Semantic Network Analysis of Islamist Sources using Time Slices as Nodes and Semantic Similarity as Link Strengths:

Some Implications for Propaganda Analysis about Jihad

James A. Danowski
Dept. of Communication
University of Illinois at Chicago
Chicago, IL USA
jdanowski@gmail.com

***Abstract*—This research analyzes Muslim nation (MN) networks associated with Jihad for the previous two years. We captured all documents from Lexis-Nexis Academic's BBC International Monitoring--which contains translated transcriptions of web pages, broadcasts, newspapers, and other content—for each of 47 Muslim nations (MNs) using the search term: jihad and MN name. We presented a new kind of semantic network time series analysis of this text. Unlike most semantic network analysis, our nodes were time segments, not words. The link strengths were similarity scores of time nodes across 779,192 word pairs. The time nodes were 105 weekly intervals. We created a two-mode matrix. Columns were the frequencies of time slices' word pairs, appearing in a three-word window. Matrix rows were three-quarters of a million word pairs extracted from the aggregate two-year text file. We converted this two-mode matrix to a one-mode matrix by computing the similarity of each pair of time slices across the rows of word pairs. This resulted in a one-mode network of 105 by 105 time units. Pearson correlations were the similarity coefficients. We conducted social network analysis of the time nodes to find the most central ones. Highly central nodes lie more often on the shortest paths between all pairs of time nodes. They therefore contain in their internal lists of highest frequency word pairs the main themes across the two-years of text. The method is highly automated and efficient. In this case only three central nodes provided the basis for an analyst's interpretations of main propaganda themes.**

Keywords-Muslim nations; jihad; web mining; semantic networks; social network analysis; time-series analysis; international networks;

I. Introduction

A. Time Concepts in Social and Intelligence Research

Linear, chronological time conceptions and time-series statistical analysis are deeply rooted in the modernist social scientific world view. Classic time-series analysis attempts to gather data from at least 120 points in time. Then the data analyst removes serial autocorrelation within variables as they change over time. After this, the data are ready for studying lagged associations between the time-series variables. Sometimes the goal is to forecast future levels of the variables, while other times a variety of hypotheses are tested. Yet, anthropological studies reveal large variations in cultural time conceptions other than the modernist perspective. Taking a different view of time in social network studies can reveal information not as readily available in traditional time-series analysis.

Nevertheless, social network research traditionally analyzes data without regard for time. An aggregate network is analyzed, combining the data from multiple time periods, or from gathering cross-sectional data at a single point in time. There are, however, increasing exceptions. Valente [1] reviews studies that examine change over time in networks. Snijders [2] has developed the Siena software for examining changes in relatively small social networks. Carley's ORA [3] handles large social network data over time. Gloor's Condor [4] easily animates visuals of changing networks over time using a slider to speed up or slow down the network movie.

Other relevant research includes Hibbs' [5] network analysis of words cooccurring in intervals across 10 years of U.S. trade policy documents. Danowski and Cepela [6] analyzed U. S. presidential cabinet member networks from Nixon's through Obama's. They did semantic network analysis of full text of all *New York Times* and *Washington Post* stories mentioning at least one cabinet member. Their procedure used an 'include list' that enables extraction of networks only among the words on the list. An include list is the opposite of a stop or drop list. In that case, the words were cabinet members' names and aliases. When any pair of them appeared within three word positions in the documents they were counted as linked. Time slices were made equal to the average frequency of the Gallup job approval poll reports, which for Obama were at two-week intervals. Cabinet members were then network analyzed within each time slice and their centrality computed. The hypothesis was supported that when Obama's own centrality was higher than the average of his other cabinet members, his Gallup job approval ratings declined six weeks later. The explanation was based on Hofstede's [7] finding that among 50 nations the

Paper presented to the European Intelligence and Security Informatics Conference [EISIC 2012], Odense, Denmark, August 22-24, 2012.

American population most dislikes social power differences, while favoring individualism rather than collectivism. When Obama's centrality is high his power distance is high, resulting in declines in job approval.

Another study did a naturalistic experiment, analyzing documents before, during, and after the Arab uprisings in Tunisia and Egypt [8]. It found that countries that became more central in the networks based on their web hyperlinks, increased the radical 'jihad' content on their websites. Later events were consistent with the expectation that jihadists would see the uprisings as opportunities to work for establishment of Islamist political states with sharia law in countries that as yet had other political systems or were in flux.

Another use is of text mining over time has been to anticipate future events from open sources. One commercial organization, Recorded Futures [9], tracks open source mentions of the time future events will occur. They are not actually predicting future events, but describing the current reports of planned future activities.

Academic researchers have been seeing to what extent they can predict financial market activity from text mining of Twitter tweets [10] [11]. Others have been attempting to predict movie box office receipts from early tweet sentiment about films soon to be released for which trailers, ads, and publicity have become available [12].

The U.S. Intelligence Advanced Research Projects Agency (IARPA), a unit under the Director of National Intelligence, is currently running a competition among research teams to see how effective they can become at predicting a wide range of events from a variety of open sources. Teams are evaluated on how close in time their predictions come to actual events that eventually occur in South America [13].

Other purposes for text mining documents over time include goals to assess propaganda, sometimes called "perception management," [14] analyzing the opposition's communication as well as one's own. Such goals can be addressed with a new approach to time-based text mining. Explicating this method and providing an example are the goals of this paper.

B. *Turning Time Outside-In*

This paper presents an over-time semantic network analysis approach that rather than treating time as an exogenous variable external to the text network, makes time an endogenous variable, internal to it. One can imagined this as turning time outside-in. Network analysis of words is central to this method, but in a new way. Instead of words being nodes in the network, we make time intervals the nodes. The link strengths among time nodes are how similar each pair of time nodes is across the word pairs in text content. We call this approach Semantic Network Time Node Analysis (SENTINA).

SENTINA makes use of an existing widely-used social network variable, node centrality. Highly central time nodes provide key links among relatively diversely linked time nodes. If the texts mined are political propaganda, by opening up only the most central time nodes and examining their most frequent word pairs we see the key propaganda frames and themes that cut across time.

Previous time-series analysis methods have examined textual elements across a linear horizontal time axis. Elements are arrayed according to chronological order, increasing from left to right on the time axis. An example is the approach taken in IBM's i2 Analyst's Notebook [15]. It uses a chronological series of time histograms representing relationships among entities' and graphing them as a network. Such an approach makes it difficult to automatically extract themes stretching over periods of time. One needs an efficient way to consume all of the text at once and systematically analyze what was in it. This is difficult to do when constructing a discrete sequential series. The overarching themes are chopped into pieces and strung out over the time line. SENTINA avoids these difficulties

Next we will present a summary overview and explanation of the SENTINA method. Second, we will describe the detailed methodological steps. Third, an example of two years of documents from 47 Muslim majority nations regarding 'jihad' illustrates the method. Fourth, we discuss the implications and applications of the analysis, its limitations, and future research possibilities.

II. Overview of Methods

A. *Identify Relevant Text Documents*

For a particular discourse domain, for example, jihadist propaganda, one conducts a search of a text database that contains translations of jihadist web pages, television news, radio broadcasts, and newspapers. BBC International Monitoring service provides such transcripts and they are available in the Lexis-Nexis database. A similar text service, World News Connection, is provided to the U. S. National Technical Information Service (NTIS) by the Open Source Center (OSC). Analysts from OSC domestic and overseas bureaus prepare this service for the U.S. government, and it is available by subscription to others through the Dialog database [16].

B. *Extract Text*

Full-text documents over approximately 100 time intervals or more are extracted. These time intervals could be daily if the domain is highly active, or more likely would be weekly, which would cover an approximate two-year time frame.

C. *Remove Duplicate Documents*

Although the commercial version of Lexis-Nexis has a command to remove duplicate documents, if one uses Lexis-Nexis Academic, as we do, there are many duplicate documents. This occurs as there may be different editions of the same source, or different sources used the same content. Duplication distorts the analysis. We created a program, DeDup [17] that removes duplicate text from a corpus.

D. *Time Slicing*

The first step is to segment the file containing all of the documents into time intervals, such as weeks. WORDij's [18]

TimeSlice program does such time segmentation then inserts codes into the large text file representing the time segments.

E. Extract Word Pairs

Next a semantic network analysis procedure, such as WORDij's WordLink, or AutoMap [19] is run to produce a file for each time segment that contains each word pair occurring within a window that is three words wide on either side of each word in the text. This creates a unique word-pair file for each time segment. It contains for each word pair found the frequency within the time segment.

F. Extract Aggreate Master File of Word Pairs

To obtain a benchmark for setting up the next phase of the analysis, the same word-pair extraction is run on the large text file but ignores the time segments. This produces a master list of word pairs for the entire corpus.

G. Create Two-Mode Matrix of Word Paris by Time Slices

The next step is to create a matrix of the master word pairs as row labels and the time intervals as columns. A program we created, Matrixer [20], rearranges each of the time segment columns so that its individual word pairs appear in the same order as in the master file. If a word pair does not occur in the particular time segment, its cell entry is zero, otherwise its entry is the frequency of that word pair.

H. Compute Similarities of Time Slices

With such a rectangular matrix of time interval word pairs, one then computes the similarities of each pair of time intervals across the master word pairs. In standard network analysis this would be equivalent to converting a two-mode network of word pairs by time intervals into a one-mode time-interval network. The size of the matrix may be too big for some network analysis programs. In the current study the file was too large to be converted by the widely used social network analysis program: UCINET [21].

As an alternative the analysis was run in SPSS v. 20 [22]. We computed the similarity scores for each pair of time intervals with the analysis procedure "Proximities" located under the "Correlation" tab. Because the values of cells were not binary but a ratio-level measurement of count values ranging from zero on up, the appropriate similarity metric was the Pearson correlation coefficient. These measured for each pair of time nodes how similar they were in their frequency distributions across 779,191 master word pairs that occurred at least 3 times. We exported the resulting one-mode matrix of time by time similarities for analysis in network analysis programs.

I. Network Analyze the Time-by-Time Node Network

We then conducted social network analysis where the nodes are the time intervals and their link strengths are their similarity scores. In this case we used the programs UCINET and NetDraw [23]. Another program that could be used is Pajek [24], favored by European network analysts.

J. Compute Centrality Scores for Time Nodes

Centrality scores for the time nodes are examined and the nodes with the highest centralities are further analyzed by opening up their respective word-pair files and examining the word pairs sorted by frequency. The highest frequency word pairs indicate the key content of the time node. They reveal, therefore, that these particular word pairs were most instrumental in creating high similarities with other time nodes. Central nodes typically have high degree centrality, which is the number of links with other nodes, and high Freeman betweenness centrality [25], which indexes on how many shortest paths between all other nodes the particular node lays. Now the analyst studies the most central time nodes' key word pairs and the structure of the overall time node network to interpret the meaning in light of the original goals.

III. Example: Jihad Documents of Muslim Majority Countries

"Jihad" has many different meanings in Muslim religion, ranging from an agenda for personal development to the meanings of violent holy war against invaders or non-believers. An earlier study [8] found that the concept of 'jihad' was important in analyzing the differences among Muslim majority nations' responses to the so-called Arab Spring. The semantic network for 'jihad' was empirically dominated by concepts related to Islamist political ideology and terrorism concepts, evidence of an Arab Winter. These findings were consistent with the propaganda of reestablishing a Caliphate of a pan-Islamic nature, and from this base, accelerating the progress of conflicts with non-believers on a global basis [26]. We selected 'jihad' as the focus for the current study because it provides a strong case for counterterrorism-based assessment of relevant propaganda. We analyzed the 48 Muslim majority countries. Nevertheless, because Kosovo does not have its own internet top-level domain it was excluded from this analysis.

A. Methods Details

1) Text Source: Text mining was done using the Lexis-Nexis source BBC International Monitoring. It provides translations of web pages, television and radio broadcasts, and newspaper documents from within these countries.

2) Search Strategy: The search phrase was repeated for 47 countries, within the time frame of June 1, 2010 to April 13, 2012. That phrase was: jihad AND country name. Lexis-Nexis automatically includes in the results lexical variants of the search terms, such as jihadi, jihadist, jihadists. Each article in the database was retrieved and downloaded.

3) Text Preparation: We combinaed these documents into a single file for the initial text analysis. It was 113 megabytes in size. There were many duplicate documents. On the one hand, it could be argued that such redundancy is a valid indicator of how important various news gatekeepers perceive the document. This would support the inclusion of all such documents in the analysis. On the other hand, some types of media sources are more likely to produce redundant documents. This may be associated with biases the representation according to their ideological perspectives. We chose the redundancy removal option, running our DeDup program. This shrunk the 113 megabyte file to 30 megabytes. Indeed, the level of redundancy was high.

4) Semantic Network Analysis: We selected a weekly time slice for segmenting the aggregate text file. WORDij's TimeSlice performed this analysis, creating 105 time intervals. WordLink in WORDij extracted all word pairs appearing three words on either side of each word in the text. A small stop word list removed functions words such as pronouns, articles, conjunctions, and other highly frequent words that do not add much semantic content. The program also dropped word pairs occurring only once or twice. This resulted in 779,191 unique word pairs in the master list. The 779,191 by 105 rectangular matrix created was input to SPSS v. 20 and it computed the proximities among each pair of 105 time intervals across the full range of master pairs using the Pearson correlation coefficient for the similarity measure.

We exported the 105 by 105 time slice matrix from SPSS and imported it to UCINET to compute the betweenness index for the 105 time nodes. NetDraw graphed the network. We chose a coefficient greater than .71 for the link strength threshold, dropping links with smaller coefficients. We selected this correlation arbitrarily, but based on some reasoning. In the finance community, when a shareholder or partner of a firm has a 51% share of ownership they have a controlling interest in the firm. For reasons such as these we chose a threshold greater than .71, which would result in shared variance of 51% or higher. (Squaring the correlation gives the percent variance explained.)

NetDraw produced the network layout with the standard graph theoretic method. It uses the Kamada and Kawai method of laying out the nodes according to the shortest paths among them, applying multidimensional scaling to the geodesics (shortest paths) [27].

IV. Results

Figure I shows the network graph. Thirty nine of the 105 time nodes were below the similarity threshold: a correlation of at least r = .71, and do not appear in the figure. Nodes are sized based on Freeman betweenness centrality. The list of betweenness centrality scores for time nodes appears in Table I.

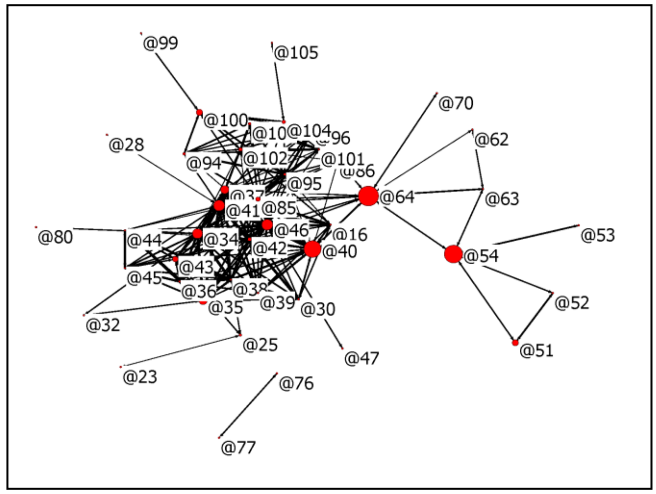

FIGURE I. Network of Time Nodes

TABLE I. Normalized Freeman Betweenness for Time Nodes: Top Ten

	Betweenness
@64	3.502
@54	3.081
@40	3.010
@41	1.803
@46	1.769
@34	1.598
@37	1.274
@35	1.155
@100	.803
@51	.803

TABLE II. Time Node @64 Most Frequent Word Pairs: Top 40

WORD 1	WORD 2	FREQUENCY
district	province	54
soldiers	killed	53
killed	wounded	52
afghan	jihad	51
area	district	42
islamic	emirate	38
american	soldiers	38
united	states	32
foreign	soldiers	30
mojahedin	islamic	28
mojahedin	emirate	28
soldiers	wounded	28
armed	attack	25
great	prophet	23
prime	minister	22
islamic	republic	22
helmand	province	22
foreign	killed	21
took	place	21
fighting	place	20
fierce	fighting	20
large	number	19
tanks	destroyed	19
american	forces	18
security	forces	18
destroyed	soldiers	18
islamic	iran	17
attack	carried	17

afghan	east	17
seriously	wounded	17
republic	iran	17
killed	afghan	16
invading	soldiers	16
local	mojahedin	16
wounded	fighting	16
tank	destroyed	16
mojahedin	area	15
mojahedin	fighters	14
foreign	internal	14
totally	destroyed	14

TABLE III. TIME NODE @54 MOST FREQUENT WORD PAIRS: TOP 40

WORD 1	WORD 2	FREQUENCY
district	province	51
soldiers	killed	51
afghan	taleban	49
afghan	jihad	49
killed	wounded	43
american	soldiers	43
area	district	43
took	place	41
security	forces	33
islamic	republic	31
islamic	iran	30
islamic	emirate	28
seriously	wounded	26
islamic	revolution	26
afghan	east	25
united	states	25
republic	iran	25
mojahedin	islamic	24
mojahedin	emirate	23
forces	afghan	22
afghan	government	22
american	forces	21
armed	attack	21
killed	seriously	20
year	jihad	20
year	economic	20
kordestan	province	20
afghan	forces	19
soldiers	wounded	19
explosion	carried	19
foreign	forces	19
supreme	leader	19
god	willing	19
afghan	people	18
soldiers	board	18
fighting	place	18
people	kordestan	18
attack	carried	16
district	today	16
afghan	security	16

TABLE IV. TIME NODE @40 MOST FREQUENT WORD PAIRS: TOP 40

WORD 1	WORD 2	FREQUENCY
afghan	jihad	53
afghan	taleban	50
soldiers	killed	49
district	province	46
killed	wounded	45
area	district	37
peace	council	37
soldiers	wounded	36
american	soldiers	32
armed	attack	31
high	peace	26
high	council	25
seriously	wounded	23
american	forces	22
took	place	22
foreign	soldiers	22
killed	seriously	21
islamic	emirate	21
safi	rabbani	21
mojahedin	islamic	20
mojahedin	emirate	20
killed	attack	19
foreign	killed	19
explosion	carried	19
district	today	18
helmand	province	18
soldiers	seriously	17
district	helmand	17
american	killed	16
mojahedin	area	16
foreign	forces	16
centre	district	16
great	game	16
powerful	explosion	15
fighting	place	14
mine	explosion	14
attack	carried	13
province	today	13
wounded	explosion	13
peace	afghanistan	13
american	wounded	12

A. *Central Time Nodes as Windows for Future Content*

Highly central time nodes in a network are linked to a number of nodes that are not linked as strongly to one another. Many links appear to radiate from them. In other words, the central node is most often on the shortest path between all pairs of nodes. When using time nodes and semantic similarity, as we did in this study, the set of links for the highly central nodes are based on having similar content to a number of other nodes. These central nodes, therefore, provide a special view of the kinds of content that will be reproduced in future time periods. For the three most central time periods, @64, @54, and @40, Tables II-IV reveal that there is a lot of similarity in casualty reports regarding American and allied troops. One time period, @40, makes reference to an Afghan peace process, a concept that is picked up in later time intervals.

Overall, the content that dominates across the two-year period is highly redundant, reinforcing a view that Taleban troops are successful on the battlefield in killing invading soldiers and destroying their equipment. There is, however, little information on the tactical or strategic implications of these actions. It may be the case that the Islamic audiences need few reminders of the motives and goals for the war. Perhaps it is sufficient to make reference to 'jihad' as a code word that is elsewhere elaborated in terms of ideology, political agenda, and fighting.

V. DISCUSSION

This research introduced a new way to treat time in semantic network analysis. This was to turn time outside-in.

In other words, instead of time being an exogenous variable in the semantic system, it was endogenous. The network nodes were time slices instead of words. The strength of links between time nodes was based on their semantic similarity across more than three-quarters of a million word pairs extracted from two-years of open-source documents. These were created by the BBC International Monitoring service that translates foreign language web sites, radio and television broadcasts, and newspaper articles into English. We extracted all such documents for 47 Muslim nations that contained the word 'jihad.' Earlier research found that the dominant use of this term was in political or terrorism senses. We considered such documents to be propaganda meant for domestic or foreign audiences or both.

When we opened the central time nodes to examine their semantic networks we observed the thematic content that was the basis for this node being centrally linked to other time nodes. The central nodes contained the word pairs that cut across time.

Three time nodes were most central. This shows that identifying themes by looking at the most frequent semantic content inside these central time nodes is easier than if one tried to find these themes with a traditional linear alignment of semantic content for chronological time periods. The typical way of treating time is as a linear axis along which discrete events are strung out. That approach requires a difficult analytical process involving more human interpretation as to how best to assemble into themes the chopped up pieces of the aggregate semantic network that reappeared at multiple times.

A popular tool such as IBM's i2 Analyst's Notebook does trend analysis for concepts this harder way. One would be trying to piece together patterns across 105 sequential periods. It would be difficult to separate the clear signals from the noise. Analyst's Notebook does implement social network analysis functionality but it is the traditional kind that mainly looks at nodes as people and links as relationships of various kinds among them, most often some type of communication. They do not fully deploy the social network algorithms for semantic network analysis, although this has been done for 30 years [28][29][30][31][32][33][34] and more recently has been taken up by an expanding number of researchers, too numerous to cite. Semantic network analysis is no longer practiced only in the communication discipline but across disciplines, both inside and outside the social sciences, such as computer science.

Turning time outside-in and doing a different kind of semantic network analysis, as we have demonstrated, is easy to do. One simply treats time slices of text as nodes and the links as node pairs' similarity across the distributions of word pairs. Following this move, one measures the centrality of the time nodes. Finally, one opens the internal word pair files of the most central time nodes and looks at the most frequent word pairs to easily locate time-transcendent themes. These are the recurring frames of narratives that are consistent with the strategic propaganda goals of organized social actors.

These rhetorical moves of text sources' aim for an agenda-setting influence on opinion leaders among the larger social networks of target audiences. Agendas based on frequency of mention of issues in the media have been found to change audiences' views about what the important issues are [35]. Communication of these agendas, where frequency of attention to an issue indicates its importance, "tell people what is important to think about." [36]. Experiments investigating causal flow have found that many people comply. Moreover, the frames of narratives presented in media have been found to "tell people how to think regarding these issues" [37].

In improving the measurement of these propaganda processes, our twist on time provides a new set of tools in this area of research. The tools are in tune with Web 3.0. Hopefully, before too long, other researchers will try such an approach and refine it, and intelligence analysts will see commercial and governmental applications of these methods that are readily accessible.

This analysis showed how this new kind of semantic network analysis, based on creating time periods as nodes and their links as similarity coefficients among them, provided an automatic way to efficiently and effectively identify propaganda themes that prevailed over multiple time periods. The approach takes advantage of the formalisms, concepts, and measures of social network analysis, as it has developed for the past 85 years.

VI. LIMITATIONS

In a demonstration of a new method, how the research was done is the chief concern. The content analyzed and the findings are irrelevant in terms of building the body of knowledge at the level of theory. Yet, the illustration of the method requires use of content and the generation of findings that makes sense to the reader. This is what gives face validity to the method. If that occurs, others may use the method in more substantively driven ways, and replicate such work by others to judge whether generalizations are predictively valid. What would follow would be research designed to address questions of causality and alternative explanations.

A limitation of the present research is that it focused on method, yet the scholarly enterprise is fundamentally driven by building theoretical knowledge through empirical support. As a result, the reader who is not interested in new methods may find this research not highly engaging and may attend more to some of the peripheral elements, the substantive assumptions and interpretations the author made in order to animate the exercise of the method. These, however, are not important in this kind of paper.

VII. Future Research Directions

The method introduced needs to be tested repeatedly with a range of different content domains. Also of value would be testing it with other kinds of media. Twitter has become recognized as the first source for news among all kinds of media, traditional and social. It increasingly is used by political actors in many countries for campaign propaganda and other propaganda purposes. A noteworthy example is the claimed deceptive use of Twitter by the Putin presidential campaign [38]. Opponents to Putin, his administration, and security operations, actively use Twitter, too. Opening up the Twitter "fire hose" and shooting it through these methods would be desirable.

VIII. Conclusion

Automated analysis of propaganda over time to aid final interpretations by analysts can be taken further than established semantic network analysis techniques. This research has demonstrated a new semantic network analysis method, SENTINA. It takes advantage of non-linear semantic time signatures. It slices up texts over time, but takes time out of the common linear sequence. Instead it uses a network approach treating time slices as nodes and the similarity of semantics inside the pairs of slices as links. This initial work suggests that this may enable more efficient and effective analysis of propaganda.

Acknowledgments

The author thanks Richard Weeks for programming work on the DeDup and Matrixer software.

References

[1] T. W. Valente, "Network models and methods for studying the diffusion of innovations," in Models and methods in social network analysis, P. J. Carrington, J. Scott, S. Wassermann, Eds. Cambridge University Press, 2005, pp. 98-116.

[2] T. A. B. Snjders, "Models for longitudinal network data," in Models and methods in social network analysis, P. J. Carrington, J. Scott, S. Wassermann, Eds. Cambridge University Press, 2005, pp. 215-247.

[3] K. M. Carley, "Destabilization of covert networks," Computational & Mathematical Organization Theory, vol. 12, no.1, pp. 51-66, 2006.

[4] P. A. Gloor, J. Krauss, S. Nann, K. Fischbach, and D. Schoder, "Web Science 2.0: identifying trends through semantic social network analysis," IEEE Proceedings of Computational Science and Engineering, 2009. CSE '09, vol. 4, pp. 215-222, August 2009.

[5] M. Hibbs, "Semantic framing of news about U.S. trade policy: 1982 to 1992," Dissertation: University of Illinois at Chicago, 1993.

[6] J. A. Danowski and N. Cepela, "Automatic mapping of social networks of actors from text corpora: time series analysis," in Data mining for social network data, Annals of information science, vol. 12, N. Memon, J. Jie Xu, D. L. Hicks, and H. Chen, Eds. New York: Springer Science+Business Media, 2010, pp. 31-46.

[7] G. Hofstede, "The cultural relativity of organizational practices and theories," Journal of International Business Studies, vol. 14, no. 2, pp. 75-89, Fall 1983.

[8] J. A. Danowski and H. W. Park, "Web hyperlink network & jihad content among Muslim nations: a comparison with networks mined from major world news," IEEE Proceedings of the International Symposium on Open-Source Intelligence and Mining, Athens, Greece, pp. 100-107, September 2011.

[9] "White paper: big data for the future – unlocking the predictive power of the web," Recorded Future, October 19, 2011 (Retrieved June 14, 2012 from https://www.recordedfuture.com/about).

[10] X. Zhang, H. Fuehres, and P. A. Gloor, "Predicting stock market indicators through Twitter: "i hope it is not as bad as i fear," Procedia - Social and Behavioral Sciences, vol. 26, pp. 55–62, 2011.

[11] J. Bollena, H. Maoa, and X. Zengb, "Twitter mood predicts the stockmarket," Journal of Computational Science, vol. 2, issue 1, pp. 1-8, March 2011.

[12] S. Asur and B. A. Huberman, "Predicting the future with social media," Proceedings of 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, vol. 1, pp.492-499, 2010.

[13] "Open source indicators (OSI) program - broad agency announcement (BAA)", Intelligence Advanced Research Projects Agency, IARPA-BAA-11-11, August 23, 2011 (Rectreived June 14, 2012 from http://www.iarpa.gov/solicitations_osi.html).

[14] P. M. Taylor, "Perception management and the 'war' against terrorism," Journal of InformationWarfare, vol. 1, issue 3, pp. 16-29, 2002.

[15] "IBM i2 Analyst's Notebook," IBM, (Retrieved June 14, 2012 from http://www.i2group.com/us/products/analysis-product-line/ibm-i2-analysts-notebook).

[16] "World News Connection: a foreign news service from the U.S. Government," National Technical Information Service, 2006, (Retrieved June 14, 2012 from http://wnc.fedworld.gov/).

[17] J. A. Danowski, "DeDup software for removing duplicate content in text mining," Naperville, IL: Communication Research Network, 2012.

[18] J. A. Danowski, "WORDij semantic network analysis software," Chicago: University of Illinois at Chicago, 2009.

[19] J. Diesner and K. M. Carley, "AutoMap1.2 – extract, analyze, represent, and compare mental models from texts," CASOS Technical Report, CMU-ISRI-04-100, Pittsburg, PA: Carnegie Mellon University, January 2004.

[20] J. A. Danowski, "Matrixer: software for creating 2-mode matrices of words or word pairs by separate files of text," Naperville, IL: Communication Research Network, 2012.

[21] S. P. Borgatti, M. E. Everett, and L. C. Freeman, "Ucinet for Windows: software for social network analysis," Harvard, MA: Analytic Technologies, 2002.

[22] "SPSS statistics faculty pack," IBM, Armonk: NY (Retrieved June 14, 2012 from http://www-01.ibm.com/software/analytics/spss/products/statistics/faculty-pack/).

[23] S. P. Borgatti, "NetDraw: graph visualization software," Harvard, MA: Analytic Technologies, 2002.

[24] V. Batagelj and A. Mrvar, "PAJEK program for large network analysis," Ljubljana, Slovenia: University of Ljubljana, 2012.

[25] L. C. Freeman, "Centrality in social networks: conceptual clarification," Social Networks, vol. 1, no. 3, pp. 215-239, 1978-1979.

[26] "Stadium crowd pushes for Islamist dream," BBC, 24 September 2011 (Retrieved June 12, 2012 from http://newsvote.bbc.co.uk/2/hi/south_asia/6943070.stm).

[27] T. Kamada and S. Kawai, "An algorithm for drawing general undirected graphs," Information Processing Letters, vol. 31, no. 1, pp.7-15, 1989.

[28] J. A. Danowski, "A network-based content analysis methodology for computer-mediated communication: an illustration with a computer bulletin board, in Communication Yearbook 5, M. Burgoon, Ed. New Brunswick, NJ: Transaction Books, 1982, pp. 904-925.

[29] J. A. Danowski, "WORDIJ: A word pair approach to information retrieval," Proceedings of the DARPA/NIST TREC Conference Washington, DC: National Institute of Standards and Technology, pp. 131-136, 1993.

[30] J. A. Danowski, "Network analysis of message content," in Progress in communication sciences XII, G. Barnett and W. Richards, Eds. Norwood, NJ: Ablex, 1993, pp. 197-222.

[31] R. Rice and J. A. Danowski, "Is it really just like a fancy answering machine? Comparing semantic networks of different types of voice mail

users," Journal of Business Communication, vol. 30, pp. 369-397, October, 1993.

[32] J. A. Danowski and R. Lind, "A computer content analysis linking gender language in presidential candidates' news to gender gaps in polls." in Applications of computer content analysis, M. West, Ed. Westport, Conn.: Ablex Publishing., 2001, pp. 157-167.

[33] J. A. Danowski, "Inferences from word networks in messages," in The content analysis reader, K. Krippendorff and M. Bock, Eds. Los Angeles, CA: Sage Publications , 2010, pp. 421-430.

[34] J. Diesner and K. M. Carley, "Using network text analysis to detect the organizational structure of covert networks," in Proceedings of the North American Association for Computational Social and Organizational Science (NAACSOS) Conference, Pittsburgh, PA, 2004.

[35] M. E. McCombs and D. L. Shaw, "The agenda-setting function of mass media," The Public Opinion Quarterly, vol. 36, no. 2, pp.176-187, 1972.

[36] J. W. Dearing and E. M. Rogers, Agenda-setting. Thousand Oaks, CA: Sage, 1996.

[37] S. D. Reese, O. H. Gandy, and A. E. Grant,. Framing public life pr: perspectives on media and our understanding of the social world. Mahwah, NJ: Lawrence Erlbaum Associates, 2001.

[38] "Russia's Putin set to return as president in 2012," BBC, September 12, 2011 (Retrieved June 13, 2012 from http://www.bbc.co.uk/news/world-europe-15045816).