Analyzing Change Over Time in Organizations' Publics with a Semantic Network Include List: An Illustration with Facebook

James A. Danowski
University of Illinois at Chicago
1007 W. Harrison St., MC 132
Chicago, IL 60607 USA
jdanowski@gmail.com

***Abstract*—This research highlights a kind of semantic network analysis based on an include list. We analyze the networks only among words on the list as they appear in a series of text corpora word pairs for an organization. The example uses documents about Facebook over a 12-month period, dividing them into 12 time-based files. In each time slice we map networks among key publics and measure the centrality of each from one time period to the next. The network of publics becomes more complex across time. Publics fluctuate in centrality. We describe other kinds of semantic network analysis for business applications using include lists.**

Keywords-semantic network analysis, network analysis in business; include lists

I. INTRODUCTION

A. Semantic Network Research in Business Organizations

Businesses can use semantic network analysis in a variety of ways. They can map the relationships among departments that appear across company documents to show the functional structure [1] and compare it to the formal structure. Analysts can network analyze email in terms of who sends messages to whom [2] and further, what the network of concepts is across these messages [3][4][5]. They can process email from customers to identify the main topics about which they are writing [6]. Using internal documents, the locus of different semantic network concepts can be overlaid onto the hierarchical levels of the organization to see how this matches with desired hierarchical communication. Positivity of semantic networks can give information about the extent to which an organizational unit or the whole organization is languishing or flourishing [7].

B. Research Focus

This paper focuses on another application, using semantic networks on documents about an organization to identify the networks of internal and external publics as these are co-mentioned in news stories and web documents. We derive a large list of potential publics from a survey of public relations practitioners. This enables the "include list" method for mapping networks of publics that appear for business.

Strategic analysts and public relations officers of a large business may find it challenging to systematically track key publics that appear in news and web documents about the organization. This paper shows one way to do it. Automated semantic network analysis of documents can efficiently do such tracking on a regular, frequent basis and apply statistical tests to identify significant changes.

To illustrate this approach, one year's worth of documents focused on Facebook are analyzed with this include list. Slicing the year into monthly intervals enables investigation of change over time in the centrality of publics for the organization.

In the sections to follow we first provide a conceptual and operational definition of 'include list' and distinguish it from 'ontology.' Following this we briefly summarize 10 steps describing how this kind of network research can be done. The detailed methods follow. Then we present results. Next we discuss them, including limitations and future research directions.

C. Include Lists

An include list is the opposite of a 'stop' or 'drop list' that contains terms typically removed from basic semantic network analysis. When analyzing the full text of a corpus of documents, semantic network researchers typically drop words that carry little meaning because they are basic grammatical function words, i.e. 'that, to, and, etc." Leaving these words in the analysis results in a network that is dominated by these function words that link to so many different other words that the network looks like a bowl of spaghetti.

The opposite of a 'stop list' is an 'include list.' It contains only those words among which the analyst wishes to define a network. All other words are dropped from the analysis. Thus, the include list is applied to the full text corpus, here the documents about Facebook, to find only the links only among the words on the list as they occur in the particular set of documents.

D. Related Work

Danowski and Cepela [8] used an include list approach to map the internal organizational structure of U.S. presidents' cabinets. The include list contained the names and aliases of all cabinet members for an administration. They ran this list on the full text of news articles in the *New York Times* and the *Washington Post* that mentioned any cabinet member. This produced a b-weekly network of cabinet members based on how frequently pairs of cabinet members appeared together in news stories. This was done for the cabinets of Nixon through Obama. Hypotheses were tested about changes in Gallup presidential approval ratings following changes in the centrality of the president.

Another study [1] identified the networks of collaborating departments of a college for an accreditation review. The include list of department names and aliases was applied year by year to local news stories mentioning the college over a four-year period.

There is considerable attention in the literature to corporate branding, corporate identity, corporate image, corporate social responsibility, corporate reputation, and crises. A study [9] tested hypotheses about these concepts in relation to corporate reputation. The researchers selected the top 30 and the bottom 30 corporations in reputation from an annual rating of 600 world corporations based on surveys. They analyzed the semantic networks of words in an include list containing crisis, the corporate communication terms, and the corporation names. Twelve months of news and web documents about the 60 organizations formed the full text corpus for the include list runs. Shortest path analysis found that top reputation corporation names were significantly closer to these communication terms and further from the crisis term, compared to the bottom-ranked corporations.

In an interorganizational study, organizations listed on the White House's web pages about the Gulf oil leak were network analyzed based on an include list run with related news stories in daily Gulf-area newspapers using a weekly time interval [10]. The hypothesis was supported that the more central BP was in the interorganizational network, the more negative was the sentiment in the news stories.

International networks have also been identified with include lists [11] run on translations of Muslim web sites, broadcasts, and newspaper stories by BBC International Monitoring, as well as on documents appearing in English in major world publications. Analysts conducted a naturalistic field experiment across three time periods based on before, during, and after the early Muslim Middle-East and North Africa uprisings. They found that political Islam concepts strengthened for countries that became more central in the network of Muslim nations. As centrality of a nation increased it had an increased presence of 'Jihad' and 'sharia' on its web pages.

E. Ontology Contrast with Include List

The term 'ontology' in computer science has some similarity to an 'include list' but is not a synonym. An ontology is a "…representational vocabulary for a shared domain of discourse — definitions of classes, relations, functions, and other objects [12]. An include list has some of these features, primarily the specification of members of a class. Rather than these elements having pre-specified relations as in an ontology, in an include list the elements are connected through a single *a priori* relation identifying the class. For example, the names of an organization's departments share the relation of co-membership in this social unit. More detailed relations among subsets of the class are typically not specified in advance of text mining. The frequencies of cooccurrence of the include list elements in the analyzed text corpus create a network that provides an empirical basis for more specific relations among elements. This contrasts with an ontology's qualitatively constructed knowledge system that specifies the particular relationships among class elements. With the include list approach, patterns of observed links give each element relational properties shared with subsets of other elements. Various structural measures, such as degree (number of links in the network), various measures of centrality, membership in clusters, groups, or communities, etc. can further elaborate class elements' relationships.

Unlike in ontological text analysis, the links are not based on some qualitative, ad hoc, "arm chair" specification of relationships among elements in some domain. For example, an ontology of terrorism might include categorical slots such as the types of terrorists, the different kinds of terror acts, the various targets of terror, objects used in the acts, and the means of implementation of the acts. Then, by searching some corpus with the ontology, the text mining software fills these slots based on each particular terror event found. Here is a narrative illustrating ontology slots filled by an event: a state-sponsored terror group kidnaps Western tourists as hostages from a hotel, then beheads them with a sabre, and sends the heads to the embassies of the tourists. Having completed filling slots for one relevant event, the software searches for another terror event to process, a procedure that is repeated until all of the text is processed.

In contrast, with an include list the relations among the class of elements are not domain specific. They are based on the generalized, context-free network formalisms from graph theory and/or social network analysis. The occurrence of the elements and their structural positioning and associated properties are instantiations of network analysis' transcendent body of assumptions and conceptual and operational definitions.

An include list could be about anything of interest to a company and built specifically for it. It could contain names of competitors, community organizations, people, issues, products, and attributes, etc. It need not be defined based on a prior survey. Business executives could create the list in some other way, such as through discussion in a meeting, interaction carried out over email, or in some other fashion. This study happened to use the survey approach because of the particular project goals for the original collection of data. Any sort of include list can be the basis for a semantic network study of the kind we focus on in this paper.

F. Steps in the Process

The basic steps of the approach are as follows:

- Develop an include list of words of interest. (In this case we used 145 names of various internal and external organizational publics found in a survey of public relations practitioner, plus the name of the organization of interest. Each public is entered into a UTF-8 file, one public per line of the file.)

- Identify the sources of full text documents to be analyzed with the include list. (Here we extracted documents from Lexis-Nexis' "Major World Publications" about Facebook over a one-year period.)

- Build a file of the full text documents. (Lexis-Nexis limits the size of download files to 500 documents each. After completing all downloading we combined the files into a single one for further analyses.)

- Remove duplicate documents. (DeDup [13] is a program we developed for this purpose.)

- Time slice the text file into standard intervals. (WORDij's [14] TimeSlice program enabled us to insert time stamp headers marking one-month intervals.)

- Run on the file a semantic network software package that incorporates the include list option. (We used WORDij's WordLink program option for specifying an include list.)

- Network analyze the include list words and their frequencies of cooccurrence in the documents in each time slice. (WordLink ran the include list against the full text of each time slice and counted word pairs within a sliding window through the text, rather than using a "bag of words" approach.)

- Compute statistical measures of the network, for example: centrality of the include list terms. (We imported the list of found word pairs and their frequencies for each time slice into UCINET [15] for computation of betweenness centrality [16].)

- Graph the networks of the include list word-pair frequencies in each time slice. (NetDraw [17] enabled us to graph the 12 monthly networks using standardized spring embedding layouts.)

- Analyze the changing positions of publics over time. (For Facebook we plotted the time series for an external public, "users," and an internal public, "employees.")

WORDij's WordLink program, in addition to having the include list functionality, has a string conversion utility. One can convert phrases containing multiple words into a one-word unigram. For example: 'European Union' could be converted to the single term: 'European_Union.' Aliases could also be converted to a common term. For example, EU could also be converted to 'European_Union.'

II. METHODS

A. Include List Construction

The include list was constructed by first surveying 343 public relations practitioners who were members of the Public Relations Society of America (PRSA). One of the items in the online survey asked for an open-ended response: "Please list your key publics." Seven boxes were available for respondents to enter text. We compiled the list of entries across all cases then removed names that occurred only once or twice. The result was a list of 145 key publics. This, plus the name of the business studied, became our include list run against the news stories about it to see what the network of publics looked like and to measure its structural properties, such as the centrality of the publics.

B. Organization Studied

To keep the illustration simple, we studied a single organization. We chose Facebook as the organization for our demonstration. Because Facebook has been in the process of recently developing an Initial Public Offering (IPO) of stock, we wished to see how this event was reflected in the changing publics discussed in the press about Facebook. The time frame chosen was one year from May 25, 2011 to May 25, 2012.

C. Text Collection

We used Lexis-Nexis Academic (http://www.lexisnexis.com/en-us/home.page), the largest database of world news and other documents, as our source for text mining. We selected all documents in the Major Publications category, which includes material produced around the world. We extracted the documents for the one year period and downloaded them to a PC.

D. Search Strategy

The search term we used was: ATLEAST10(facebook), repeated for each month. This yielded 3,402 documents comprising 17 megabytes. The first part of the search term specified that we only wanted to extract documents in which Facebook was mentioned at least 10 times. The reason for this was that the resulting documents would be likely to focus considerable attention on Facebook, rather than discussing a variety of organizations and mentioning Facebook only once or a few times. For example, our search term excluded a common type of document that lists the daily stock prices of a list of organizations, in this case that happened to mention Facebook, too. Even though it did not begin selling stock until the 12th period, there was considerable brief discussion of its potential share value and other comments in larger reports focused on diverse topics across the year. It would be more valid for the analysis to select documents focused primarily on Facebook,

hence the choice of at least 10 mentions in a selected document. This was arbitrary threshold based on previous exploration. Ten mentions consistently produced the desired focus, although there is no particular logic for 10 being the preferred number.

E. Removing Duplicate Documents

We downloaded the articles for the 12 months but put all of the text into a single file for redundancy removal. Because there is usually redundancy of articles found in Lexis-Nexis Academic, due to some sources picking up the same wire service text or other sources publishing the same story in multiple editions, this can distort the analysis. In the commercial version of Lexis-Nexis there is a command to remove redundant documents retrieved for a search. Nevertheless, the academic version does not have this feature. We therefore constructed a program, DeDup, to remove redundant text. This was run on the Facebook text file. The original file of 17 megabytes was reduced to 11.48 megabytes.

F. Time Slicing Documents

We used the TimeSlice program in WORDij. It allowed us to automatically insert time code headers into the large aggregated file. We chose a monthly time interval. Each inserted header indicated where a new month of documents began. This way we could set the parameters for one network analysis run, instead of doing 12 separate runs.

G. Word Pair Extraction and Network Creation

Following time slicing, we next used the WordLink program in WORDij. It counts word pairs appearing close together, preserving the order in which they occur (We have found that a sliding word window 3 words wide is optimal for extracting word pairs). We inserted into WordLink the include list that contained the word 'facebook' and each of 145 words indicating names of publics. Table I shows the log file containing parameters.

TABLE I. LOG FILE

```
Text file name: C:\Users\jad\Downloads\ASONAM BASA\facebookm.txt
Configuration:
Drop list file name: none
Include list file name: C:\Users\jad\\Publics\Public include list.txt
Character filter file name: none
Select list file name: C:\Users\jad\Downloads\BASA\facebooksel
Drop words less frequent than: 3
Drop word pairs less frequent than: 3
Preserve word pair order: yes
Include numbers as words: no
Link until sentence end: yes
Link steps: 3
Linkage Strength Method: CONSTANT
Remove punctuation inside words: yes
Compound words: combine
Using Porter stemming algorithm: no
Using Chinese filter: no
Replace English contracted forms: no
Replace 's ending by is word: no

The program processed the file in 1.15135 minutes.
```

III. RESULTS

WordLink output a list of word pairs and frequencies among the words on the include list. The most frequent of these appear in Table 2. The creation of 12 monthly time slices resulted in sub files ranging in size from 843 kilobytes to 2.8 megabytes. While it would be interesting to present each of the 12 network graphs, space does not permit it. So, we included graphs for the first month (Figure II), the 6th month (Figure III), and the last month's (Figure IV). Nevertheless, following these figures are the results for each of the 12 month's statistical computations on centrality of publics. In creating each month's graph, we dropped word pairs appearing less than 10 times to increase the clarity of the illustrations. Darker links indicate stronger links.

One can readily observe in comparing Figures II, III, and IV that the first two networks are simpler in structure than the third. The first two networks contain fewer publics and the overall structure is quite centralized. The network at time two contains the same number of publics but 3 are in a separate component. Examining the publics in Figure IV reveals that a number of investment-related ones, which was during the time that Facebook made its Initial Public Offering of stock. The first network contains no such publics. The network in Figure III has more publics and denser linkages.

A. Centrality of Publics

In UCINET we computed the Freeman Betweenness [16] scores for each public in each of the 12 networks. Table II shows the top 20 most central publics for each of the two networks across the 12 time periods. In Table II one can observe the publics that remained relatively stable in centrality over time and those that varied more. As an example of varying publics, Figure I shows the centrality of 'users' and 'employees.' Table III contains the most frequent word pairs from the aggregate text file across all months so that the reader can see how the include words are paired.

IV. DISCUSSION

This paper has demonstrated how a business can map the semantic networks of entities of interest. While we could have analyzed the network of all words appearing close to one another, this would have produced a very large network containing many word pairs not relevant to identifying networks of publics. In fact, it would have required extensive manual processing of such results to locate the publics. Instead we used a more efficient method. We did the semantic network analysis of only those words appearing on an 'include list.'

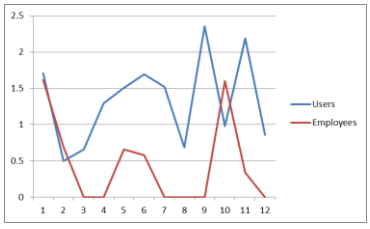

FIGURE I. Centrality of Users and Employees over Time

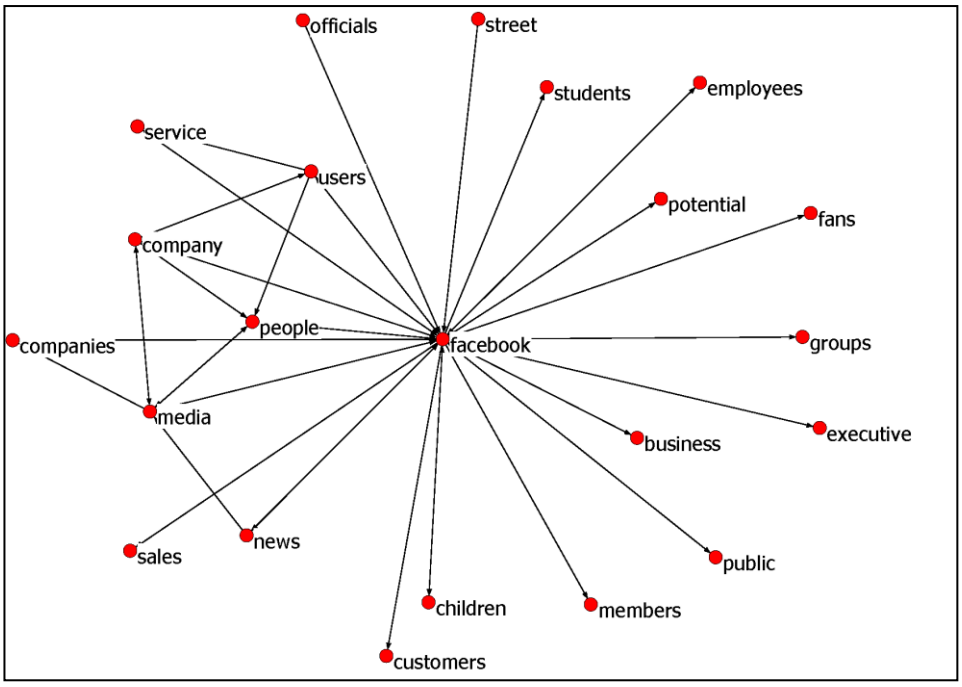

FIGURE II. NETWORK OF PUBLICS FOR APRIL 25-MAY 25, 2011

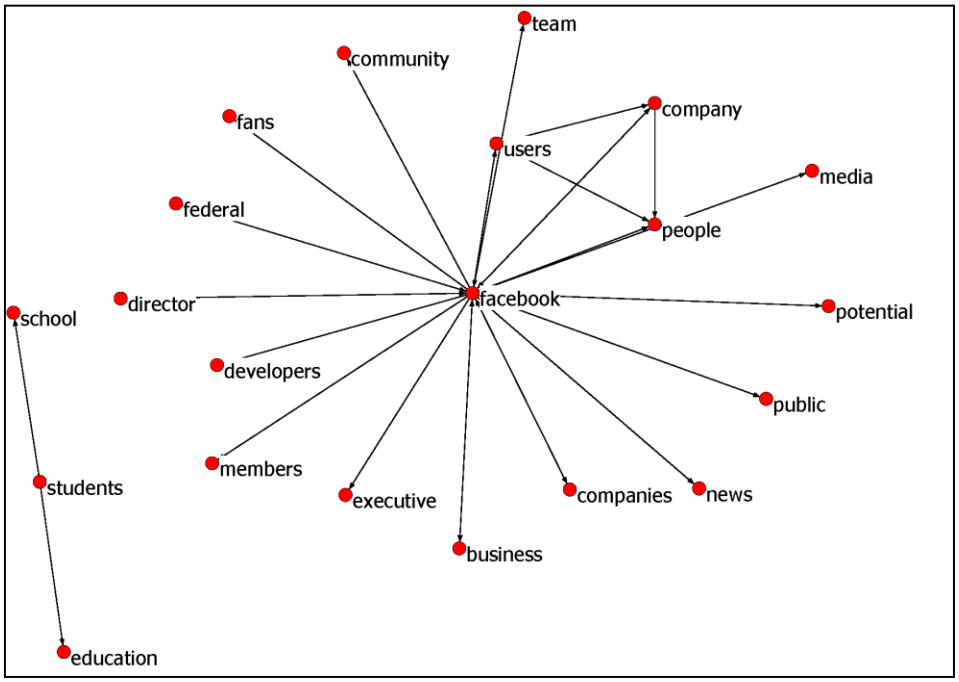

FIGURE III. NETWORK OF PUBLICS FOR SEPTEMBER 25-OCTOBER 25, 2011

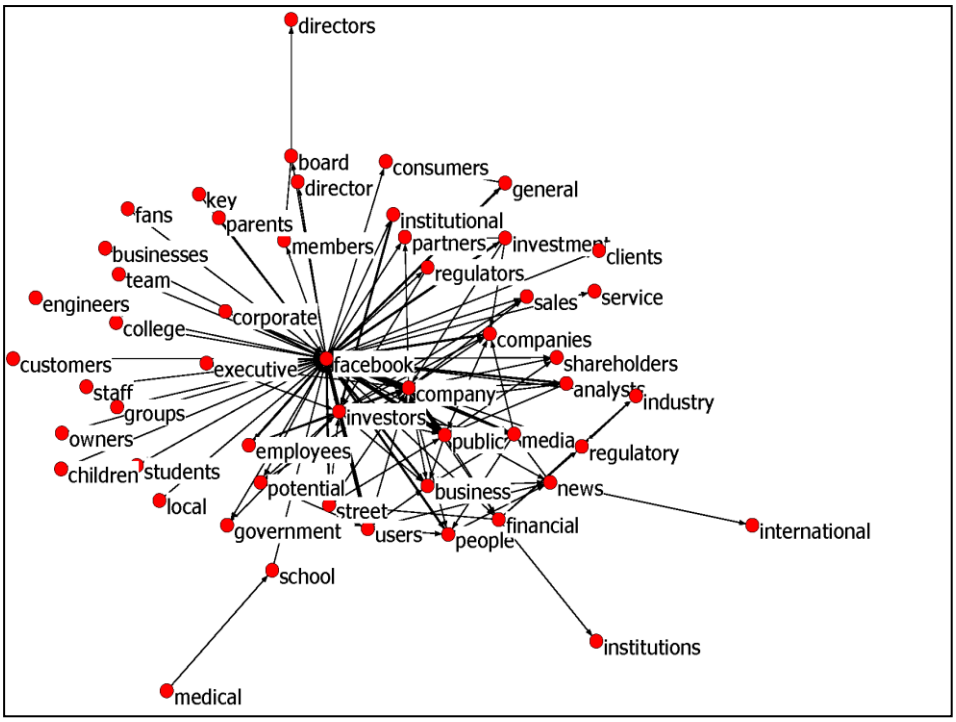

FIGURE IV. NETWORK OF PUBLICS FOR APRIL 25-MAY 25, 2012

TABLE II. CENTRALITY OF TOP 20 PUBLICS PER MONTH FOR FACEBOOK

1		2		3		4		5		6	
facebook	26.997	facebook	23.345	facebook	22.320	facebook	30.088	facebook	31.699	facebook	21.178
media	4.731	media	4.582	media	4.932	media	7.512	media	5.615	people	5.197
people	2.592	company	2.051	people	2.894	people	3.181	people	4.682	company	3.902
company	2.118	people	1.859	news	1.906	users	1.294	company	3.759	students	2.661
users	1.714	potential	1.733	companies	1.883	public	0.909	companies	2.137	children	2.190
employees	1.612	public	1.100	business	1.839	members	0.838	news	1.533	users	1.694
news	1.246	companies	1.080	executive	1.783	leaders	0.727	users	1.509	media	1.668
business	1.008	investment	0.998	public	1.460	service	0.517	business	1.229	government	1.431
executive	0.896	business	0.850	sales	1.216	company	0.440	service	1.005	community	1.430
college	0.824	service	0.719	members	0.750	city	0.426	fans	0.978	public	1.392
consumers	0.724	news	0.692	users	0.662	parents	0.291	members	0.904	education	1.090
general	0.639	customers	0.650	director	0.591	local	0.285	federal	0.689	business	0.859
federal	0.634	employees	0.629	agencies	0.557	executive	0.258	international	0.678	companies	0.751
companies	0.632	citizens	0.614	children	0.546	companies	0.180	partners	0.678	financial	0.718
financial	0.624	consumers	0.556	travel	0.546	school	0.105	corporate	0.668	service	0.651
investors	0.408	city	0.522	customers	0.474	government	0.087	employees	0.659	investors	0.640
school	0.393	industry	0.511	company	0.442	children	0.079	executive	0.622	college	0.623
customers	0.306	education	0.504	potential	0.120	students	0.068	groups	0.463	employees	0.582
national	0.302	users	0.501	service	0.048	news	0.057	public	0.301	politicians	0.530
public	0.297	government	0.419	students	0.034	board	0.048	local	0.214	parents	0.492

7		8		9		10		11		12	
facebook	35.678	facebook	27.379	facebook	31.901	facebook	28.248	facebook	25.887	facebook	32.400
company	3.063	media	3.609	company	6.902	media	8.243	media	5.740	company	6.867
media	3.012	public	2.348	people	4.190	public	2.115	company	2.418	investors	6.301
people	2.921	director	1.247	media	3.068	companies	2.030	users	2.187	public	3.414
business	1.883	children	0.763	business	2.646	company	1.681	business	2.091	business	2.524
users	1.522	financial	0.752	users	2.356	employees	1.600	government	1.725	news	1.847
news	1.515	company	0.715	companies	2.019	students	1.165	people	1.517	media	1.564
service	1.453	agencies	0.691	public	1.679	school	1.151	investors	1.315	companies	1.395
executive	1.093	users	0.683	leaders	1.387	executive	1.118	employers	1.221	people	1.273
public	0.997	executive	0.645	businesses	1.288	people	1.089	service	0.801	financial	1.226
city	0.890	people	0.594	investors	0.829	users	0.981	news	0.647	users	0.860
military	0.761	health	0.585	customers	0.777	federal	0.829	citizens	0.604	key	0.840
sales	0.733	opinion	0.556	policy	0.758	service	0.654	engineers	0.604	regulatory	0.759
college	0.694	leaders	0.537	sector	0.696	local	0.587	international	0.573	policy	0.754
director	0.307	groups	0.537	members	0.680	financial	0.587	officials	0.485	service	0.753
companies	0.212	senior	0.526	director	0.491	directors	0.587	staff	0.409	sales	0.751

consumers	0.087	departments	0.400	financial	0.491	team	0.587	companies	0.353	medical	0.728
financial	0.078	business	0.399	government	0.411	prospective	0.567	employees	0.337	children	0.724
members	0.042	companies	0.301	executive	0.374	board	0.563	executive	0.203	members	0.723
general	0.041	members	0.298	service	0.356	national	0.563	public	0.175	analysts	0.473

TABLE III. AGGREGATE DATA ACROSS THE 12 MONTHS: WORD PAIRS AND FREQUENCIES

WORD PAIR		FREQUENCY													
facebook	users	2773	media	business	114	team	facebook	63	facebook	individuals	42	agencies	ffacebook	31
facebook	company	1214	facebook	investment	110	groups	facebook	63	facebook	agencies	42	schools	ffacebook	31
facebook	people	1074	facebook	consumers	110	facebook	parents	62	public	street	42	public	sshareholders	30
users	facebook	1040	general	facebook	109	service	people	62	officials	facebook	42	media	ssales	30
people	facebook	1011	users	service	108	industry	facebook	61	facebook	city	41	consumers	mmedia	30
facebook	public	931	facebook	students	105	facebook	clients	60	financial	industry	41	engineers	ffacebook	29
facebook	media	618	facebook	financial	97	facebook	college	60	clients	facebook	41	industry	mmedia	29
company	facebook	542	facebook	analysts	95	partners	facebook	60	business	customers	40	investment	iinvestors	29
media	facebook	494	company	media	95	companies	business	60	regulators	facebook	40	investment	ppartners	29
facebook	business	482	facebook	director	94	companies	customers	60	executive	people	39	companies	ppotential	29
facebook	investors	402	facebook	policy	93	facebook	state	58	executive	media	39	senior	mmedia	29
facebook	companies	397	facebook	federal	92	public	users	58	potential	company	39	media	tteam	29
facebook	service	377	facebook	developers	92	news	people	58	federal	users	39	media	ffans	29
public	company	332	street	facebook	92	policy	facebook	58	media	businesses	39	students	mmedia	29
investors	facebook	332	facebook	businesses	90	government	facebook	58	media	service	39	service	pproviders	29
facebook	news	330	facebook	school	90	consumers	facebook	58	facebook	officials	38	fans	mmedia	29
companies	facebook	329	business	users	90	people	companies	57	investors	analysts	38	public	eemployees	28
company	users	323	public	investors	89	company	companies	57	analysts	investors	38	people	ggroups	28
facebook	fans	314	executive	company	89	public	companies	56	board	facebook	38	news	ccompany	28
public	facebook	289	facebook	staff	88	investors	companies	56	public	people	37	company	aanalysts	28
facebook	members	282	facebook	industry	87	media	public	56	users	companies	37	street	ppublic	28
company	public	280	companies	media	86	local	facebook	56	media	director	37	media	iindustry	28
business	facebook	276	facebook	local	85	facebook	regulators	55	government	people	37	media	ccommunity	28
news	facebook	249	facebook	community	84	users	members	55	corporate	facebook	37	service	ccustomers	28
users	people	237	investors	public	84	college	students	55	company	shareholders	36	community	ppeople	28
service	facebook	216	potential	investors	84	facebook	facilities	54	company	financial	36	visitors	ffacebook	28
facebook	executive	199	facebook	team	82	users	business	54	company	customers	36	facebook	vvisitors	27
facebook	employees	198	business	company	81	business	companies	54	investors	business	36	investment	ccompanies	27
investors	company	185	companies	people	81	business	people	53	companies	company	36	companies	cconsumers	27
company	people	180	facebook	general	79	employees	company	53	public	business	35	service	ccompany	27
media	people	179	businesses	facebook	79	public	media	52	company	street	35	fans	ppeople	27
facebook	potential	177	facebook	government	78	people	business	52	street	company	35	people	cchildren	26
executive	facebook	173	facebook	partners	77	investment	company	52	school	students	35	news	sservice	26
users	company	165	institutional	investors	77	students	school	52	public	financial	34	company	nnews	26
facebook	street	159	media	news	77	analysts	company	51	public	policy	34	executive	bbusiness	26
facebook	sales	156	members	facebook	76	children	parents	51	company	partners	34	center	ffacebook	26
media	companies	146	users	public	75	company	service	50	financial	regulatory	34	companies	sservice	26
facebook	children	145	companies	users	75	users	groups	50	employees	investors	34	director	ccompany	26
parents	facebook	142	company	employees	73	developers	facebook	50	companies	employees	34	service	mmedia	26
parents	children	142	users	media	73	national	facebook	49	groups	people	34	facebook	ccenter	25
company	investors	139	key	facebook	73	people	public	48	facebook	television	33	general	ppublic	25
companies	public	139	customers	facebook	72	news	users	48	company	investment	33	investors	ppotential	25
facebook	customers	138	sales	facebook	72	company	sales	48	company	potential	33	companies	bbusinesses	25
fans	facebook	136	people	news	71	state	facebook	48	shareholders	facebook	33	politicians	ffacebook	25
analysts	facebook	134	people	company	71	financial	company	48	business	school	33	director	ppeople	25
facebook	groups	133	facebook	shareholders	70	facebook	internationa	47	businesses	customers	33	director	ppolicy	25
people	media	130	people	service	70	facebook	press	46	employers	facebook	33	television	ffacebook	25
media	users	128	employees	facebook	69	college	facebook	46	facebook	regulatory	32	people	ppotential	24
service	users	128	director	media	69	community	facebook	46	investors	users	32	executive	ppublic	24
children	facebook	127	executive	director	68	executive	users	45	industry	regulatory	32	financial	iinvestors	24
potential	facebook	123	federal	facebook	68	press	facebook	45	facebook	schools	31	business	ppublic	24
students	facebook	121	staff	facebook	68	facebook	senior	44	people	school	31	business	ppotential	24
director	facebook	120	facebook	key	67	financial	public	44	business	investors	31	school	bbusiness	24
people	users	118	facebook	corporate	67	companies	investors	44	businesses	media	31	media	aagencies	24
users	news	117	news	media	67	media	consumers	44	sales	company	31	developers	uusers	24
school	facebook	117	media	customers	67	facebook	engineers	43	city	facebook	31	schools	sstudents	24
media	company	117	business	media	65	street	investors	43	board	directors	31	internationalf	facebook	24
investment	facebook	116	facebook	board	64	potential	customers	43	media	potential	31	facebook	iinternal	23
company	business	114	financial	facebook	63	senior	facebook	43	team	people	31	executive	ccompanies	23

The include list technique ignored all of the other words in the text. It mapped the cooccurrence of words only on the include list that were within the specified word window, here 3 words wide. Our example used an include list developed from an earlier study in which 343 PRSA public relations practitioners were asked an open-ended online survey question to identify their key publics. We used words that occurred at least 3 times, resulting in an include list containing 145 names of publics. We could have refined this list by using the 'string conversion' functionality of WORDij. This would convert phrases into a single unigram. For example, 'investment community' would be converted into the single term of 'investment_community.' This would avoid the error that would occur if community sometimes appeared by itself or appeared with other words, such as 'local community.' Because our goal was not to focus on the specific content of the networks but instead to illustrate a set of procedures, we did not go back to the original raw survey data and construct a 'string conversion' file.

Some semantic network analysts like to stem words and reduce them to their roots, removing plural forms, gerunds, adverbial forms, and other lexical variants. We typically do not stem. Although stemming is useful in information retrieval, it removes nuances from semantic networks. When the intent is to explore social meanings rather than retrieve information efficiently, the more nuances that are retrained in the analysis, the more valid the network results in reflecting the collective meanings of text in documents. Information about the content of the network is not removed, provided that proximity rather than bag of words approaches are used and word order in pairs is maintained.

Businesses typically have developed lists of key terms for different purposes that could be used in an include list depending on the goals of the analysis. They may have a list of key individuals, perhaps investment analysts whose statements they may wish to track over time in relation to their

business. Competitor organizations may be of interest. They may wish to map the competitors in relation to their business as a whole; or its products; markets; product features; policy issues, relevant groups and individuals; or other such terms that would be formed into an 'include list.'

The text analyzed here were documents appearing in the database Lexis-Nexis Academic. Other sources of text could be used. For example, one could use emails received from customers and analyze these with a relevant include list. Or, one could analyze collections of internal documents. Social media such as Facebook and Twitter could be the source of texts mined. The time frame for analysis could be optimally set, perhaps to days, weeks, or even smaller intervals such as hours.

V. Limitations

This research had no purpose other than demonstration. The organization studied, the analysis of a single organization, the time frame chosen, the use of selected software, the analysis of the changing centrality of two publics among many, the exclusive use of open-source documents, most of which were news stories often written by external observers writing for various purposes rather than only being an organization's own press releases, were therefore all arbitrary choices. This study provides no knowledge claims. Hopefully, however, it stimulates the reader to do future research of value.

VI. Conclusion

We demonstrated semantic network analysis using an include list to map the publics of a business, as represented in full-text documents extracted about a business from Lexis-Nexis Academic for a 12-month period. We extracted documents that mentioned the company Facebook at least 10 times. We mapped the monthly networks of publics using an include list derived from a survey of PRSA public relations practitioners. We computed centrality of each public that appeared in a particular time period, and looked more specifically at changes over time in the centrality of users and employees. This demonstrated that using include lists is a feasible way to do one type of semantic network analysis. In future research it can be used for a variety of purposes. For management goals one could analyze a set of competitors using an include list. Or, for scientific purposes, the investigator could select a large sample of organizations, analyze them with the same include list, and test hypotheses about differences.

Acknowlegments

The author is grateful to Richard Weeks for programming work on the DeDup software.

Conflicts Of Interests

The author has no conflicts of interest.

References

[1] J. A. Danowski, "Identifying collaborative innovation networks at the inter-departmental level," Procedia - Social and Behavioral Sciences, 2, no. 4, pp. 6404–6417, 2010.

[2] N. Pathak, S. Mane, and J. Srivastava, "Who thinks who knows who? socio-cognitive analysis of email networks," Technical Report TR-06-023. Minneapolis, MN: Department of Computer Science and Engineering, University of Minnesota, 2006.

[3] J A. Danowski and P. Edison-Swift, "Crisis effects on intraorganizational computer-based communication, Communication Research, vol. 12, pp. 251-270, 1985.

[4] J. A. Danowski, K. Riopelle, and J. Gluesing, "The revolution in diffusion models caused by new media: the shift from s-shaped to convex curves," In The diffusion of innovations: a communication science perspective, G. A. Barnett and A. Vishwanath, Eds.. New York: Peter Lang Publishing, 2011, pp. 123-144.

[5] J. Diesner, T. L. Frantz, and K. M. Carley, "Communication networks from the Enron email corpus: it's always about the people. Enron is no different," Computational & Mathematical Organization Theory, vol. 11, no. 3, pp. 201-228, 2005.

[6] K. Coussement and D. Van den Poel, "Improving customer complaint management by automatic email classification using linguistic style features as predictors," Decision Support Systems, vol. 44, no. 4, pp. 870–882, 2008.

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

[8] 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, pps. 31-46.

[9] M. V. Carrillo, J. A. Danowski, A. Castillo, and J. L. T. Jimenez, "Semantic networks for corporate communication concepts and crisis: differences based on corporate reputation," Observatorio, vol. 6, no. 2, pp. 127-145, 2011.

[10] J. A. Danowski, "Mining organizations' networks: multi-level approach. In Mining analysis and research trends: techniques and applications I. Ting, T. P. Hong, and L. S. L. Wang, Eds. IGI Global, 2012, pp. 205-230.

[11] J. A. Danowski and H. W. Park, "Web network and content changes associated with the 2011 Muslim Middle-East and North African early uprisings: a naturalistic field experiment," IEEE Intelligence and Security Informatics Conference (EISIC), European, pp. 100-107, August 2011.

[12] T. R. Gruber, "A translation approach to portable ontology specifications," Knowledge Acquisition, vol. 5, no. 2, pp. 199-220. 1993.

[13] J. A. Danowski, "DeDup: A program for removing duplicate text from text mining results." [computer program]. Chicago: University of Illinois at Chicago, 2012.

[14] J. A. Danowski, "WORDij version 3.0: semantic network analysis software," [computer program]. Chicago: University of Illinois at Chicago, 2010.

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

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

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