Searching the Net for Differences of Opinion*

Warren Sack (wsack@ucsc.edu),
John Kelly (kjw1@columbia.edu),
Michael Dale (dale@ucsc.edu)

0.0 **Introduction**

Political theorists, at least since John Stuart Mill in his book *On Liberty* (1859), have repeatedly asserted that exposure to conflicting viewpoints is beneficial for democracy. Through exposure to political viewpoints contrary to their own, citizens are said to gain political tolerance and an understanding of opposing rationales. Recent empirical work (e.g., Fishkin, 1991; Mutz, 2002) has confirmed these assertions. However, there is no clear means by which a citizen can find opposing opinions. Factors such as the consolidation of media ownership (e.g., Bagdikian, 2004), neighborhood segregation (by, for example, race and class), lack of weak ties in personal and cross-community-oriented social networks (Putnam, 1999; Granovetter, 1973), proliferation of ideological-exclusive weblogs and radio and television talk shows, and recent technological developments that allow the "filtering" of Internet-distributed news (e.g., Sunstein, 2002) all make it difficult for individual citizens to find significantly different opinions. Contrary to Negroponte (1996), we posit the development of a software technology to facilitate the construction of a "Daily Not Me," i.e., a sort of "search engine" that,

---

* This paper will be presented at the *Second Conference on Online Deliberation: Design, Research, and Practice / DIAC-2005* (Stanford, CA: Stanford University, 20-22 May 2005).

when given a topic (e.g., “abortion”), will return a range of diverse opinions about the topic (e.g., “pro-choice” and “pro-life”).

In this paper we present some preliminary results towards this long-term goal. Our work bootstraps recent, prior work in which one of the co-authors (Kelly, 2004) used qualitative content analysis to characterize the political leanings of one hundred and twenty, prolific, Usenet newsgroup authors. Software was developed to automatically download, from a Usenet newsgroup archive, tens of thousands of discussion threads containing over one million individual messages. Within these threads of discussion we were able to find several thousand “mixed exchanges” in which known discussants (i.e., two or more discussants identified by Kelly) of differing political opinion exchanged messages. We have performed an empirical analysis of the structural characteristics (e.g., size, branching factor) of the discussion threads surrounding these mixed exchanges. Our goal is to identify a set of computable, search heuristics that might be employed in a “Daily Not Me” technology for finding opposing, political viewpoints as expressed in the archives of online discussion groups.

We understand this work to be complementary to the design and innovation work of Fishkin (Fishkin, 1991) and others to create new environments and situations where deliberative discussion can take place. We hypothesize that in vast, online discussion spaces – like the space of Usenet newsgroups – there must exist places, or at least moments, when deliberative discussion already takes place “in the wild.” We envision a search engine that, when given a topic, will find likely threads of discussion where opposing opinions are being or have been expressed. In this paper we report on our initial efforts to implement the first pre-processing step of such a search engine. We need to identify one or more quick and relatively-accurate heuristics that can be used to comb through a large database of newsgroup or weblog postings to identify likely places of political exchange. The output of the mechanisms we describe here will be the input to further pre-processing steps of the search engine that perform a detailed analysis of the contents of the messages. In short, the heuristics described here are “triage” techniques intended to narrow down which message threads should be given more detailed analyses.

First, we give an overview of the newsgroup messages we have examined and shortly describe the results of a previous study by one of the co-authors (Kelly, 2004) upon which we rely for the present work. Second, we describe a set of independent variables associated with the discussion threads. Our dependent variable concerns whether or not liberals and conservatives exchanged messages in a discussion thread. We seek a model in which some combination of easily-measured, independent variables can be used to predict the likelihood that a thread contains an exchange of views between at least one liberal, discussion participant and one conservative, discussant. Third, we present such a thread classification model as a simple discriminant function. We further simplify this model by eliminating some of the independent variables that are closely correlated with others. Fourth, we attempt to verify our model by testing it against one hundred, hand-tagged discussion threads. Finally, we present our conclusions and shortly discuss future work.

1.0 **Messages, Discussion Threads and Newsgroups**

The present work builds on previous work done by one of the co-authors (Kelly, 2004). Kelly read several thousand posts made to six Usenet newsgroups: (1) alt.fan.noam-chomsky, (2) alt.politics.bush, (3) alt.politics.democrats.d, (4) alt.politics.economics, (5) talk.abortion, (6) talk.politics.mideast. All of these newsgroups are public, online discussions archived on thousands of newsgroup (NNTP) servers throughout the Internet. One of the most extensive archives of Usenet newsgroup postings can be found at the Google Groups website (http://groups.google.com). Summary statistics about newsgroups and the newsgroup participant's posting activity can be found at Microsoft Research's Netscan site. It is a so-called "Usenet Social Accounting Search Engine." Using statistics from Netscan, Kelly was able to identify about twenty high-frequency posters from each of the six newsgroups. Extensive study of the messages posted by these 119 frequent participants led to the articulation of a set of political categories and the identification of the political category associated with 97 of the 119 posters. The political point of view of twenty-two of the posters was uncharacterizable. Most of these undefined positions were occupied by posters (13 of the 22) participating in the talk.politics.mideast newsgroup; the rest were distributed more-or-less equally across the other groups.

We will not review Kelly's results in this paper but rather only explain how we have incorporated a simplified version of some of his results into the present study. While Kelly identified twelve different political positions occupied by newsgroup participants (including conservative/Republican, liberal/Democrat, independent/uncommitted, libertarian/right, anarchist/left, Christian Right, and others) we have (perhaps too insensitively) coerced these twelve positions into just three categories: left and right and unrecognized. Thus, each of the 119 participants studied by Kelly has been labeled as unrecognizable or political left or right.

Recall that a discussion thread is constituted from an initial message, all of the replies to the initial message, all of the replies to these replies, etc. Our aim has been to study the structure of those discussion threads in which at least one known person of the left exchanged a message with at least one known person of the right. We call an exchange of messages between posters of opposing political positions a "mixed exchange." We are interested in threads that contain one or more mixed exchanges because they are potentially deliberative exchanges. We hope to be able to formulate heuristics to automatically detect threads that are likely to contain a mixed exchange.

Because Kelly made known to us the political position of one hundred and nineteen frequent Usenet newsgroup posters, it was a straightforward task to download another set of threads – from the same time period -- that contain messages posted by one or more of the these known participants. We downloaded, from an NNTP (Usenet newsgroup) server over one million messages, but chose to focus on a subset of about sixteen hundred (1664 out of 25,590) discussion threads in which at least two messages were posted to the thread and in which at least two-thirds (66%) of the messages posted in the threads were posted by one or more of our known participants. A total of 13,156 messages were posted to these 1664 discussion threads.

Our choice to concentrate on these threads (in which messages from known, high-frequency participants constitute at least two-thirds of the messages) is motivated by three considerations. First, Fiore, Teirnan and Smith (2001) have shown that newsgroup readers are more likely to value the messages contributed by frequent, long-term participants (i.e., in comparison with messages contributed by occasional or one-time participants). Second, although it is possible that an occasional or one-time participant might contribute a brilliant and insightful point of view to the discussion, it is more likely that the frequent contributors will be the participants who contribute, within the confines of the newsgroup, a detailed and well-developed argument. Third, ideally, we would analyze only threads in which we knew the political position of all of the participants, but this subset of threads does not adequately represent the diversity of thread sizes and shapes. Rather, it tends be restricted to threads of small size. To work with a sufficiently diverse group of threads we have chosen to examine threads where at least two-thirds of the messages have been contributed by participants' whose political positions are known to us (using Kelly's data).

2.0 **Variables: On the Structure of Discussion Threads**

From a graph theoretic perspective, discussion threads are trees because each thread has one, and only one initial message, each message can have multiple replies, but none of messages are replies to multiple messages and a message posted earlier in the discussion thread cannot be a reply to a message posted later. These formal properties are due to the standard definition of the email message header format (RFC822; see its definition at, for example, http://www.rfc-editor.org/) and are maintained by all common email client programs. This is most easily understood by remembering that when one hits the "reply" button in an email program, one is replying to one and only one other message.

Because discussion threads are formally trees, we can define a set of variables that characterize the size and shape of the discussion threads. We are especially interested in the following variables:

M: the number of messages posted to a thread;

L: the number of leaves in a thread tree; leaves are messages that received no replies;

I: the number of interior nodes in the thread tree; ***I = M – L***;

I/L: the number of interior nodes divided by the number of leaves;

P: the number of people who posted a message to a thread;

maxMp/M: the maximum number of messages posted by one person to a thread divided by the number of messages posted to a thread;

maxD: the maximum depth from the root of the tree (i.e., the initial post) to one of the leaves of the thread tree;

meanD: the mean depth from the root of the tree to the leaves of the tree;

maxB: the maximum branching factor in the tree; we understand the thread tree to branch when a message receives one or more replies; thus the message in the thread with the greatest number of replies has the maximum branching factor;

meanB: the mean branching factor in the tree;

meanMp: the mean number of messages posted by a person participating in the thread;

meanT: the mean amount of time (in seconds) between messages posted to the thread;

In addition we assigned a ***score*** to each thread where a score of 1 indicates a mixed exchanged (as defined above): a person of the left replied to a message from a person of the right or vice versa. A score of 0 indicates that no such exchange happened in the thread. Scores of greater than 1 occurred when more than one mixed exchange occurred. We calculated a Spearman's ***r*** correlation between each of our independent variables and the score.

variable	*M*	*L*	*I*	*I/L*	*P*	*maxMp/M*	*maxD*	*meanD*	*maxB*	*meanB*	*meanMp*	*meanT*
correlation with score	0.65	0.82	0.72	0.42	0.66	0.18	0.71	0.56	0.79	0.69	0.68	0.14

Figure 1*: Correlations between independent variables and thread score (N=1664; 2≤M)*

Figure 2 illustrates the first correlation in more detail. It shows the variable ***M***, the number of messages posted to a thread, plotted against the threads' ***score***s. A linear regression model clearly works quite well for threads with twenty-five or fewer messages (correlation coefficient = 0.72).

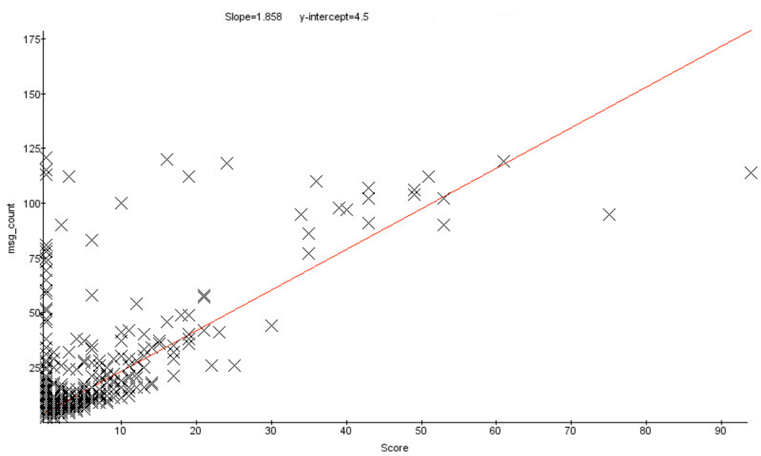

***Figure 2**: **M** (number of messages) plotted against **score** (number of mixed exchanges) for all threads containing two or more messages.*

But, the linear model does not seem to fit as well for threads of size larger than twenty-five messages. Examination of the correlations for this subset of threads (25 ≤ M) shows them to be weaker with exclusively larger threads; see figure 3.

variable	*M*	*L*	*I*	*I/L*	*P*	*maxMp/M*	*maxD*	*meanD*	*maxB*	*meanB*	*meanMp*	*meanT*
correlation with score	0.37	0.32	0.39	0.22	0.21	0.26	0.34	0.34	0.12	0.01	0.42	-0.19

***Figure 3**: Correlations between independent variables and thread score (N=98; 25≤M)*

Figure 4 can be compared to figure 2. Figure 4 shows the same plot (thread size versus score) for a subset of the threads; i.e., the threads that contain 25 or more messages. Clearly the linear regression model for this subset of threads is not a very good fit (correlation coefficient = 0.48).

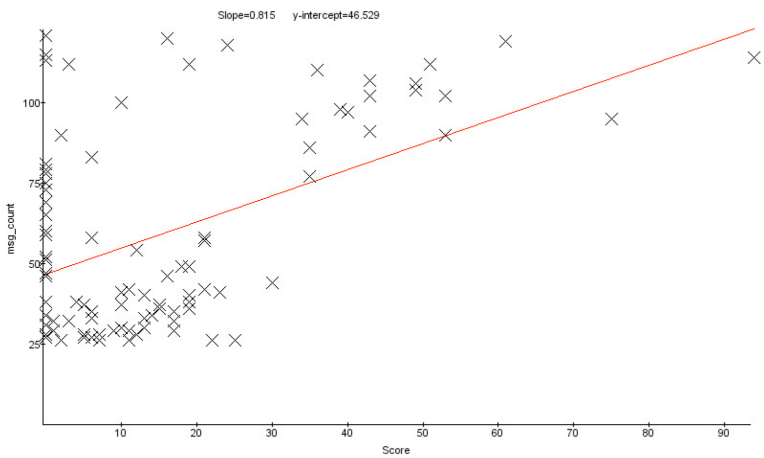

***Figure 4**: **M** (number of messages) plotted against **score** (number of mixed exchanges) for all threads containing twenty-five or more messages.*

Figures 5 and 6 plot thread size against frequency for, respectively, all 25,590 discussion threads in our sample and for the 1664 threads we used to compute the correlations in figure 2. Both figures 5 and 6 approximate a Zipf distribution with the great majority of threads containing fewer than twenty-five messages.

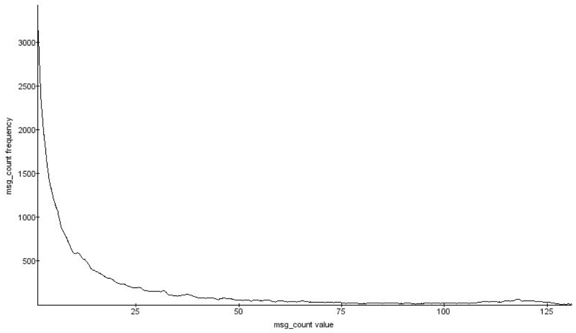

***Figure 5**: Frequency distribution for all threads: number of messages per thread*

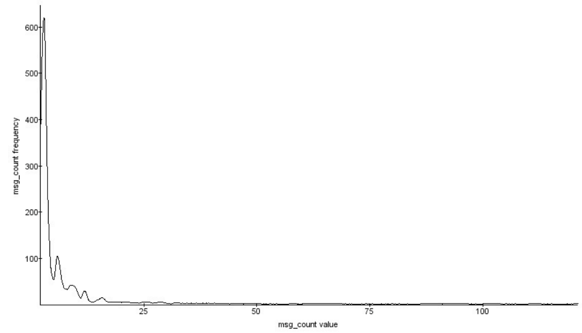

***Figure 6**: Frequency distribution for the 1664 threads that are the focus of our study: number of messages per thread*

Consequently, knowing this – that the majority of the threads contain fewer than twenty-five messages – it is tempting to conclude that a very simple model would suffice. For example, for we might just use the size of the thread to predict whether or not a thread is likely to contain a mixed exchange. After all, such a model will work pretty well for the majority of the threads we will encounter. Furthermore, a model like this is commonsensical: the more messages posted to a thread, the more likely a mixed exchange will occur.

But, there at least two reasons why we need a model that will work for large threads as well as for relatively small threads. Large threads are of interest because they are more likely than small threads to contain a deliberatively-elaborated point of view. While small threads containing one or more long messages might contain a detailed explanation of someone's point of view, it is only through an extended back-and-forth with an interlocutor that the strengths and weaknesses of a point of view can be unpacked and explored in detail. So, we assume that long threads are more likely to be representative of some sort of deliberative exchange than short, small threads.

Second, recall that our immediate goal is to find a set of quick and computationally-inexpensive heuristics for predicting if a thread is likely to contain a mixed exchange and thus for determining if more computational resources should be devoted to analyzing the thread in detail. A linear model (like the one shown in figure 2) would roughly predict that we should look at all of the large threads and none of the small ones. But, simply because they contain a large number of messages, large threads are computationally expensive to analyze in detail. If we can eliminate even some of the large threads, then we are likely to save many computational resources in the subsequent phases of analysis. Consequently, we desire a model that works for small and large threads, but especially for large threads.

3.0 **A Thread Classification Model: Search Heuristics**

To create a model that will work for small and large discussion threads we first simplify the problem. Rather than attempting to predict the number of mixed exchanges in a thread we will be satisfied with sorting threads into one of two categories: (1) those containing mixed exchanges; and, (2) those containing no mixed exchanges. Consequently, the problem we now face is this: Can a discriminant function be designed such that, given a thread, when it is greater than zero it is more likely that the thread contains one or more mixed exchanges; and, when it is zero or less-than zero it is more likely that the thread does not contain a mixed exchange?

This can be formalized as follows. Associated with each thread is a vector of twelve, independent variables, as detailed in section 1.0 of this paper: (*M*, *L*, *I*, *I/L*, *P*, *maxMp/M*, *maxD*, *meanD*, *maxB*, *meanB*, *meanMp*, *meanT).* We are exploring 1664 discussion threads, thus we can order the thread trees from 1..1664. For a given thread tree j, in this order, we will denote the vector of independent variables simply as $\mathbf{v}_j$, where the letter in bold indicates that the variable denotes a vector and the subscript indicates the discussion thread that the vector is associated with. Since we are examining thread trees in which 66% of the messages posted were contributed by participants with a political position known to us (via Kelly's hand analysis), also associated with each of these threads is a score. But we are restricting our interest to the distinction between those threads with scores greater than zero (score>0) versus those threads with scores of zero (score=0).

Using the associated vectors and scores for our 1664 threads, we estimate the following two sets of conditional probabilities:

$P(\mathbf{v}_j|\text{score}=0) = \prod_{k=1..12} p(v_k|\text{score}=0)$; and,

$P(\mathbf{v}_j|\text{score}>0) = \prod_{k=1..12} p(v_k|\text{score}>0)$.

Thanks to Bayes formula we can convert the estimated *prior* probabilities (i.e., in which we know the score) into *posterior* probabilities (in which we want to predict the score). So, our estimated discriminant function is this:

$g(\mathbf{v}_j) = P(\text{score}>0|\mathbf{v}_j) - P(\text{score}=0|\mathbf{v}_j)$

where if $g(\mathbf{v}_j) \geq 0$ then the score is more likely to be positive;

else the score more likely to be zero.

But, as is, this estimated discriminant function can not be applied to a discussion threads outside our original set of 1664 thread trees unless the unseen discussion thread has a vector associated with it that *exactly matches* the vector of some tree in our original set of trees. We, rather crudely, address this problem by dividing the values for each variable into equally populated quartiles that we call *small*, *medium*, *large* and *extra large*. For instance, the quartile divisions for M, the number of messages in the thread trees are small (M<3); medium (M=3); large (3<M<6); and, extra large (M≥6). This allows one to see, for example, that if a thread has an *extra large* number of messages, then it is more likely to have a positive score (i.e., more like to contain one or more mixed exchanges) than to have a score of zero.

Given these definitions and this simplification of values into quartiles a discriminant function can be calculated and then tested against the same 1664 thread trees to get some idea of how accurate it might be. Using all twelve variables in the function we find that it predicts the correct category (either score=0 or score>0) for 1156 out of the 1664 thread trees (accuracy of 69%). But, since we are searching for threads likely to contain a mixed exchange, the power of the model is better measured according to the usual criteria of information retrieval where *recall* denotes the completeness of the retrieval and *precision* denotes the purity of the retrieval.[1] Using Kelly's (2004) analysis we

[1] "Consider an example information request *I* (of a test reference collection) and its set *R* of relevant documents. Let |*R*| be the number of documents in this set. Assume that a given retrieval strategy (which is being evaluated) processes the information request *I* and generates a document answer set *A*. Let |*A*| be the number of documents in this set. Further, let |*Ra*| be the number of documents in the intersection of the sets *R* and *A*. … The recall and precision measures are defined as follows. **Recall** is the fraction of

know that 660 of the threads had a positive score; 1004 had a score of zero. The model miscategorized 508 threads; 122 of them were miscategorized as having a mixed exchange (when their actual score was zero); and, 386 of them were mistakenly assigned a score of zero. Consequent, the results for this model (with all twelve variables) are precision = 69% and recall = 42%.

As is always the case in information retrieval tasks, there is a tradeoff that must be made between precision and recall (Buckland and Gey, 1994). In our case, a low – but non-zero -- recall rate is fine because discussion threads are not a scarce commodity. For example, Google Groups has hundreds of millions of newsgroup messages indexed. But, we would like a precision score that is as high as possible so that fewer threads without mixed exchanges are given further scrutiny.

We can refine this estimated discriminant function by first simplifying it. Our initial discriminant function -- which includes all twelve variables -- is based on the assumption that each of the variables that describe the size and shape of the thread trees is independent from all of the other variables. This is clearly not the case. For example, the mean depth of the tree (meanD) is likely to be correlated with max depth (maxD) and the number of messages in the tree (M); and, such is the case: r(meanD,maxD) = 0.97; r(meanD,M) = 0.93. So, a refined discriminant function need not contain all twelve variables.

---

the relevant documents (the set *R*) which has been retrieved; i.e., *Recall* = |*Ra*| / |*R*|. **Precision** is the fraction of the retrieved documents (the set *A*) which is relevant; i.e., *Precision* = |*Ra*| / |*A*|." (Baeza-Yates and Ribeiro-Neto, 1999: 75).

As can be seen in Figures 1 and 3, the mean amount of time between message postings in a thread (meanT) is only loosely correlated (r = 0.14 and r = -0.19) with the score of a thread tree. Also, we note, now that archives of Usenet newsgroup messages persist for years due to archives like Google Groups (rather than, as was previously the standard, for days or weeks), one message posted to a thread months after the thread was active can significantly change this value. Consequently, this variable (meanT) will not be incorporated into the refined model.

Although it would be possible to systematically simplify and optimize the discriminant function for better precision and non-zero recall we have not done so. If our training data was complete – i.e., if we knew the political positions of 100% of the participants – then automatic optimization would make sense. However, since this is not the case, we have instead manually selected variables for inclusion and exclusion and then run the resultant functions against the data to compare results. Our hope is to find a function that has high precision and non-zero recall and that can be explained intuitively.

Our resultant, simplified, discriminant function contains three almost-independent variables: g(maxMp/M, maxB, meanMp). (The following Spearman correlations indicate the degree of (in)dependence of these variables: r(maxMp/M,maxB) = -0.14; r(maxMp/M,meanMp) = 0.35; r(maxB,meanMp) = 0.64.) This function of three parameters, g(maxMp/M, maxB, meanMp), accurately categorizes 68% of the threads with precision of 75% and recall of 29%.

When maxMp/M is extra large, mixed exchanges are unlikely. Intuitively one can understand the logic of this: when maxMp/M is large one participant has posted many more messages than the other participants in the thread; thus, the thread is dominated by one voice and more likely to be monological rather than dialogical in nature.

When maxB is small, the score for the thread is more likely to be zero. This too is relatively intuitive: threads containing at least one message that received a lot of replies are more likely to incorporate many engaged discussants than threads containing only messages with few replies.

Finally, we are interested in threads in which meanMp is relatively large as this is an indication that several people are contributing substantially to the discussion.

Here is the exact definition of the simplified, discriminant function in pseudo code:

```
g(maxMp/M, maxB, meanMp) {
    // Calculate probability that score>0
    if (maxMp/M  < 0.33) then p11 = 0.57;
    else if (maxMp/M  < 0.6) then p11 = 0.47;
    else p11 = 0.13;
    if (maxB == 1) then p12 = 0;
    else if (maxB == 2) then p12 = 0.27;
    else p12 = 0.52;
    if (meanMp < 1.5) then p13 = 0.38;
    else if (meanMp < 2) then p13 = 0.3;
    else p13 = 0.53;
    p1 = p11 * p12 * p13;
    // Calculate the probability that score==0
    p0 = (1 – p11) * (1 – p12) * (1 – p13);
    // Return positive number if score likely to be
    // greater than zero.  Return zero or negative
    // number if score likely to be zero.
    return(p1 – p0);
```

}

4.0 **Verification of the Model**

A set of one hundred discussion threads were randomly selected from the same six newsgroups. To approximate the size distribution of our original collection of 1664 threads, 25 threads with 2 messages, 25 threads with 3 messages, 25 threads with 4 or 5 messages, and 25 threads with 6 or more messages were selected. Each of the authors of this paper independently read and tagged the threads as either containing or not containing a mixed exchange. Our purpose was to verify our discriminant function on a manually tagged corpus of discussion threads.

It is noteworthy that even the three of us did not always agree on which threads did or did not contain a mixed exchange. All three of us were in agreement only 58% of the time. To test our model we used a majority vote: if two of us agreed that a mixed exchange had taken place in the thread, then the thread was marked as having a mixed exchange. This difficulty in manual tagging indicates a much deeper problem: Can even a well-educated, interested and motivated person recognize a deliberative discussion when he or she sees one? While we'd like the computer to recognize such an exchange, it's not clear what criteria people use to recognize such an exchange.

Our manual tagging revealed that 8 of 25 threads containing only two messages did incorporate a mixed exchange. Approximately this number of mixed exchanges were also found in the next two quartiles (threads with three messages; and, threads with four or five messages). But, the refined (i.e., the model with only three variables) discriminant function did not predict any of these. (Although the unrefined discriminant function – i.e., the model containing all twelve variables – did find 77% of the mixed exchanges in threads of size 3 or 4 messages. The downside to this was that this unrefined model had only 45% precision. And, the unrefined model predicts – incorrectly – that every thread containing 6 or more messages will contained a mixed exchange.)

However, in the fourth quartile – wherein discussion threads contain six or more messages – the refined discriminant function performed with 71% recall and 94% precision. While this certainly is not ideal – we would like a model that works for all thread sizes – it is practical since it is most crucial that the model can work on large thread sizes with a fair amount of precision. Moreover, the use of different discriminant functions for different thread sizes would be easy enough to implement.

5.0 **Conclusions, Discussion and Future Work**

The approach demonstrated in this work, to attempt to automatically identify mixed, possibly deliberative, exchange in discussion threads by examining the thread trees' structures – their topologies and morphologies – might strike some as quixotic. Or, perhaps, at least as quixotic as the enterprise of Chomskyan linguistics in its attempts to tell us something about language and the human mind by closely reading syntax trees. Nevertheless, even outside of Chomskyan linguistics there is a long history of employing structural characteristics in order to define and distinguish cultural or literary genres or discourses. One might begin such a history with Russian Formalist Vladimir Propp's *Morphology of the Folktale*, published in 1928, and extend it right up to the present work in structural analysis, i.e., social network analysis as practiced in sociology and elsewhere (see, for example, the journal of *Poetics*, Volume 27, Issues 2-3, Pages 57-231 (March 2000), "Relational analysis and institutional meanings: Formal models for the study of culture," edited by John Mohr).

For the purposes of this project we are tactical – not committed – structuralists. When writing computer programs to ferret through large amounts of data, it is tactically effective to devise a set of quick methods to narrow one's search down to a smaller corpus that potentially contains items worthy of closer examination. Such computational methods depend upon structural attributes of the data. In this case our work essentially boils down to this: if one wants to find a mixed – potentially deliberative -- exchange in a large set of Usenet newsgroup threads, look for those threads in which (a) no one person dominates the discussion; (b) everyone participating in the thread has posted at least a couple of messages; and, (c) there is at least one message with multiple replies. This paper details our search for this heuristic and presents the heuristic in a more precise form; i.e., as what one might call, in the discipline of pattern classification, a "discriminant function" (Duda et al., 2001).

If we understand the problem of identifying deliberative discussion threads as a problem of content analysis (Krippendorf, 2004), then it is perhaps easier to see why we, the co-authors, did not always agree in our independent judgments to tag some threads as containing a mixed exchange and others as containing no exchange of political opinion. (All three of us agreed only 58% of the time.) Some researchers have proposed necessary criteria for categorizing an exchange as deliberative (Coleman and Goetze, 2001). But, these criteria are not well-defined enough to make deliberative discussion a genre that can be easily, i.e., procedurally, tagged using methods of content analysis.

However, a co-author of this paper built a system, the Conversation Map (Sack, 2000), for automatically summarizing and visualizing very large-scale conversations (VLSC), like high-volume Usenet newsgroups. In a recent paper Sack demonstrates how the deliberative criteria of Anthony Wilhelm (Wilhelm, 2000) following James Fishkin (Fishkin, 1992) can be automated using a mixture of techniques from social network analysis and computational linguistics, as they are incorporated into the Conversation Map system (Sack, 2005). We foresee extending the work reported in the present paper to incorporate various computational linguistic techniques now implemented in the Conversation Map system. This will be done by extending our discriminant function to include variables measuring aspects of the content of the messages. For example, understanding the various quotation patterns employed by the participants in the thread provides us with an interesting measure of interactivity between the participants. We have carried out empirical analyses of quotation practices with our colleagues at INRIA in France (see Barcellini, Détienne, Burkhardt and Sack, 2005).

In future work we plan to extend these simplest of models for identifying discussion threads containing mixed (political) exchanges – as reported in this paper – to include a set of linguistic and social network criteria; criteria that we have already implemented in computational form in the Conversation Map system (Sack, 2000). This, we hope, will bring us closer to achieving our long-term goal to implement a search engine that, when given a topic, will find likely threads of discussion where opposing opinions have been expressed.

Acknowledgements

This work has been supported by Award #0416353 (“An Interface and Search Engine for Deliberation”) from the National Science Foundation, Directorate for Computer and Information Science and Engineering; and by National Science Foundation Award #IIS-0441999. Any opinions, findings and conclusions or recommendations expressed in this website are those of the authors and do not necessarily reflect those of the National Science Foundation.

6.0 **References**

Ricard Baeza-Yates and Berthier Ribeiro-Neto, *Modern Information Retrieval* (New York: Addison-Wesley, 1999).

Ben H. Bagdikian, *The new media monopoly* (Boston: Beacon Press, 2004).

Flore Barcellini, Françoise Détienne, Jean-Marie Burkhardt, Warren Sack, "A study of online discussions in an Open-Source Software Community: Reconstructing thematic coherence and argumentation from quotation practices," to appear in the *Proceedings of the Communities and Technologies Conference*, Milan, Italy, June 2005.

Michael K. Buckland and Fredric Gey, “The relationship between recall and precision,” *Journal of the American Society for Information Science* 45, no. 1 (Jan. 1994): 12-19.

Stephen Coleman and John Goetze, “Bowling Together: Online Public Engagement in Policy Deliberation,” (London: The Hansard Society, 2001).

Richard O. Duda, Peter E. Hart and David G. Stork, *Pattern classification* (New York: Wiley, 2001).

Andrew T. Fiore, Scott Lee Tiernan, Marc A. Smith “Observed Behavior and Perceived Value of Authors in Usenet Newsgroups: Bridging the Gap,” in the *Proceedings of ACM SIGCHI*, Minneapolis, MN, 20-25 April 2002: 323-330.

James S. Fishkin, *Democracy and deliberation: new directions for democratic reform* (New Haven: Yale University Press, 1991).

James S. Fishkin, “Beyond Teledemocracy: ‘America on the Line’,” *The Responsive Community* 2 (3), 1992: 13-19.

Mark Granovetter, "The Strength of Weak Ties", *American Journal of Sociology* (1973) Vol. 78, pp.1360-1380.

John Kelly, “Community and Selectivity in Unstructured, Online, Political Discourse,” paper presented at *Internet Research 5.0* (Association of Internet Researchers), Sussex, September 2004.

Klaus Krippendorff, *Content analysis: an introduction to its methodology* (Thousand Oaks, CA: Sage, 2004).

John Stuart Mill, *On liberty* (New York: Bantam, 1993).

John Mohr (editor) “Relational analysis and institutional meanings: Formal models for the study of culture,” *Poetics*, Volume 27, Issues 2-3 (March 2000): 57-231.

Diana C. Mutz, “Cross-cutting Social Networks: Testing Democratic Theory in Practice,” *American Political Science Review* Vol. 96, No. 1 March 2002: 111-126.

Nicholas Negroponte, *Being digital* (New York: Knopf, 1995)

Vladimir Propp, *Morphology of the Folktale* (Austin, TX: University of Texas Press, 1968)

Robert D. Putnam, *Bowling alone: the collapse and revival of American community* (New York: Simon & Schuster, 2000).

Warren Sack, “Conversation Map: An Interface for Very Large-Scale Conversations,” *Journal of Management Information Systems*, Vol. 17 No. 3, Winter 2001: 73 – 92.

Warren Sack, “Discourse Architecture and Very Large-Scale Conversations,” in *Digital Formations*, Robert Latham and Saskia Sassen, Editors (Princeton, NJ: Princeton University Press/Social Science Research Council, 2005).

Cass R. Sunstein, *Republic.com* (Princeton, NJ: Princeton University Press, 2001).

Anthony Wilhelm, *Democracy in the Digital Age: Challenges to Political Life in Cyberspace* (New York: Routledge, 2000).