


Changes in Muslim Nations' Centrality Mined from Open-Source World Jihad News:

A Comparison of Networks in Late 2010, Early 2011, and Post-Bin Laden 

James A. Danowski, Ph.D.
Dept. of Communication
University of Illinois at Chicago
Chicago, USA
jimd@uic.edu

Abstract—This research analyzes the changes in Muslim nation (MN) networks and semantic networks associated with Jihad linked with three recent periods: 1) the late 2010 period, the early 2011 Muslim Middle East and North Africa uprisings and 3) the takedown of Osama Bin Laden. Mined were transcripts of web sites, broadcasts, newspapers, and other content captured for 46 Muslim nations. Results show that Somalia made the largest move upward across the three periods, increasing 21 times in network centrality. Iran is consistently in the top 2 positions. The network increased in link strength and indegree but became less structured in the early uprising period, and continued the decline in structure in the post-Bin Laden period, results consistent with crisis effects. Words paired with ‘jihad’ that increased and decreased in the early uprising and post-Bin Laden periods revealed messages that reflected major changes in substantive content in the three periods. The results appear to have face validity, and demonstrate how mining open-sources for inter-networks and for semantic networks about a topic of interest, in this case: ‘jihad,’ can provide quantitative evidence with statistical tests that have intelligence and security implications.

Keywords—Muslim nations; jihad; web mining; semantic networks; social network analysis; international networks; Middle-East uprisings, Osama Bin Laden takedown; cyber security

I. INTRODUCTION

The web contains a wide variety of open sources of intelligence material regarding Jihad and Muslim nations. One source of interest contains transcripts of broadcasts, newspaper stories, and other documents, Lexis-Nexis. This research examines two main events against a benchmark of prior documents, the early 2011 uprisings of the Middle East and North Africa (MMENA), and the takedown of Osama Bin Laden. Of particular interest is how the concept of Jihad may have changed in these sources, and how the relationships among nations with majority Muslim populations may have changed in terms of their interrelationships represented in these sources. Analysts can evaluate these changes for their security significance.

II. BACKGROUND AND LITERATURE REVIEW

Social networks in relation to Muslim Middle-East uprisings have been of interest for at least nearly twenty years [1]. Like most social network (SNA) research, that study analyzed networks among individuals as nodes. Nevertheless, network analysis of the MMENA region using nations as nodes has also been useful. Studies have been done of the MMENA network of telephone call traffic [2]. The Dark Web project has used data mining methods to gather web content from around the internet that has a jihadi focus [3]. While these projects have been useful open sources of intelligence about jihadi activity on the web, additional research, such as presented here, using SNA methods to map the links among the 46 MNs and profile them both in terms of network structure and jihad content in major news sources such as in transcriptions of broadcasts, newspaper stories, and other web content, is also of potential value.

The recent wave of MMENA uprisings beginning earlier this year, and continuing, have been so recent that scholarly research has yet to catch up with them. Mainly news and opinion in the popular press are the only sources available. While some of these documents are analytical treatments with intelligence value, it is perhaps useful to conduct research using quantitative scholarly methods.

Each nation of the world has a unique top-level domain name, such as .id for Indonesia and most intra-country web pages and other internet information is coded for that domain name. This is one way in which the science of Webometrics [4] represents domain interlinkage.. The research questions include: 1) how has the centrality of Muslim nations extracted from news documents changed from a baseline period to the early 2011 MMENA uprising period, and from that period to the post-Bin Laden period? 2) how has the concept of ‘jihad’ changed in association to other words, comparing these three periods?

III. RESEARCH DESIGN AND METHODOLOGY

The period from January 1, 2011 to April 30, 2011 is treated as the period of early 2011 Muslim Middle-East and North Africa uprisings. The period of May 1, 2011, when the takedown of

Osama Bin Laden was announced, until May 16, when the current data collection was completed because of the paper submission deadline, constitutes the second major event comparison period. Given that the uprising period comprised four months, a benchmark set of data were collected for the last four months of 2010, from September 1 to December 31.

The data for this research was extracted from full text documents downloaded from Lexis-Nexis Academic's "major world publications" source, which contains transcripts of broadcasts, newspaper stories, web documents, and other textual material. Within each of the three periods the same search strategy was repeated. It obtained the full texts of all documents that contained two terms: jihad and each of 46 Muslim nations' names. The term 'jihad', the Boolean connector 'and', followed by the individual country name formed the queries. These countries were 46 nations containing majority Muslim populations. One country was eliminated because it has no top-level domain, Kosovo. The reason for this exclusion is that another paper mines web hyperlinks among these MNs tlds, as shown in Table 1, and also searches for the word 'jihad' in web content. In other words, 46 searches were repeated three times, once for each comparison period. Documents were obtained in plain text format and combined into a single file for each period. There were 26.7 mb of text for period one, 41.5 mb for period two, a 55% increase from the 4-month period benchmark period to the four month early 2011 uprising period, and 23.2mb for period three, 17 days after the takedown of Bin Laden.

The software suite WORDij [5] was used for five types of analysis of these documents: 1) all words appearing within three positions of one another were counted as word pairs and their frequencies cumulated, using the WordLink program in WORDij; 2) the words within two steps of the word 'jihad' were extracted from the total word pair lists using the program NodeTric, for node-centric analysis, in WORDij; 3) an include list of country names (converted to top-level domain codes (tlds)) was used to count country pairs that appeared within three country name positions in each sentence of each document, creating a country network data set for each period, using different options of the WordLink program; 4) country pairs that significantly increased and decreased from period one to two and two to three were identified using Z-Pairs software in WORDij; and 5) the word pairs in the node-centric network surrounding 'jihad' were also identified that significantly increased and decreased from one period to the next.

Next, two kinds of network analysis were performed, one of country-country networks and the second of word-networks surrounding 'jihad.' Although WORDij has a graphing utility that animates changes over time in networks, the features of Ucinet [6] and NetDraw [7] are more robust for computing network statistics such as flow betweenness centrality [8][9] and for graphing the networks with flexible coding options, such as resizing nodes according to flow betweenness centrality.

Figs. 1 shows the breakdown of kinds of documents extracted. Figs. 2-4 show the county networks identified in each period, and Figs. 5-7 show the semantic networks for 'jihad' in each period. To foster visual utility, word pairs less frequent than 50 occurrences were dropped in each word network. A small stop word list containing basic grammatical function words was used to drop such words during the WordLink analysis. For the country networks identified in WordLink using the include list (as opposed to stop list) option all frequencies were used.

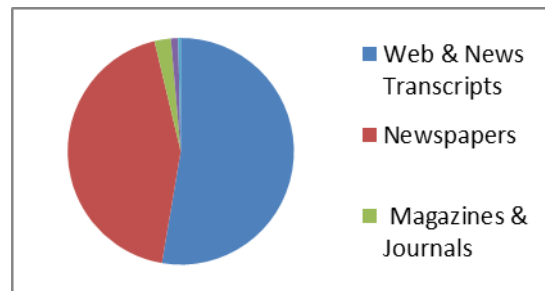


Figure 1. Distribution of Lexis-Nexis Sources Used

Table 1. Key to Top-Level Domains (TLDs)

TLD	Country	TLD	Country
af	Afghanistan	my	Malaysia
al	Albania	mv	Maldives
dz	Algeria	ml	Mali
az	Azerbaijan	mr	Mauritania
bh	Bahrain	ma	Morocco
bd	Bangladesh	ne	Niger
bn	Brunei	ng	Nigeria
bf	Burkina Faso	om	Oman
td	Chad	pk	Pakistan
km	Comoros	qa	Qatar
dj	Djibouti	sa	Saudi Arabia
eg	Egypt	sn	Senegal
gm	Gambia	sl	Sierra Leone
gn	Guinea	so	Somalia
id	Indonesia	sd	Sudan
ir	Iran	sy	Syria
iq	Iraq	tj	Tajikistan
jo	Jordan	tn	Tunisia
kz	Kazakhstan	tr	Turkey
kw	Kuwait	tm	Turkmenistan
kg	Kyrgyzstan	ae	U.A.E.
lb	Lebanon	uz	Uzbekistan
ly	Libya	ye	Yemen

The measure of centrality most appropriate to communication data is flow betweenness centrality, rather than betweenness centrality [10]. In its formal definition, betweenness centrality has assumptions that each link is coded as 0 or 1, present or absent, that positions of nodes on only the single shortest path is of interest, the geodesic. The number of times that each node appears on a shortest path linking two other nodes is aggregated and normalized to produce the betweenness measure. The Freeman betweenness centrality measures each actors' positional advantage, actors "between" other actors, and on whom other actors are presumed to be dependent on to conduct communication with one another through the more central intermediary nodes with high betweenness.

In contrast, imagine two actors want to exchange messages, but the geodesic path between them is blocked by an uncooperative or undesirable intermediate node. If another path exists around this blockage, the two nodes will be inclined to use it even though it is not the shortest path. Nodes may also chose to communicate multiple messages through different paths at the same time or close to the same time. This assumption fits the reality of human communication and other communication behaviors more than the restriction of communication only to one shortest path. Nodes are more often free to use any path connecting them, and may find this more fitting to the realities of the situation than being restricted only to the geodesic paths. Flow betweenness assumes that nodes will use multiple pathways, proportional to their length and valued strength. As

implemented in Ucinet and described in its online help section [6]: “flow betweenness is formally computed as follows: “Let m_{jk} be the amount of flow between vertex j and vertex k which must pass through i for any maximum flow. The flow betweenness of vertex i is the sum of all m_{jk} where i, j and k are distinct and $j < k$. The flow betweenness is therefore a measure of the contribution of a vertex to all possible maximum flows. The normalized flow betweenness centrality of a vertex i is the flow betweenness of i divided by the total flow through all pairs of points where i is not a source or sink. For a network with vertices v_1, \dots, v_n and maximum flow betweenness centrality c_{max} , the network flow betweenness centralization measure is $S(c_{max} - c(v_i))$ divided by the maximum value possible, where $c(v_i)$ is the flow betweenness centrality of vertex v_i .

IV. RESULTS

Table 2 shows for each country in the network based on cooccurrences of countries in the documents analyzed in each period, the flow betweenness values, ranking the countries within each period. The values shown are the normalized ones, so comparisons can be directly made in the values even though the numbers of countries present are different in each period.

TABLE 2. FLOW BETWEENNESS CENTRALITY BY TLD FOR 2010, EARLY 2011 UPRISINGS, AND POST BIN LADEN TAKEDOWN

LATE 2010	UPRISING	POST BIN LADEN
af	12.33	pk 9.28
ir	9.90	ir 7.41
uz	7.86	ng 6.88
sd	7.57	sd 6.28
pk	6.28	uz 6.25
dz	6.16	af 5.37
sa	5.00	eg 5.19
ne	4.65	ml 4.84
ye	4.26	dz 4.65
ng	3.52	gn 4.65
tr	3.06	tr 4.49
tj	2.97	sa 3.80
mr	2.65	tj 2.98
iq	2.59	ye 2.90
az	2.51	kz 2.61
td	2.36	ne 2.59
bf	2.30	so 2.46
kw	1.91	kg 2.45
eg	1.85	ly 2.33
sn	1.81	bf 2.14
id	1.75	sl 2.13
sy	1.68	id 2.06
ma	1.68	gm 1.93
ml	1.25	sn 1.86
lb	1.20	bh 1.75
tn	1.16	iq 1.73
ae	1.10	jo 1.69
bh	0.91	mr 1.64
gn	0.82	td 1.47
jo	0.71	lb 1.44
al	0.64	az 1.22
ly	0.62	ma 1.16
kg	0.61	sy 1.09
so	0.52	tn 0.92
kz	0.32	qa 0.71
my	0.27	bd 0.68
qa	0.27	kw 0.66
tm	0.21	ae 0.46
om	0.19	my 0.21
bd	0.07	om 0.11
gm	0.00	tm 0.02
mv	0.00	dj 0.00

dj	0.00	al	0.00
sl	0.00	mv	0.00
		km	0.00

Figs. 2-4 show the networks among countries in the jihad documents. If a country appeared together with another within three country positions on either side within each sentence (recall that all but country words are dropped), it is counted as linked. The graphics give an overall visualization of country flow betweenness centrality differences with the size of the country nodes scaled according to this centrality.

Table 2 shows the numerical values for flow betweenness centrality for each country in each of the three periods. We note the following results.

Somalia makes the largest moves, jumping from rank 35 and a normalized centrality score of .52 to rank 17 and a score of 2.46 from baseline period to the early uprising. That is a 4.7 times increase. It continues to move upward in the post-Bin Laden period to a rank of 2 and a score of 10.87, an additional 4.4 times increase.

Egypt makes a large move in centrality. It moves from rank 19 and a normalized value of 1.85 to rank 7 at a normalized value of 5.19, increasing by a factor of 2.8 times, then moves to rank 3 and a value of 8.83, a 78% increase.

Mali goes from rank 24 in the benchmark period to rank 8 in the early uprising period, centrality increasing by 3.9 times, then moving down to rank 10, and dropping in centrality by 67%.

Libya begins at rank 32 in the benchmark period and moves to rank 19 in the early uprising period, a change in centrality of 3.8 times higher. Then in has a minor rank change to 18 in the post-Bin Laden period, with a 64% decline in centrality.

Bahrain moves from a benchmark rank of 28 to 24, with an increase in centrality of 92%, and continues upward to rank 14, gaining 34% in centrality.

Iran consistently is in the top two in each period but drops 34% during the uprising period but later increases 61% in the post-Bin Laden period.

Sudan was relatively high in period 1 and 2 at rank 4 in each, then dropped to rank 15 in the post-Bin Laden period, dropping by 67%

Pakistan moves up from rank 5 in the benchmark period to rank 1 in the uprising period, going from 6.28 to 9.28, a 48% increase, but drops to rank 6 in the post-Bin Laden period, a decrease of a factor of 2 times.

Nigeria makes a move up in centrality from rank 10 in the benchmark period to rank 3 in the uprising period, nearly doubling in centrality, but drops to zero in the post-Bin Laden period.

Kazakhstan makes a big move up from rank 36 in the benchmark period to rank 15 in the uprising period, centrality changing from .32 to 2.51, an 8.2 times increase, but it drops to rank 39 with a centrality score of 0.

Kyrgyzstan shows a similar pattern, going from rank 34 in the benchmark period to rank 18 in the uprising period, a four-fold move in centrality, but drops to rank 45 and centrality of 0 in the post-Bin Laden period.

Syria moves down from rank 22 in the benchmark period to rank 32 in the early uprising period but jumps to rank 8 in the post-Bin Laden period, an 2.8 times increase. This may be associated with Syrian uprisings gaining momentum later than the early uprising period.

Guinea makes a big move up from rank 29 in the benchmark period to rank 10 in the uprising period, an 5.5 times increase in centrality, but disappears in the post-Bin Laden period.

The next type of SNA is of the words in the documents. Words within 3 word positions on either side of each word were counted in terms of cumulative frequencies within each period, then the node-centric network around the word 'jihad' and moving two steps out from it was extracted. These networks are shown in Figs. 5-7. Note that because there are more total words in the early uprising period, the jihad-centric network is larger.

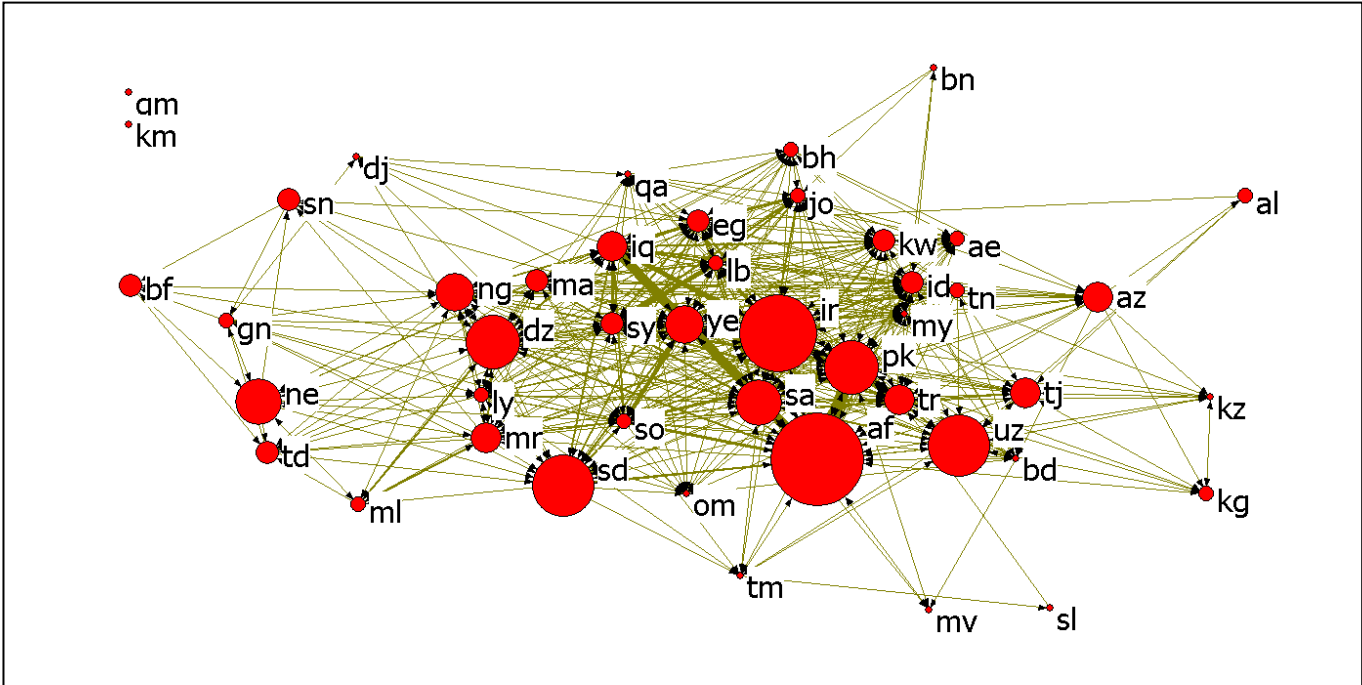


Figure 2. Benchmark Period: Late 2010

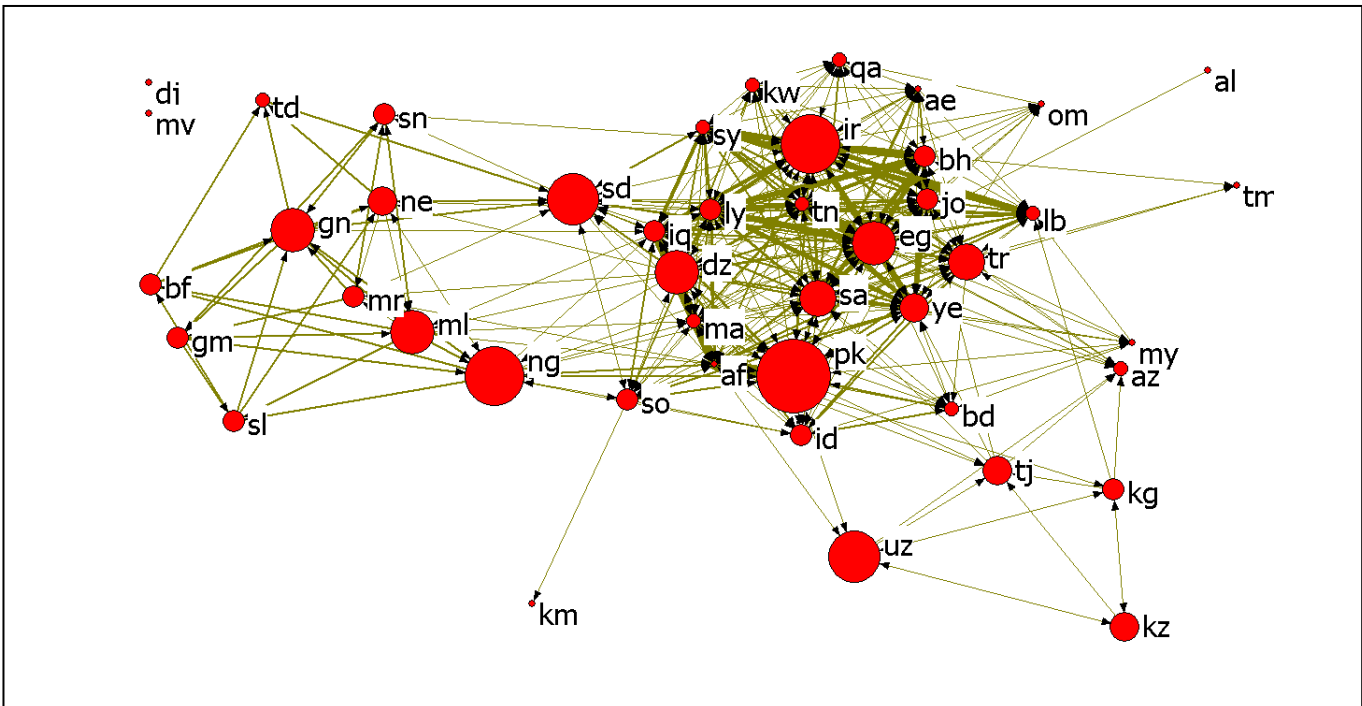


Figure 3. Early 2011 Uprising Period MNs Network Flow Betweenness Centrality

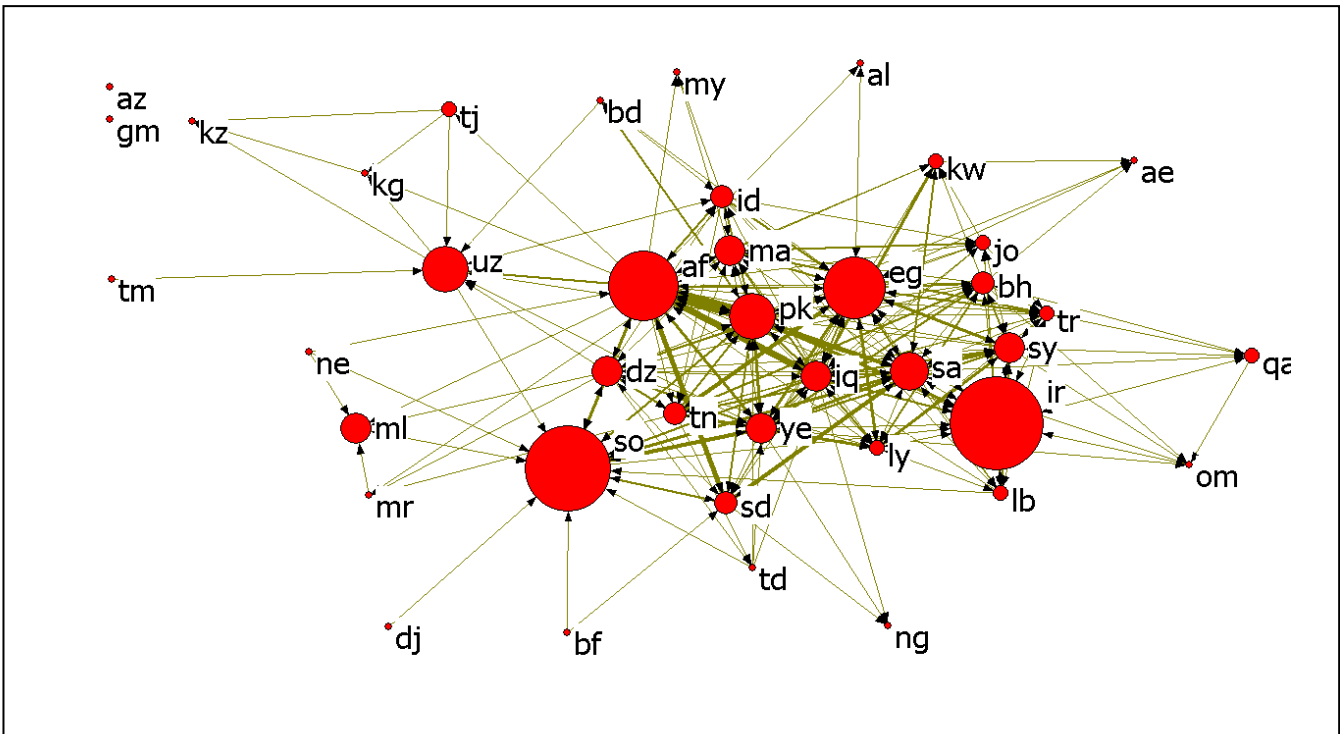


Figure 4. Post Bin Laden Period MNs Network Flow Betweenness Centrality

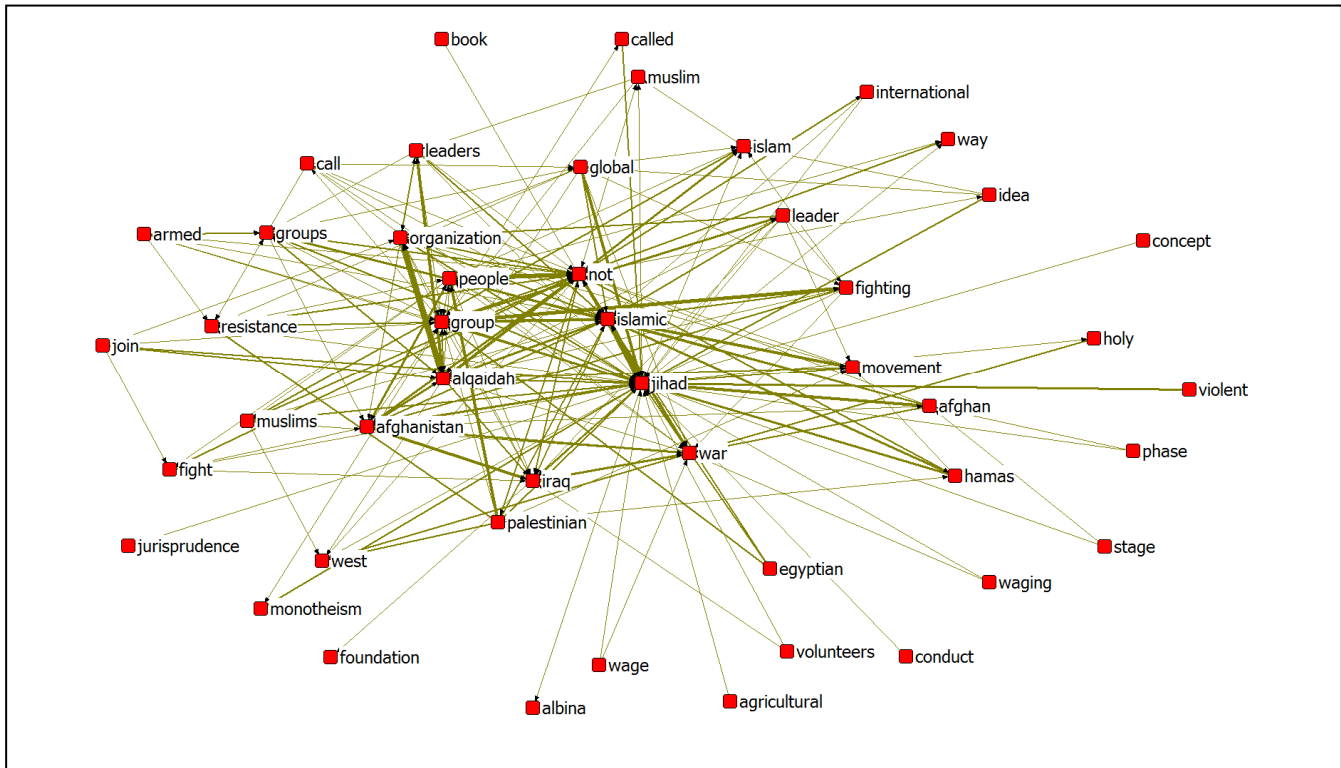


Figure 5. Jihad-Centric Network in Late 2010 Period

TABLE 3. WORD PAIRS INCREASING FROM LATE 2010 TO EARLY UPRISING

Word Pair	Z-Score
"jihad" "website"	25.09
"voice" "jihad"	25.09
"economic" "jihad"	21.97
"Taleban" "jihad"	17.79
"jihad" "pashto"	17.70
"year" "jihad"	16.45
"jihad" "leader"	8.62
"jihad" "supreme"	8.35
"jihad" "year"	7.81
"jihad" "alkhazin"	6.68

TABLE 4. WORD PAIRS DECREASING FROM LATE 2010 TO EARLY UPRISING

"jihad" "group"	25.20
"jihadist" "movements"	25.21
"group" "jihad"	24.37
"islamic" "jihad"	23.91
"jihadist" "groups"	23.25
"monotheism" "jihad"	22.75
"jihadist" "alqaidah"	21.99
"jihad" "organization"	21.80
"jihad" "afghanistan"	21.04
"alqaidah" "jihad"	19.61

TABLE 5. WORD PAIRS INCREASING EARLY UPRISING TO POST-BIN LADEN

"jihad" "jews"	13.97
"jihad" "crusaders"	13.90
"front" "jihad"	13.75
"egyptian" "jihad"	13.11
"group" "jihad"	12.14
"jihad" "bin"	12.14
"international" "jihad"	11.32
"world" "jihad"	11.05
"laden" "jihad"	10.67
"jihad" "al"	10.43

TABLE 6. WORD PAIR DECREASES: EARLY UPRISING TO POST-BIN LADEN

"voice" "jihad"	39.31
"jihad" "website"	39.31
"economic" "jihad"	35.94
"year" "jihad"	29.10
"afghan" "jihad"	25.95
"palestinian" "jihad"	12.89
"hamas" "jihad"	11.93
"fight" "jihad"	11.35
"jihad" "resistance"	18.15
"jihad" "country"	11.31

Visually examining the graphs for the three 'jihad' semantic networks, it is observed that the simplest overall structure is for the benchmark period; mean indegree is 55.0 and the network centralization is .68%. The indegree of the network increases in the early uprising period to a mean of 105.5, nearly doubling, with centralization dropping more than down to a third of the value at .20%. In the post-Bin Laden period, mean indegree is nearly the same at 93.1 while centralization drops nearly in half to .12%. These results

indicate that the early uprising period has network results consistent with crisis effects, an increase in communication but less structured [11][12]. The crisis effect continues with the post-Bin Laden period with a further major decline in structure in the network.

Tables 3-6 show the word pairs connected directly with 'jihad' that significantly increased and decreased proportional terms, which is relative frequency which controls for the number of total words. The top 10 changing pairs are shown from benchmark to early uprising, then from early uprising to the post-Bin Laden period of observation.

The words paired with 'jihad' that increased from the benchmark period to the early uprising period, shown in Table 2, were: website, voice, economic, Taleban, Pashto, year, leader, supreme, and alkhazin. Accounting for most of these references were numerous documents from the website Voice of the Taleban and from a publication, Jihad Al-Khazin. As well, Iran declared an "Economic Jihad Year."

Those decreasing from the benchmark, displayed in Table 3, were: group, movements, groups, Islamic, monotheism, Alqaidah, organization, Afghanistan. There is considerable material from London-based newspaper Al-Quds al-Arabi in a report on Jihadist Salafism which made numerous references to the Egyptian Islamic Jihad Group led by Ayman al-Zawahiri, and various other Egyptian jihadist movements. This, however, was a significantly reduced relative frequency. There is also a long Pan-Arab daily serialization: "other face" of Al-Qa'idah, which makes frequent reference to links between Al-Qa'idah and al-Zarqawi's organization, the Group of Monotheism and Jihad, which Al-Qa'idah eventually absorbed. These references, however, were in significant decline from the benchmark period.

Table 4 shows that from the early uprising period to the post-Bin Laden period the words paired with 'jihad' that had the top 10 largest significant increases were: Jews, crusaders, front, Egyptian, group, Bin, Laden, international, world, al. There are many historical references to the group Bin Laden formed: "The International Islamic Front for Jihad Against Jews and Crusaders," an umbrella group that called for the killing of Americans including civilians. The many references to 'Egyptian' and 'group' are primarily in the context of stating that Dr. Zawahiri joined his group Egyptian Jihad with Bin Laden's Al Qa'idah organization.

The 10 most decreasing words, presented in Table 5, were: voice, website, economic, year, Afghan, Palestinian, fight, resistance, country. There were fewer reports from the Voice of Jihad Website of the Afghan Taleban. The 'economic jihad' references are connected with Iran's President's conflict with the Majlis, forming tensions that are not considered in the spirit of the "Economic Jihad Year." After the Cairo Accord meetings on Palestine on May 10, there was less reference to Palestinian and Hamas jihad against Israel. There were fewer references to statements of Prof Borhanoddin Rabbani, the leader of jihad and resistance of the people of Afghanistan and chairman of the High Peace Council, who has talked about a lasting peace in Afghanistan in the previous few months.

V. CONCLUSIONS

This research analyzed the changes in Muslim nation networks and semantic networks associated with Jihad linked with three recent periods: 1) the late 2010 period, the early 2011 Muslim Middle East and North Africa uprisings and 3) the takedown of Osama Bin Laden. Texts were mined including transcripts of web sites, broadcasts, newspaper stories, and other content captured. Searches were done using the term 'jihad' and the Boolean operator 'and' with the name of each country.

Results show that Somalia made the largest move upward across the three periods, increasing 21 times in network centrality. Iran is consistently in the top 2 positions. The network increased in link strength and indegree but became less structured in the early uprising period, and continued the decline in structure in the post-Bin Laden period, results consistent with crisis effects.

Words paired with 'jihad' that increased and decreased in relative frequency in the early uprising period and post-Bin Laden periods were associated with changes in substantive content of the documents.

These findings illustrate what can be observed from the open-source documents from various websites, broadcasts, newspapers, newsletters, and other such sources, mined in terms of international networks and semantic networks associated with Jihad, particularly when partitioning the documents in time slices that have some event significance.

The methods appear to have face validity. Analysts may find value in deploying these methods for security work. Further research could develop predictions from one point in time to another point in time regarding changes in discourse themes and their prominence, and in country network centrality. Staying in the natural language of the sources is useful for framing strategic messages.

The author declares that he has no competing interests.

VI. REFERENCES

- [1] G. Denooux, *Urban unrest in the Middle East: A comparative study of informal networks in Egypt, Iran, and Lebanon*. Albany, NY: State University of New York Press, 1993.
- [2] J. A. Danowski, "Arab countries' global telephone traffic networks and civil society discourse," in *Civic discourse and digital age communications in the Middle East*, H. Amin and L. Gher, L, Eds. Westport, Conn.: Ablex Publishing, 2000, pp. 93-108.
- [3] H. Chen, W. Chung, J. Qin, E. Ried, M. Sageman, and G. Weimann, "Uncovering the dark Web: A case study of Jihad on the Web," *Journal of the American Society for Information Science and Technology*, vol. 49, no. 8, pp. 1347-1359, June.
- [4] H. W. Park and M. Thelwall, "Hyperlink analyses of the World Wide Web: A review," *Journal of Computer-Mediated Communication*, vol. 8, no. 4, pp. 0-0, July 2003.
- [5] J. A. Danowski, "WORDij: Semantic network analysis software suite." Chicago: University of Illinois at Chicago 2009.
- [6] S. P. Borgatti, M. E. Everett, and L. C. Freeman, "Ucinet for Windows: software for social network analysis. Harvard, MA: Analytic Technologies, 2002.
- [7] S. P. Borgatti, "NetDraw: Graph visualization software. Harvard, MA: Analytic Technologies, 2002.
- [8] L. C. Freeman and S. P. Borgatti, "Centrality in valued graphs: A measure of betweenness based on network flow," *Social Networks*, vol. 13, no. 2, pp. 141-154, June 1991.
- [9] S. P. Borgatti, "Centrality and network flow," *Social Networks*, vol. 27, no. 1, pp. 55-71, January 2005.
- [10] L. C. Freeman, "Centrality in social networks: Conceptual clarification." *Social Networks*, vol. 1, no. 3, pp. 215-239, 1979.
- [11] J. A. Danowski and P. Edison-Swift, "Crisis effects on intraorganizational computer-based communication." *Communication Research*, vol. 12, pp. 251-270. 1985.
- [12] B. S. Greenberg, "Diffusion of news of the Kennedy assassination," *Public Opinion Quarterly*, vol. 28, no. 2, pp. 225-232, 1964.