Sentiment Network Analysis of Taleban and RFE/RL Open-Source Content About Afghanistan

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

***Abstract*—Analysis of sentiment expressed in political systems' communication over long periods of time has been difficult. This research illustrates a method based on network analysis, the Sentiment Network Analyzer (SNAZ). It identifies weighted shortest paths between seed words and 3,500 target sentiment words as these occur in semantic networks extracted from open-source documents sliced into time intervals. Computing the normalized intensity ratios of positive and negative sentiment for each time slice enables application of the "Losada Line." For a system to be flourishing there must be at least 2.9 times more positive than negative communication. Below that ratio the system is languishing. Excessive positivity above a ratio of 11.6 marks the disintegration of a system into chaotic oscillation. We collected and analyzed five years of documents propaganda mentioning the Taleban from Afghani and Pakistani sources transcribed by BBC International Monitoring. Likewise, we extract and analyze stories communicated by Radio Free Europe/Radio Liberty (RFE/RL) connected with Afghanistan over the same five-year period. Semantic network and sentiment network analysis is coupled with the computation of positivity ratios in each time slice during this period. Taleban content is generally evident of flourishing, except for a period of oscillation between flourishing and excessive positivity beginning in the third quarter of 2010. RFE/RL is consistently languishing, reaching the 2.9 flourishing level in only one period. We discuss possible reasons. We also consider some implications for perception management and counterterrorism strategy.**

Keywords- web mining; semantic networks; sentiment network analysis; time-series analysis; positivity ratios.

I. INTRODUCTION

When conducting text mining of open-source documents such as available in large databases, e.g. Lexis-Nexis, one has access to a wealth of over time information. Current patterns identified can be analyzed with respect to earlier time periods to provide evidence of reliability and predictive validity. Documents can be effectively analyzed over time with different methods, depending on the purpose of the investigation.

For example, a recent study [1] analyzed documents before, during, and after the Arab uprisings in Tunisia and Egypt. It found that countries that became more central in the networks of documents increased in the prevalence of 'jihad' concepts on their websites. This supported the expectation that the uprisings in the region would be seen as opportunities for jihadists to work for establishment of Islamist states and sharia law in countries that as yet had other political systems. Subsequent events in the region have been consistent with this assertion in a number of cases, most notably, Tunisia, Egypt, and Libya. Such a before/after quasi-experimental design can be easily constructed by open-source text mining over time.

Other purposes for different approaches to text mining documents over time include goals to assess propaganda, sometimes called "perception management," analyzing the opposition's communication as well as one's own, for use in evaluating counter-propaganda and counterterrorism information campaigns [2]. This paper illustrates that such goals can be addressed with a semantic network approach to measuring sentiment in time-based text mining.

First this paper will describe the methods. Second, it will explain the steps in their use. Third, we illustrate the application of the methods in analysis of five years of documents from transcriptions of websites, radio and television broadcasts, and news stories originated by the Taleban. We compare this to from Radio Free Europe/Radio Liberty Afghanistan Reports. Fourth, the paper discusses the implications of the results and potential applications.

II. RELATED RESEARCH

A. Sentiment Analysis

Measurement of sentiment in open-source information is currently an active area of research. Approaches are varied. (See Pang and Lee for a detailed review [3].) In the recent literature, some researches take entire Twitter tweets and give them to a panel of judges in Amazon's Mechanical Turk and have them make a binary judgment as to whether the entire tweet is positive or negative in sentiment [4]. They then use the coded sentiment values to predict different data. Other researchers have broken tweets down into individual words and used statistical methods to see which words are associated with increases in stock indices and with decreases. They use this as a means of inferring which words are positive and which are negative in sentiment based on whether regression coefficients are positive or negative [5]. Thus, predictive validity determines the sentiment value of words.

Paper presented to the Open-Source Intelligence and Web Mining conference [OSINT-WM 2012], Odense, Denmark, August 22-23, 2012.

An earlier method of sentiment analysis was implemented in the Linguistic Information and Word Count program, LIWC [6]. It uses a dictionary of stemmed words and includes among its 72 categories of measurement a positive emotion and a negative emotion category. The positive emotion stems dictionary contains 217 entries and the negative 350. Anytime one of these words appears it contributes to the cumulative sentiment score, which is the number of such words in each category per 100 words of text. Such a method is for the most part not sensitive to the linguistic context for the word. There are a relatively small number of contextual rules that determine sentiment based on occurrence of other words in the same sentence.

In a subsequent section this paper will describe a network-based method for sentiment analysis. It uses the entire semantic network from the text analyzed. A program, WORDij's [7] WordLink option first network analysis the entire text file. Then it automatically locates the network positions in this whole network of sentiment words from a list of 3,521. The analyst selects a seed word and the program traces the shortest paths across all words in the entire network to each of the sentiment words found in the network. The program computes the shortest path summary weights, based on the total of the link frequencies along the paths to the sentiment words. It then constructs a set of sentiment indices including the summary shortest path link frequencies for all of the positive and for all of the negative words. The index of interest in this paper is the positivity ratio, described in a subsequent section that explains in detail this Sentiment Network Analyzer (SNAZ) [8]. Before we describe it more fully and illustrate its use, we consider the relevant research-based theory about positivity ratios.

B. Positivity Ratios

Prior research on the implications of positive and negative emotion expressed in communication is relevant to our paper. Fredrickson and Losada [9] have studied the ratios of positive to negative affect expressed across all communication in a time segment measured for various system levels: individual, dyad, group, and organization. They find consistent evidence across these levels that when there is 2.9 times more positive than negative affect expressed in messages, the social unit is performing optimally over time. They call this state flourishing. "To flourish means to live within an optimal range of human functioning, one that connotes goodness, generativity, growth, and resilience." [10][11] When the ratio is below 2.9 the system has been found to operate less effectively and considered "languishing." Languishing fosters distress, impairment, and limitations in activities [12]. There is an upper limit on positivity above the threshold of 2.9. Social system units flourish with ratios upwards to ratio of 11.6. Above this level increasingly positivity leads to system destabilization [9].

Losada [13], Losada & Heaphy [15], Fredrickson and Losada [9] have used Chaos Theory and Lorenz equations [14] to further hypothesize that

> "1. Flourishing is associated with dynamics that are nonrepetitive, innovative, highly flexible, and dynamically stable; that is, they represent the complex order of chaos, not the rigidity of limit cycles and point attractors. 2. Human flourishing at larger scales (e.g., groups) shows a similar structure and process to human flourishing at smaller scales (e.g., individuals). 3. Appropriate negativity is a critical ingredient within human flourishing that serves to maintain a grounded, negentropic system. 4. The complex dynamics of flourishing first show signs of disintegration at a positivity ratio of 11.6. 5. Human flourishing is optimal functioning characterized by four key components: (a) goodness, indexed by happiness, satisfaction, and superior functioning; (b) generativity, indexed by broadened thought–action repertoires and behavioral flexibility; (c) growth, indexed by gains in enduring personal and social resources; and (d) resilience, indexed by survival and growth in the aftermath of adversity." [9]

In periods of languishing, the social unit loses behavioral and conceptual flexibility and the ability to question; it is stuck in self-absorbed advocacy [9]. Losada & Heaphy [15] have found that within organizations, positive experiences engender broader information processing strategies and greater variability in perspectives across organizational members. Sutcliffe & Vogus, 2003 [16] have found positivity predictive of organizations' resilience during periods of threat.

These results have led to naming of a "Losada Line" [13]. This line is drawn across a time series plot of positivity ratios over time. The value of the line is a 2.9 ratio. These particular ratios underlying the Losada Line, although tested with samples of individuals, dyads, groups, and organizations, have not been tested at higher levels when the organizations are large enough to represent a political system such as a nation or a large insurgency that has created shadow government features. It is interesting, therefore, to consider the case of propaganda from such systems in this research. The consistency of the cross-level findings leads to the expectation that the same processes occur at these higher levels.

I. TEXT SOURCES

One source this study used is content mined from the web service Lexis-Nexis which contains texts from the BBC International Monitoring Service that covers foreign language web sites, radio and television, and newspapers, translating them into English. This study mined the BBC translations of messages containing the word 'taleban' in its websites, radio broadcasts, news reports, and press conferences over a five-year period.

We compared these documents to a second source: the "Afghanistan Report" published by the U. S. government supported Radio Free Europe/Radio Liberty (RFE/RL). Formerly funded by the C.I.A. it has since been funded directly by the U.S. Congress. RFE/RL broadcasts news, information, and analysis to 21 countries in 28 languages in Eastern Europe, Central Asia, and the Middle East wherever "the free flow of information is either banned by government authorities or not fully developed" [17].

Radio Azadi, is RFE/RL's Afghan service was the most popular radio station in Afghanistan, in 2009, according to RFE/RE. Each month Afghan listeners mailed station hundreds of hand-written letters [18].

On January 15, 2010, RFE/RL began broadcasting Radio Mashaal in Pashto to the Pashtun tribal areas of Pakistan [19]. The station's goal is to counter the increasing number of Islamist extremist radio stations operating near the border between Pakistan and Afghanistan. These stations broadcast pro-Taliban messages and fatwas produced by pro-Taliban clerics [19]. Radio Mashaal says that it broadcasts local and international news with in-depth reports on terrorism, politics, women's issues, and health care (with an emphasis on preventive medicine). Included are roundtable discussions and interviews with tribal leaders and local policymakers. There are also regular call-in programs [20].

Unfortunately the RFE/RL transcriptions of all broadcasts are not available through open sources. But, we found 379 English stories about Afghanistan written between May, 2007 and May, 2012 in the RFE/RL online archives.

II. Research Questions

The basic research questions of this study are:

1) To what extent has Taleban propaganda shown evidence of flourishing or languishing from May 2007 to May 2012?

2) To what extent has U.S. propaganda directed to Afghanistan shown evidence of flourishing or languishing from May 2007 to May 2012?

3) How consistent are the positivity/negativity ratios over time for each system? If there is deviation, what is the direction of deviation: too much positive or not enough positive content?

III. Methods

A. *Census of Documents*

This research did not draw a sample of documents. Rather, we did a census of all documents that were from the BBC International Monitoring service for a five year period that contained the term: (jihad and taleban) or (jihad and afghanistan) or (jihad and pakistan) or (jihad and taleban and afghanistan) or (jihad and taleban and pakistan) or (jihad and taliban). There were 1,877 such documents. This was 29.6 megabytes of text.

For RFE/RL we found 379 English documents concerning Afghanistan in its archives over the same five-year period. This amounted to 3.2 megabytes of text. The procedures we used on these two data sets normalize then to remove the effects of different size files and content elements. This is accomplished by converting raw numbers to proportions based on the total size of each corpus. Normalization will be further described in later section.

The extraction of the complete available documents from each system eliminates selection bias. In addition, the various levels of social systems that Losada has studied, through organizational level are not fundamentally different from the larger system level we studied. According to systems theory, a fundamental assumption is that all human systems are open rather than closed systems and therefore subject to the same basic entropy and negative entropy processes.

B. *Time Slicing*

Our next step segmented the two files containing all of the respective documents into time intervals. For the Taleban documents there was sufficient text to construct a weekly time series of 262 weeks. Nevertheless, the RFL/FE documents were only numerous enough to construct a quarterly (3-month) time series. For comparison purposes we analyzed the Taleban texts with this interval as well. WORDij, a text analysis program developed over the last several decades [21][22][23][24][25][26][27], contains a TimeSlice program that does time segmentation the user chooses and inserts codes into the large text file representing the time segments. Then when WordLink extracts word pairs from a document, in a single run it processes all of the time segments, or whatever else in other research might be the basis for creating subsets of text. This eliminates the need to repetitively run the program in such cases. For example, the weekly Taleban analysis would have required 262 separate runs, had there not been this efficient WordLink feature.

C. *Extracting Word Pairs*

Following time segmentation, we ran a semantic network analysis procedure to produce a word pair file for each time slice. Word pairs are the basis for creating networks. The words are the nodes and the frequency of cooccurrence of the word pair is the link strength. Overlaying all of these pairs, produces a network connecting them according to which pairs occur closely and how often. WordLink extracts word pairs. Figure I shows basic options of the program, while Figure II shows advanced options. Actual parameter settings used in this paper appear in Table I. Next we discuss the main options of WordLink and how we used some of them in this research. This will give the reader an understanding of what kind of semantic network analysis we have done.

1) *Select Text File*

The first WordLink run used a text file that contained all of the documents for the Taleban texts. These were marked with time slice headers inserted in the TimeSlice utility, so that WordLink would run in a single pass separate analysis for each of the slices using the same parameters. The RFE/RL text was separately treated the same way.

2) *Drop List*

A drop list, often called a “stop list” in the Information Retrieval community, contained grammatical function words with little substantive meaning, prepositions, pronouns, conjunctions, etc. If these had been left in, the network links would be so highly dominated by these words that the network would be like a bowl of spaghetti, with very little differentiation of structures, and not usable.

3) Pajek Output

The WordLink program was set to output a .net file containing the node labels, and edge list in a format that is compatible with a variety of network analysis and graphing programs. This is the Pajek [28] input file format. We have it checked by default but did not use it in this analysis.

4) Drop Low Frequency Words and Pairs

In natural language processing one typically drops words and pairs that occur only once or twice. There are many such words, but they do not contribute much to the later network construction. In fact, they detract from it because with them the graph practically unreadable. The default setting is to drop words and pairs with frequencies less than 3.

5) Link Strength Weighting Method

The default is set to treat all links the same, no matter how far apart the words in pairs are. Nevertheless, there are options to weight the links by squaring them or by using an exponential function, based on how far apart the words are. We rarely use this feature because the default word pairing window is set to three words wide on either side of each word in the text. Such short distances do not merit weighting.

6) Window Size for Extracting Word Pairs

The word proximity criterion can be thought of as a word window that slides through the text, counting all word pairs inside as it moves from word in the full text. It captures pairs 3 positions before each focal word, as well as those within 3 words after it. The window size can be increased, which may make weighting desirable. There is no practical limit on how large the window can be. In early experimentation, however, we used a reference network and then systematically incremented the word window size from 1 on up to 32 and ran a network analysis for each increment. The networks at 1 and 2 were significantly different from the reference network. At size 3 up to 32, however, the results all matched it. So, we chose 3 as the default because it minimizes output files sizes yet delivers a valid network representation.

This windowing approach extracts pairs of close words. This retains contextual information. As well, word order is typically preserved. This results in many grammatical rules and pragmatic usage being imbedded among the word pairs. For example, in English, “white horse” is more likely in discourse than “horse white.” Ordering makes it easier to interpret possible meanings of the networks. This is unlike the “bag of words” approach that treats all words in a document as linked.

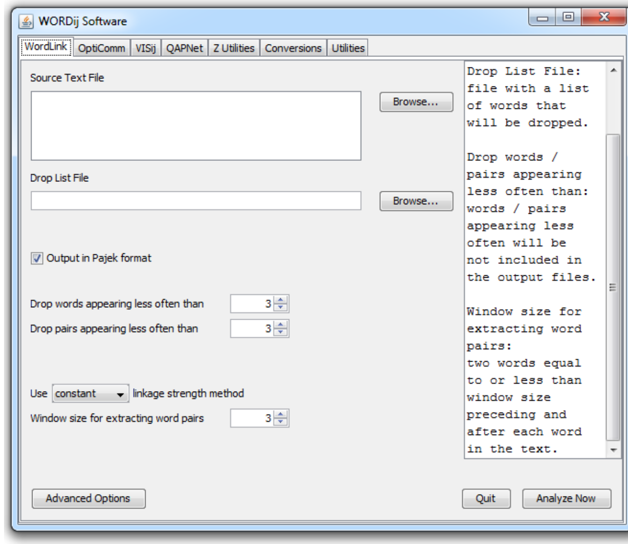

FIGURE 1. WORDLINK BASIC OPTIONS

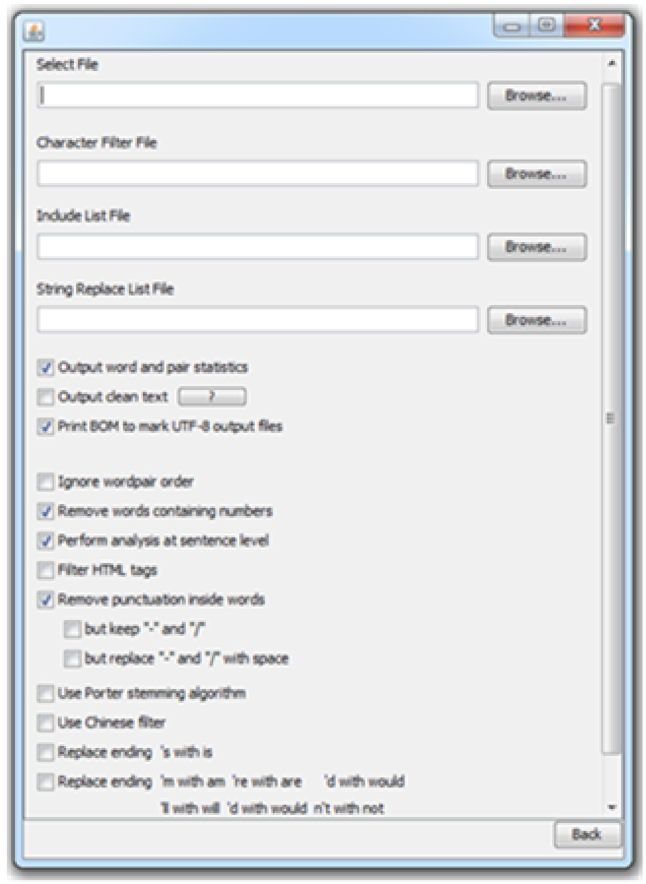

FIGURE II. WORDLINK ADVANCED OPTIONS

7) Select file

The select file is produced by automatic time slicing in the TimeSlice program. It tells WordLink which header codes that are inserted throughout the text to use in what order. Codes need not be time-based or automatically inserted. One can insert them manually means to separate differ files, such as those produced by attribute data such as characteristics of the social units that originally produced the texts. We used the TimeSlice created select file.

8) The Character Filter File

WORDij processes UTF-8 texts, so it handles the various languages that use graphic characters such as Arabic Chinese, Japanese, and Korean. Nevertheless, sometimes text contains special format characters that are not in the UTF-8 format. The character filter contains a list of such characters and removes them from the analysis.

9) The Include List

Opposite to a stop or drop list is the include list. One places words in the include list that one wishes to see connections among in the text files. All other words are removed from the analysis. For example, in one study [26] we were interested in analyzing the cooccurrence of Muslim countries in news documents so we could map an international network containing only those countries.

The other options listed on the second half of the Advanced options page of the WordLink program are hopefully self-explanatory and will not be described in the interest of space limitations.

TABLE I. LOG FILE FROM WORDLINK RUN

```
Text file name: C:\Users\jad\Downloads\BBC
Taleban Jihad\
bbctalebanQ.txt
Configuration:
    Drop list file name:
C:\Users\jad\Downloads\WORDij\droplist.txt
    Include list file name: none
    Character filter file name:
C:\Users\jad\Downloads\WORDij\WORDij\
    character_filter_UTF_10082011.txt
    Select list file name:
C:\Users\jad\Downloads\ \BBC Taleban
Jihad\bbctalebanSQ.txt
    Drop words less frequent than: 3
    Drop word pairs less frequent than: 3
    Preserve wordpair order: yes
    Include numbers as words: no
    Link until sentence end: yes
    Link steps: 3
    Linkage Strength Method: CONSTANT
    Remove punctuation: no
    Compound words: keep
    Using Porter stemming algorithm: no
    Using Chinese filter: no
    Replace English contracted forms: no
```

A. Network Sentiment Analysis

The SNAZ [8] program takes a set of positive (n=1,201) and negative sentiment target words (n=2,323), and given a seed word, traces the weighted shortest paths from the seed to each of the sentiment words, then computes the sum of the weights of the positive and of the negative paths. We discuss specifically how we do this in a subsequent section. The original target sentiment words were selected in two ways. One was to take the LIWC sentiment dictionary items and de-stem them unto all lexical variants. Two was to select from a comprehensive list of English language modifiers all words and lexical variants that had clear positive or negative meanings. The source we used for this was the Automatically Generated Inflection Database (AGID), fourth revision, January 3, 2003. The author, Kevin Atkinson, describes it as an automatically created database of inflections from a large word list [29].

1) The basic process – step 1: finding the shortest path

Walk the network to find the minimum depths of all nodes from the seed node:

- initialize all node depths to zero (unassigned);
- each node keeps a list of its predecessor notes (for use in Step 2): initialize these lists;
- set the depth of the seed node to 1;
- descend into the network breadth first, giving any unassigned node a depth one greater than that of its predecessor(s);

2) Step Two: The basic process – step 2: extracting the shortest paths to each target

Back-track from each target to the seed node:

- repeatedly work back from the target node via each of its predecessors to the seed, keeping a list of the nodes encountered;
- when the seed node is reached, reverse the list and calculate its score.
- add the node number of each node to the predecessor lists of it successor(s);
- the process terminates at the depth where no unassigned nodes are found;

All the shortest paths from the seed node will now have been found.

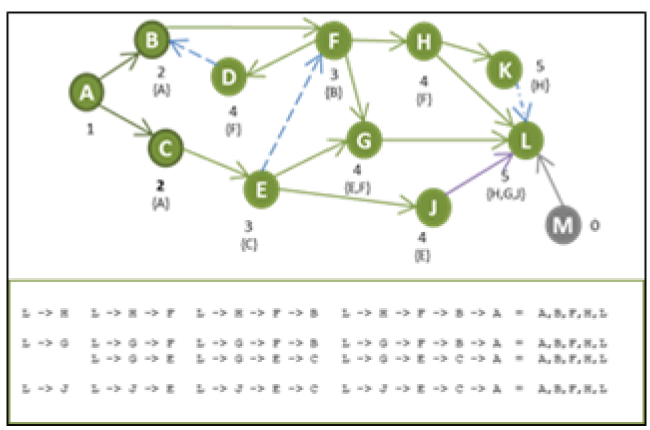

FIGURE III. EXAMPLE OF SHORTEST PATH COMPUTATIONS

In figure III, M is the target semantic word and A is the seed word.

B. *Computing Positivity Ratios*

SNAZ has the user select a seed word and it then traces the shortest path in the actual network from the text being assessed for sentiment to each of the positive and negative target words. It computes the summary word pair strengths for each shortest path from seed to target.

1) *Normalization of Text Variables*

We normalized the sentiment elements to remove the effects of different text file sizes and different occurrences of sentiment words. This was performed when computing the positive to negative sentiment ratios in each of the files. The values for the positive and negative weights found were not divided by a demoninator that is the total possible sentiment words, those contained in the 3,5121 word sentiment target file. Instead we divide the respective sentiment weights found by the number of unique sentiment target words found. We also normalized at a second level when the first normalized values for positive sentiment are divided by the normalized values for negative sentiment, as computed in the previously described step.

2) *Run Time*

SNAZ makes these sentiment calculations efficiently in a single run for a large number of time slice networks or networks based on other criteria. For example, the program computed sentiment values for the 262 time slice Taleban networks in 14.6 seconds on a PC. (These data were not used in the present analysis but the run time is reported for your information.) SNAZ ran the 21 quarterly semantic networks used in 0.7 seconds.

IV. RESULTS

Figure IV shows the positivity ratio for each quarterly period for both the Taleban and RFE/RL sources. Out of 21 time periods the RFE/RL documents were below the optimal 2.9 ratio in all slices except for one when the value was 3.0. This shows a pattern of languishing over 20 periods across five years. In contrast, the Taleban positivity scores were always above the threshold of 2.9. In three time slices the values exceed the top boundary of the optimal positivity ratio of 11.6 with a score of 26.3, and two others at a two quarter lag of approximately 15 each.

At these higher levels extreme positivity work in Losada's work with Chaos Theory has found that the system destabilizes. Nevertheless, for 19 of the 21 time slices the Taleban data is in the flourishing zone. For RFE/RL Figure IV shows that in quarter 12 the single time that RFE/RL reaches the flourishing level. In the next quarter the Taleban score shoots into the highly excessive level of 26.3, and then oscillates for the remaining 8 quarters. Prior to that move the Taleban values are quite consistently evidence of flourishing.

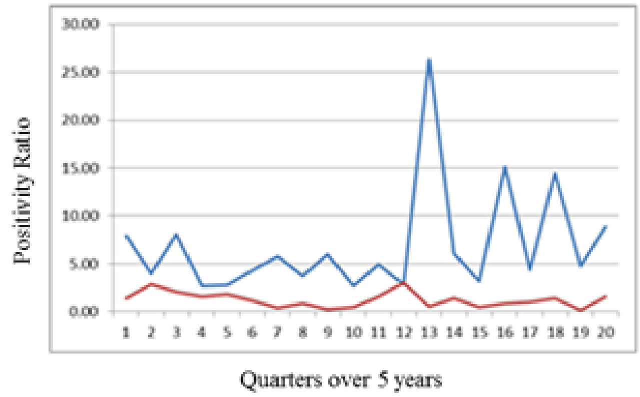

FIGURE IV. POSITIVITY RATIO FOR TALEBAN (BLUE LINE) AND FOR RFL/RE (RED LINE) OVER 21 QUARTERS

Afterwards, the Taleban system destabilized into a period of oscillation which continues to the present this study was completed. Although it never drops into the languishing levels, it bounces beyond the flourishing range in three more regular cycles that are two quarters long.

V. DISCUSSION

This paper focused on computing the positivity ratio for the Taleban Afghani and Pakistani propaganda and did the same for U.S.'s RFE/RL propaganda about Afghanistan over a five-year period. A quarterly time interval was used for segmenting semantic networks analyzed with WORDij's WordLink, and with the Semantic Network Analyzer (SNAZ). Based on Losada's work we computed positivity ratios at each time slice for each of the two systems compared. We found that the RFE/RL positivity was consistently very low, in the floundering range, except for one time slice when it reached the flourishing level.

On the other hand, the Taleban content was consistently above the threshold for flourishing. This threshold is a positivity ratio 2.9 times more positive than negative. Nevertheless, all was not well with the Taleban. In the 13th of 21 quarters the Taleban content became excessively positive with a ratio of 26.3, exceeding the flourishing top boundary of 11.6 by more than double. After that there was evidence for chaotic system destabilized oscillation occurring on a two-quarter cycle, at which times positively exceeded the top threshold for a healthy system.

Too much positive emotion may indicate a period of systemic mania. Some of the features observed at the individual level, such as: loss of social constraints, preoccupation with grand plans or schemes, hyper religiosity, hyper vigilance, and suspiciousness [30] may generalize to the macro system level. Earlier findings at the group and organizational levels may also have parallels with macro-level system mania. In periods of languishing, the social unit loses behavioral and conceptual flexibility and the ability to question; it is stuck in self-absorbed advocacy [9]. Within organizations, positivity outside the flourishing range engenders narrower information processing strategies and less variability in perspectives across organizational members. Organizations' resilience during periods of threat is reduced. These effects may generalize to the larger system level, for which evidence from future research is needed.

For the Taleban, the initial excessive positive spiking may relate to the resignation during this period of General Stanley McCrystal as Commander, International Security Assistance Force (ISAF) and Commander, U.S. Forces Afghanistan (USFOR-A). A manual reading of the documents to see if this was evident showed that his resignation generated considerable commentary including positive excitement regarding how this was linked with evidence that the Americans were losing the Afghan war as Taleban successes were increasing. While the semantic evidence supports this interpretation, there may be other unknown causes.

For the U.S., is interesting to note what occurred in period 12 in which the RFE/RL documents reach their highest positivity score. One event was the attempted Time Square New York bombing by Faisal Shahzad. He stated that this had been planned and supported by the Taleban in Pakistan, and evidence reported from Pakistan confirmed this. This event and its treatment in the press may account for a significant portion of the positivity of the RFE/RL materials in this period. Nevertheless, this is only speculative.

It is not clear who the intended audiences are for RFE/RL's Afghanistan Reports. Those exposed to this information received content that portrayed the U.S., at least with respect to Afghanistan, but perhaps more generally, as a political system that was languishing rather than flourishing. It is not known what strategy, if any, drives such a presentation. It may result from the journalistic norms in the U.S. that characterizes news more in negative terms than positive. It is possible that these norms are operating without conscious strategizing by RFE/RL information managers. Or, they may be consciously following U.S. journalistic norms for how to write stories. A further speculation may be that for domestic political reasons the administration wants to show low enthusiasm for the activity in Afghanistan because this is appealing to a large proportion of domestic constituents who feel this way. Nevertheless, this research offers no evidence to enable a defensible proposition about the reasons for the low levels of U.S. Afghanistan Reports positivity.

To the extent the languishing portrayal of the U.S. in Afghanistan reaches Afghan audiences, they may become more negative about the U.S. and allies' activities in their homeland. This may make it more difficult to recruit and train Afghan army, police, and security forces, a key component of the U.S. exit strategy. Whether this is the case requires further research.

At the same time, the more positive, flourishing representations by the Taleban may gain extra traction among Taleban audiences in an environment fostered by the activities of the foreign troops. The Taleban members themselves may be particular energized to engage in insurrectionist acts by the positive portrayals these receive. Perhaps the periods of excessive positivity further promote terrorist acts. Recruitment of members may also flourish, perhaps even for suicide bombers. These possibilities may be supported by further research, or rejected in favor of other interpretations.

For U.S. counter-propaganda strategies it is interesting that if one projects the Taleban pattern of undamped oscillation of the positivity to the next quarter, it would appear to offer an opportunity to enhance the instability. If the pattern, which would then have repeated four times over a two year-period, continues, further opportunities may be expected on a predictable basis, provided the effects of excessive positivity are generalizable to the Taleban case. If so, the timing of activities such as peace negotiations and reconciliation talks can be accordingly optimized along with message content strategies. Given the characteristics of mania, avoiding contact during periods of Taleban excessive positivity may be particularly effective in minimizing obstacles to progress. Such possibilities require an empirical grounding.

VI. CONCLUSION

This paper has demonstrated the Semantic Network Analyzer (SNAZ), a novel network-based method for quantifying sentiment in collections of open-source documents, In addition, theories of sentiment based on Fredrickson & Losada's theory of positivity ratios have been applied to the macro-system level in our analysis of the propaganda of the Taleban and of Radio Free Europe/Radio Liberty. The Taleban content generally showed evidence of system flourishing, while the RFE/RL content generally evinced languishing. The most notable exception is the spike in Taleban positivity which began an unstable period of oscillation from the third quarter of 2010 to the present.

The methods illustrated in this example for analysis of five years of open-source documents connected with Afghanistan may prove useful in analyzing other content to assess positivity ratios of groups, organizations, and larger political entities. These ratios provide a new application of Losada's principles to assessing propaganda. Its positivity may influence receivers' perceptions of system flourishing or languishing. Counter-propaganda and counterterrorism strategy may be informed by considering these characteristics of opponents' communication. Further research is required to gauge the various speculative interpretations we have made about the findings, perhaps supporting their validity but also providing alternative explanations.

REFERENCES

[1] J. A. Danowski, "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,". IEEE Proceedings of the Intelligence and Security Informatics Conference (EISIC), 2011 European, Athens, Greece, pp. 314-321, September 12-14, 2011.

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

[3] Bo Pang and Lillian Lee, "Opinion mining and sentiment analysis," Foundations and Trends in Information Retrieval, vol. 2, issue 1-2, pp. 1-135, January 2008.

[4] 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.

[5] 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, The 2nd Collaborative Innovation Networks Conference - COINs2010, vol. 26, pp. 55–62, 2011.

[6] J. W. Pennebaker, R. J. Booth, and M. E. Francis, M. E. Linguistic Inquiry and Word Count (LIWC). Austin, TX, USA: liwc.net, 2007.

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

[8] J. A. Danowski, "SNAZ software for semantic network analysis," Naperville, IL: Communication Research Network, 2012.

[9] B. Fredrickson and M. Losada, "Positive affect and the complex dynamics of human flourishing, American Psychologist, vol. 60, no. 7, pp. 678–686, 2005.

[10] C. L. M. Keyes, "The mental health continuum: from languishing to flourishing in life," Journal of Health and Social Behavior, vol. 43, pp. 207–222, 2002.

[11] S. Lyubomirsky, L. King, and E. Diener, E., "The benefits of frequent positive affect: does happiness lead to success?" Psychological Bulletin, vol. 131, no. 6, pp. 803-855, 2009.

[12] B. L. Fredrickson, M. M. Tugade, C. E. Waugh, and G. Larkin, "What good are positive emotions in crises? A prospective study of resilience and emotions following the terrorist attacks on the United States on September 11th, 2001," Journal of Personality and Social Psychology, vol. 84, pp. 365–376, 2003.

[13] M. Losada, "The complex dynamics of high performance teams," Mathematical and Computer Modelling, vol. 30, no 9–10, pp. 179–192, 1999.

[14] E. N. Lorenz, E.N., The essence of chaos, Seattle: University of Washington Press, 1993.

[15] M. Losada and E. Heaphy, "The role of positivity and connectivity in the performance of business teams: a nonlinear dynamics model," American Behavioral Scientist, vol. 47, no. 6, pp. 740–765, 2004.

[16] K. M. Sutcliffe and T. J. Vogus, Organizing for resilience. In Positive organizational scholarship: Foundations of a new discipline, K. S. Cameron, J. E. Dutton, and R. E. Quinn, Eds. San Francisco: Berrett-Koehler, 2003, pp. 94–110.

[17] Radio Free Europe/Radio Liberty, FAQ (Retrieved May 12, 2012 from http://www.rferl.org/section/FAQ/777.html#section5).

[18] "Poetry from Praktia to Prague," Off-Mic Blog, Radio Free Europe/Radio Liberty (Retrieved May 12, 2012 from http://www.rferl.org/content/poetry_from_paktia_to_prague/1811546.html.)

[19] "Holbrooke at RFE event: 'Deal with Taliban propaganda head-on'," Radio Free Europe/Radio Liberty. FLASHBACK, Sep. 18, 2009. (Retrieved May 12, 2012 from http://www.rferl.org/content/Holbrooke_Reception/1826309.html).

[20] "RFE/RL Launches Radio Station in Pakistan's Pashtun Heartland," Radio Free Europe/Radio Liberty, 2010. (Retrieved May 12, 2012 from http://www.rferl.org/content/Press_Room_Press_Release_Radio_Mashaal_Launch_/1929887.html).

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

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

[29] K. Atkinson, "GNU Aspell.," aspell.net, 2011.

[30] W. Mansell and R. Pedley, The ascent into mania: A review of psychological processes associated with the development of manic symptoms," Clinical Psychology Review, vol. 28, no. 3, pp. 494-520, 2008.