

# Automatic Mapping of Social Networks: Time Series Analysis of News Sentiment and Presidential Job Approval

James A. Danowski  
Noah Cepela  
University of Illinois at Chicago

**Abstract.** To test hypotheses about presidential cabinet network centrality and presidential job approval over time and to illustrate automatic social network identification this research mined the social networks among cabinets of Presidents Reagan through G.W. Bush based on the members' co-occurrence in news stories. Each administration's data was sliced into time intervals corresponding to Gallup presidential approval polls. It was hypothesized that when the centrality of the president is lower than that of other cabinet members, job approval ratings are higher. This is based on the assumption that news is generally negative and when the president stands above the other cabinet members in network centrality, he or she is more likely to be associated with the negative press coverage in the minds of members of the public. The hypothesis was supported. Nevertheless, when the positive and negative sentiment of news stories is added, each administration was found to have different effects.

## 1 Introduction

Political and communication science has long valued a network analysis approach to conceptualizing and measuring phenomena. Among the earliest to map the networks of political actors were the studies of political communication among voters (Katz & Lazarsfeld, 1955). At the level of community, others have investigated networks of political power (Hunter, 1953, Dahl, 1961, Polsby, 1965). Organizations have been conceptualized in political economy terms using social network analysis frameworks (Tichy, Tushman, Fombrun, 1979; Galaskiewicz, 1979). A sweeping explication of political networks ranging from individual through international levels has placed network concepts at the center of political processes (Knocke, 1994). Of particular relevance to the current study, presidential cabinets have been seen in network terms (Feeno, 1959), although have yet to be measured from this perspective.

Here we introduce a method of automatic identification of the networks among presidential cabinet actors. Mining large volumes of news and web documents for evidence of the identities of social actors and their relationships is increasingly feasible. Moreover, because most online information has time stamps, it is possible to construct time-series analyses of how social networks change over time, and how the network variables are associated with other kinds of variables over time. This can give two of the three necessary conditions for causality: 1) association, and 2) time order, leaving for the analyst's further examination: 3)

Danowski, J. A., & Cepela, N. (2010). Automatic mapping of social networks of actors from text corpora: Time series analysis, in Memon, N., Jie Xu, J., Hicks, D. L., & Chen, H. (Eds.) *Data mining for social network data*. *Annals of information science*, 12, (pps. 31-46). New York: Springer Science+Business Media.

ruling out rival explanations as potential causes of the response variables of interest.

To illustrate the measurement of association and time-order from social network mining we use as an example the relationships among members of the U.S. presidential administration cabinets for Presidents Reagan, G.H.W Bush, Clinton and G.W. Bush. We identify the networks among the cabinet members based on their co-occurrence in news stories. The network centrality of each actor and of the entire network is indexed and examined in association with the job approval ratings of the president over the course of the administrations.

One of the features of data mining for such networks is that the time slices can be readily set according to the situational conditions of the processes being studied. For example, for Clinton, the Gallup job approval ratings were measured on average 30 days apart, while for the G.W. Bush administration they were 22 days apart. Network time slices can be set according to the time intervals of the response variable, as is done in this study, increasing the interval validity of the research design.

Political scientists and communication scholars have studied predictors of presidential job approval and favorability for several decades. In the most recent wave of research, media variables are increasingly examined as predictors of presidential job approval and favorability (Gronke & Newman, 2000). The current research is an example of this. Rather than looking only at the amount of coverage within nominal categories of content, we take a more refined approach of automatic content analysis of the networks among cabinet members portrayed in the press within time slices.

Regarding our response variable, presidential job approval, prior research has found that job approval and favorability, measured separately in the Gallup polls, is very highly correlated. Few respondents hold inconsistent attitudes, such as reporting that a president is doing a bad job yet that they strongly like him (Cohen, 2000). In the current research, therefore, we use only the job approval variable.

In theorizing and measuring the effects of news coverage on public opinion, researchers have taken a variety of approaches. A network approach was central to the two-step flow model 66 years ago when researchers proposed that opinion leaders in social networks mediated news coverage effects on public opinion of the electorate (Lazarsfeld, Berelson, & Gaudet, 1944). Since that time, research on news coverage and political attitudes has mainly set aside concern with social networks and conceptualized the agenda-setting process of news coverage, investigating the extent to which the amount of news coverage of an issue is associated with how important the public perceives the issue to be (McCombs & Shaw More). Recently, still in an individualistic framework, news has been studied in terms of narrative framing and its effects (Entman, 1993), although investigators have returned to conceptualizing network variables in modeling news coverage effects on sentiment (Entman, 2007). Nevertheless, the focus of attention is on communication networks of elites in influencing media framing, rather than on networks contained within the content itself. We propose that networks among presidential cabinet members represented in the news mediate framing's influence on public attitudes toward the presidency.

The negative information orientation of the press is well documented (Entman, 2007). Positive events are not nearly as likely to be considered “news” as are negative events or negative characterizations of processes or personalities. Given a generally negative valence to news, our theoretical argument derives from the research on “divided presidencies” (Nicholson, Segura, & Woods, 2002), where one party holds the White House and another holds the Congress. Studies have found that this results in higher job approval ratings for the president. Investigators have reasoned that this is because of added uncertainty resulting from less ability to assign blame to the president for lack of political progress or failed initiatives. This uncertainty weakens the normal situation in which there is a negative bias in the media toward political actors and therefore increases the chances that the population perceives the president more positively, absent the normal flow of negative information being specifically tied to the chief executive.

We further reason that when the president is portrayed as a more central figure in the administration, he is more clearly the “lightning rod” for the generally negative information orientation of the press. Absent countervailing information, this negative coverage of the president results in audience members perceiving the president to have lower job performance, and thus they give him lower job approval ratings. On the other hand, when the president is less central in the administration network, as other cabinet members are more central, this structural dispersion makes it more difficult for the media to successfully tag the president in a negative manner. The negative “lightning” of media messages is more diffuse with multiple smaller bolt strikes. As a result, when the average centrality of cabinet members is higher, which lowers the centrality of the president, the job approval of the president is higher. The president is more likely to “fly beneath the radar” from the perspective of media audiences and with less connection of negative news with the identity of the president himself, job approval will increase. So, in addition to expecting that when average network centrality of cabinet members is higher presidential approval is higher, we also expect to find that when the president has higher network centrality, job approval is lower. Stating this more succinctly, the following hypothesis is offered:

*H1: The greater the similarity of the centrality of the president and his cabinet members, the higher the job approval ratings for the president.*

The contemporary speed with which these effects can be expected to occur is related to the substantial shortening of the news cycle since the growth of online news sources. These have affected all media such that the media cycle is no longer weekly or daily, it is hourly and even minutes/seconds in periodicity. The cycle was longer during the Reagan, G.H.W. Bush, and Clinton administrations. The “web” was launched in the middle of Clinton’s two terms and was still in early stages of development through the late 1990s. For the G.W. Bush presidency, on the other hand, the online news cycle had shortened (Dezső, Almaas, Lukács, Rácz, Szakadát, & Barabási, 2006), political blogs grew rapidly in his first term (Adamic, & Glace, 2005) and exerted increasing power on media framing through his second term (Farrell, & Drezner, 2008). We therefore expect that the relationship between administration network centrality and presidential job approval is

longer for the Reagan through Clinton administrations than it is for the G.W. Bush administration.

*H2: The lag between centrality similarity of president and cabinet members and an increase in job approval ratings for the president is shorter for the G.W. Bush administration than for the Reagan, G.H.W. Bush, and Clinton administrations.*

## 2 Methods

We have network analyzed the cabinets of each of the presidencies since Nixon aggregated across their respective terms (Cepela & Danowski, 2009). All Lexis-Nexis news stories in the *New York Times* and the *Washington Post* during the administrations that mention at least one of the cabinet members were captured in full text form. A separate search was done for each cabinet member during the time of the presidency. We then aggregated all of these files into one text file for each administration. There were 26mb of text and 30,194 stories for Nixon and his cabinet, 16.7mb and 18,432 stories for the short Ford term, 163mb and 46,586 stories for Carter, 653mb of text and 135,996 stories for Reagan's two terms, 93mb of text and 17,265 stories for G.H.W Bush, 674mb of text and 114,511 stories for the two Clinton administrations, and 504mb and 89,810 stories for the two G.W. Bush administrations. Although G.H.W Bush's administration generated relatively little press coverage, considering that the president is the single most covered news source (Jones, 2008) this would appear to be the result of a press strategy or of some other systematic variation.

For the current analysis four separate text files were created in UTF-8 format. Each file was then analyzed using WORDij 3.0 (Danowski, 2009a; 2009b) to measure the co-occurrence of actors mentioned within 400 words of each other. Each output file was then put into UCINET (Borgatti, Everett, & Freeman, 2002) to measure average network centralization, and individual centrality scores for each member and NetDraw (Borgatti, 2002) to create the static network visualizations. WORDij 3.0 produces time-series movies of networks, but these cannot be shown in such a paper.

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. Additionally, some features to aid in multi-node type analysis including people, organizations, places, and formal concepts and objects is enumeration of proper nouns and automatic creation of include lists from them. For this paper, using WORDij 3.0's string replacement and include list functions, all aliases we created for each cabinet member's name were converted to a single string and then proximity-based co-occurrences were computed.

## **2.1 Link Coding with Proximities not “Bag of Words”**

The extent to which cabinet members co-occur within large numbers of documents is the basis for defining the link in the social network with more co-occurrences indicating higher link strength. This co-occurrence indexing is proximity based so it avoids the problems of the simplistic “bag of words” approaches common from Information Science and Information Retrieval. Those treat all words in an entire document as having a link with one another, regardless of how far apart these terms might be in the document. While such word bags are useful for document retrieval they blur social meaning by ignoring or masking the more precise relationships of social units within the texts, whether these units are words, people, or other entities. More reflective of social structure are methods using a proximity criterion to define co-occurrence of links within a relatively short distance as introduced by Danowski (1982; 1993a; 1993b).

## **2.2 Optimal Window Size for Actor Social Networks**

We performed tests with different window sizes, using as a criterion the overall network structure. Windows of 3, 10, 25, 50, 100, and 200 were used. The window size of 3 produces the same network results as the other windows sizes, for example, having a QAP correlation coefficient of .997 with the network resulting from a window size of 200. The reason for this is because with the include file approach (opposite of a stop list or drop file) the WordLink subroutine of WORDij drops all words from the initial text except for the strings on the include list. So, the window of 3 when using an include list has face and predictive validity.

## **2.3 Actor Co-occurrence Segmentation Software**

We use the software package WORDij 3.0 that has a graphical user interface in java but runs fast on Windows, Mac, and Linux/Unix operating systems because all of the network analysis computations are done in C++ not java, unlike some other similar software such as AutoMap (Diesner & Carley, 2004). Among the many options for analysis in WORDij 3.0, relevant here is the option to produce Pajek (Batagelj, 1998) formatted output files because we wished to import these to UCINET for centrality computations.

## **2.3 Network Centrality Measures**

The most often used centrality measure in social network analysis is “betweenness centrality” (Hanneman, & Riddle, 2005). Nevertheless, use of this measure developed by Freeman (1977) has been shown to be very often inappropriate in

terms of its operational assumptions in relation to conceptual definitions (Borgatti, 2005). The betweenness measure assumes that messages flow through a network along a single shortest path, moving sequentially from one node on this path to the next. No weighting is assumed for the strength of each link in terms of its overall frequency of activation, bandwidth, or channel capacity. Accordingly it treats only dichotomous links, so even if strength or vertex valued data is available, the measure removes such information and codes each link as either present or absent. Sometimes analysts do this dichotomization by binarizing the data with a split at the median into 0 and 1 codes for each link. In the betweenness centrality model it is not possible for a node to receive messages from more than one other node nor can a node send out the same message to multiple nodes. These assumptions do not fit well with our conception of presidential cabinet networks. We assume that our representations of networks have an association with the actual communication among the actors.

Organizational communication is such that some relationships are stronger than others, with these stronger links more frequently activated. In addition, it is common for individuals to communicate the same basic message to more than one node, sometimes simultaneously as would occur in situations such as group meetings, or email copied to multiple recipients. In such a model, individuals may receive the same basic content from more than one source. The centrality most appropriate for these assumptions is “flow betweenness” (Borgatti, 2005). It indexes the degree to which each node is present on all possible paths among the nodes in the network, weighted for link strength values. Thus flow betweenness centrality was used in the current study.

Alternatively, eigenvector centrality (Bonacich, 1972) is consistent with a social influence model similar to that described as guiding this investigation. Nevertheless, with the relatively small numbers of nodes that are in a cabinet network the eigenvector solutions can produce anomalous results as indicated by features including negative eigenvectors, which occur when the triangular inequality principle of node relations in Euclidean spaces is violated. In such distance models if A is linked with B at a particular distance based on vertex strength, and A is linked with C at a particular distance, then the B and C link is determined, yet does not empirically correspond to the formulation. Flow betweenness is the best measure of centrality for the current research.

## 2.4 Time Segmentation

TimeSlice is a utility in the WORDij 3.0 software package that allows one to segment the larger corpus into sections of any width, such as by number of days, weeks, quarters, or years. The time sliced files are in turn input into the WordLink program in WORDij that generates basic information for each time slice in eight output files for words, for word pairs and for statistics such as entropy and mutual information, and in various formats, such as the .net Pajek format.

## 2.5 Creating the String Replacement and Include Lists

The first step in preparing the list of names for the network analysis in WORDij 3.0 is to create a string replacement list, an advanced option. This converts aliases for each name into a unigram. Table 1 shows an example.

**Table 1. Examples of String Replacement (Partial File for Nixon Cabinet)**

```
Richard Nixon->richard_nixon
richard nixon->richard_nixon
nixon-richard_nixon
President->richard_nixon
president->richard_nixon
Vice President->spiro_agnew
vice president->spiro_agnew
Spiro Agnew->spiro_agnew
spiro agnew->spiro_agnew
agnew->spiro_agnew
Gerald Ford->gerald_ford
gerald ford->gerald_ford
ford->gerald_ford
William Rogers->william_rogers
william rogers->william_rogers
rogers->william_rogers
Henry Kissinger->henry_kissinger
henry kissinger->henry_kissinger
kissenger->kissinger
David Kennedy->david_kennedy
david kennedy->david_kennedy
kennedy->david_kennedy
John Connally->john_connally
john connally->john_connally
connallh->john_connally
George Shultz->george_shultz
george shultz->george_shultz
shultz->george_shultz
```

The new strings created from the string replacement file are input as an include file in the WordLink identification of co-occurring words, in this case name unigrams. The include list is the opposite of a stop or drop list. Rather than removing certain words from the network analysis, the include list contains all of the words to be network analyzed and co-occurrences are indexed only for these terms, in this case name unigrams.

**Table 2. Example Include List for Nixon Cabinet**

```
caspar_weinberger
```

claude\_brinegar  
clifford\_hardin  
clifford\_hardin  
david\_kennedy  
earl\_butz  
elliott\_richardson  
frederick\_dent  
george\_romney  
george\_shultz  
gerald\_ford  
henry\_kissinger  
james\_hodgson  
james\_schlesinger  
john\_connally  
john\_mitchell  
maurice\_stans  
melvin\_laird  
peter\_brennan  
peter\_peterson  
rich\_kleindienst  
richard\_nixon  
robert\_finch  
rogers\_morton  
spiro\_agnew  
walter\_hickel  
william\_rogers  
william\_saxbe  
william\_simon  
winton\_blount

## 2.6 Post-Processing of Link Data for Centrality Measures

The WORDij 3.0 program has the option of producing a network file in the .net Pajek format. This is one of the import file types that UCINET accepts and converts to its system files. We chose UCINET because it is widely accepted in the social network analysis community and we wished to use common, validated centrality indices to profile the structures of the cabinets. Given the status of UCINET and the ease of output importing we felt no need to incorporate centrality measures into WORDij. This Pajek format is one of the import file types in UCINET, which is where we compute the centrality statistics for each time slice.

In the current study, within each time slice we compute the flow betweenness centrality statistics for each of the cabinet members appearing in the social network in the time period and also compute the average of such centralities across

all administration members. To index the extent to which the president stood apart from the cabinet in centrality we divided the president's centrality by the average centrality of cabinet members so that higher values indicate the president more likely stands above the other cabinet members and serves as a media "lightning rod" to which negative stories connect to the identity of the president. Lower values indicate it is more likely that the president is "below the radar" of negative public opinion formation as there is less likelihood of audiences linking the negative stories to the president per se.

Presidential job approval ratings were obtained from the Roper Center archives. For each presidency we computed the average time interval between such measurements and used this as the time slice criterion to produce time-series data.

## **2.7 Time-Series Statistical Analysis**

For Reagan the interval of job approval rating was 22 days, resulting in 132 time intervals; for G.H.W. Bush it was 13 days and 121 intervals; for Clinton it was 30 days and 98 time intervals, and for G.W. Bush it was 22 days and 133 time intervals. To test the hypotheses we created for each administration a data file in SPSS where we entered columns of data for president centrality, average administration (cabinet) centrality, and job approval from the polls. We used the statistical function of differencing adjacent time-series to remove serial autocorrelation ( $d=1$ ) and computed cross-correlations to examine seven lags before and after each centrality variable period to see what the relationship between centrality and job approval might be.

## **2.8 Combining Visualization with Statistical Centrality of Actors**

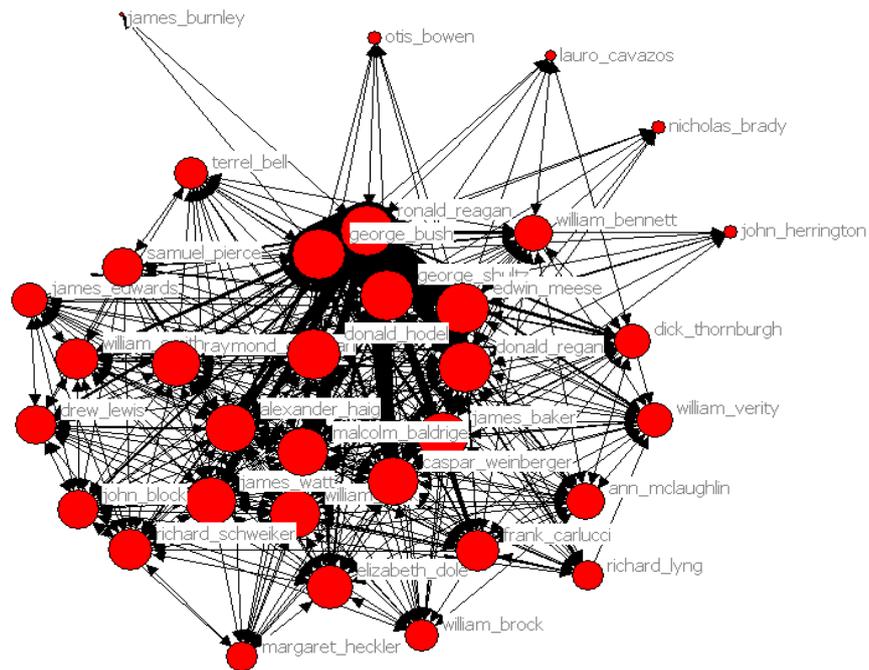
A fundamental tenet of data analysis is to first visualize it. WORDij 3.0 has VISij for creating static or time-series movies of changes in network composition and structure, although NetDraw has more options for rendering static networks such as having larger circles for more central nodes. We used node centrality to visually render the nodes' network size. For link strength we used the maximum available range of thickness of links, from 0 to 12. Our larger array of strengths was converted to this scale.

## **2.9 Sentiment Analysis**

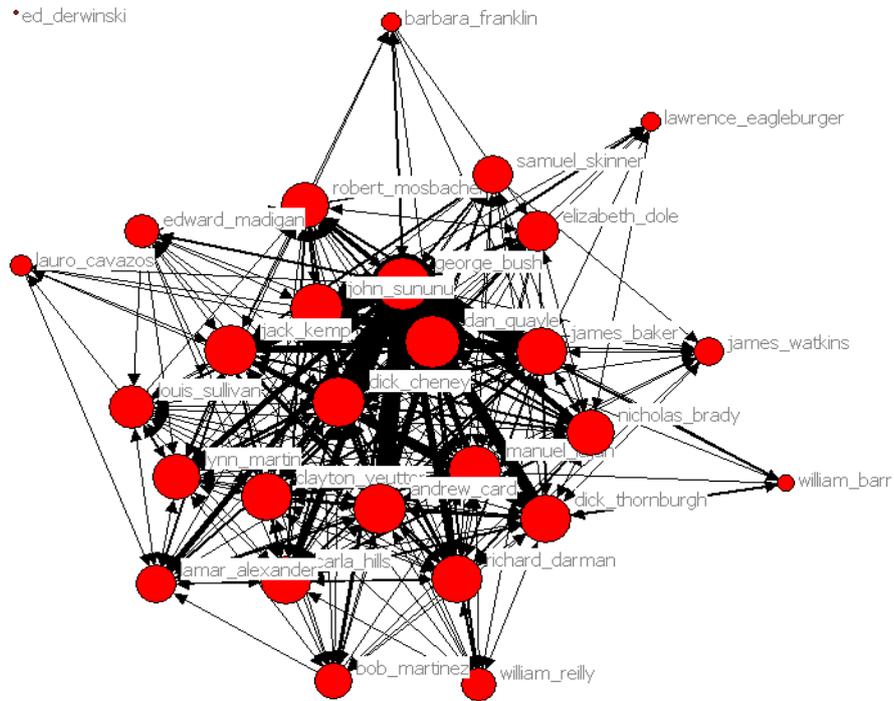
The program LIWC2007 was used to compute sentiment statistics. Each time slice for each presidency was analyzed for the relative frequency of positive and negative words. These values were added to the over time data on presidential centrality and job approval using SPSS.

### 3. Results

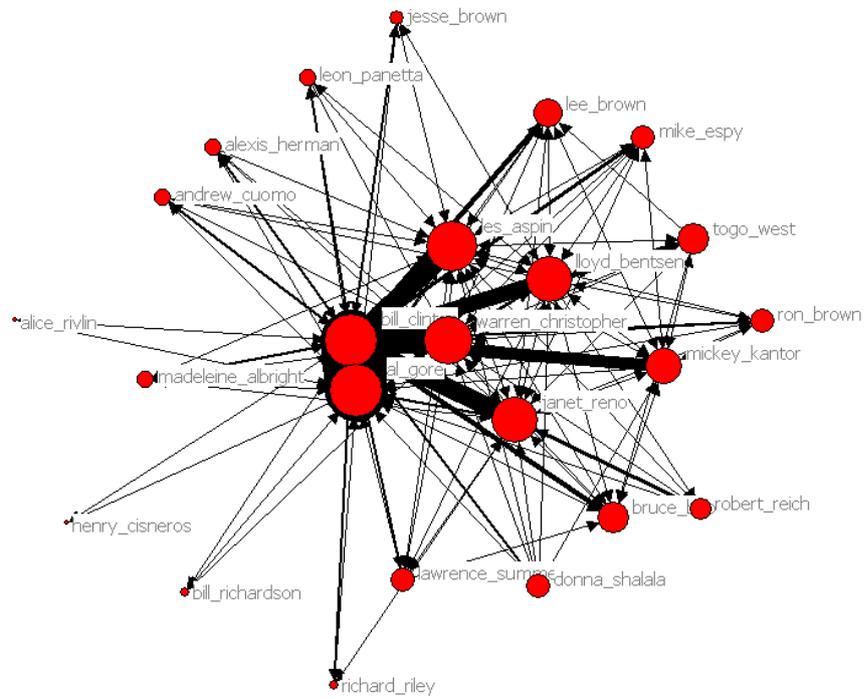
To give the reader a sense of the differences between the four cabinets in overall network structure, Figures 1 through 4 show the aggregate cabinet social networks for the cabinets of Reagan, G.H.W Bush, Clinton and G.W. Bush administrations.



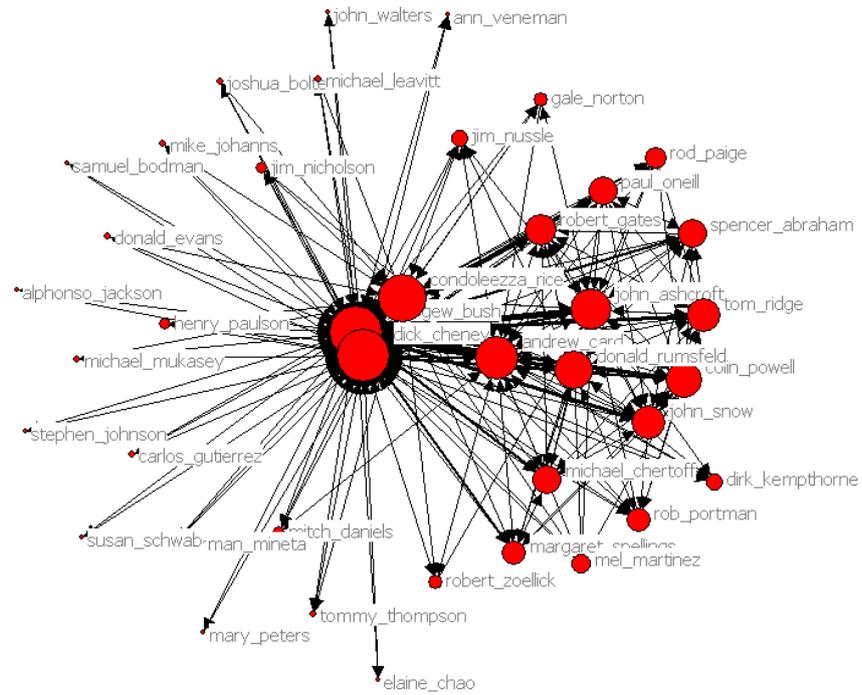
**Figure 1. Aggregate Reagan Cabinet**



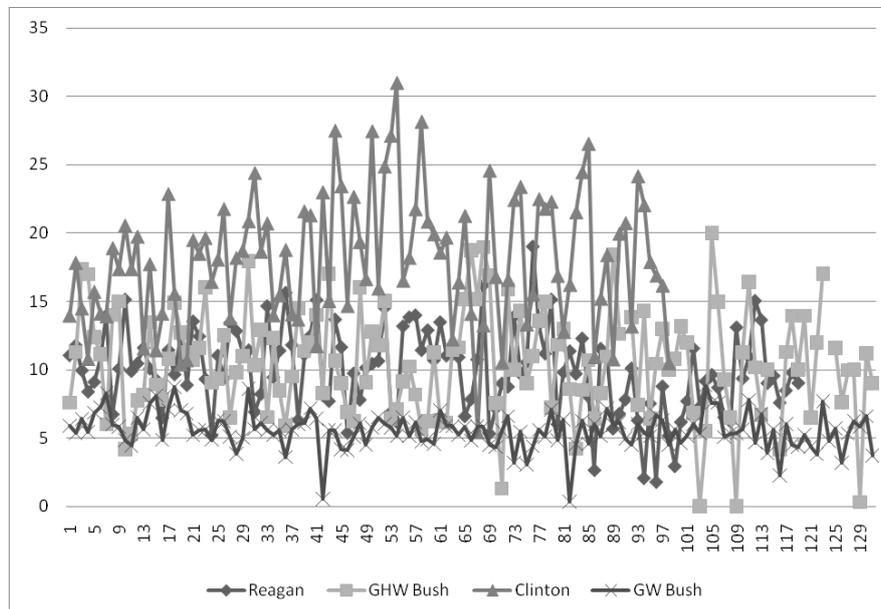
**Figure 2. Aggregate G.H.W. Bush Cabinet**



**Figure 3. Aggregate Clinton Cabinet Network**



**Figure 4. Aggregate G.W. Bush Cabinet Network**



**Figure 5. Ratio of President Centrality to Cabinet Average**

The centrality of the president in the cabinet network was divided by the average centrality. This ratio represents the extent to which the president stands out in centrality compared to the others. Figure 5 shows this ratio for each president. It was noteworthy that the G.H.W. Bush administration was unique in having a high proportion of zero centrality time periods. Because centrality can be computed only on a connected network, this indicates that in these periods isolated pairs of cabinet members or isolated individuals were treated in the news stories within the observation window. This finding was confirmed by systematic visual examination of the senior Bush administration's zero centrality periods. Because the White House itself is the primary source of cabinet news it would appear that this recurring deviation from the norm may have been strategic, in that presidential-level political communication is unlikely left to chance. Across the series, G.H.W. Bush is most similar to his cabinet members in centrality. Clinton stands out as generally being the most central compared to his cabinet.

### 3.1 Hypotheses Tests for Centrality and Job Approval

The cross-correlations with differencing to remove serial autocorrelation based on +7 through -7 lags found that a lag of 1 produced the highest coefficient at  $-.13$  ( $p < .05$ ) for the ratio of president to cabinet centrality in relation to job approval. The more the president stood apart in terms of centrality the lower the job approval rating three time periods later. The same lag pattern was found for G.H.W. Bush, a coefficient of  $-.34$  ( $p < .05$ ) at lag 1. Similarly, Clinton data showed a coefficient of  $-.16$  ( $p < .05$ ), but at lag 4. G.W. Bush centrality ratio had a coefficient of  $-.14$  ( $p < .05$ ) at a lag 2. These results support hypothesis one. Hypothe-

sis two, that the G.W. Bush administration would have a shorter lag, cannot be statistically tested but the results are not consistent with the hypothesis. Although the lag for G.W. Bush is one half that for Clinton, the lags for Reagan and G.H.W. Bush are half that of G.W. Bush.

### **3.2 Hypotheses Tests with Sentiment, Network Centrality, and Job Approval**

With sentiment variables added to network centrality of the president as predictors of approval, the dependent variable was approval differenced by 1 to remove serial autocorrelation. When the undifferenced and unlagged sentiment and network variables were added to the equation, the Durbin Watson coefficients were very close to 2.0 indicating that error terms were not correlated and hence no further serial autocorrelation remained that required removal. Ordinary least squares regression was run. Checks were made whether lagging the network variables improved the regression results but this did not have a significant effect once sentiment variables were added to the equation. Durbin Watson coefficients indicated nearly dead on values showing no need to remove further serial autocorrelation or lag variables. The Reagan administration showed the least influence of media sentiment and network structure on job approval, with none of the variables entering the equation as significant. The Clinton administration showed the highest effects of media sentiment, but other positive and negative sentiment at the same time was associated with greater job approval as was the president being more central in the cabinet network. The G.H.W. Bush administration was the only one to have the hypothesized relationship among network centrality, negative sentiment and reduced job approval. The G.W. Bush administration had the opposite relationship. When president centrality has higher, and coverage was more negative, job approval was higher.

## **4. Discussion**

The findings support the hypothesis that as the president's centrality is closer to the average cabinet centrality there is a positive association with job approval only when news sentiment is not considered. For two presidents, Reagan and G.H.W. Bush, when the president's centrality drops closer to the average, by the next polling period job approval is higher. The same pattern occurs for the other two presidents but their lags are longer. For Clinton the lag is 4 periods, and for G.W. Bush the lag is 2 time periods.

It was theorized that when the president stands above the rest of the cabinet in network centrality, negative press information is more likely to be associated with the president in the minds of members of the public and they will report lower presidential job approval ratings. The president identity is like a lightning rod absorbing the negative press sentiment. On the other hand, when the president's centrality is closer to that of the other cabinet members, he is less likely to be as-

sociated with negative press information, “flying below the radar” of public opinion processes. It is as if the negative press information fragments and diffuses into smaller bolts attached to other cabinet members. As the president is not being as directly connected with negative press content, job approval ratings increase. Once sentiment was added to the equations the most appropriate multiple regression models showed that only the G.H.W. Bush administration showed support for the hypothesis. The fact that the Reagan administration job approval was uncorrelated with either sentiment or network centrality suggests that other factors may be associated with his job approval, perhaps the direct effects of the “Great Communicator” messages are operative. Clinton showed the most variability in network centrality. When he was more central his news sentiment was also both more positive and more negative at the same time and his job approval ratings were highest. Perhaps his reputation as the “Comeback Kid” was associated with such an effect. As former director of the CIA, G.H.W. Bush could be expected to follow the most proven methods for management of media effects. Overall, he exhibited the least overall centrality among his cabinet and when he did stand above the cabinet his approval declined and he at the same time had more negative press coverage, nearly twice as negative as positive. His administration was the only one to support the hypothesis. Later G.W. Bush did show a significant effect of network centrality but in the opposite direction. When he was more central he also had more negative coverage as did his father, but the effects on job approval were opposite. They were higher. Perhaps the strong and steady criticism from the left rallied his more loyal segment of the population to attempt to counter this by rating him with higher job approval.

Taken together, the results show that making generalizations about presidencies, network structures, and media sentiment is not supported by the evidence. Taking sentiment alone, there were no consistent patterns across presidencies with job approval. Moreover, the network structures of the administrations were likewise highly variant across the four administrations. Pundits might be cautioned not to assume as they often do that they can make interpretations of a particular presidency’s activities, media coverage, and effects by reference to past presidencies.

The hypothesis about the shortening of the lag cycle associated with the internet’s shortening of the news cycle, was not clearly supported by the evidence. While G.W. Bush’s lag was one half that of Clinton’s, both were longer than for the two prior presidents.

Nevertheless, studying only four administrations introduces problems of external validity. One would not want to generalize the findings of this study to the population of presidencies, certainly those that occurred prior to the establishment of the *New York Times* and *Washington Post* as elite political newspapers. Even if this were feasible, it would not be desirable given an assumption widely held by political observers that the introduction of television changed presidential politics, and now the observation that the internet has changed presidential politics in yet other ways.

It is interesting to note, however, that for the two presidencies that took place before the internet, the time lag is the same and is only one period long, while for the two presidencies during and after internet development the lag is

longer. While this may be mere coincidence, it suggests a possible hypothesis for future research that elite newspaper coverage, in the *New York Times* and *Washington Post*, takes longer to have an effect as alternative online news sources have proliferated. It would be useful to conduct a future study that would comparatively test the basic hypothesis examined here by mapping networks separately in both the elite press and in internet content.

We sought to present an illustration of time-series analysis of social networks obtained from data mining, yet used political communication theory to generate some hypotheses to add further meaningfulness to the analysis. The study is fruitful on three counts. One is that it reveals substantive variation that future research can address; second, it demonstrates that the WORDij 3.0-based methods produces useful variation; and, third, it suggests that attempts to empirically support generalizations about presidential politics and media may not be a productive pursuit.

When data mining for social networks, the representations one obtains are based on the nature of the source data and on the assumptions made for the operationalization of data extraction and the network representation. These are the networks as portrayed in the medium from which data is mined. Questions about the extent to which such networks compare to the “real” or “actual” network are misplaced, for there is no real or actual network that is independent of data collection and extraction procedures or the subjective impressions of the participants or observers of them. The latter may be the basis for an “inter-subjective” network, but this is as close as one could come to the actual network independent from the instrumentation extracting network data.

Rather than the reality of the networks, the key is the validity of the networks. Face validity is the first level of validity and least measurable, but when linking network representations to other data, such as to job approval and news sentiment in this study, predictive validity becomes relevant. Although external validity is limited because of studying only four administrations, the statistical results provide predictive validity for the investigation.

Some philosophers (Baudrillard, 1994) argue that the media are a simulacrum with a completely self-contained representation of information that bears no correspondence to that outside of it. This is an extreme position that is invalidated by the fact that one can link media representations of such things as social networks among actors to data from outside the media system, in this study to Gallup job approval poll data. Such cross-system predictive validity is important to data mining for social networks as it is in any sort of social research. When extracting a social network from data mining one should move quickly to establish whether and how much predictive validity such constructions have, otherwise mining for “toy” networks is merely for analytical playmates. For example, some research has linked networks of message content from president’s letters to stockholders to stock price performance (Danowski, 1988; Swales, & Yoon, 1992; Zhai, Hsu, A., & Halgamuge, 2007). In the present study, we illustrate not only a procedure for mining of social network data but we link these data over time to independently obtained side data of Gallup presidential approval polls through which the results of mining can be validated.

Work is underway to more fully automate social network mining to move beyond the use of *a priori* lists of actors to ontological categories of actors for which software can automatically add new actors.

### Acknowledgement

The authors are grateful for Rafal Radulski and his timely, efficient, and effective programming assistance and for his excellent communication skills. The authors are thankful for the insightful comments of two anonymous reviewers. This research is supported by National Science Foundation Grant 0527487, Human and Social Dynamics Program.

## 5. References

- Adamic, L., & Glance, N. (2005) The Political Blogosphere and the 2004 U.S. Election: Divided they blog, In LinkKDD '05: *Proceedings of the 3rd international workshop on Link discovery*, 36-43.
- Batagelj, V. (1998). Pajek: Program for large network analysis. *Connections*, 21(2):47.
- Baudrillard, J. (1994). *Simulacra*. Translated by S.F. Glaser. Ann Arbor: University of Michigan Press.
- Bonacich, P., 1972. Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, 2, 113–120.
- Borgatti, S.P. 2002. NetDraw: Graph visualization software. Harvard: Analytic Technologies.
- Borgatti, S.P. (2005). Centrality and network flow. *Social Networks*, 27, 55–71
- Borgatti, S.P., Everett, M.G. and Freeman, L.C. (2002). UCINET for Windows: Software for social network analysis. Harvard, MA: Analytic Technologies.
- Cepela, N., & Danowski, J.A. (2009). Automatic mapping of political networks of actors from large collections of news stories. *Proceedings of the 1<sup>st</sup> ASONAM conference*.
- Cohen, J.E. (2000). The polls: The components of presidential favorability. *Presidential Studies Quarterly*, 30(1), 169-177.
- Dahl, R. (1961). A critique of the ruling-elite model. *American Political Science Review*, 52, 463-469.
- Danowski, J. A. (1982). A network-based content analysis methodology for computer-mediated communication: An illustration with a computer bulletin board, in M. Burgoon (Ed.), *Communication Yearbook 5* (pp. 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 and G. Barnett (Eds.) *Handbook of organizational communication* (pp. 335-384). Norwood, NJ: Ablex.
- Danowski, J. A. (1993a). Network analysis of message content. G. Barnett & W. Richards (Eds.). *Progress in communication sciences XII* (pp. 197-222). Norwood, NJ: Ablex.
- Danowski, J. A. (1993b). WORDij: A word pair approach to information retrieval. *Proceedings of the DARPA/NIST TREC Conference* (pp. 131-136). Washington, DC: National Institute of Standards and Technology.
- Danowski, J.A. (2009a). WORDij 3.0 [computer program]. Chicago: University of Illinois at Chicago. <http://WORDij.net>.
- Danowski, J.A. (2009b). Inferences from word networks in messages. In Krippendorff, K. & Bock, M.A (eds.) (2009). *The content analysis reader* (pp. 421-429). Sage Publications.
- Dezsö, Z., Almaas, E., Lukács, A., Rácz, B., Szakadát, I., & Barabási, A.L. (2006) Dynamics of information access on the web. *Physical Review E*, 73, 066132-1-066132-6.
- Diesner, J., & Carley, K. (2004). *AutoMap 1.2: Extract, Analyze, Represent, and Compare Mental Models from Texts*. reports-archive.adm.cs.cmu.edu
- Entman, R.M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43(4), 51-58.
- Entman, R.M. (2007). Framing bias: Media in the distribution of power. *Journal of Communication*, 57, 163–173.
- Farrell, H., & Drezner, D.W. (2008). The power and politics of blogs. *Public Choice* 134: 15–30.
- Fenko, R.F. (1959). *The President's cabinet: An analysis in the period from Wilson to Eisenhower*. Harvard University Press.
- Freeman, L.C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, 40 (1): 35-41.
- Galaskiewicz, J. (1979). The structure of community organizational networks. *Social Forces*, 57(4), 1346-1364.
- Gronke, P., & Newman, B. (2000). FDR to Clinton, Mueller to ? : A field essay on presidential approval. *Political Research Quarterly*, 56(4), 501-512.
- Hanneman, R. A. & Riddle, M. (2005). *Introduction to social network methods*. Riverside, CA: University of California, Riverside (<http://faculty.ucr.edu/~hanneman/>)
- Hunter, F. (1953) *Community power structure*. Chapel Hill: University of North Carolina Press.
- Jones, S. (2008). Television news: Geographic and source biases, 1982-2004. *International Journal of communication*. 2, 223-250.
- Katz, E., & Lazarsfeld, P. (1955). *Personal influence*. New York: Free Press.
- Knoke, D. (1994). *Political networks: The structural perspective*. Cambridge University Press. Lazarsfeld, B., Berelson, B., Gaudet, H. (1948). *The people's choice*. New York: Columbia University Press.

- McCombs, M.E., & Shaw, D.L. (1972). The agenda-setting function of mass media, *The Public Opinion Quarterly*, 36(2), 176-187.
- Nicholson, S.P., Segura, G.M., & Woods, N.D. (2002). Presidential approval and the mixed blessing of divided government. *The Journal of Politics*, 64(3), 701-720
- Polsby, N. (1963). *Community power and political theory*. New Haven: Yale University Press.
- Roper Center Public Opinion Archives, <http://www.ropercenter.uconn.edu/>
- Rousseau, R., & Zhang, L. (2008). Betweenness centrality and Q-measures in directed valued networks. *Scientometrics*, 75(3), 575-590.
- Swales, Jr., G.S., & Yoon, Y. (1992). Applying artificial neural networks to investment analysis. *Financial Analysts Journal*, 48(5), 78-80.
- Tichy, N. M., Tushman, M., Fombrun, C. (1979). Social network analysis for organizations. *The Academy of Management Review*, 4(4), 507-519.
- Zhai, Y., Hsu, A., & Halgamuge, K. (2007). Combining News and Technical Indicators in Daily Stock Price Trends Prediction, In Liu et al. (Eds.): ISNN 2007, Part III, LNCS 4493, pp. 1087-1096. Springer-Verlag Berlin Heidelberg.