Social Network Mining, Analysis, and Research Trends:

Techniques and Applications

I-Hsien Ting
National University of Kaohsiung, Taiwan

Tzung-Pei Hong
National University of Kaohsiung, Taiwan

Leon S.L. Wang
National University of Kaohsiung, Taiwan

Information Science
REFERENCE

Managing Director: Lindsay Johnston
Senior Editorial Director: Heather Probst
Book Production Manager: Sean Woznicki
Development Manager: Joel Gamon
Development Editor: Myla Harty
Acquisitions Editor: Erika Gallagher
Typesetters: Lisandro Gonzalez
Print Coordinator: Jamie Snavely
Cover Design: Nick Newcomer, Greg Snader

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com

Library of Congress Cataloging-in-Publication Data

Social network mining, analysis, and research trends: techniques and applications / I-Hsien Ting, Tzung-Pei Hong and Leon S.L. Wang, editors.
p. cm.
Includes bibliographical references and index.
Summary: "This book covers current research trends in the area of social networks analysis and mining, sharing research from experts in the social network analysis and mining communities, as well as practitioners from social science, business, and computer science"--Provided by publisher.
ISBN 978-1-61350-513-7 (hardcover) -- ISBN 978-1-61350-514-4 (ebook) -- ISBN 978-1-61350-515-1 (print & perpetual access) 1. Online social networks. 2. Data mining. I. Ting, I-Hsien. II. Hong, Tzung-Pei, 1963- III. Wang, Leon Shyue-Liang, 1955-
HM742.S6288 2012
302.30285--dc23
2011038347

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

Chapter 13
Mining Organizations' Networks:
Multi-Level Approach[1]

James A. Danowski
University of Illinois at Chicago, USA

ABSTRACT

This chapter presents six examples of organization-related social network mining: 1) interorganizational and sentiment networks in the Deepwater BP Oil Spill events, 2) intraorganizational interdepartmental networks in the Savannah College of Art and Design (SCAD), 3) who-to-whom email networks across the organizational hierarchy the Ford Motor Company's automotive engineering innovation: "Sync® w/ MyFord Touch", 4) networks of selected individuals who left that organization, 5) semantic associations across email for a corporate innovation in that organization, and 6) assessment of sentiment across its email for innovations over time. These examples are discussed in terms of motivations, methods, implications, and applications.

OVERVIEW

When you think of social network analysis you probably visualize individuals as nodes. This is quite natural, given the "social" aspect, and more specifically because the origins of social network analysis, going back some 89 years (Freeman, 1996), are in the relations among individuals. Nevertheless, 'social' is also considered at levels of analysis in which the focal nodes are groups, subunits of organizations, organizations, or more macro-level human systems. In this chapter I center on organizational social network analysis, focusing on interorganizational, organizational, departmental, and also individuals as they communicate with organizations.

After this overview I will discuss why such a focus is desirable in light of the literature. First, however, let me point out that this chapter also includes mining for and automatic identification of an organization's networks at several levels based on textual elements of documents available on the web such as news story databases, blogs, reports, and other electronic text content, and also

DOI: 10.4018/978-1-61350-513-7.ch013

available from internal organization documents, such as email, or other kinds of electronic text accessible in real time or as archived internally or in the clouds.

Imbedded in electronic texts are layers of social networks that can be unpeeled with the miner's tools. The highest-order type of network mining I focus on is identification of *interorganizational* networks. This has been an area of research that previously used manual procedures other than mining (Scott, 1988, 1991, 2000; Galaskiewicz, & Shatin, 1981; Galaskiewicz, 1985; Mizruchi, 1996). Here I use an automated procedure based on the co-appearance of organizations across a corpus of web documents. In particular, each time a focal organization appears together with another organization within a proximity window in a document, the pair of organizations is automatically counted. Taking the aggregated collection of pairs, I network analyze them. This enables indexing the organizations' positions in the network and computation of various network structure measures common to social network analysis (SNA). Moreover, by slicing the collection of documents into time segments one can analyze the network structure as a time series. Time-sequenced associations are one of the necessary conditions for establishing possible causal relationships among variables. For example, we could also measure the sentiment expressed in the documents about the organizations and identify any synchronous or lagged associations between sentiment and network structure (Danowski & Cepela, 2010). As an example of this, Noah Cepela and I measured the organizational networks among U.S. presidents' cabinets over time, from Nixon through G.W. Bush, automatically indexing co-appearance of cabinet members in documents. We examined the time-lagged relationships between the presidents' centrality in the administrative network and the link between news sentiment and job approval as measured by the Gallup polls. We tailored the time slicing to the frequency of the Gallup polls for each presidency. This shows how one can build some of the necessary conditions for causality evidence: identify a sequence of networks, index its attributes, and add measures of other attributes of the organizational actors and contextual factors of theoretical interest. What remains is ruling out rival explanations for observed time-ordered associations.

In this chapter my first several examples illustrate such a new approach to interorganizational social network analysis and data mining. I compile a list of organizations of interest and search across large text corpora for the co-appearance of pairs of organizations in news documents, blogs, reports, and related venues. In the first example I mine for an interorganizational network. I also slice time segments across the larger mining time frame to enable time-series analysis of these network structures. The example examines the interorganizational networks associated with coverage of the Deepwater BP Gulf Oil Spill of 2010. I then examine the relationship between news story sentiment about the most central organization, BP, and its position in the network overtime using a new network-based sentiment measuring approach, based on average shortest paths from BP to each of several thousand possible sentiment words, generating theoretically interesting findings about the sequencing of sentiment and centrality.

A second kind of organizational network mining is for the departments within an organization that co-appear across news stories and other text documents. For example, consider the departments within a university. In many cases the departments are proxies for disciplines. By identifying each pair of departments co-mentioned across a corpus of news and/or blog content about the university one can automatically map the representation of the collaborative network of the university's departments over time in the mining corpora. My example is from the Savannah College of Art and Design (SCAD) that was under accreditation review in 2009 and needed evidence of collaborative networks across departments, surrogates for disciplines.

Third, one may wish to measure the networks of individuals associated with the organization. Based on individuals' co-appearance in organizational documents, one could measure the structural properties of the network at the macro level, such as its size, density, and centralization, and also measure the network positions of the individuals, for example their flow betweenness centrality (Freeman, Borgatti, & White, 1991). The mapping of presidential cabinet networks is an example of this level of organizational analysis.

A fourth kind of organizational text mining is to exact semantic networks from intra-organizational documents, such as email, reports, or other textual materials. There are many ways one could peel the intraorganizational network onion, besides the interdepartmental level network analysis. For example, here I demonstrate mapping the network of individuals based on who sends email to whom, mining email captured automatically in the engineering function of the Ford Motor Company as they work on development of a cluster of innovations for vehicles, referred to as the "Sync® w/ MyFord Touch" product. I consider this as one variation of semantic network analysis because we mine for individuals' names, which are words.

Other research, although not numerous, has measured networks from email for various purposes (Danowski & Edison-Swift, 1985; Gloor, 2006; Gloor and Zhao, 2006; Diesner, Frantz, and Carley, 2005). Based on work with my colleagues Ken Riopelle and Julia Gluesing and I, with associates within the Ford Motor Company, I show how one can make two new uses of organizational email network analysis. One use is representing the network across levels of the organizational hierarchy in terms of the number of steps in the chain of command from the CEO on down. A second use, overlaid onto who-to-whom email hierarchical networks, is to examine the network structures of individuals who have left the organization. This can be useful for management to inform replacement individuals with whom they will be expected to communicate in what kind of pattern. After management reviews this network and approves of particular links, they can give the recruit a detailed overview of their predecessor's ego-centric network, specifying the names of particular persons with whom it is probably important to communicate.

The other kind of semantic analysis I demonstrate is of the words (other than people's names) that are associated with one of the innovations of "Sync® w/ MyFord Touch." We sliced the networks into weekly intervals to track the evolution of the meanings associated with a cluster of innovations for use of the automobile driver. Here, however, given space constraints, I will show the semantic network of the earliest time slice, and for only one of the innovation concepts, the cockpit.

In sum, by considering the examples of mining for organizational networks at different levels of analysis I hope through this chapter to illuminate your thinking about ways in which you might conceptualize related analyses, or new clusters of links among your conceptual light bulbs, or shall we say 'semantic nodes.'

My intent is to clarify how one can test hypotheses about organizations at various levels or to develop management applications for strategic public relations, intelligence functions, and business analytics. You will also likely think of future directions for development of organizational mining methods and analysis, sparking new ideas about applying and evolving the techniques. In sum, the examples will trigger new thoughts at both the conceptual and methodological levels.

Overview of Social Network Analysis Organizational Mining Literature

Let's put into context organizational social network mining in terms of the populations of related studies in the literature about data mining and social network analysis combined with various interorganizational, organizational, and intraorganizational terms using Google Scholar.

The first pattern of note is that SNA has exploded in number of studies conducted (Borgatti & Foster, 2003; Borgatti, & Molina, 2003) since its beginnings 89 years ago (Freeman, 1996). On January 13, 2011 Google Scholar (http://scholar.google.com) returns 38,500 hits on for the phrase 'social network analysis.' Combining 'data mining' and 'social network analysis' results in 5,300 hits, or 14% of the total pool of SNA hits. Most of the prior research on SNA studies individuals as nodes in a network. In contrast, SNA using organizations as nodes, or using sub-organization units larger than the individual yields 14,000 hits, although many of these use individuals as nodes within organizations as they conceptualize or measure organizational social networks (Scott, 1988, 1991, 2000).

When adding 'data mining' to the 'organization' and 'social network analysis' terms there are 1,320 hits. 'Interorganizational' coupled with 'social network analysis' returns 3,580 hits and when both are coupled with 'data mining' in only 128 hits.

Consider that many of these hits are not directly about using methods of data mining and SNA in interorganizational or organizational contexts but these terms are mentioned somewhere in the Google Scholar records. These search results show that there is a not much literature about interorganizational networks and data mining, while there are more hits about 'organizations' and 'social network analysis' (Tichy, Tushman, & Fombrun, 1979) and 'data mining,' although most of the studies use individuals as nodes in the organizational network. The primary data mined is electronic mail, much of it with a single corpus, from Enron. When considering 'semantic networks' in the 'organization' and 'data mining' there are 1,320 hits, while adding 'social network analysis' returns only 128.

In short, given only this simple search of the literature, there is evidence for the need for more focus on interorganizational, organizational, and organizational subunit levels with respect to SNA and data mining. There is also need for more attention to semantic networks in organizations using SNA approaches. A more lengthy review of the literature is not within the scope of this chapter.

Example 1: Interorganizational Networks Associated with the Deepwater BP Oil Spill

In thinking of what example to use to illustrate interorganizational network analysis and data mining, it occurred to me that organizations were one of the main focal points of much news coverage about the 2010 Deepwater BP Gulf Oil Spill. In particular, along with the three major private sector organizations, federal, state, and local government organizations appeared to be active during this period.

I decided to examine the interorganizational networks from a particular vantage point, that of the White House. Accordingly, I went to the website http://www.whitehouse.gov/deepwater-bp-oil-spill/ and copied the names of organizations listed as working on the problems. Some of the links on the web site were to pages of states in the affected areas of the Gulf: Louisiana, Mississippi, Alabama, and Florida, so I extracted as well the names of any organizations appearing on those related web sites. My final list, shown in Table 1, consisted of 81 organizations, approximately half of them being various volunteer organizations.

I used one of the features of WORDij 3.0 (http://wordij.net) (Danowski, 2010), the specification of a string conversion of n-grams to unigrams, and therefore converted the multi-word names of the organizations to a single acronym to aid in the display of the networks. This avoided the problems resulting from many of the governmental organizations having long names that would make difficult visual comprehension of network graphs.

Many applications of natural language processing to large textual corpora, particularly in information retrieval, but in other areas as well, use a

Table 1. Unigrams Key for Organizations in Figure 2, Figure 4, Figure 5, Figure 6, Figure 7, and Figure 8

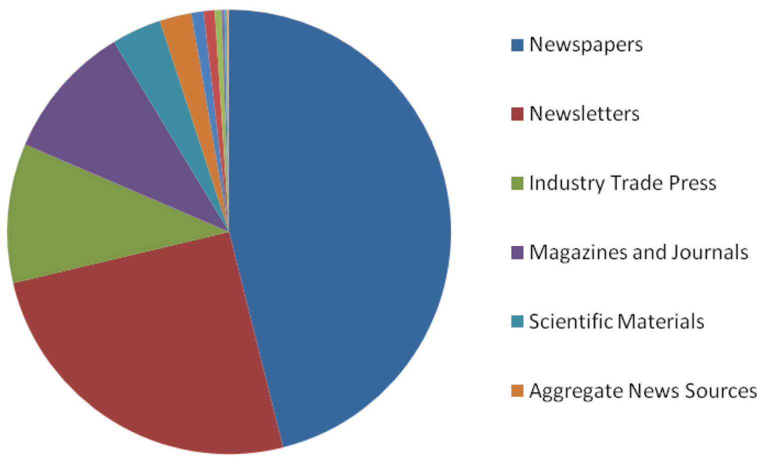
ALa Americorp Louisiana
ACS Adventist Community Services
AGOFBCI Governor's Office of Faith-Based and Community Initiaties
AlAC Alabama AmeriCorps
ARC American Red Cross
ASHFA America's Second Harvest/Feeding America
BP BP, British Petroleum
CathC Catholic Charities
CCA Christian Contractors Association, Inc.
CCC Caribbean Conservation Corporation
CCST Community Crisis Support Team
CDR Christian Disaster Response
CNCS Corporation for National and Community Service
COH Convoy of Hope
CompAll Compassion Alliance
CRWRC Christian Reformed World Relief Committee
CWS Church World Service
DHS Department of Homeland Security
DOD Department of Defense
DOF Defenders of Wildlife
DOI Department of the Interior
DOIFWS Department of the Interior's Fish and Wildlife Service
DOINPS Department of the Interior's National Park Service
DOL Department of Labor
EPA Environmental Protection Agency
ERD Episcopal Relief and Development
ESF15 Emergency Support Function 15
FAFB Florida Association of Food Banks
FarmSh Farm Share
FAud Florida Audubon
FAVC Florida Association of Volunteer Centers
FBDR Florida Baptist Disaster Relief
FCaCon Florida Catholic Conference
FCSDA Florida Conference of Seventh-Day Adventists
FDEA Florida Department of Elder Affairs
FGCVCS Governor's Commission on Volunteerism and Community Service
FIND Florida Interfaith Networking in Disaster
FIRST Florida Immediate Response Stress Team
FJC Florida Jaycees
FLVOAD Florida Voluntary Organizations Active in Disasters

continued on following page

Table 1. Continued

FSERT Florida State Emergency Response Team
FUMC Florida United Methodist Conference
GCCF Gulf Coast Claims Facility
Halli Halliburton
HON Hands On Network
HumSoc Humane Society of the United States
KFB Keep Florida Beautiful
LDCRT Louisiana Department of Culture, Recreation, and Tourism
LDR Lutheran Disaster Response
LOF Lions of Florida
LSC Louisiana Serve Commission
MCVS Mississippi Commission for Volunteer Service
MDS Mennonite Disaster Service
NatComm National Commission on the BP Deepwater Horizon Oil Spill and Offshore Drilling
NDR Nazarene Disaster Response
NFG Network for Good
NOAA National Oceanic and Atmospheric Administration
NRMCS Night Runners Mobile Crisis Services, Inc.
OpBless Operation Blessing
OSHA National Institute for Occupational Safety and Health
PDA Presbyterian Disaster Assistance
PFC National Pollution Fund Center
POLI Points of Light Institute
RAla Ready Alabama
RTG RestoreTheGulf. gov
SBA Small Business Administration
SBCDR Southern Baptist Convention Disaster Relief
SERVOF State Emergency Responders and Volunteers of Florida
SNAT Service Nation
SVM Scientology Volunteer Ministers
TEWF The Eagles Wings Foundation
TransO Transocean
TSA The Salvation Army
UCCDRM United Church of Christ Disaster Response Ministries
UMCR United Methodist Committee Relief
USCG U.S. Coast Guard
UW The United Way
UWaySMS United Way of South MS
VolFLA Volunteer Florida
VolLA Volunteer Louisiana
VolMS Volunteer Mississippi

Figure 1. Document Sources in the Lexis-Nexis Text Collection



stop word list 'stoplist' to drop frequent function words that appear quite uniformly across documents. In English the most common words at the top of a stop list are: 'the, of, that, and,' and other such grammatical function words that do not carry a large field of social meaning. WORDij has an option to use a stoplist or droplist of words, but for the current example, we make use of the opposite, an *include list*. WORDij enables one to specify a list of words, in this case organization name unigrams, for analysis of their cooccurrences. All other words not on the include list are removed from the analysis. This enables an efficient means of mapping highly-focused networks, in this case, the networks among organizations mentioned in news stories about the oil spill.

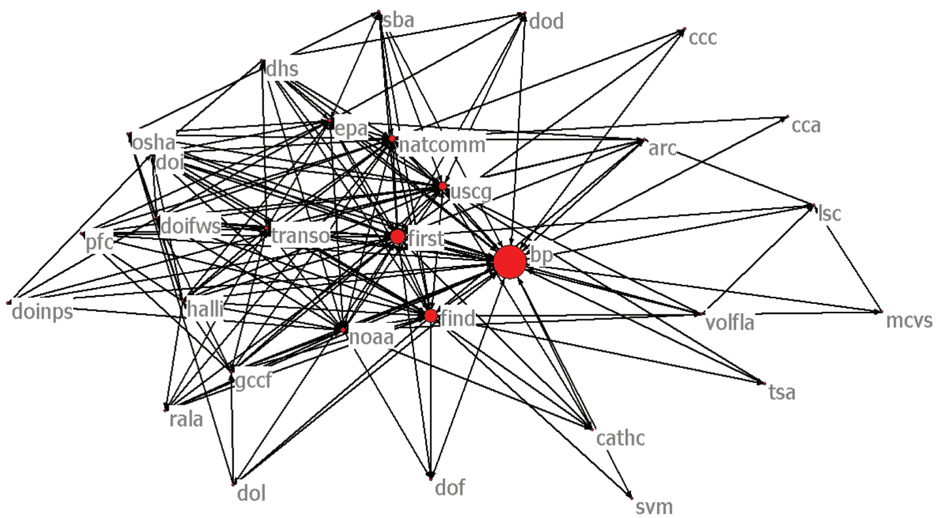
Another feature of WORDij is the TimeSlice procedure that allows for taking a larger time frame and dividing the textual corpus into smaller equal time intervals. In this example, the larger time span ran from April 20, 2010 to when this chapter was written in first week of January, 2011. We experimented with setting a weekly or bi-weekly time interval and found that the weekly interval had portions of the span that were sparse in numbers of organizations. A two-week interval was chosen for illustrative purposes. The TimeSlice function prepares the textual corpus so that the basic WordLink co-occurrence counter quickly runs through each time interval producing the same output files for each slice.

I collected a corpus of 4,728 documents in Lexis-Nexis Academic from the *New York Times*, *Washington Post, USA Today, Baton Rouge Advocate* (Louisiana), *St. Petersburg Times* (Florida), *Birmingham News* (Alabama), and the *Jackson Clarion-Ledger* (Mississippi). Figure 1 shows the breakdown of the kind of documents in the corpus.

TimeSlice divided the overall corpus into 38 two-week intervals. WORDij WordLink produced the network output file with the optional .net Pajek format (Batagelj & Mrvar, 1998) for each interval and I input this network file to UCINET 6 for conversion to its own format.

This enabled me to run computations of flow betweenness centrality (Freeman, Borgatii, & White, 1991) for each time period, producing centrality scores for each organization as well as providing an overall centralization score for the whole network. Borgatti (2005) has pointed that while betweenness centrality (Freeman, 1979) is most often used in SNA, almost all such uses are inappropriate because the assumptions of the measure do not fit well with the nature of the data. Betweenness centrality assumes that each link has the same strength. Betweenness is

Figure 2. Interorganizational BP Oil Spill Network Aggregated Across Time



computed on dichotomized, linked/not-linked relationships. Our data are continuous and highly varying. Flow Betweenness (Freeman, Borgatti, & White, 1991) eliminates the need for binarizing the network, which discards valuable variance on link strengths. Betweenness also assumed is that messages flow between nodes based on the single shortest possible path. Nevertheless, in actual communication situations there are varying link strengths, such as based on the frequency of communication. Nodes can also send messages to multiple nodes, varying in path lengths, perhaps even avoiding the node with highest betweenness centrality. Flow betweenness centrality (Freeman, Borgatti, & White, 1991) has assumptions which best meet these conditions. If one is using news stories, email, or other forms of communication as input texts, then it is more appropriate to use it rather than betweenness centrality.

Conversion to UCINET format was also useful to produce NetDraw graphic depictions of each network, which works with the UCINET system files. WORDij has a network visualization function, VISij, which animates the time-series of network structures, but it is not possible to present a network movie in a printed chapter. NetDraw has more options than VISij for displaying the network in terms of sizes of nodes, labels, links, and sizing nodes and links by variables. For node sizing we chose the closest centralization measure to flow betweenness available in NetDraw, classic betweenness centrality. Figure 2 shows the network of organizations aggregated over time.

For other computations, shown in Figure 3 we used flow betweenness for overall network centralization. Looking at this graph one sees that the lowest centralization in the first half of the time series is for the first time slice. A peak is reached at the sixth slice, followed by a drop to a local low at the ninth slice. The next peak occurs at the fifteenth slice with the lowest in the final slice, the nineteenth one. As a result I thought it would be informative to display the network graphs for each of these noteworthy slices, shown in Figure 4 and Figure 8.

An interesting question with theoretical implications is how is the sentiment expressed in the documents associated with the network structure? To explore this question, I used the same two-week time-slice interval but this time used no include list of organizations when running WordLink, as a result analyzing the full-text of

Figure 3. Interorganizational Centralization Over Time

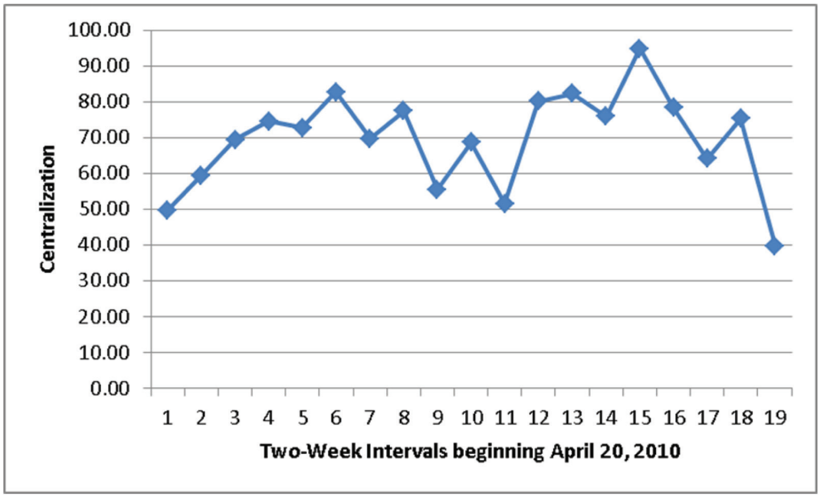
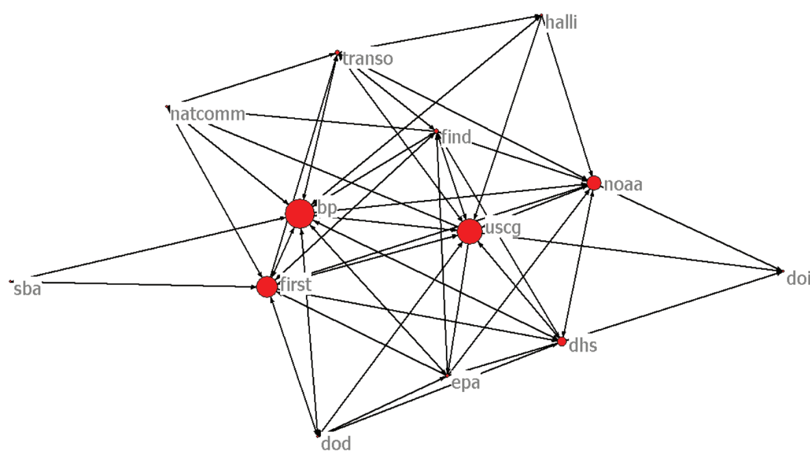


Figure 4. Period 1/Interval 1: Low Centralization



the documents' for semantic networks. After that I input the word pair files to a shortest path network analysis program in WORDij, OptiCommReport. This program allows the specification of a seed word, in this case BP, and then traces the shortest paths from the seed to each of 3,457 lexical variants of positive and negative words taken from the Linguistic Inquiry and Word Count (LIWC) (Pennebaker, Booth, & Francis, 2007). LIWC is a dictionary-based content analysis software package that indexes the occurrences of some 72 categories of words. I took their dictionary entries for two of the categories: positive emotion and negative emotion. There are 266 positive emotion word stems and 346 negative emotion stems. To build a more robust sentiment analysis system I expanded these word stems to all lexical inflected forms of each using the AGID comprehensive list of inflected forms of words on Kevin Atkinson's Word List Page (http://wordlist.sourceforge.net/). As a result, I have 1,219 positive words and 2,238 negative words. My OptiCommReport

Figure 5. Period 1/Interval 6: High Centralization

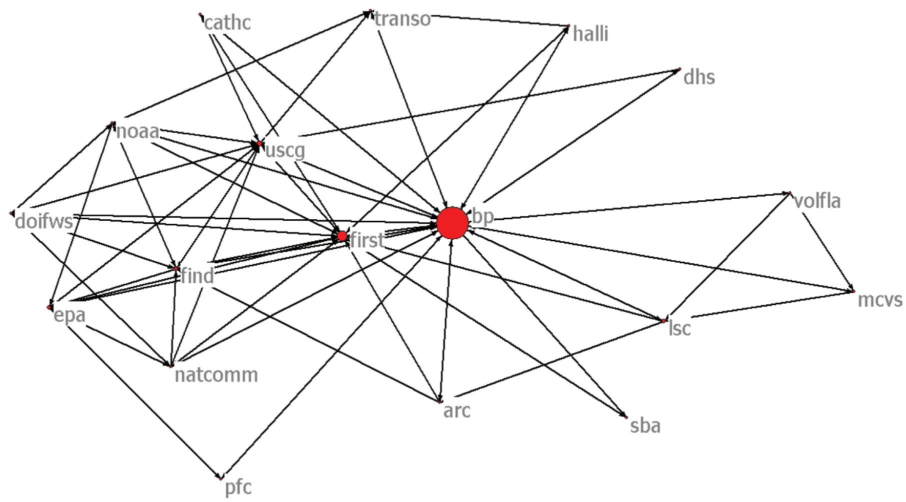
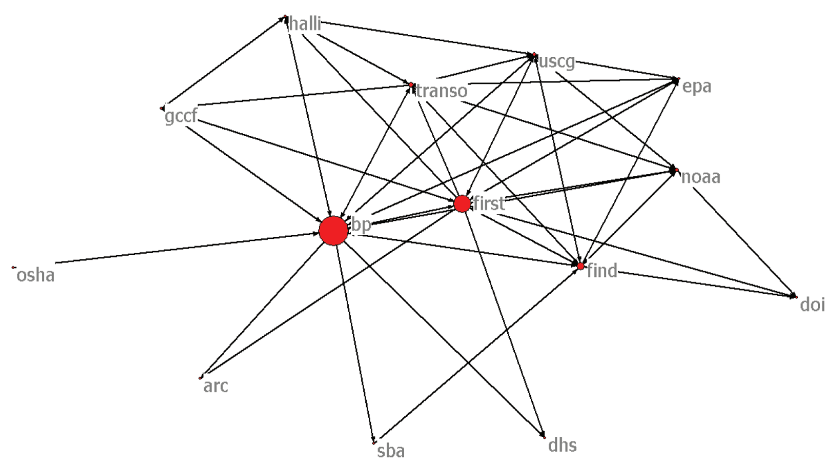


Figure 6. Period 1/Interval 9: Low Centralization



program identifies the shortest weighted path from the seed word to each of the semantic target words in each of the word pair output files. I then compute from the results the total normalized link strengths for the shortest path to each of the positive words and divide by the total link strengths for the shortest path to each of the negative words, and divide the positive by the negative normalized strengths to produce a ratio of positive to negative network threads. This is the positivity index.

For the current example, so you could effectively see the plot of the centrality of BP in relation to positivity for BP, I multiplied the positivity ratio by a constant of 50 to scale it close to the centrality score range. Figure 9 shows the BP centrality values over time as the top line and the positivity ratio as the lower line in the graph. In doing further statistical analysis, I first examined the autocorrelation of the variables in SPSS Forecasting and found that there was no significant autocorrelation

Figure 7. Period 1/Interval 13: High Centralization

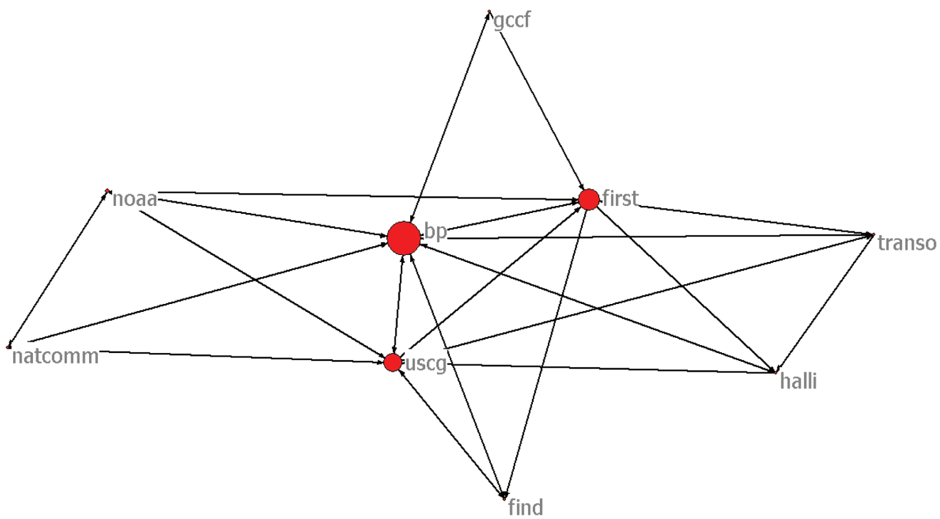
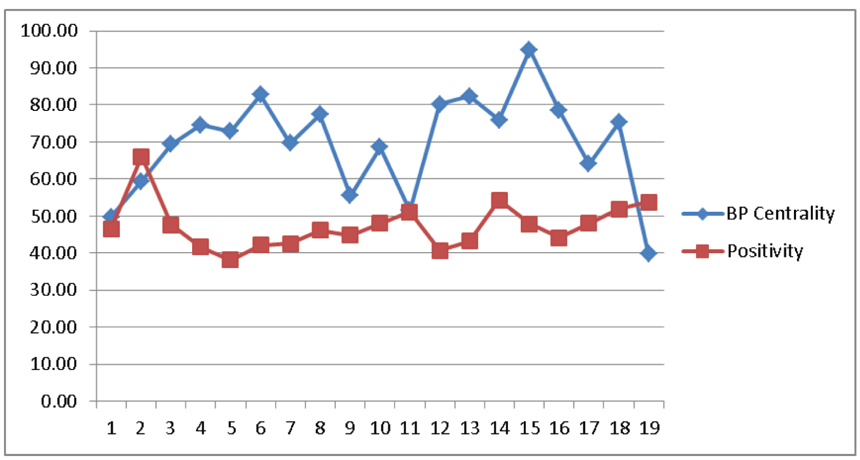


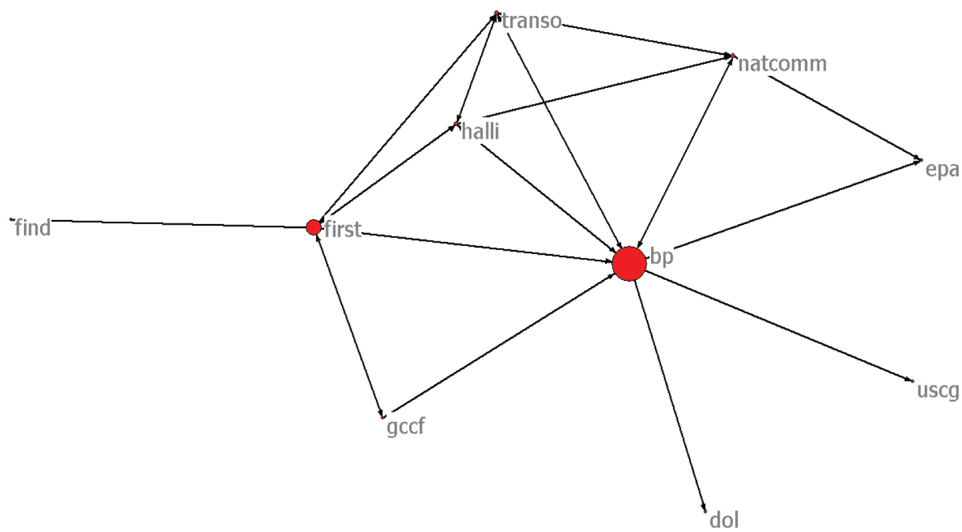
Figure 8. Period 2/Interval 19: Low Centralization



at any lag. It was therefore appropriate to compute the Pearson correlation across the time-series for the two variables. The correlation was $r = -.42$ ($p < .04$). This indicates that when BP is in a less central position in the network, the sentiment is more positive, and conversely when BP is in a more central position the sentiment is more negative. Theoretically, it is valuable to examine in future research the generalizability of this centrality/negativity association and to what extent it is generalizable and more importantly, under what conditions, and why.

Consider that here at the interorganizational level, in the context of news documents, there is an association between network centrality and sentiment that differs from the general findings at the ego-centric network level. As individuals' networks are more centralized there is more of an instrumental focus and less of an emotional focus suggested by Granovetter (1973) and Burt (1995). In our case, however, centralization around

Figure 9. Interorganizational Centralization & Positivity Over Time



BP is associated with more negative sentiment while less centralization is associated with more positive sentiment. This pattern is probably different from the individual level because of the context of news documents being different from that of individual-level social network behaviors. There are institutional-level factors for journalism related to the macro-level societal functions of news that privilege negative news for societal surveillance purposes. The most noteworthy implication of these findings is that one should be cautious when attempting to generalize associations across network levels. As the level of analysis shifts there may be contextual changes that render cross-level generalization more complex than the core associations subject to potential generalization. In other words, what happens in networks where the individuals are nodes may be quite different than where organizations are nodes. I will leave further contemplation of the implications of this analysis to you and move on to the second example.

Example 2: Identifying COINs at the Inter-Departmental Level

Overview

Collaborative Innovation Networks (COINS) (Gloor, 2006) are typically defined using individuals as nodes (Cross, Borgatti, & Parker, 2002; Cross, & Parker, 2004). This is a bias of the social network literature more generally, as we discussed at the beginning of this chapter. Departments in organizations or higher order social units can, however, be considered as forming COINs of interest.

My goal here is to demonstrate the basic techniques of doing a type of automatic social network analysis at the interdepartmental level. The application goal was to produce data on interdisciplinary/interdepartmental collaboration as evidence in support of an accreditation review for the Savannah College of Art and Design (SCAD).

There are four key points this example makes: 1) collaborative innovation often takes place in organizational settings with resources and constraints shaped by system components at levels higher than individuals; 2) The individual level of analysis used in most COINS research ignores the departmental level in organizations, where departments are often considered by participants as the key collaborating units; 3) The betweenness centrality measures used in most COINS research to identify innovative groups of individuals are not appropriate given the assumptions that such collaborations involve communication processes in which messages need not always flow through the shortest path, can be distributed through more than one path either synchronously or asynchronously, and may be increased in frequency of communication by the social actors.

As Borgatti (2005) points out, betweenness centrality is inconsistent with these assumptions because it is based on finding one shortest path linking each pair of actors, treats links as present or absent rather than having valued strengths, and assumes one message is disseminated down each shortest path. Instead, flow betweeenness is the measure that is consistent with the theoretical and practical assumptions about communication discussed in the previous example. This is the measure for identifying COINs used in this research; and 4) How media represent collaborations can be important both for mapping COINs and for observing how these are portrayed in media. Theoretically interesting audience loops back to the COINs may communicate perceptions that influence both the participants and their social observers, communicate changes in resources, or impose constraints on the innovation.

Methods

A list of the department names in the Savannah College of Art and Design was obtained from college personnel. My approach to assembling corpora for mining was to search Lexis/Nexis Academic to identify stories about the college from 2005 to 2008. I obtained every story containing the college's full name or acronym (SCAD), resulting in a census of the relevant text universe. I then aggregated all of these files into one text file and used the TimeSlice utility in WORDij 3.0 to segment the file into four annual files. Each of the four text files I automatically analyzed using WORDij 3.0's WordLink program to measure the co-occurrence of the department names.

As I pointed out in example 1, WORDij was originally designed to analyze large numbers of co-occurring words to create semantic networks. Nevertheless, social actors' names are indeed words and mining for their co-occurrence is no different. WORDij 3.0 not only has a stop-word list or droplist, it also has its opposite, an 'include list' that will map the network only among words on it. For this example, using WORDij 3.0's string replacement and include list functions, all aliases I created for each department's name were converted to a single unigram of letters and I then computed proximity-based co-occurrences in WordLink.

Automatic Link Coding with Proximities not with 'Bag of Words'

A key point I saved until now is that proximity co-occurrence indexing (Danowski, 1982, 1993a, 1993b, 2009, Diesner & Carley, 2004) avoids the problems of the simplistic 'bag of words' approaches common from Information Science and Information Retrieval. While word bags are useful for document retrieval they blur social meaning by ignoring the relationships of social units within the texts, whether these units are words, people, or other entities. It is more analytically precise, however, to use a proximity criterion in defining relationships among entities network analyzed. I used a three-word window, based on empirical testing of window size and network strucrture validity (Danowski, 1993b) operating on the text file after all words except the names of departments were automatically removed by the use of the WordLink 'include list' of department names. Table 2 shows an example of a portion of a larger include file.

Table 2. Examples of String Replacement

Department of Advertising and Design->Advert_Design
Dept. of Advertising and Design->Advert_Design
Dept of Advertising and Design->Advert_Design
Advertising and Design Department->Advert_Design
Advertising and Design Dept. ->Advert_Design
Advertising and Design Dept->Advert_Design
Advertising and Design->Advert_Design
Department of Accessory Design->Accessory
Dept. of Accessory Design->Accessory
Dept of Accessory Design->Accessory
Accessory Design->Accessory
Accessory Design Department->Accessory
Accessory Design Dept. ->Accessory
Accessory Design Dept->Accessory

Post-Processing of Link Data for Centrality Measures

As you recall from the first example, the WORDij 3.0 program has the option of producing a network file in the. net Pajek (Batagelj & Mrvar, 1998) format. This is one of the import file types that UCINET (Borgatti, Everett, & Freeman, 2002) accepts and converts to its system files. I chose UCINET because it is widely accepted in the social network analysis community and I wished to use common validated centrality indices to profile the structures. Given the status of UCINET and the ease of output importing I have felt no need to incorporate centrality measures into WORDij.

Combining Visualization with Statistical Network Centrality of Actors

A fundamental tenet of data analysis is to first visualize it. WORDij 3.0 has VISij for creating static or time-series movies of changes in network composition and structure. My interest in this example is in profiling the aggregate networks of departments, year by year, therefore a series of static representation results. While VISij has time-series animated visualizations that NetDraw does not have, NetDraw has more options for rendering static networks such as having larger circles for more central nodes. I used eigenvector centrality (Bonacich, 2007) to visually render the nodes' network position in the graphs, because the NetDraw program does not compute flow betweenness, while UCINET does. For link strength I used the maximum available range of thickness of links, from 0 to 12. The larger array of strengths was converted to this scale.

Although visualizing data is essential to help place further statistical analysis in context, it has its limitations, beyond the lack of rules for analysts to use in assessing network visualizations. Spring-embedded layout procedures may present the analyst with a different vantage point on the network each time it is run on the same data, because the stable structure can rotate as a whole, which can result in differing interpretations. Using statistical information following visual inspection of networks affords the analyst with the best of each mode.

While when there are small numbers of social actors visualization may have sufficient face validity to support action with respect to the network, it becomes increasingly less useful as the number of nodes and links increases above 30. How intensively and extensively greater numbers of nodes are linked can add to visual information overload, rendering interpretation of networks of questionable validity. It is difficult to make effective interpretations when the network looks like a cross-cultural accident of a big bowl of spaghetti with jambalaya on top, as is usually the case with visual output from semantic network analysis programs such as Crawdad (Corman, Kuhn, McPhee, & Dooley, 2002), rather than like a plate of sushi.

Results

There were 1,946 full text documents from Lexis/Nexis Academic for 2005 through 2008 for the college. Table 3 shows departmental flow betweenness centrality on a yearly interval. Figures 10-13 show the interdepartmental networks by year. Figure 14 is the network aggregated across the four years.

Interpretation

This example focused on showing an additional variation of a new method for identifying the social network structures that emerge in analyzing co-occurrences of organizational departments in news stories using automated text mining. The method used and described in this example appears to have face validity for their accreditation review according to college officials and is worthy of further refinement

Future research of potential interest would be simultaneous mapping of concepts and objects (tools, resources, places, etc.) along with the interdepartmental links, representing all of these in the same network. In this way, one could observe large numbers of social networks automatically viewing the word-networks with which the social actors are associated and with what objects and geographic locations they are linked. Some scholars (LaTour, 2005; Diesner & Carley, 2008) prefer to treat as nodes in the same network the social units, the words they use or are used to describe them, place names, and other proper nouns. WORDij 3.0 also enables this kind of 'actor network theory' mixture of nodes in the same networks identified. Further research might find exploration of these features valuable.

Example 3: Mining Email for Analysis of Intraorganizational Innovations

Email networks within organizations has been studied by researchers such as Danowski and Edison-Swift (1985), Diesner, Frantz, and Carley (2005), and Gloor and Zhang (2006). Example 4 was conducted among the Ford Motor Company's product engineering staff across its global network as it developed the "Sync® w/ MyFord Touch" innovation. This was a package of six new products with a single overall name that were created to be new vehicle control features for drivers across a range of the company's vehicles. The number of engineers involved was approximately 1,900.

Communication about the innovations occurred in a variety of modes but the most tractable was email about the innovation. Monitoring all relevant electronic mail over time overcomes limitations of cross-sectional, self-report data which include considerable error introduced by respondents' memory processes and their ability to report only gross summary features of their communication messages about an innovation. To avoid such problems, this study created a procedure by which the over time email content of individuals was monitored both historically and in real time over a two-year period. The organization exclusively used Microsoft Outlook for email with approximately 400 servers with up to 4,000 users each. The first step in establishing the monitoring procedure was to create a new Outlook rule that a participant would apply once on their Windows pc so that it would search all historical emails stored by individuals for key words associated with the innovations and forward these emails to a dedicated researchers' server, and also forward all relevant emails in real time during the course of the study, from January 2005 to December 2007.

A major discussion about deploying this procedure was among the executives and corporate attorneys to ensure confidentiality, respect for the personal privacy of the participants, and compliance with the legal systems of employees from different countries, particularly the Eurpoean ones. After an in-depth review by the company's lawyers from multiple countries the process was approved as an opt-in choice, an approach more characteristic of Europe than the USA. In all cases, participation was voluntary and employees could stop participating at any time. Participants were asked to run rules in Microsoft Outlook to enable automatic forwarding of email to a 'dummy email

Table 3. Interdepartmental Normalized Flow Betweenness by Year c8v9*

2005
nFlowBtn

Mean 2. 23
Std Dev 3. 99
Network Centralization Index = 10.8%
nFlowBtn
----
fashion 12.62
animation 12.54
interior 11.89
filmtv 5.05
architec 3.62
performing 3.62
painting 2.56
print 1.81
sequential 0.92
photog 0.58
illus 0.40
2006
nFlowBtn

Mean 1.41
Std Dev 3.10
Network Centralization Index 10.3%
nFlowBtn
----
painting 11.22
teaching 9.34
performing 4.86
photog 2.28
animation 0.85
writing 0.78
foundation 0.13
fashion 0.06
jewelry 0.06
2007
nFlowBtn

Mean 3.78
Std Dev 5.67
Network Centralization Index 14. 04%
nFlowBtn
----
painting 17.25
fashion 15.60
animation 14.42
performing 12.40
interior 11.72
architec 7.32
print 6.97
filmtv 3.35
jewelry 1.69
urban 1.38
writing 1.28
accessory 0.95
teaching 0.18
2008
nFlowBtn

Mean 4.87
Std Dev 7.89
Network Centralization 23.9%
nFlowBtn
----
photog 27.82
painting 26.39
interior 17.50
performing 10.67
architec 10.25
teaching 9.83
fashion 4.16
sequential 3.56
animation 3.05
filmtv 2.89
jewelry 2.49
foundation 1.74
urban 1.25
sculpture 0.16

Figure 10. 2005 Interdepartmental Network

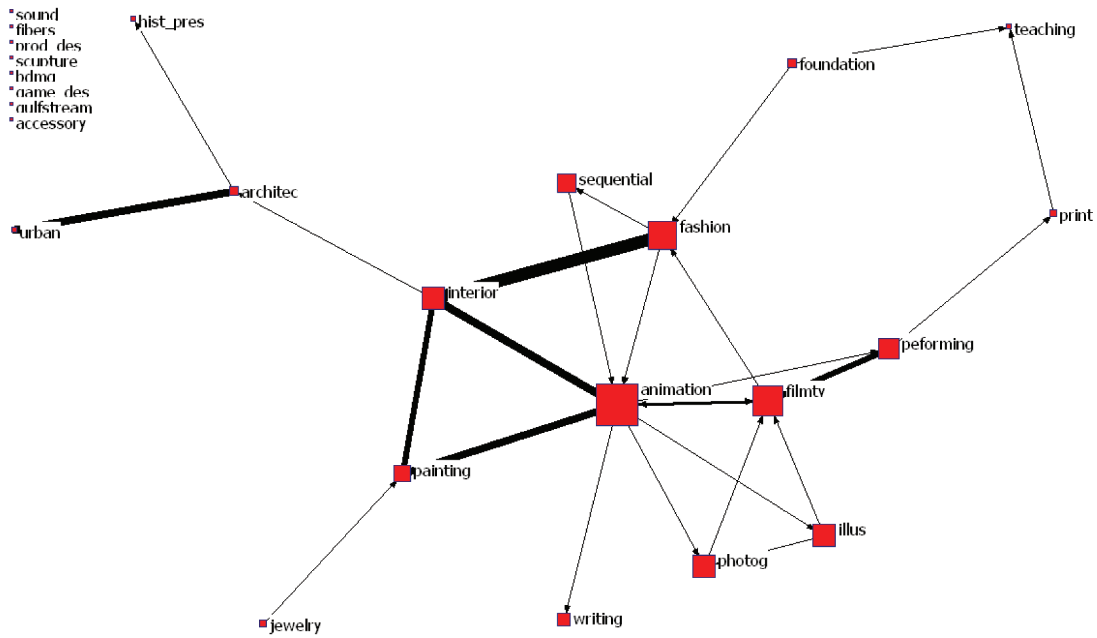


Figure 11. 2006 Interdepartmental Network

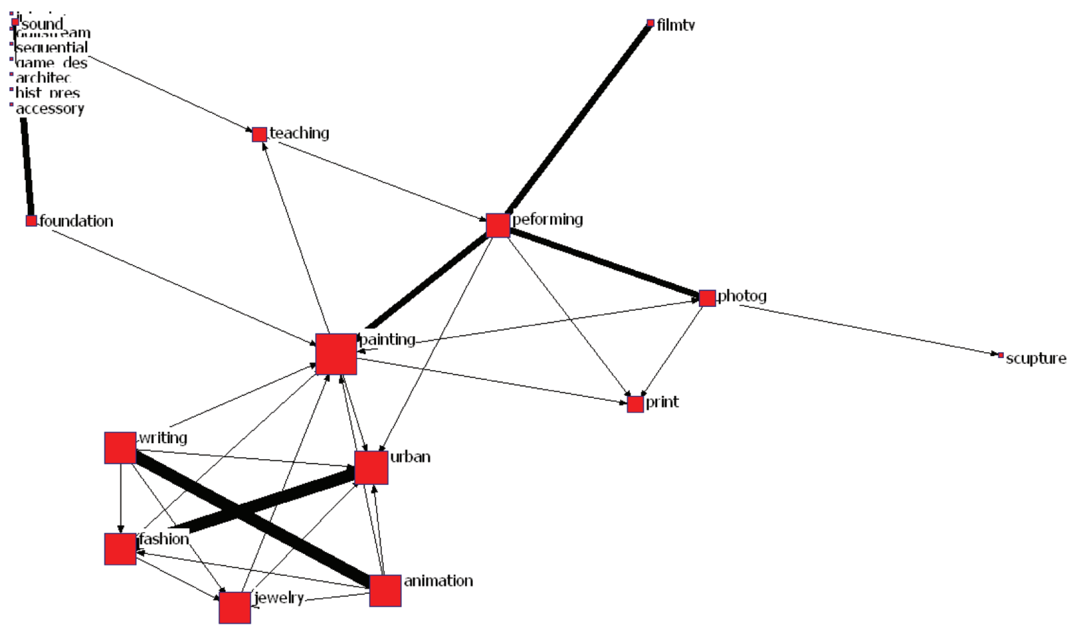


Figure 12. 2007 Interdepartmental Network

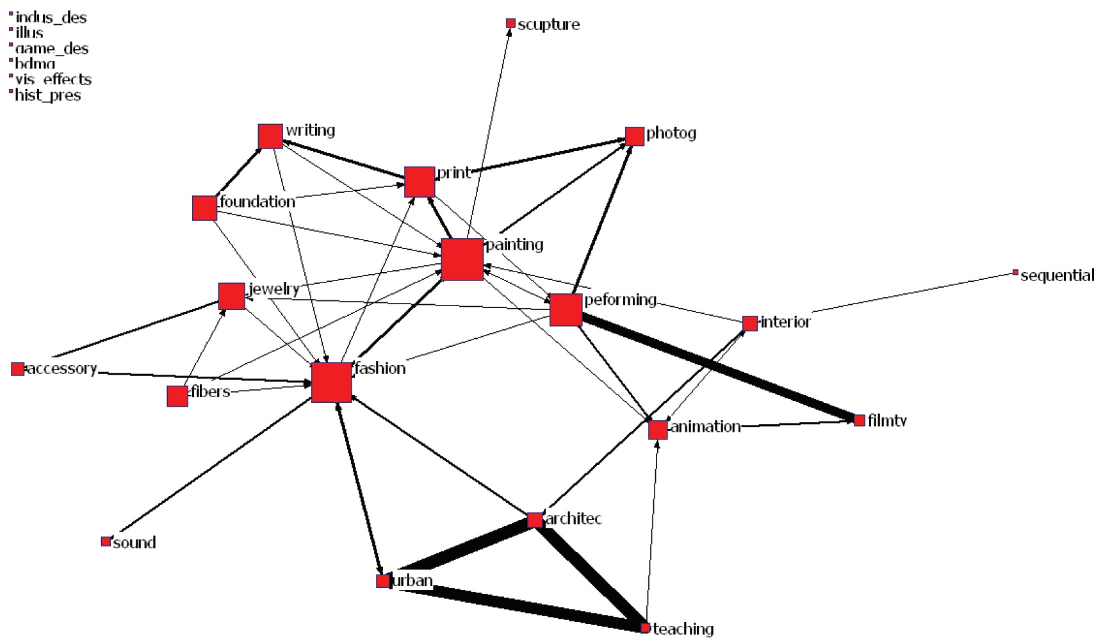


Figure 13. 2008 Interdepartmental Network

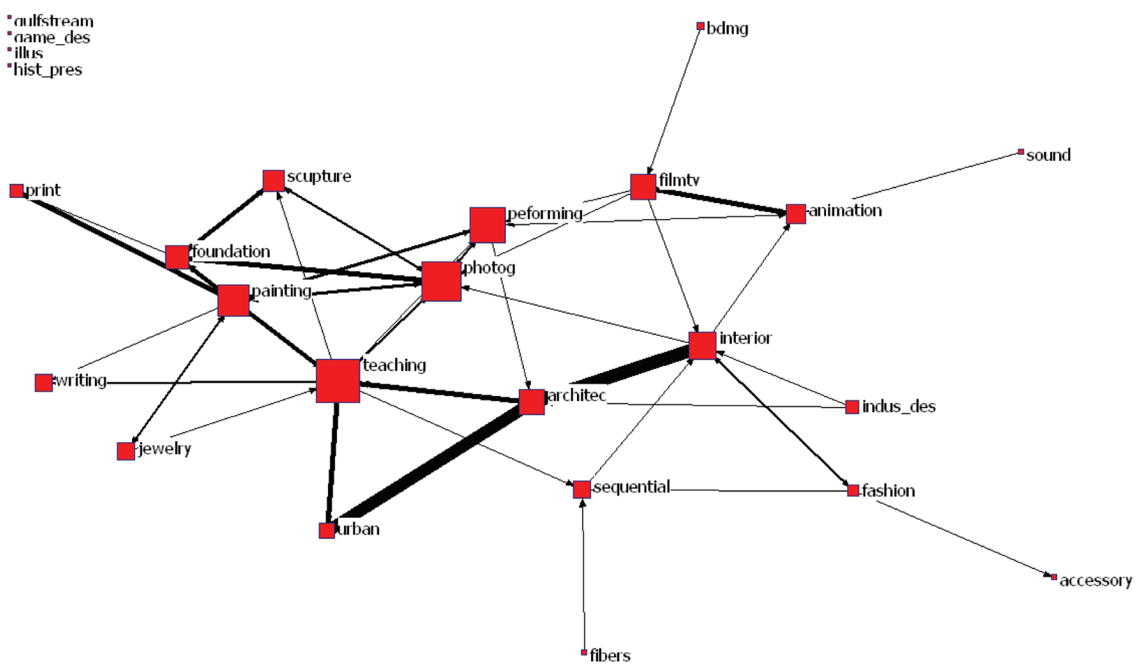
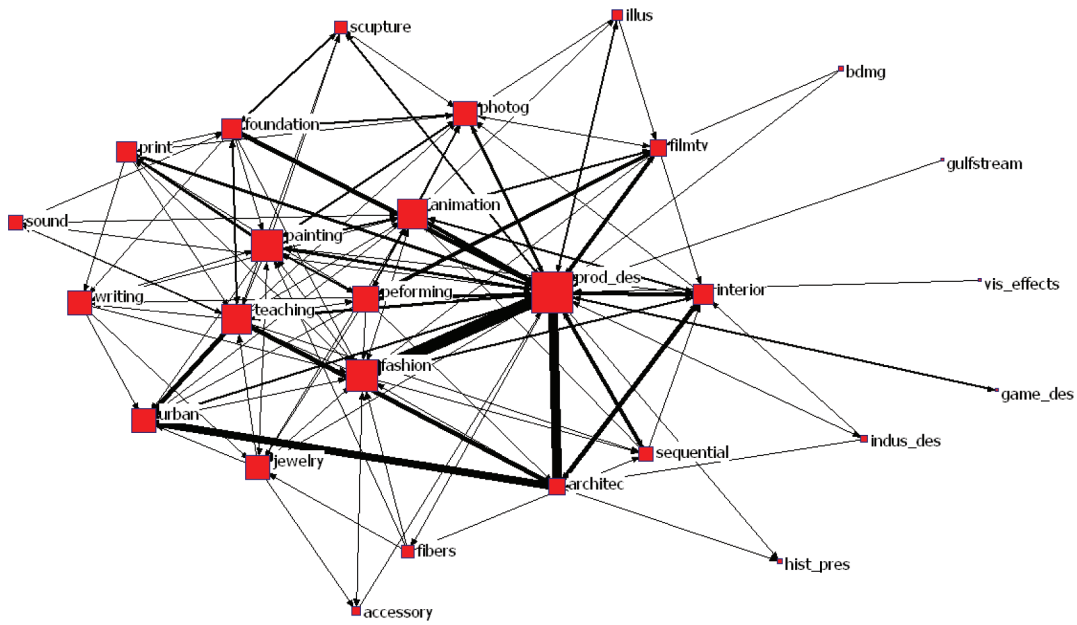


Figure 14. Aggregate Interdepartmental Network



address' on a secure server designated to store the study data.

Once the email data collection process was approved, the academic researchers for this National Science Foundation grant received IRB approval from the two universities at which the PI (Jullia Gluesing at Wayne State University and Co-PI Ken Riopelle also at WSU), and I as Co-PI at the University of Illinois at Chicago were located. I based the protocol on a procedure to protect privacy used in an email network analysis study by Danowski and Edison-Swift (1985).We converted all names in the emails to numbers whose key was known only to the corporate staff who collaborated in this anonymizing activity. While the corporate research team knew the identities of all named individuals in the emails, the academic researchers had no knowledge of these identities. IRB applications were submitted for Exempt status for the research, meaning that because no identifying information was known to the academic researchers, the research was approved as exempt from the need to obtain informed consent.

The first analysis step was to deploy the Outlook rule to the project manager's email going back 9 years. The academic researchers then conducted a network analysis of the who-to-whom network from these emails to identify the most central individuals. The Negopy network analysis program (Richards, 1985) found one large group, evidence of a negentropic center/periphery structure. Because Negopy computes the geodesic distance scores among all pairs of nodes in each group, in this case one, comparable centrality information was available for all nodes in the network. This information was presented to the corporate researchers who used this information along with other considerations to choose 298 target individuals for the email harvesting.

The 298 targets were initially sent an email requesting their participation and application of the Outlook rules. Thirty eight chose to voluntarily participate, for a participation rate of 13%. Follow-up by project staff noted that potential users did not wish to have their email monitored in this fashion because of uncertainty around what it would find and how it would be used, even though most were aware that all corporate email was legally monitored in real time for 'illegal' activity words, as authorized by the Electronic Communication Privacy Act in the U.S. Nevertheless,. Forwarding and Replying were very extensively used in this organization, creating long chains of email going back over two years. Because of this mining of threaded email we found that capturing all innovation emails from only 1% of the 1,900 project engineers was sufficient to capture emails of approximately 1,900 individuals exchanging approximately 45,000 emails. As a result we constructed a two-year time series of emails about the innovations.

There were many facets to the automated email analysis, tracking the semantic networks associated with innovations over time, measuring the who-to-whom networks in relation to this message content, and the four examples shown here: 1) overlaying the who-to-whom email network on the formal organizational hierarchy; 2) highlighting the networks of people who left to provide communication training data for replacement people; 3) semantic associations to an innovation in the organization; and 4) sentiment analysis based on ratios of positive to negative email content over time.

Typically networks in organizations are shown in a flat horizontal plane, ignoring the formal organizational structure. In our example the corporate managers of the project wanted to see how the innovation's who-to-whom email network related to levels in the chain of command. Colleagues Ken Riopelle, Andrew Seary, and Julia Gluesing of the research team used MultiNet (Seary, 2005; Riopelle, Danowski, & Gluesing, 2008) for this purpose. The CEO was defined as 'level 1" and each level below, for the 100 most active nodes, was indexed down to level 9, as is seen in Figure 15. A related interest of the corporate managers was to see the networks of those who had left during the previous downsizing, to help orient new replacements to their expected communication networks. Figure 16 shows one example for four individuals who left, showing levels ranging from 3 down to 15.

Active Nodes

To show an example of semantic analysis of innovations from the content of the emails, we

Figure 15. Email Communication Who-to-Whom Networks Across Levels: 100 Most

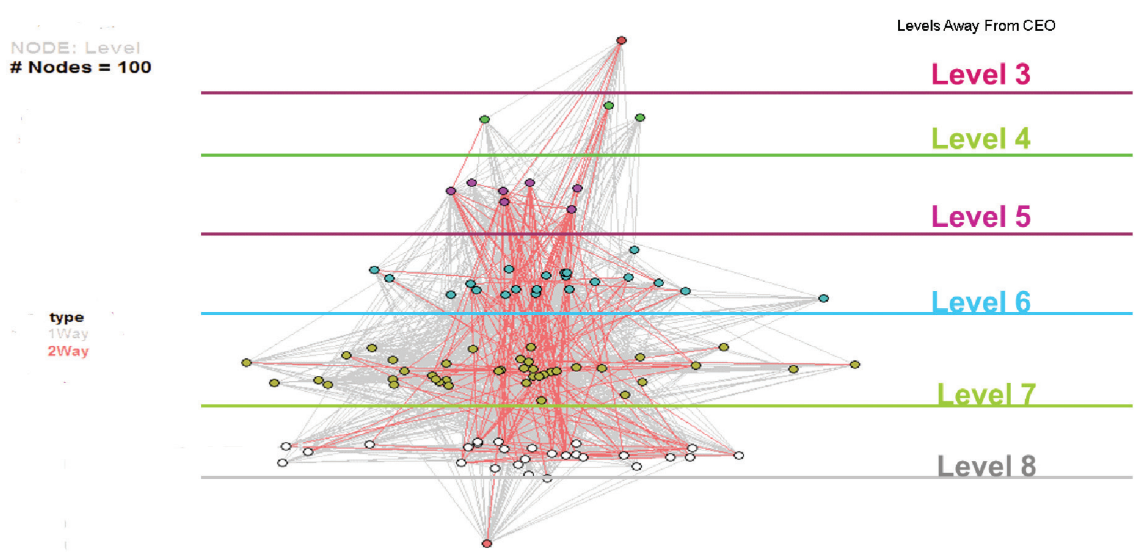
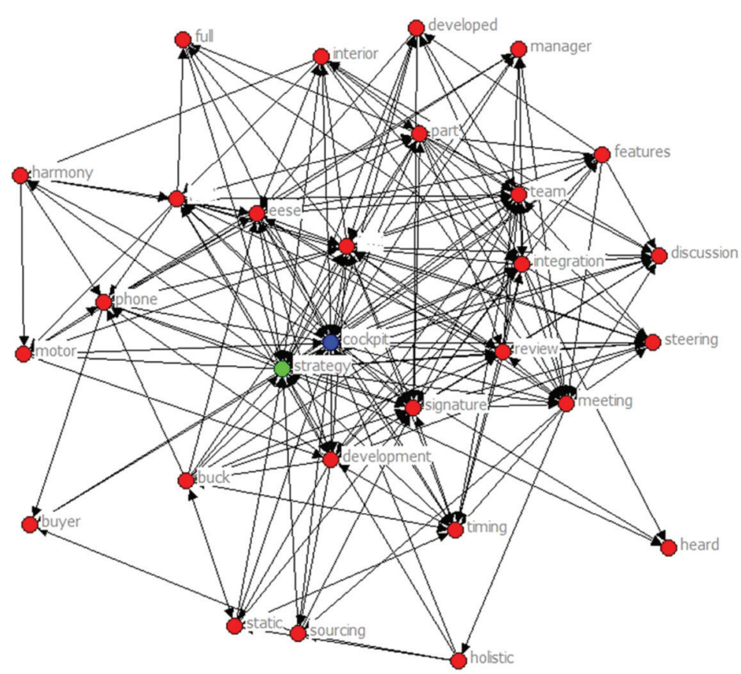


Figure 16. Who-to-Whom email networks of Some Individuals Who Left the Organization

present in Figure 17 one network associated with the concept of automobile 'cockpit,' for which innovations were developing. Several concept labels are redacted

Lastly, we present a graph of the ratio of positive to negative sentiment over time (Danowski, Riopelle, Gluesing, 2008). We indexed sentiment of email texts using the LIWC diction-

Figure 17. Semantic Network Associated with the 'Cockpit' Innovation

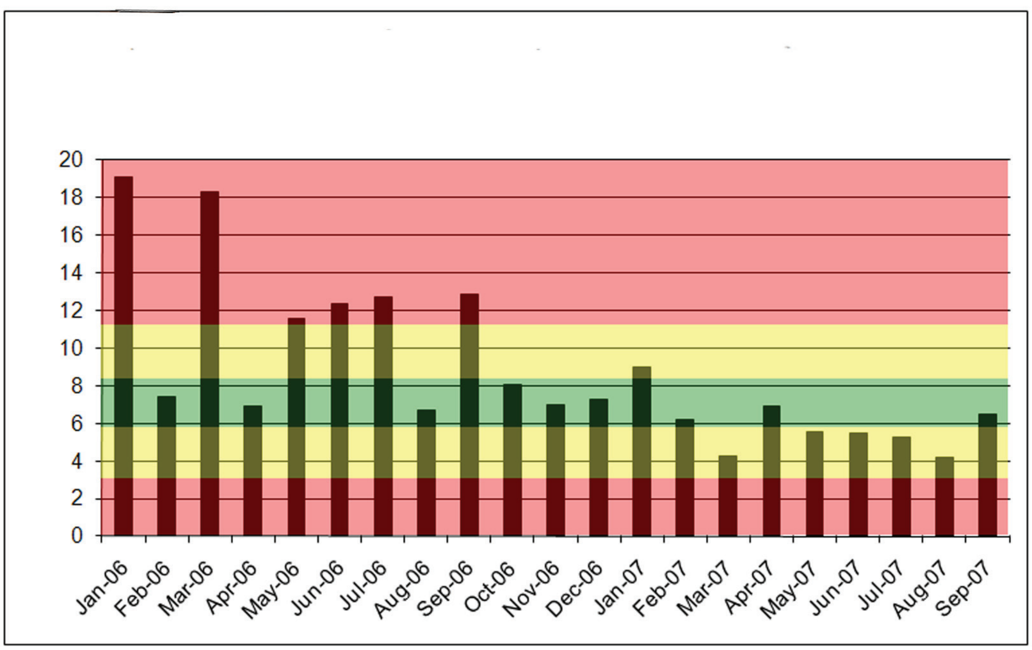


ary-based content analysis software (Pennebaker, Booth, & Francis, 2007). This work was done prior to the later development of the shortest-path netwok analysis of positive and negative inflected forms of sentiment words illustrated by OptiCommReport in example 1. The analysis of the ratio of positive to negative communication is based on the work of Losada (Fredrickson, & Losada, 2005) which has found an optimal range of this ratio for healthy system performance, from 3.0 up to 11.0. Under that ratio level there is likely ineffective performance. This chart in Figure 18 shows the Losada Line graph for all of the engineers working on the innovations. Not shown here is a similar chart for one of the elements of the package of innovations. At one point the positive/negative ratio dropped significantly below the 3.0 level and stayed low over the next six months. We did not have a chance to share this information with corporate executives, but six months later they decided that the unit was not meeting project objectives and terminated its innovation and the engineers. After seeing our results the executives exclaimed that they wished they had known of them because they could have terminted six months earlier and saved millions of dollors.

CONCLUSION

I have illustrated in this chapter six examples of organization-related social network mining: 1) interorganizatinal networks in the Deepwater BP Oil Spill events and sentiment analysis over time, 2) intraorganizational interdepartmental networks in the Savannah College of Art and Design (SCAD) over time, 3) who-to-whom email networks across the Ford Motor Company hierarcy in an automative engineering function, 4) networks of selected individuals who left that organization, 5) semantic associations across email for a corporate innova-

Figure 18. Positive/Negative Email Sentiment Ratio Over Time



tion the "Sync® w/ MyFord Touch" product, and 6) assessment of sentiment across its email for innovations over time.

I have attempted to give you sufficient detail on the motivations for these examples, their methods, and possible scientific and management applications to stimulate your own ideas along these or other lines of research. While much social network analysis mining uses individuals as nodes, the first two examples use organizations as nodes. The third and fourth examples makes use of formal organizational structure in relation to mining communicatioin networks among individuals. Examples 5 and 6 analyze words as nodes mined from organizational email. This array of organizational mining examples may stimulate you to pursue your own.

ACKNOWLEDGMENT

This research was supported in part by the National Science Foundation's Human and Social Dynamics (HSD) Award #SES-527487: Accelerating the diffusion of innovations: A "digital diffusion dashboard" methodology for global networked organizations.

Special thanks to Julia Gluesing and Ken Riopelle of the Department of Industrial and Manufacturing Engineering at Wayne State University, Detroit, Michigan, USA for their management of researchers at the Ford Motor Company.

REFERENCES

Batagelj, V., & Mrvar, A. (1998). Pajek - Program for large network analysis. *Connections*, *21*(2), 47–57.

Bonacich, P. (2007). Some unique properties of eigenvector centrality. *Social Networks*, *29*(4), 555–564. doi:10.1016/j.socnet.2007.04.002

Borgatti, S. P. (2002). *NetDraw: Graph visualization software.* Harvard, MA: Analytic Technologies.

Borgatti, S. P. (2005). Centrality and network flow. *Social Networks*, *27*, 55–71. doi:10.1016/j.socnet.2004.11.008

Borgatti, S. P., Everett, M. G., & Freeman, L. C. (2002). *UCINET for windows: Software for social network analysis*. Harvard, MA: Analytic Technologies.

Borgatti, S. P., & Foster, P. C. (2003). The network paradigm in organizational research: A review and typology. *Journal of Management*, *29*(6), 991–1013.

Borgatti, S. P., & Molina, J. L. (2003). Ethical and strategic issues in organizational social network analysis. *The Journal of Applied Behavioral Science*, *20*(10), 1–13.

Burt, R. S. (1995). *Structural holes: The social structure of competition*. Cambridge, MA: Harvard University Press.

Carley, K. M., Diesner, J., Reminga, J., & Tsvetovata, M. (2007). Toward an interoperable dynamic network analysis toolkit. *Decision Support Systems*, *43*(4), 1324–1347. doi:10.1016/j.dss.2006.04.003

Chaffee, S. H., & Metzger, M. J. (2001). The end of mass communication? *Mass Communication & Society*, *4*(4), 365–379. doi:10.1207/S15327825MCS0404_3

Corman, S., Kuhn, R., Mc, T., Phee, R. D., & Dooley, K. J. (2002). Studying complex discursive systems: Centering Resonance Analysis of communication. *Human Communication Research*, *28*(2), 157206. doi:10.1093/hcr/28.2.157

Cross, R., Borgatti, S. P., & Parker, A. (2002). Making invisible work visible: Using social network analysis to support strategic collaboration. *California Management Review*, *44*(2), 25–46.

Cross, R., & Parker, A. (2004). *The hidden power of social networks*. Boston, MA: Harvard Business School Press.

Danowski, J. A. (1982). A network-based content analysis methodology for computer-mediated communication: An illustration with a computer bulletin board. In Bostrom, R. N. (Ed). *Communication Yearbook, 6*, 904-925. New Brunswick, NJ: Transaction Books.

Danowski, J. A. (1988). Organizational infographics and automated auditing: Using computers to unobtrusively gather and analyze communication. In G. Goldhaber & G. A. Barnett (Eds.) *Handbook of Organizational Communication*, (pp. 335-384). Norwood, NJ: Danowski, J. A. (1993a). *Automatic narrative generation via statistical content analysis of large-scale textual collections.* Paper presented at Conference on Computing for the Social Sciences, National Center for Supercomputing Applications, University of Illinois at Urbana-Campaign, May 18-21.

Danowski, J. A. (1993b). WORDij 3.0: A word-pair approach to information retrieval. *Proceedings of the DARPA/NIST TREC Conference*, Washington, DC.

Danowski, J. A. (1993c). Network analysis of message content. In Rice, R. E., Richards, W., & Barnett, G. (Eds.), *Progress in communication sciences XII* (pp. 197–222). Norwood, NJ: Ablex.

Danowski, J. A. (2007). *Comparing semantic network analysis approaches.* Paper presented to the International Association for Social Network Analysis, Sunbelt XXVII annual meetings Corfu, Greece, May.

Danowski, J. A. (2009). Inferences from word networks in messages. In Krippendorff, K., & Bock, M. A. (Eds.), *The content analysis reader* (pp. 421–429). Sage Publications.

Danowski, J. A. (2010). *WORDij 3.0. [Computer program].* Chicago, IL: University of Illinois at Chicago.

Danowski, J. A., & Cepela, N. (2010). Automatic mapping of social networks of actors from text corpora: Time series analysis. *Annals of Information Systems*, *12*, 31–46. doi:10.1007/978-1-4419-6287-4_3

Danowski, J. A., & Edison-Swift, P. (1985). Crisis effects on intraorganizational computer-based communication. *Communication Research*, *12*, 251–270. doi:10.1177/009365085012002005

Danowski, J. A., & Edison-Swift, P. (1985). Crisis effects on intraorganizational computer-based communication. *Communication Research*, *12*(2), 251–270. doi:10.1177/009365085012002005

Danowski, J. A., Gluesing, J., & Riopelle, K. (2009). *ICT's effects on diffusion curves, email networks, and semantic networks in an organization: The emergence of herding, bandwagons, and information cascades with less interpersonal communication.* Paper presented to the Communication and Technology Division of the International Communication Association, Chicago, May 21-25.

Danowski, J. A., Riopelle, K., & Gluesing, J. (2008). *Using network analysis of who-to-whom email networks to structure semantic network analysis of email content and manage it in innovation contexts.* Paper presented at annual meetings of the International Network for Social Network Analysis, Sunbelt XXVIII annual meetings, St. Pete Beach, FL, January, 2008.

Diesner, J., & Carley, K. M. (2004). *Using network text analysis to detect the organizational structure of covert networks.* Pittsburg, PA: CASOS, Carnegie-Mellon. Available from: http://www.contrib.andrew.cmu.edu/ ~jdiesner/publications/NAACSOS_2004_ Diesner_Carley_Detect_Covert_Networks.pdf

Diesner, J., & Carley, K. M. (2004). *AutoMap 1.2: Extract, analyze, represent, and compare mental models from texts.* Pittsburg, PA: CASOS, Carnegie-Mellon University. Available from: http://reports-archive.adm.cs.cmu.edu/ anon/anon/usr/ftp/isri2004/ CMU-ISRI-04-100.pdf

Diesner, J., & Carley, K. M. (2008). Conditional random fields for entity extraction and ontological text coding. *Computational & Mathematical Organization Theory*, *14*(3), 248262. doi:10.1007/s10588-008-9029-z

Diesner, J., Frantz, T. L., & Carley, K. M. (2005). Communication networks from the Enron email corpus 'it's always about the people. Enron is no different'. *Computational & Mathematical Organization Theory*, *11*(3), 201228. doi:10.1007/s10588-005-5377-0

Fredrickson, B. L., & Losada, M. F. (2005). Positive affect and the complex dynamics of human flourishing. *The American Psychologist*, *60*(7), 678–686. doi:10.1037/0003-066X.60.7.678

Freeman, L. C. (1979). Centrality in social networks: Conceptual clarification. *Social Networks*, *1*, 215–239. doi:10.1016/0378-8733(78)90021-7

Freeman, L. C. (1996). Some antecedents of social network analysis. *Connections*, *19*(1), 39–42.

Freeman, L. C., Borgatti, S. P., & White, D. R. (1991). Centrality in valued graphs: A measure of betweenness based on network flow. *Social Networks*, *13*(2), 141–154. doi:10.1016/0378-8733(91)90017-N

Galaskiewicz, J. (1985). Professional networks and the institutionalization of a single mindset. *American Sociological Review*, *50*, 639–658. doi:10.2307/2095379

Galaskiewicz, J., & Shatin, D. (1981). Leadership and networking among neighborhood human service organizations. *Administrative Science Quarterly*, *26*, 434–4. doi:10.2307/2392516

Gloor, P. (2006). *Swarm creativity: Competitive advantage through collaborative innovation networks*. Oxford University Press US.

Gloor, P., & Zhao, Y. (2006). *Analyzing actors and their discussion topics by semantic social network analysis.* Tenth International Conference on Information Visualization, 5-7 July 2006, 130-135. London, England. Available from: http://www.ickn.org/documents/IV06_G2341_Gloor.pdf

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, *78*(6), 1360–1380. doi:10.1086/225469

Latour, B. (2005). *Reassembling the social: An introduction to Actor-Network-Theory*. London: Oxford University Press.

Mizruchi, M. S. (1996). What do interlocks do? An analysis, critique, and assessment of research on interlocking directorates. *Annual Review of Sociology*, *22*, 27129. doi:10.1146/annurev.soc.22.1.271

Monge, P. R., & Eisenberg, E. M. (1987). Emergent communication networks. In Jablin, F. M., Putnam, L. L., Roberts, K. H., & Porter, L. W. (Eds.), *Handbook of organizational communication: An interdisciplinary perspective* (p. 304342). Newbury Park: Sage.

Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). *Linguistic Inquiry and Word Count 2007 (LIWC2007)*. Austin, TX: LIWC, Inc.

Richards, W. D. Jr. (1995). *NEGOPY 4.30 Manual and user's Guide*. Burnaby, Canada: School of Communication, Simon Fraser University.

Riopelle, K., Danowski, J. A., & Bishop, A. (2010). *Expression of sentiment by different node positions in email networks.* Paper presented to annual meetings of the International Network for Social Network Analysts, Riva Del Garda, Italy, June 29-July 4.

Riopelle, K., Danowski, J. A., & Gluesing, J. (2008). *Organizational hierarchy predictors of symmetric and asymmetric dyad and triad distributions in a large global organization*. Paper presented at annual meetings of the International Network for Social Network Analysis, Sunbelt XXVIII annual meetings, St. Pete Beach, FL, January.

Scott, J. (1988). Social network analysis. *Sociology*, *22*(1), 109–127. doi:10.1177/0038038588022001007

Scott, J. (1991). Networks of corporate power. *Annual Review of Sociology*, *17*, 181203. doi:10.1146/annurev.so.17.080191.001145

Scott, J. (2000). *Social network analysis: A handbook* (2nd ed.). Thousand Oaks: Sage.

Seary, A. J. (2005) *MultiNet: An interactive program for analysing and visualizing complex networks.* Doctoral Dissertation, Burnaby, CA: Simon Frazier University. Available from: http://people.math.sfu.ca/~goddyn/ Theses/searyPhDThesis.pdf

Tichy, N. M., Tushman, M. L., & Fombrun, C. (1979). Social network analysis for organizations. *Academy of Management Review*, *4*(4), 507–519.

Valente, T. W. (1996). Social network thresholds in the diffusion of innovations. *Social Networks*, *18*, 69–89. doi:10.1016/0378-8733(95)00256-1