).

Research indicates, however, that estimating networks and their properties is plagued with uncertainty (Butts 2003;Robins et al 2004; Tsvetovat and Karley 2007; Kossinets 2008).

Particular concern has been given to the estimation of networks in the face of missing data. And since social networks are created as a result of interactions between participants, the effects of missing data, or unobserved respondents in a network limits our ability to estimate the network. For example, in order to assess the impact of missing data in a social network, Kossinets (2008) performs a sensitivity analysis to account for this problem. There are three important problems associated with missing data in a social network: boundary specification, survey non-response, and fixed choice design. With sensitivity analysis, Kossinet assessed which source of the missing data problems—network boundary specification, survey non-response, or fixed choice design--weighs more prominently in the network estimates. His results show that network boundary specification and fixed choice designs are two of the problems that more prominently alter the estimates of the network.

Some scholars have developed new ways to increasing confidence intervals and decreasing error measurements in network analysis. For example, in addressing the problems associated with the technique of snowball sampling, Tsvetovat and Carley (2007), develop a new probabilistic sampling technique —the Simulated Annealing technique--as a better way of accounting for the hidden nodes within a social network, in their study, a covert terrorist organization. With this new sampling technique every participant that has been accounted for in the network has an equal chance of being sampled; not only the ones with high levels of SNA metrics. This sampling algorithm eliminates participants (or nodes) that do not communicate a lot, therefore reducing the likelihood of mining unnecessary data, and reducing node bias within a network.

In a similar study, but with a different research question, Butts (2003) utilizes a Bayesian method of inference to measure the informant's accuracy and the network structure in the presence of measurement error and missing data. With the Bayesian method, Butts (104-105) addresses four measurement problems associated with social network analysis: the extent of error in the existing data, the mechanism by which error is produced, finding new ways in producing higher quality data, and in developing new ways in accounting for the lack of data in a social network, which this paper attempts to do by using posterior sampling techniques. All of these problems are important to social network analysis.

Additionally, Robins et al (2004) develop an exponential random graph in order to account for the problem of non-responding network members in a social network. By modeling respondents and non-respondents as separate nodes in a social network structure, and by treating the nodes of non-respondents separately, they find that they could develop estimates for non-respondents similar to the estimates developed for the respondents in the social structure. Using random graph models, Robins et al. (2004) analyze different structural effects within a social network and find that comparisons between the two models with different nodes may contribute to the predictive capacity of a social network model.

Similarly, and most relevant for this paper, Borgatti, Carley, and Krackhardt (2006) analyze the robustness of measures of centrality on networks with missing data. The authors use varying sizes of networks ranging from very sparse to very dense and subject them to error tests commonly associated with missing data in a social network. Their findings suggest that networks that are denser were the most robust in the face of errors associated with the missing data in a social network. The most occurring errors associated with missing data in a social

network are: node deletion, edge deletion, node addition and edge addition (Borgatti, Carley and Krackhardt 2006).

Bootstrapping a network

The range of methodological issues confronting the analysis gives rise to the need to broaden our capacity to use network analysis with an inferential tool. If we know that networks are sensitive to missing data, or subject to the incorporation of too much unnecessary information from snowball sampling, then we find ourselves in the position of needing more robust estimation methods with means of making statistical inferences. Our use of classical hypothesis tests for most quantitative research has become ubiquitous, while SNA seems to avoid the topic entirely. And most critically, we need methods to make these tests of inference. Current SNA methods are largely descriptive models, with no attempt to produce statistical inference about important hypotheses involving the network. Currently, we seek to ascertain the most central actor(s) in a network, but not attempt to determine if they are significantly more central than other actors.

The reason for this is rather simple: tests of inference about network parameters have seen little development, and in the limited instances where they exist, they have not been implemented in the SNA software that is currently available. When we perform tests of inference in standard statistical analysis, we rely on the wealth of theoretical development of sampling distributions of various parameters of interest. The normal and t-distributions are so ubiquitous, that we seldom stop to consider that our means of inference rest upon a well developed body of statistical theory about the parameters in which we are interested.

Social networks have little such statistical foundation with which to appeal. While we can calculate various centrality scores for each node or actor in the network, we have little way to ascertain if that most central actor is really significantly "closer", or "more between" than the second most central node (or any other!). And in the analysis of terrorist groups based upon collections of interactions among actors under surveillance, distinguishing the most critical actors or nodes is a highly desirable goal. To be able to add statistical confidence should add as much value to SNA as it does to the analysis of sample means. If we must base an intelligence, or counterintelligence, decision based upon our limited knowledge of a network, then to be able to place statistical confidence limits about the relative importance of the actors in the network should have substantial value. While we confine ourselves to the analysis of terrorist networks here, with particular attention to the identification of the most central actors, the relevance of tests of inference to the broader analysis of more theoretical issues is apparent.

We will therefore apply bootstrapping to this question because bootstrapping as a paradigm seems quite appropriate to the key concerns we have here:

- There are no known theoretical sampling distributions for many of the parameters we use in SNA, and bootstrapping can provide a variety of confidence intervals when parametric tests are absent..
- Bootstrapping, by allowing the data to provide the empirical density function of the population in question, will ameliorate the impact of extraneous isolates that might result for over-sampling nodes due to snowball sampling.

- It will provide tests of inference that let us conclude the relative importance of actors in the face of potentially limited data, much like the conventional bootstrapping of statistics is most applicable when the samples are too small to let theoretically assumed sampling distributions apply. Missing network linkages are like unsampled data points. If they are random, they generally serve to weaken inference tests.

Bootstrapping a network presents a few theoretical issues to articulate. The bootstrapping paradigm assumes that the best estimate of the theoretical density function is the empirical density function provided by the data. When there is no theoretical density function (e.g. there is no known test of inference for a statistic), then the utility of bootstrapping becomes significant as an alternative means of providing a test of inference. For an example of this use of the bootstrap in examining confidence intervals about the difference in two sample medians, see Mooney and Duval (1993).

It should be pointed out that bootstrapping does assume that the data generating process is produces an independent and identically distributed set of data. While there is some cause to reflect on whether a network sample/resample is **iid**, there are a few considerations that suggest we continue to explore this application in any event:

- Bootstrapping is likely to be robust even in the face of this assumption. This is a topic for further research.
- Most data analysis we do assumes that our data are **iid**, and they almost never are. Common examples would include the z-test and the t-test.
- The resample with replacement moves the resample to greater independence.
- The inherent "thinning" of the network due to sampling with replacement and dropping duplicate values likely reduces statistical power, and makes our tests more conservative.

This question bears much more investigation, but at this point, we take the position that a bootstrapped confidence interval is likely to be a conservative estimate of inference, in that we are more likely to committee Type II errors than type I errors. And a low power, or even biased tool may be better than no tool at all.

Bootstrapping a social network requires that we treat the existing network as a sample, and that we perform resamples from the network. This is the customary manner for taking bootstrap resamples, but it requires a preliminary decision on the nature of the sample, and it has an unusual effect for Social Network Analysis.

Unlike conventional bootstrapped estimation, the resample data can be of two different types. We could resample either nodes or edges (links). Given the limitations of this study, we choose to evaluate only link resampling. Our assumption for intelligence analysis is that the network data collected is a series of contacts made by an initial actor, or set of actors and our data collection is based upon ascertaining who these individuals talked to. The data network is developed by collecting links: hence we will resample links. Resampling nodes will prove an equally important task for evaluation, but will not be examined here. The Krebs 9/11 data provided here was indeed collected by starting with two individuals and expanding the network as links became know. Therefore our resampling frame replicates the stochastic component of the data we are examining.

Since we will often be examining an undirected network, based on a set of interactions between the members of the network, the resampling methodology guarantees that each resampled network will be different and have lower density, for some links (edges in graph theory) or actors (nodes in graph theory) will be omitted while others will be selected more than once. If we are doing an unweighted graph, as the simplest case, then the duplicate edges are superfluous, and are discarded. Hence the network resamples will of necessity be lower density than the original. In resampling nodes, the omission of any given node in the resample will also likely produce a graph or network of lower density. Should a node have very few or no links then it may also occasionally be omitted from the resample, even when we are resampling links, and thus reduce the size of the network as well.

Since the network density will in all probability be smaller than, and not identical to, the original sample, the same will likely be true for network characteristics such as centralization. The omission of any link may reduce, or possibly increase, the centralization scores for other actors, as potential paths between nodes will be affected by the presence or absence of the resampled links. The omission of a link for a node with one tie to the rest of the network will reduce the size, and increase its density. But in addition, the removal of a link to more heavily connected nodes will reduce the total path distances to be counted, and result in decreased centrality. In general, we will expect the reduction in sampled links to outweigh the reduction in singleton nodes, and our parameters to be weaker as a result.

In resampling a network we will find that we can develop confidence intervals for a variety of network characteristics. In this preliminary study we will focus on centralization, and in particular betweenness and closeness, as these characteristics have great interest in intelligence assessment of terrorist networks. ***[Betweenness and closeness measures not available at conference time due to programming difficulties!]***

Analysis

To demonstrate the effectiveness of bootstrapping a network, we will use two well known existing networks: the Jemaah Islamiyah network collected by Stuart Koschade, and the 9/11 hijacker network assembled by Valdis Krebs. Both networks were collected from open source information, and hence represent a likely sample of all the possible links that exist or that could have been found by investigators. The presentation of these networks in this paper gives us two somewhat different examples. Jemaah Islamiyah is a small network with two distinct cliques, with all extraneous contacts excluded. Koschade presents only the known configuration of contacts in a post hoc fashion. Such evidence is noteworthy for prosecution, when all possible "persons-of-interest" have been investigated, and the remaining suspects are arraigned or charged. This makes for a smaller size network of higher density. Centrality scores are also higher, due to the density of the network.

The Jemaah Islamiyah network is comprised of the 17 members of the group that participated in the December 2002 bombing of the Bali, Indonesia nightclub. The network is a good example of a tight network with two clear clusters (cliques) and a significant actor with strong measures of betweenness and closeness. Samudra and Idris are the two most centrally placed actors in the network, as can be seen from the network structure provided by Koschade's network diagram in Figure 1. Samudra, in fact, provides the only link between the

bomb makers and Team Lima, the group setting off the bombs, and scores rather high on betweenness.

Figure 1.
Jemaah Islamiyah Network for the Bali Nightclub Bombing in December 2002
(from Koschade 2006)

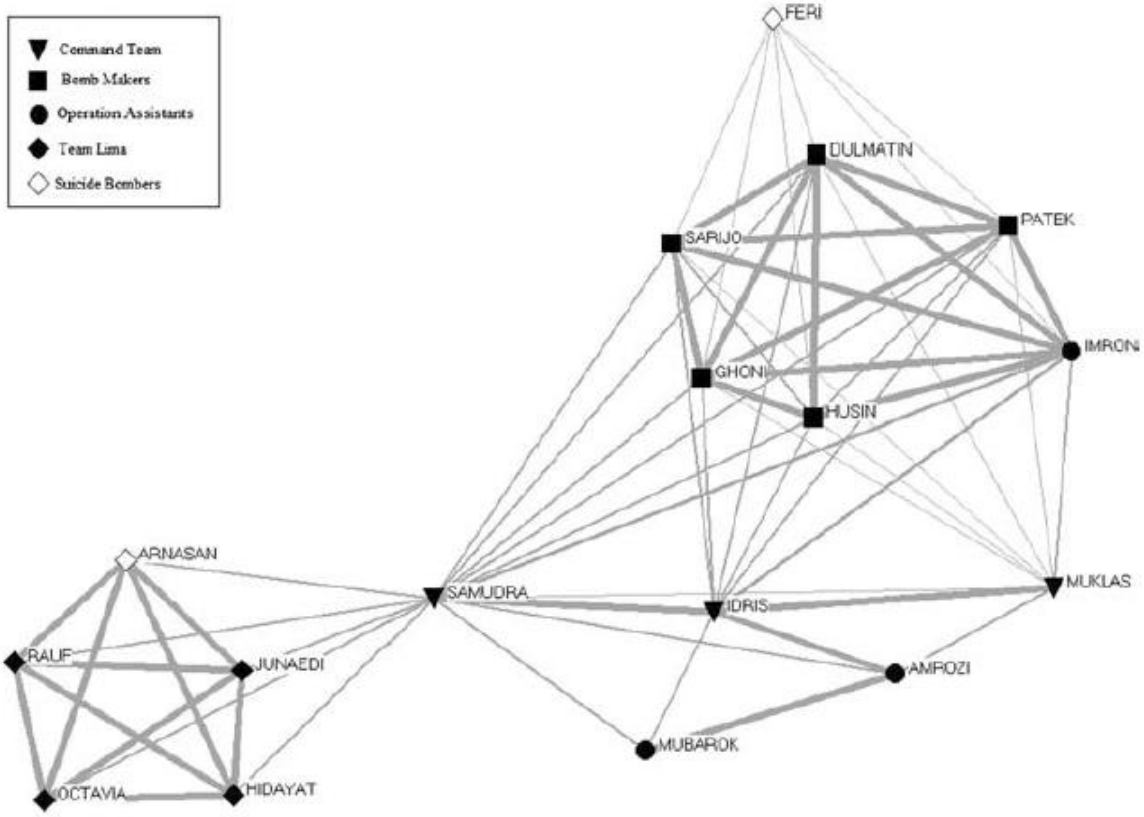


Figure 1. Jemaah Islamiyah Graph—Bali Operation October 6, 2002–October 11, 2002.

In this graph we would clearly expect Samudra to be the most between actor, since he is the only link between the bomb makers and the actual bombers. Koschade's centrality measures are provided in Table 1.

Table 1

Ranked Centrality measures for Major actors in Jemaah Islamiyah Network (from Koschade 2006)

	Std Centrality	Betweenness	Closeness
Samudra	93.750	50.972	94.118
Idris	62.500	5.139	72.727

The original centrality scores of these two actors are somewhat different, but being mindful of the relatively small network size, and the absence of isolates and loosely affiliated nodes with single contact with the primary cliques, we need to perform tests of inference to see to if these scores are actually significantly different from each other. The essential research question is: "How likely it is that central actor 'one' is significantly more central than central Actor 'two.'"

Table 2 provides bootstrapped confidence intervals for the standardized centrality scores. Bootstrapping does indeed result in resamples of lower density. The original network has a density of .434, while the average of the 1000 resampled networks is .311. We also need to realize that reduced density means reduced centrality scores. Both actors have slightly lower centrality across the 1000 replications, but their difference is comparable: about 30%. Also provided are 95% confidence intervals based upon the bootstrapped resamples. We provide the percentile CI, since it is the easiest to calculate, being determined by the 25^{th} and 975^{th} values of the sorted bootstrapped centrality scores. These confidence intervals indicate that both Idris's actual as well as bootstrapped average standardized centrality score lie below Samudra's 95% confidence interval. Given a known network of limited external affiliation, we can make clear inferences about the relative importance of actors within the network. Such information may be quite useful in well known networks, for post hoc inference in support of prosecution.

Table 2
Bootstrapped 95% Confidence Intervals (Percentile)
Ranked Centrality measures for Major actors in Jemaah Islamiyah Network

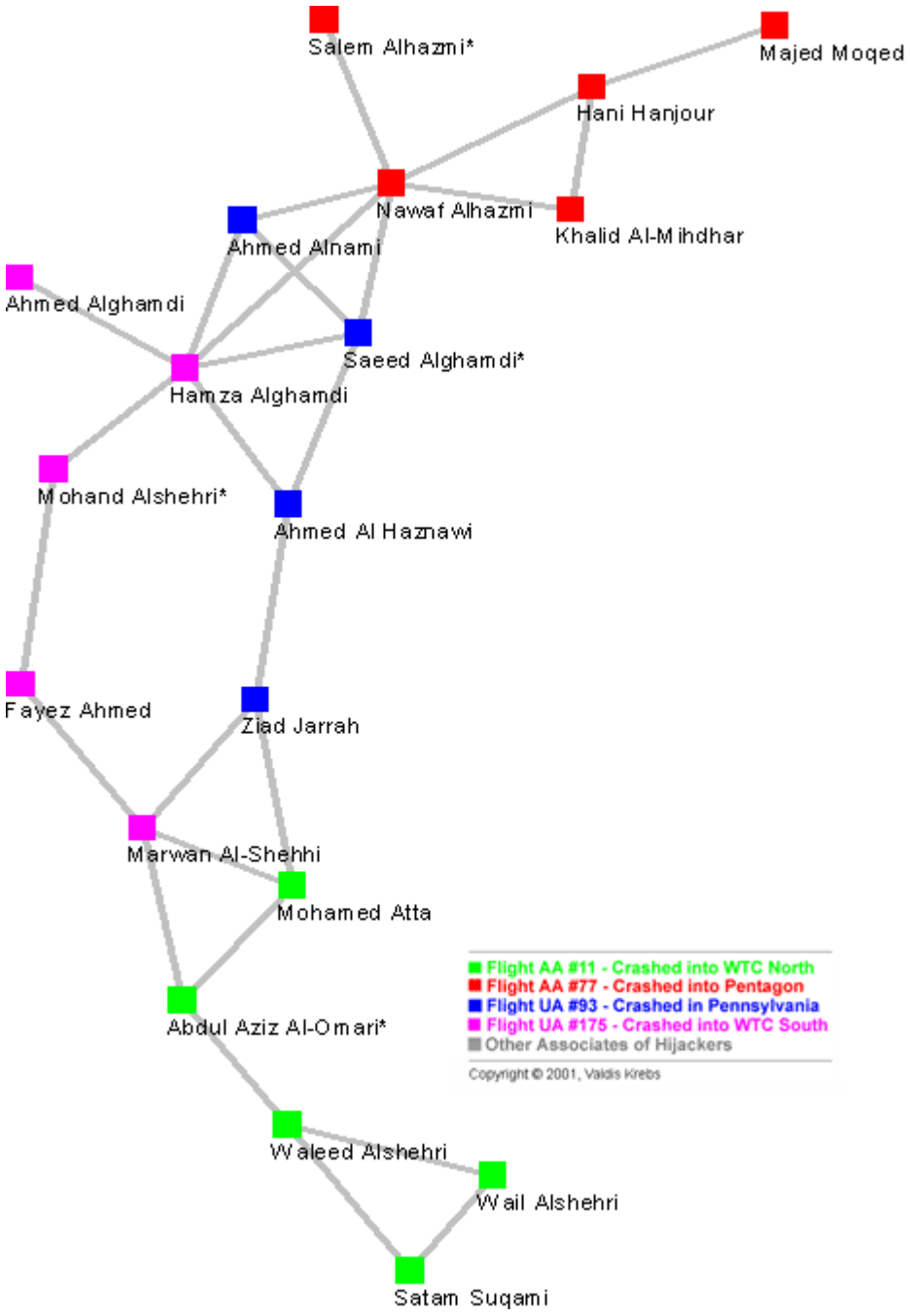
	Std Centrality	Upper 95% CI	Lower 95% CI
Samudra	85.575	93.75	68.75
Idris	56.194	62.50	43.75

Yet we need to examine the utility of bootstrapping as a counterintelligence tool in making inferences about a larger more diffuse network with larger numbers of actors of less central importance. One of the most widely discussed examples of network analysis of terrorist activity is Krebs' now classic analysis of the 9/11 hijackers (2002). Through the use of public sources that identified contacts between an initial pair of the hijackers known to the FBI in 2000, Nawaf Alhazmi and Khalid Almihdhar, a picture of the network emerges that indicates the

prominence of Mohammad Atta as the local leader and broker in communications within the group. In the process of expansion of the network from all contacts within one step of the initial pair, it is with the inclusion of secondary steps or contacts that Atta emerges as having maximum closeness and betweenness for the network. Atta's closeness score indicates that he had the greatest access to others in his network, and his betweenness score indicates that he had the greatest ability to influence the flow of information in the network. In addition, all 19 hijackers are within two steps of the initial known pair. Krebs' work is also *post hoc* analysis, yet it has pointed the way since it not only identifies the social structure of the hijacker's network, it also reveals the channels of communication flows, prototyping our understanding of terrorist network behavior.

Figure 2 shows Krebs' initial network of the 9/11 hijackers and their contacts. This graph resembles Koscade's JI graph in that all nodes are confirmed hijackers and no external actors are included. This graph is not particularly dense, and no substantially central actor is apparent. Certainly, Mohammed Atta, the known major actor in the operation does not stand out, based upon his connections to the other 18 members of the operation.

Figure 2
The 9/11 Hijackers
(From Krebs 2002)



Yet if we expand to look at all actors within 3 steps of the initial pair, then we obtain a much large, if less dense network, as seen in Krebs' representation provided in Figure 3. This graph is much richer, from the standpoint of counterintelligence analysis and provides somewhat more insight into the structure of the network.

Figure 3

The 9/11 Hijacker Extended Network
(From Krebs 2002)

Table 3 provides the Standardized Centrality scores for the three most central of the 9/11 hijackers: Mohammed Atta, Marwan Al-Shehhi, and Hani Hanjour. It is noteworthy that none of the most central actors is one of the initial pair that initiated the data collection. Mohammad Atta is clearly the most central, with Al-Shehhi, also being somewhat more centrally located than Hanjour. Yet these scores provide us no real feeling for whether these differences are significant. The 9/11 network data is substantially less dense, the scores are substantially lower, and the key actors are much closer together than the Jemaah Islamiyah terrorists. Mohammed Atta seems clearly more "between" and hence the conduit of information between cliques, but we cannot substantiate even this large gap without some form of inference.

Table 3

Ranked Centrality measures for Major actors in 9/11 Hijacker Network (from Krebs 2002)

	Std Centrality	Betweenness	Closeness
Mohammed Atta	36.1	58.8	58.7
Marwan Al-Shehhi	29.5	25.2	46.6
Hani Hanjour	21.3	12.6	44.5

Table 4
Bootstrapped 95% Confidence Intervals
Standardized Centrality measures for Major actors in 9/11 Hijacker Network

	Std Centrality	Upper 95% CI	Lower 95% CI
Mohammed Atta	21.90	26.67	16.00
Marwan Al-Shehhi	19.21	23.94	13.70
Hani Hanjour	14.10	18.06	9.33

Table 4 gives us the bootstrapped confidence intervals for the three most central 9/11 hijackers. While Marwan Al-Shehhi cannot be determined to be significantly less central than Mohammed Atta, both Atta and Al-Shehhi are significantly more central than the third most

central actor. Bootstrapped confidence intervals let us see that two of these actors are clearly significant players in the network.

Conclusion

Were this analysis to be run in "real time" as new data comes in, key actors in such a network would likely emerge, and resources for intelligence could be diverted to the appropriate targets. Had a monitoring program existed that looked at significant differences in the role or placement of these actors, it is possible that enough attention might have been paid to Atta (and others) that 9/11 might have been preventable. While these Confidence Intervals do not tell us the complete story, that allow inference to be applied, and for researchers to be more confident in decisions. While it is by no means certain that such network analysis, monitoring, and bootstrapped inference based on a network affiliated with this initial pair (Nawaf Alhazmi and Khalid Almihdhar) would have revealed the plot in time to disrupt it, the findings are suggestive that network analysis as a research tool provides promise as a monitoring and detection tool as well. As noted, the use of networks to examine the personal relationship structures in Sunni tribes played a significant role in the capture of Saddam Hussein and is being implemented in tracking the development of the use of IEDs in Iraq (McFate, 2005). And in an earlier post-hoc analysis of the transnational activities of Aum Shinrikyo, Picarelli (1998) suggests that network analysis could have revealed the increasing WMD threat that the cult eventually realized.

This preliminary foray into bootstrapping networks for inference has produced rather promising results. Clearly, a network may be bootstrapped to provide statistical inference concerning parameters of interest for individual nodes or actors. We can indeed perform tests of inference on such parameters, even in the face of no known statistical tests or sampling distributions. And in the case of the analyses presented here, those tests of inference affirm the graphic representation of the networks examined, and the post hoc analysis of these well known cases. Mohammed Atta's central role is affirmed, even if we cannot distinctly conclude that he is the single most central actor, we know that only one other actor is within range of that claim.

Such statistical confidence is of substantial value. If we envision the collection of the 9/11 network as a process, much as is likely in counter-terrorism monitoring of known terrorists, we speculate that the collection of the network affiliations of the actors lets the network structure as more data is added to the network. With concomitant analysis based upon bootstrapped inference, we can see where to increase assets, and make assessments for interdiction. Decisions made on network disruption that can be made with confidence – statistical confidence – are likely to be more credible, defensible, and accurate than those based on intuition and conjecture. The bootstrap analysis of networks should prove an invaluable tool for the conduct of social network analysis, whether it be the specific examinations of suspected terrorist cells, or the testing of theoretical propositions that focus on the centrality of the actors being examined.

References

Adams, J. & Roscigno, V. (2005). White Supremacists, Oppositional Culture and the World Wide Web. *Social Forces,* 84,759-778.

Apostolakis, G. and Lemon, D. (2005). "A screening methodology for the identification and ranking of infrastructure vulnerabilities due to terrorism." *Risk Analysis,* 25(2): 361-376.

Bertlet, C. (2004). Christian Identity: Apocalyptic Style, Political Religion, Palingenesis and Neo-Fascism. *Totalitarian Movements and Political Religion,* 5, 459-506.

Blee, K. (2007). Ethnographies of the Far Right. *Journal of Contemporary Ethnography,* 36:119-127.

Borgatti, Stephen, P. , Carley, Kathleen, M., Krackhardt, David (2006). On the Robustness of Centrality Measures under Conditions of Imperfect Data. *Social Networks,* 28:124-136.

Brandes, U., Kenis, P., & Raab, J. (2006). Explanation through Network Visualization. *Methodology,* 2, 16-23.

Breiger, R., Carley, K., et al. (2003). *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, National Academy Press.

Butts, Carter, T. (2003). Network Inference, Error, and Informant (in) Accuracy: A Bayesian Approach. *Social Networks* 25: 103-140.

Cao, X. "Convergence, Divergence, and Networks in the Age of Globalization: A Social Network Approach."

Carley, K. M., Lee, J., and Krackhardt, D. (2002). Destabilizing Networks. *Connections*, 24 (3):79-92.

Chen, H., Chung, W., Qin, Y., Chau, M., Xu, J., Wang, G., Zheng, R., and Atabakhsh, H. (2003). "Crime Data Mining: An Overview and Case Studies." *ACM International Conference Proceedings Series,* Volume 130.

Chen, H., Quin, J., Reid, E., Chung, W., Zhou, Y., Xi, W., Lai, G., G., Bonillas, A., and Sageman, M. (2004). "The Dark Web Portal: Collecting and Analyzing the Presence of Domestic and International Terrorist Groups on the Web". *Proceedings of the 7th International Conference on Intelligent Transportation Systems (ITSC)*, Washington D.C.

Clark, C. R. (2005). *Modeling and Analysis of Clandestine Networks*, Masters Thesis, Air Force Institute of Technology, Wright-Patterson AFB, Ohio.

Coffman, T., Greenblatt, S., and Marcus, S. (2004). "Graph-Based Technologies for Intelligence Analysis." *Communications of the ACM,* 47:45-47.

Conway, M. and McInerney, L. (2008). "Jihadi Video & Auto-Radicalisation: Evidence from an Exploratory YouTube Study" in *Intelligence and Security Informatics: European Conference, EuroISI 2008 Proceedings,* Ortiz-Arrovyo, D., Larsen, H., Zeng, D., Hicks, D., and Wagner, G. (Eds.). Heidelberg: Spring Berlin.

Cunningham, D. (2003). "The Patterning of Repression: FBI Counterintelligence and the New Left" *Social Forces,* 82:209-240.

Dorussen, H. and H. Ward (2008). "Intergovernmental organizations and the Kantian peace: A network perspective." *Journal of Conflict Resolution* 52(2): 189.

Duval, R. and K. Christensen.(2010) "Network Analyses of Violent Non-State Actors" in L. Fenstermacher, L. Kuznar, T. Rieger, and A. (Eds) *Protecting the Homeland from International and Domestic Terrorism Threats: Current Multi-Disciplinary Perspectives on*

Root Causes, the Role of Ideology, and Programs for Counter-radicalization and Disengagement. Air Force Research Laboratory White Paper, (January)

Ferrara, L., Mårtenson, C., Svenson, P., Svensson, P., Hidalgo J., Molano, A., and Madsen A. (2008). "Integrating data sources and network analysis tools to support the fight against organized crime." *Proceedings of the IEEE ISI 2008 PAISI, PACCF, and SOCO international workshops on Intelligence and Security Informatics, pp*: 171-182.

Gerstenfeld, P., Grant, D., & Chang, C. (2003). Hate Online: A Content Analysis of Extremist Internet Sites. *Analysis of Social Issues and Public Policy,* 3, 29-44.

Golbeck, J., Mannes, A., and Hendler, J. (2005). "Semantic Web Technologies for Terrorist Network Analysis" in *Emergent Technologies and Enabling Policies for Counter Terrorism*, New York: IEEE Press.

Hafner-Burton, E., M. Kahler, et al. (2009). "Network analysis for international relations." *International Organization* 63(3).

Hafner-Burton, E. and A. Montgomery "Globalization and the Social Power Politics of International Economic Networks."

Hafner-Burton, E. and A. Montgomery (2006). "Power positions: International organizations, social networks, and conflict." *Journal of Conflict Resolution* 50(1): 3.

Hoff, P. and M. Ward (2004). "Modeling dependencies in international relations networks." *Political Analysis* 12(2): 160.

Kahler, M. *Networked politics: agency, power, and governance*, Cornell University Press.

Koschade, S. (2006). "A social network analysis of Jemaah Islamiyah: The applications to counterterrorism and intelligence." *Studies in Conflict and Terrorism*, 29(6): 559-75.

Kossinet, G. (2008). Effects of Missing Data in Social Networks. *Research Paper.*

Krebs, V. (2002). Uncloaking Terrorist Networks. *First Monday*. Volume 7 Number 4 - 1 April.

Maoz, Z. (2001). "Democratic Networks: Connecting National, Dyadic, and Systemic Levels of Analysis in the Study of Democracy and War." *War in a Changing World*: 143–182.

Maoz, Z. (2006). "Network polarization, network interdependence, and international conflict, 1816-2002." *Journal of Peace Research* 43(4): 391.

Maoz, Z. (2009). "The Effects of Strategic and Economic Interdependence on International Conflict Across Levels of Analysis." *American Journal of Political Science* 53(1): 223-240.

Maoz, Z., R. Kuperman, et al. (2006). "Structural Equivalence and International Conflict: A Social Networks Analysis." *Journal of Conflict Resolution* 50(5): 664.

Maoz, Z., L. Terris, et al. (2005). "International relations: A network approach." *New directions for international relations: confronting the method-of-analysis problem*: 35-64.

Maoz, Z., L. Terris, et al. (2008). "What is the enemy of my enemy? Causes and consequences of imbalanced international relations, 1816–2001." *The Journal of Politics* 69(01): 100-115.

McCulloh, I., and Carley, K. (2008). Social Network Change Detection. Institute for Software Research, Carnegie Mellon University, Pittsburgh, PA. Retrieved on March 26th, 2009,from http://www.casos.cs.cmu.edu/publications/papers/CMU-CS-08-116.pdf McFate, M., (2005). Iraq: The Social Context of IEDs. *Military Review,* (May-June).

Memon, N., and Larsen, H. (2006). "Practical Approaches for Analysis, Visualization, and Destabilizing Terrorist Networks." *Proceedings of the First International Conference on Availability, Reliability and Security (ARES '*06).

Milward, H., and Raab, J.. (2006). "Dark networks as organizational problems: Elements of a theory." *International Public Management Journal,* 9(3): 333-360.

Milward, H., and Raab, J. (2005). "Dark Networks as Problems Revisited: Adaptation and Transformation of Islamic Terror Organizations since 9/11." Paper presented at the 8th Public Management Research Conference at the School of Policy, Planning and Development at University of Southern California, Los Angeles, September 29 - October 1, 2005.

Mooney, C. and R. Duval. (1993). *Bootstrapping: Nonparametric Statistical Inference*. Sage Publications.

Newman, M. (2004). "Analysis of Weighted Networks" *Physical Review*, 70:2-9.

Qin, J., Zhou, Y., Reid, E., Lai, G., & Chen, H. (2007). Analyzing Terror Campaigns on the Internet: Technical Sophistication, Content Richness, and Web Interactivity. *International Journal of Human Computer Studies,* 65, 71-84.

Picarelli, J. (1998). Transnational Threat Indicators and Warning: The Utility of Network Analysis. Retrieved on March 26, 2009 from http://kdl.cs.umass.edu/events/aila1998/picarelli.pdf

Raab, J. and H. Milward. (2003). "Dark networks as problems." *Journal of Public Administration Research and Theory*, 13(4): 413-439.

Reed, B. (2006). *Formalizing the Informal a Network Analysis of an Insurgency*, University of Maryland, College Park.

Reed, B. (2007). A Social Network Approach to Understanding an Insurgency. *Parameters*, Summer, 19-30.

Renzi, F. (2006). Networks: Terra Incognita and the Case for Ethnographic Intelligence. *Military Review.* (September-October) pp: 16-22.

Robins, Garry, Pattison, Philippa, Woolcock, Jodie (2004). Missing Data in Networks: Exponential Random Graph (p*) Models for Networks with Non-Respondents. *Social Networks* 26: 257-283.

Ronfeldt, D., & Arquilla, J. (2001). Networks, Netwars, and the Fight for the Future. *First Monday,* Vol. 6. Number 10 (October 1st).

Sageman, M. (2004). *Understanding Terror Networks*. Philadelphia: University of Pennsylvania Press.

Sageman, M. (2008). *Leaderless Jihad: Terror Networks in the Twenty-First Century*. Philadelphia: University of Pennsylvania Press.

Scott, J. (2000). *Social Network Analysis: A Handbook.* (2nd ed.). London: Sage Publications.

Surface, J. (2007). *Six Degrees of Bin Laden: The FBI's Use of Link Analysis for Counterterrorism Investigations*. National Defense Intelligence College.

Travers, J. and S. Milgram. (1969). An Experimental Study of the Small World Problem. *Sociometry*, Vol. 32, No. 4. pp. 425-443.

Tsvetovat, Maksim and Carley, Kathleen, M. (2007) On Effectiveness of Wiretap Programs in Mapping Social Networks. *Computational and Mathematical Organizational Theory* 13: 63-87.

Van Duyn, Donald. (2009). "Statement Before the Senate Committee on Homeland Security and Governmental Affairs" January 8, 2009. retrieved from the Web on My 22, 2009 from http://www2.fbi.gov/congress/congress09/vanduyn010809.htm

Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. New York: Cambridge University Press.

Wesler, H., Gleave, E., Fisher, D., & Smith, M. (2007). Visualizing the Signatures of Social Roles in Online Discussion Groups. *Journal of Social Structure* Volume 8. Retrieved from http://www.cmu.edu/joss/content/articles/volume8/Welser/

Williams, P. (2008). "Violent Non-state Actors and National and International Security" *International Relations and Security Network*. Retrieved on March 26th from http://www.isn.ethz.ch/isn/Digital-Library/Publications

Zhou, Y., Reid, E., Qin, J., Lai, G., & Chen, H., (2005). U.S. Domestic Extremist Groups on the Web: Link and Content Analysis. *IEEE Intelligent Systems, Special Issue on Artificial Intelligence for National and Homeland Security*, September/October, 44-51.

Xu, J. and Chen, H. (2003). "Untangling criminal networks: A case study." *Lecture notes in computer science*: 232-248.