

Bayesian Analysis of Decision Making in Technical Expert Committees

David A. Broniatowski¹, Christopher L. Magee², Joseph Coughlin¹ and Maria C. Yang²

¹ Engineering Systems Division, Massachusetts Institute of Technology, USA

² Department of Mechanical Engineering & Engineering Systems Division, USA
david@mit.edu

Abstract

All sufficiently-complex engineered systems require oversight by committees of technical experts. Decision-making by these expert committees is poorly understood. A number of social dynamics might impact the sharing of information among expert specialists. Some of these could be beneficial to the decision process but some could lead to decisions that are not well-informed by all of the specialties represented. This research presents a quantitative empirical methodology for the study of technical expert committees based upon computational linguistic analysis of meeting transcripts. The Food and Drug Administration advisory panels are used as a case study. Output results include meaningful social network data that might potentially be used to gain insight into how the social dynamics of expertise interact with technical device attributes, ultimately leading to a committee decision.

Keywords – Group Decision-Making, Bayesian Analysis

1 Introduction

The aggregation of information from multiple human sources is critical to many disciplines. Large-scale engineered systems must successfully pool knowledge and expertise from many domains if they are to be effective. Furthermore, any large-scale engineered system must receive the approval of several stakeholders, many of whom have differing requirements, and hence different perceptions, of the system and its functionality. Examples include design reviews that large-scale engineered systems must pass (consider, for example, the PDR and CDR cycles within the aerospace domain). These approval activities are aimed at bringing in additional expertise, and, ideally, improving the ultimate design. In general, committee decision making can have a large impact upon a device or system of devices, affecting the system's architecture before it enters the market.

In the committees that concern us in this paper, information must be aggregated from multiple expert specialists. In such situations, we may expect concerns about bias, conflict of interest and deep uncertainty. The fact that different experts hold different perspectives & values makes it more likely that all aspects of a problem will come under consideration. Nevertheless, this also makes it difficult to generate consensus on the interpretation of data and even the meaning of keywords and requirements. Experts' interpretations of data are likely influenced by institutional frames (e.g., a particular profession, specialty, or organization of which they are a part). This motivates three main questions driving the research endeavor of this paper:

1. How can we study, in a quantitative, consistent manner, the impact of institutional backgrounds of technical experts on committee decisions?

2. How do technical experts' decisions change as they learn change and interact during the decision-making process?
3. How might we design committee processes so as to enable desirable behaviour on the part of technical expert committees?

In order to answer these questions, this work is aimed at developing a deeper understanding of how differential perceptions, e.g., due to different training, impact upon multi-stakeholder decision-making. Furthermore, the question of how to design decision-making processes that successfully leverage different perspectives is one that is extensible to a range of technology and policy activities across both public and private sectors. We differ from previous analyses in our use of an empirical quantitative methodology based upon a computational linguistic analysis of meeting transcripts. Such a methodology can be extended to similar studies in other domains of interest to engineers.

Dong [1] took the first steps in the direction of quantitative empirical analysis of group decision-making by using Latent Semantic Analysis [2] to measure cognitive convergence in design teams. Earlier versions of this work used a version of Dong's methodology to analyze group decision-making within the Food and Drug Administration [3]. This paper extends the work begun in [3] through the use of the Author-Topic (AT) model [4], a Bayesian inference tool commonly used in the field of information retrieval. The resulting output may be used to build social networks, whose analysis can provide insight into the resulting committee dynamics.

2 Why the FDA?

The empirical analysis mentioned above requires data in the form of committee meeting transcripts. These are often not recorded in textual form, or are proprietary to the organization that commissioned the committee. We therefore turn to transcripts of expert committees that are a matter of public record. The ideal data source must have the following attributes:

1. Analysis or evaluation of a technological artifact
2. Participation of multiple experts from different fields or areas of specialization
3. A set of expressed preferences per meeting (such as a voting record)
4. Multiple meetings, so as to enable statistical significance

These requirements are met by the Food and Drug Administration's medical device advisory panels. The most uncertain, and therefore difficult, medical devices are reviewed by expert advisory panels prior to their exposure to the American market. A device's approval, and future diffusion, often rests upon the panel's assessment of the device's safety. These panels are aimed at producing a recommendation, informed by the expertise and knowledge of panel members, which can supplement the FDA's "in-house" decision process. Multiple experts are consulted so that the group decision's efficacy can take advantage of many different institutional perspectives. Panel members' values and institutional contexts may differ, leading to different readings of the evidence, and therefore different recommendations [5]. It is difficult to determine whether panel recommendations promote FDA's mission of promoting both safety and innovation, particularly since there is no alternative against which a panel's decision may be compared. This suggests that, in order to determine the FDA's ability to differentiate between devices, we must be able to distinguish how specific innovations are perceived by actors at the regulatory level of the health care system.

In health care, Gelijns et al. note that strictly evidence-based decisions are often not possible for the following reasons [5]:

1. A given data-set may be interpreted differently by different experts, especially in the presence of high uncertainty. Unless these experts can learn from one another, good decision-making might be impaired.
2. Patterns of technological change are difficult to predict, particularly when innovations are ultimately used for different purposes than originally intended.
3. Even in the case of clear evidence, decision-makers may disagree on its implications due to differing value systems.

This suggests that a device's determination as safe or efficacious depends strongly on factors that are not within

the purview of strictly "evidence-based" decision-making. [6] argues that these are largely shaped by the perceptions, and hence, the knowledge and expertise, of risk assessors. Institutions that might impact decision-making include membership in a particular profession, specialty, or bureaucratic organization [5, 7-9].

3 How to analyze institutional factors?

The literature on decision-making by groups, as well as by expert committees, is vast. Mathematical theories of group decision-making abound within the rational choice and game-theory literature (see, for example, [10]). Within political science, committee decision-making has been studied extensively from the perspective of social-choice theory. This work often assumes that preferences are fixed in advance and that decisions are made on purely strategic grounds. Therefore, the differences between actors are modelled by different individual preference profiles [11]. Similar work in economics has differentiated between actors by "type" (i.e., expert or non-expert), with the intention of identifying an optimal allocation of experts [12]. These models have contributed much to our understanding of the rational-choice components of group decision processes. Our work examines the institutional antecedents of preference formation. If preferences have been determined successfully, it is conceivable that the considerable machinery developed in the rational-choice framework could be brought to bear.

The empirical counterpart to the rational choice analysis of group decisions is to be found within the organizational psychology literature. Here, the dominant approach is to model such groups as information processing mechanisms. Therefore, these publications focus on measuring the extent to which previously unshared information becomes shared, and on the basis of what sorts of information the ultimate decision is made [13-15]. It was often found that shared information drives decisions, suggesting that sources of unshared information (or unique expertise) may be ignored by groups [16]. An analysis by Kameda et al. attempts to identify members in a group who are "cognitively central" – such members are able to translate otherwise unshared information into terms that other group members are familiar with [17]. These analyses have provided deep insights into team dynamics within a laboratory setting and outline a number of general interpersonal dynamics that one might expect in groups and committees. This research is aimed at extending this line of work into real-world committees, with the aim of identifying which situations best correspond to which laboratory experiments.

Work within the anthropology and Science, Technology and Society (STS) literatures is perhaps most relevant to this inquiry. In particular, the penetrating analyses of Mary Douglas note that group membership may affect perception of data [18]. This is reflected in the fact that each institution, and indeed, each specialty, possesses its own unique language and jargon. This is particularly true in medical and academic disciplines, where conceptual

precision is required to communicate within the specialty. Nelson notes the importance of written and oral language as a means of encapsulating and transferring tacit knowledge [19]. On the other hand, an outsider to the institution may be unable to understand the discourse. The STS literature extends this notion by noting that language is used as a cognitive mechanism to delineate professional boundaries. This directs the attention of experts within a specialty toward a given interpretation of a problem that is consistent with that expert's training, while simultaneously directing that attention away from other possible interpretations [20-24]. Casting "organization [as] the mobilization of bias", [25] recognizes institution-specific symbolism in language, noting that the choice of terminology in defining a problem may be seen as a means of mobilizing support. Furthermore, the linguistic definition of a problem dictates, to some extent, its solution. Choosing to use specialized technical words serves to narrow the range of subjective meaning of otherwise ambiguous terminology (such as "safety" or "efficacy" in FDA's context) thereby implicitly redefining the problem according to a given speaker's particular interest. Determining the speaker's intention in using both precise and "symbolic" language can allow insight into their institutional frame of reference. Therefore, we turn to an analysis of language in order to be able to examine institutional factors in group decision-making.

The work cited above suggests that the determination of institutional and other interpersonal affinity might be identified through the use of common language and jargon. This paper utilizes language, and in particular, Bayesian modelling, to determine whether stakeholders within a committee meeting are discussing similar topics.

An earlier iteration of this work used Latent Semantic Analysis, based upon the work of [1], to study the same corpus of FDA transcripts. Results of this analysis may be found in [3]. In practice, LSA works well for identifying and separating major trends within a discourse, as in evaluating the divergence between two groups of stakeholders over time (e.g., exploring device sponsor versus committee-member trajectories within the latent semantic space). LSA has some well-known limitations inherent in the Singular Value Decomposition (SVD) used to reduce the associated term-document matrix. Among these is the assumption that words are embedded within a Euclidean "concept-space". This particular assumption breaks down when comparing words that are polysemous – i.e., having the same spelling but different meanings (compare "bat" the animal vs. "bat" in the context of baseball). LSA represents the location of these words in the Euclidean semantic space as the average over the two separate locations – a clearly incorrect representation. Furthermore, LSA assumes that the noise around each word's location in the Euclidean space is normally-distributed, an assumption that introduces increasingly more distortion into the analysis as a given speaker uses fewer words. These limitations make it difficult to resolve the linguistic attributes of individual speakers, particularly in the absence of extensive speaker data within a given

meeting. Furthermore, the latent dimensions of the LSA feature space, which nominally correspond to latent concepts of a discourse, are often difficult to interpret. These limitations motivate the use of a Bayesian model (for an excellent comparison of LSA to Bayesian models of text analysis, see [26]).

One form of analysis of these transcripts uses a variant of Rosen-Zvi et al.'s Author-Topic (AT) Model [4]. Like LSA, AT also uses a term-document matrix as input. The AT model is implemented as follows:

3.1 Construction of a word-document matrix

As with LSA, we must parse a transcript into a word-document matrix. This method is exactly the same as that followed in [3]. Namely, we first consider a corpus of documents, \mathcal{D} , containing n documents $d_1 \dots d_n$. Consider, as well, the union of all words over all documents, W . Suppose there are $m < n$ words, $w_1 \dots w_m$. We may therefore construct a "word-document matrix", \mathbf{X} , with dimensions $m \times n$, where each element in the matrix, x_{jk} , consists of a frequency count of the number of times word j appears in document k . For the analyses reported in this paper, a word-document matrix was constructed using the Python 2.5 programming language. Non-content-bearing "function words", such as "is", "a", "the", etc., were pre-identified and removed automatically. In addition, words were reduced to their roots using the PyStemmer, a Python implementation of Porter's Snowball algorithm [27].

3.2 AT Model Structure and Implementation

The Author-Topic model provides an analysis that is guided by the authorship data of the documents in the corpus, in addition to the word co-occurrence data used by LSA. Each author (in this case, a speaker in the discourse) is modelled as a multinomial distribution over a fixed number of topics that is pre-set by the modeller. Each topic is, in turn modelled as a distribution over words. A plate-notation representation of the generative process underlying the Author-Topic model is found in Figure 1. The Author-Topic model is populated using a Markov-Chain Monte Carlo Algorithm that is designed to converge to the distribution of words over topics and authors that best matches the data. Information about individual authors is included in the Bayesian inference mechanism, such that each word is assigned to a topic in proportion to the number of words by that author already in that topic, and in proportion to the number of times that specific word appears in that topic. Thus, if two authors use the same word in two different senses, AT will account for this polysemy. Details of the MCMC algorithm implementation are given in [4]. The AT model was implemented in MATLAB using the Topic Modelling Toolbox algorithm [28].

3.3 Model Parameters

The AT model requires the selection of two parameters. Each author's topic distribution is modelled as having been drawn from a uniform Dirichlet distribution, with parameter

α . A Dirichlet distribution is used because it is the conjugate prior of the multinomial distribution. One may think of α as a smoothing parameter. Values of α that are smaller than unity will tend to more closely fit the author-specific topic distribution to observed data. If α is too small, one runs the risk of overfitting. Similarly, values of α greater than unity tend to bring author-specific topic distributions closer to uniformity. A value of $\alpha=50/(\# \text{ topics})$ was used for the results presented in this paper. This is the values suggested by the creators of the Topic Modelling Toolbox after extensive empirical testing. Similar to α is the second Dirichlet parameter, β , from which the topic-specific word distributions are drawn. β values that are large tend to induce very broad topics with much overlap, whereas smaller values of β induce topics which are specific to small numbers of words. Following the empirical guidelines set forth by Griffiths and Steyvers [28], and empirical testing performed by the authors, we set the value of $\beta = 2000/(10 * \# \text{ words})$.

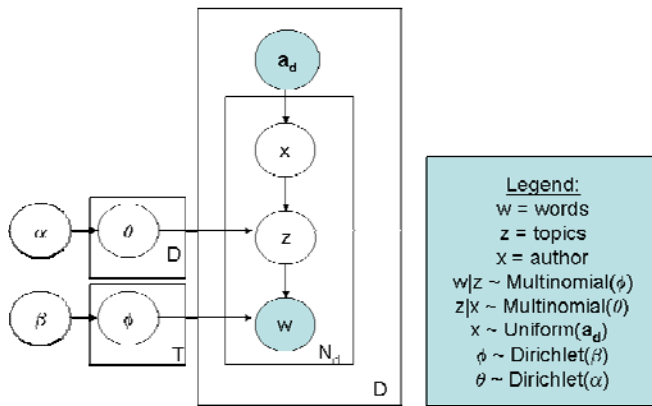


Figure 1 - A plate notation representation of the Author-Topic model from [4]. Authors are represented by a multinomial distribution over topics, which are in turn represented by a multinomial distribution over all words in the corpus.

3.4 Committee Filtering

As demonstrated in [3], it is often difficult to differentiate between panel members. This is partially because they share procedural language in common. Therefore, a large proportion of the words spoken by each committee member would all be assigned to the same topic. This problem is solved using the AT model by creating a “false author” named “committee”. Prior to running the AT model’s algorithm, all committee voting members’ statements are labeled with two possible authors – the actual speaker and “committee”. Since the AT model’s MCMC algorithm randomizes over all possible authors, words that are held in common to all committee members are assigned to “committee”, whereas words that are unique to each speaker are assigned to that speaker. In practice, this allows individual committee members’ unique topic profiles to be identified, as demonstrated below.

3.5 Sample Output

Figure 2 shows sample output of the Author-Topic model applied to the FDA Meeting held on March 4th, 2002.

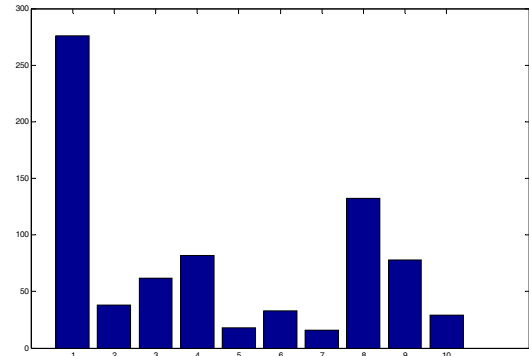


Figure 2 - Sample output from the Author-Topic model run on the FDA Circulatory Systems Devices Advisory Panel Meeting for March 4th, 2002. This chart is the per-speaker topic distribution for Dr. Konstam, one of the panel members.

Each topic in Figure 2 may be identified by its most probable words.

Table 1 displays the top five most probable word stems for each topic:

Table 1 - The top five word-stems for one run of the AT model on the corpus for the Circulatory Systems Devices Panel Meeting of March 4, 2002.

Topic Number	Top Five Word-Stems
1	'clinic endpoint efficaci comment base'
2	'trial insync icd studi was'
3	'was were sponsor just question'
4	'patient heart group were failur'
5	'devic panel pleas approv recommend'
6	'think would patient question don'
7	'dr condit vote data panel'
8	'effect just trial look would'
9	'lead implant complic ventricular event'
10	'patient pace lead were devic'

Within a clinical trial administered by the FDA, a device manufacturer must meet a certain set of clinical “endpoints”, often manifested as a proportion of a population that is free from disease or adverse events (e.g., device failure). Such trials typically have different endpoints for device safety and efficacy, both of which must be met. From this table, we can see that Dr. Konstam’s major topic of interest involved questions of what was the appropriate clinical endpoint for the study in question (often a common debate on these panel meetings). It seems that he was particularly interested in the efficacy endpoints (as opposed to the safety endpoints).

4 Generation of social networks

The above methodology can give us insight into the topics of interest for each speaker. Nevertheless, topics, on their own, provide little direct information about how individual speakers might relate to one another. Instead, we would like to use the topic information provided by the AT model to generate a social network.

4.1 Network construction

We would like to link together speakers who commonly use the same topics of discourse. In particular, we would like to be able to construct an Author-Topic matrix, **A**, with entries equal to 1 if that author uses that topic, and entries equal to 0 otherwise. This matrix, when multiplied by its transpose (**A * A'**) yields a linkage patterns among speakers. This may be interpreted as a social network [29].

4.2 Author-topic matrix determination

The AT model outputs an Author-Topic matrix that gives the total number of words assigned for each author to each topic. This information must be reduced to the **A** matrix identified above. Given no prior information about a given author's topic distribution, we might assume that that such a distribution is uniform over all topics. Therefore, we might expect *a priori* that the total number of words assigned by that author to each topic would be equal to the total number of words spoken by that author divided by the total number of topics. For example, if there are ten topics and a given author uses 200 words, we would expect there to be 20 words in each topic *a priori*. We consider an author to have a meaningful component within a given topic if that author uses significantly more words than would be expected under the *a priori* assumption outlined above. Early versions of this scheme used a threshold value of 2/(# topics) – thus, given ten topics, if more than 20% of an individual author's words were assigned to a given topic, that topic was assigned to that author. A resulting social network from this scheme is shown in Figure 3.

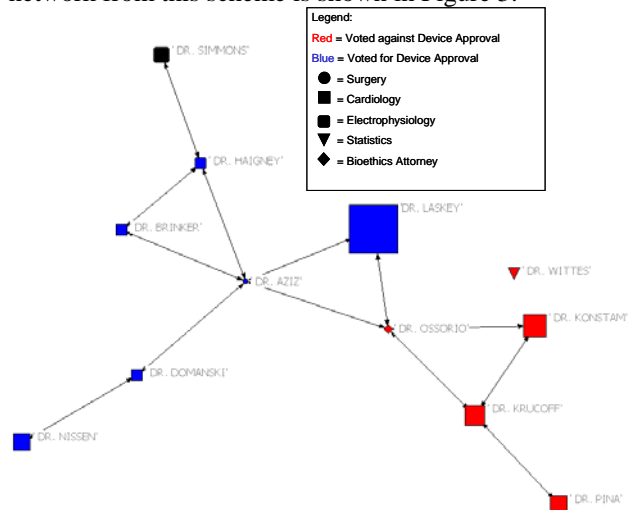


Figure 3 – Social network of the FDA Circulatory Systems Devices Advisory Panel Meeting held on March 5, 2002. Threshold value is constant for each of 10 topics. Node size is proportional to the number of words spoken by the

corresponding speaker. Dr. Simmons is a panel member who left the meeting before the vote occurred. Random seed = 3.14. Graphs were generated using UCINET. This iteration shows a potential clean cut between voters.

Current work is ongoing using the binomial statistical test as a more principled way of determining topic membership. The binomial statistical test operates by examining the cumulative distribution of the binomial probability mass function, given by

$$\Pr(K = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Figure 4 - Equation for the binomial probability mass function. Using this approach, *n* is the total number of words spoken by a given author, *k* is the total number of words spoken by that author within the topic under investigation, and *p* is 1/(# topics).

Under this scheme, an author is assigned to a topic if the cumulative probability that that author used *k* out of *n* words in a given topic is less than 0.05/*b*, where *b* is the Bonferroni significance level correction factor. Given *a* authors, *b* = *a* * (*a*-1) / 2, since one comparison is being made for each pair of authors. A sample social network from this scheme is shown in Figure 5.

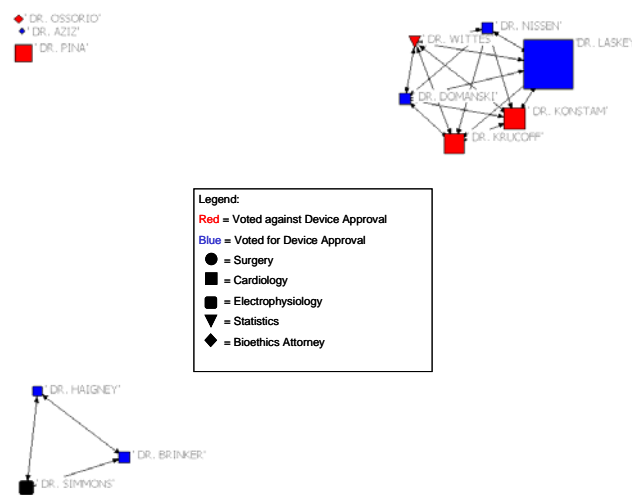


Figure 5 - Social network of the FDA Circulatory Systems Devices Advisory Panel Meeting held on March 5, 2002. Threshold value is determined using the binomial test described above. Node size is proportional to the number of words spoken by the corresponding speaker. Dr. Simmons is a panel member who left the meeting before the vote occurred. Random seed = 201.657. 2100th draw from MCMC algorithm Graphs were generated using UCINET. This iteration shows the presence of two separate discussion groups. Note that Drs. Ossorio and Aziz, both bridging members in Figure 3, are disconnected. This is due to their small number of words contributed.

Each social network generated using this scheme is the result of one MCMC iteration. Multiple iterations, when taken together, form a probability distribution over a set of possible Author-Topic assignments, and therefore, connectivity patterns. We can expect that different iterations of the MCMC algorithm will yield drastically different graphs. For example, the results a second draw from the same MCMC chain that yielded Figure 3 is shown in Figure 6, whereas the results of a second draw from the MCMC chain that yielded Figure 5 is shown in Figure 7.

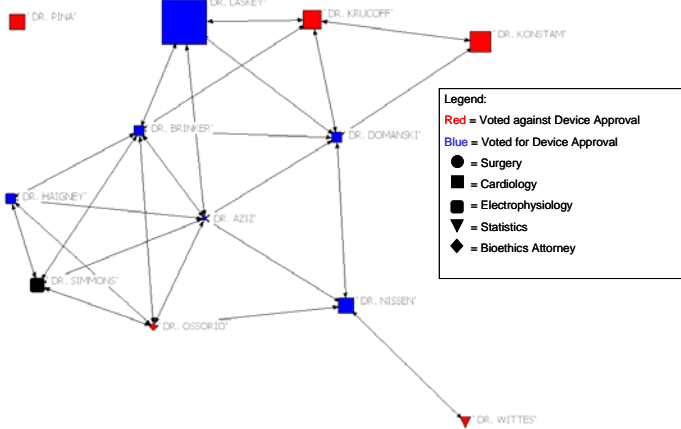


Figure 6 - A second iteration of the meeting of the FDA Circulatory Systems Devices Advisory Panel Meeting held on March 5, 2002. Threshold value is determined using the binomial test. A Node size is proportional to the number of words spoken by the corresponding speaker. Random seed = 201.657, 2200th draw from MCMC algorithm. Graphs were generated using UCINET.

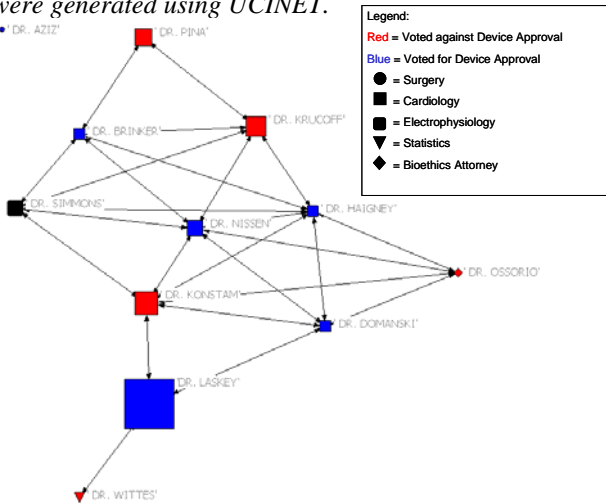


Figure 7 - A second iteration of the meeting of the FDA Circulatory Systems Devices Advisory Panel Meeting held on March 5, 2002. Threshold value is constant for each of 10 topics. A Node size is proportional to the number of words spoken by the corresponding speaker. Random seed = 613. Graphs were generated using UCINET

The high variability among draws from the MCMC algorithm suggests that links should be differentially weighted – some links appear in virtually all iterations, whereas other links appear in relatively few iterations. Averaging over multiple MCMC iterations enables a social

network to be created with weighted links, where the weight of each link is proportional to its frequency of occurrence among iterations. Examples of this may be found in Figure 8 and Figure 9, corresponding to constant and binomial threshold conditions, respectively.

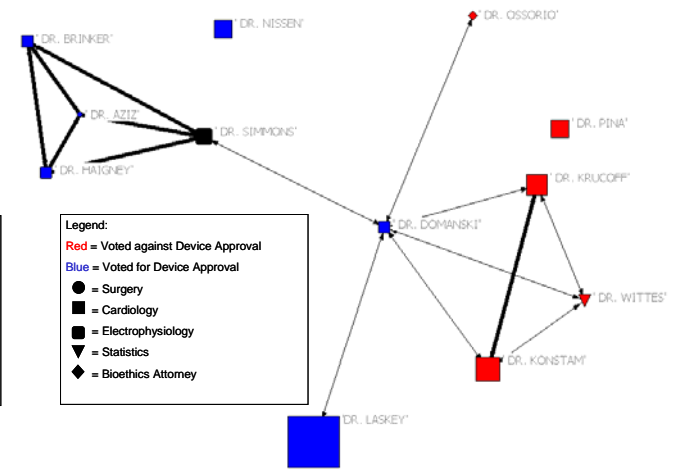


Figure 8 - Average of 200 iterations for the meeting of the FDA Circulatory Systems Devices Advisory Panel Meeting held on March 5, 2002. Iterations use a constant threshold value for each of ten topics. A heavy line indicates a strong link (linked in >100 iterations). A light line indicates that the speakers are linked more than the global average of all speakers. Remaining links have been deleted.

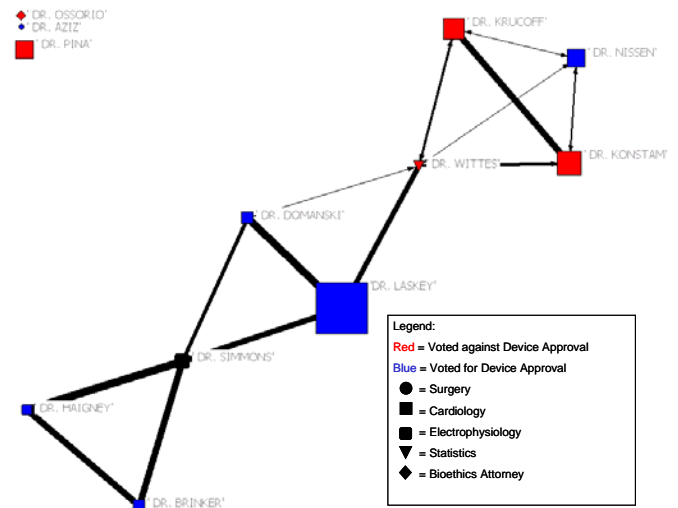


Figure 9 – Average of 200 iterations for the meeting of the FDA Circulatory Systems Devices Advisory Panel Meeting held on March 5, 2002. Iterations use a binomial threshold value for each of ten topics. Heavier lines indicate stronger links, whereas lighter lines indicate weaker links. All links shown are stronger than the global average of all speakers. Remaining links have been deleted.

The above two figures show a common structure that evolves for this meeting. For example, both figures display a structure that tends to group together those speakers who voted similarly. This is encouraging for our hypothesis that voters who speak the same way tend to vote the same way.

Furthermore, it seems that those cardiologists who voted against the device's approval might have done so for statistical reasons (both are linked to the panel's statistician).

The difference the two figures highlights the differences between the two threshold conditions. The constant threshold condition tends to overly favour speakers who talk infrequently, such as Dr. Aziz and Dr. Ossorio. Because of their relatively small numbers of words, it is harder for these speakers to achieve statistical significance using the binomial test, and so they are less likely to be linked. On the other hand, the constant threshold condition requires more words to establish a link to a frequent speaker, compared to a binomial threshold.

5 Comparison of multiple cases

The previous case demonstrated how social networks can be built. The following cases begin a preliminary analysis of the capabilities of the methodology outlined in this paper.

5.1 Grouping by medical specialty?

Some meetings display voting along institutional lines more clearly than do others. For example, Figure 10 and Figure 11 show a strong grouping by medical specialty. In particular, surgeons and internal medicine experts (cardiologist and pharmacologists) seem to form two different parts of the same graph. It is noteworthy that the statistician on the panel voted with the cardiologists, perhaps indicating that surgical decision-making might have been driven by clinical assessment, as compared to internal medicine decision-making, which may have been driven more by statistical reasoning. This is consistent with the observation that the device under analysis in this meeting was associated with a minimally-invasive surgical procedure.

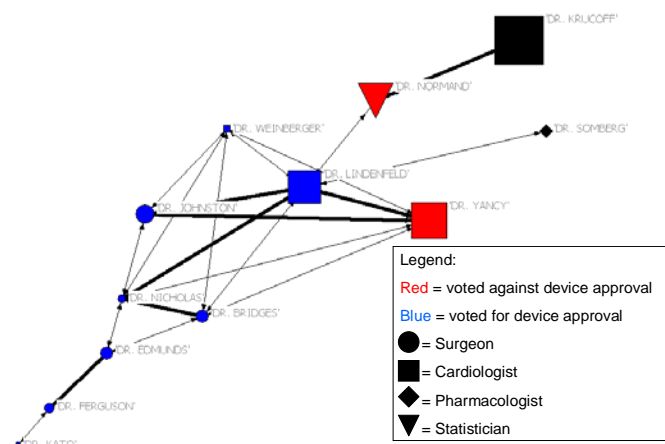


Figure 10 - Average of 200 iterations for the meeting of the FDA Circulatory Systems Devices Advisory Panel Meeting held on January 13, 2005. Iterations use a constant threshold value for each of ten topics. A heavy line indicates a strong link (linked in >100 iterations). A light line indicates that the speakers are linked more than the

global average of all speakers. Remaining links have been deleted.

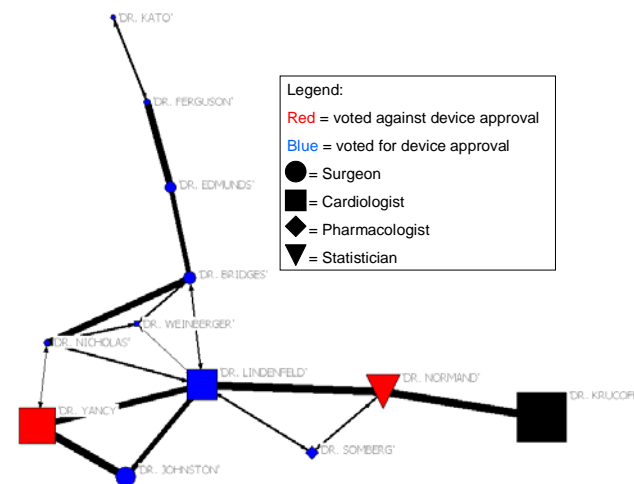


Figure 11 - Average of 200 iterations for the meeting of the FDA Circulatory Systems Devices Advisory Panel Meeting held on January 13, 2005. Iterations use a binomial threshold value for each of ten topics. Heavier lines indicate stronger links, whereas lighter lines indicate weaker links. All links shown are stronger than the global average of all speakers. Remaining links have been deleted.

Both of these figures place Dr. Lindenfeld in the most central position on the graph of committee voting members. Furthermore, Figure 11 recognizes strong links between Drs. Yancy and Norman to Dr. Lindenfeld. This is consistent with a reading of the meeting transcript that indicates that Dr. Lindenfeld shared many of the concerns of her colleagues, despite ultimately voting with the surgeon majority. Dr. Krucoff, who abstained from voting, is strongly linked to Dr. Norman, consistent with his background as a clinical trial designer who would be interested in both the clinical and the statistical elements of the analysis. It is interesting to note that both figures also display long “tails” of surgeons, who seem to have voted as a bloc.

The above results indicate that, at least in some cases, medical specialty might have some predictive value for voting outcomes. Further analysis will be aimed at attempting to confirm or deny this hypothesis.

5.2 Comparison across time

Each FDA meeting is may be divided into sections that coincide with natural breaks in the meeting. Examples of such include lunch, and coffee breaks. These breaks provide natural stopping points for an analysis. In addition, it is precisely during these breaks that committee members may share information off-the-record that would otherwise remain unshared. Thus comparing pre- and post-break graphs might provide insight into the evolution of committee decisions. Figure 12, Figure 13, and Figure 14 show the social networks of the January 13, 2005 meeting for the amount of time between each break:

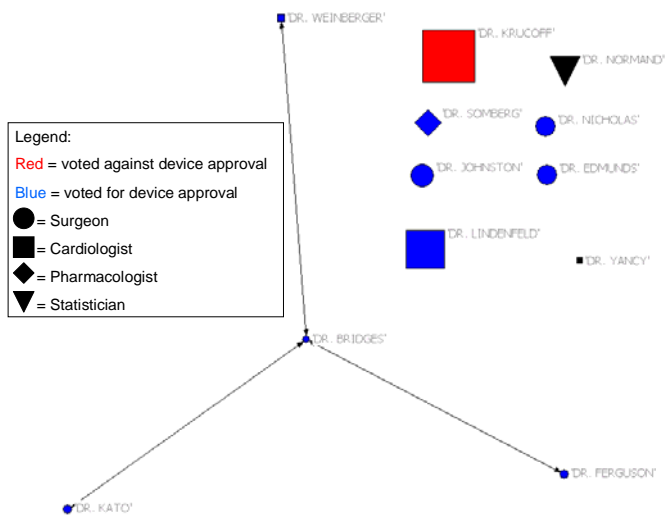


Figure 12 - First segment of the January 13, 2005 Circulatory Systems Devices Panel Meeting. This graph shows that, at this point in the meeting, voting members had not yet expressed any preferences regarding voting. Rather, committee members were listening to the open public hearing and sponsor presentations. Data include utterances 1-377 of 1671 total utterances.

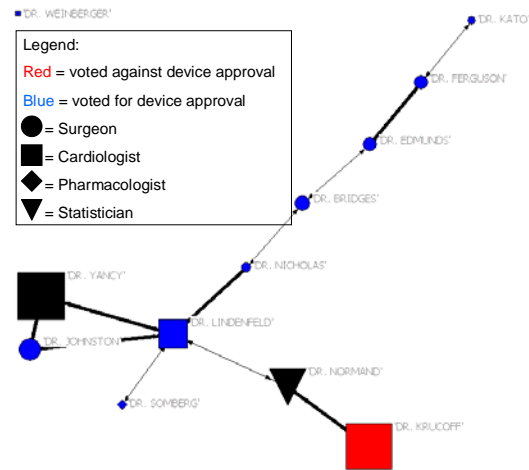


Figure 14 –Third, and final, segment of the January 13, 2005 Circulatory Systems Device Panel Meeting. This graph shows that, after lunch, the surgeons in the room, who were previous silent, seemed to align in favour of device approval. Drs. Lindenfeld, Normand, Krucoff, and Somberg seemed to maintain their relative positions between the second and third segments. Data include utterances 590-1671.

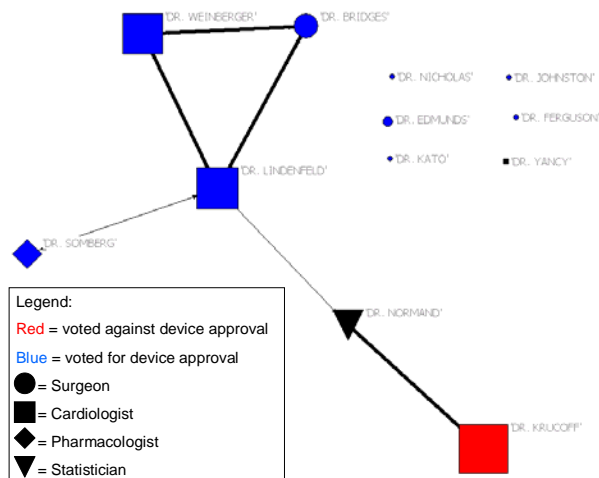


Figure 13 - Second segment of the January 13, 2005 Circulatory Systems Devices Panel Meeting. This graph shows that, at this point in the meeting, Drs. Lindenfeld, Somberg, Bridges, Weinberger, Normand and Krucoff had begun discussing the statistical elements of the clinical trial design. Five of the six surgeons present have not yet expressed opinions. Data include utterances 378-589 of 1671 total utterances.

The above figures show a small group of voters engaging in a discussion of interest – forming a coalition, as it were, while those who remain silent eventually come to dominate the voting dynamics through strength of numbers. It is particularly interesting that these two groups may be roughly divided by medical specialty.

We may perform a similar analysis on the meeting analyzed previously – i.e., the meeting of the Circulatory Systems Devices Panel of March, 5th, 2002. This meeting is divided into “before lunch” and “after lunch” segments, as shown in Figure 15 and Figure 16.

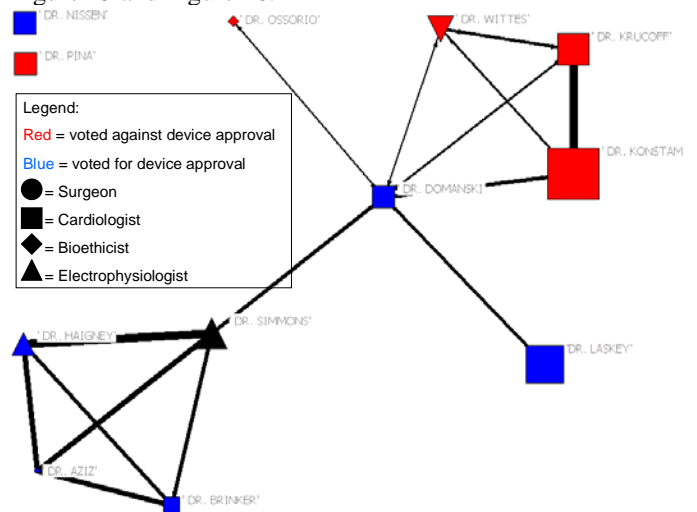


Figure 15 – Before-lunch segment of the March 5th, 2002 Circulatory Systems Devices Panel Meeting. This graph shows that, at this point in the meeting, voting members had largely aligned themselves into blocs that would later vote similarly. Data include utterances 1-703 of 1250 total utterances.

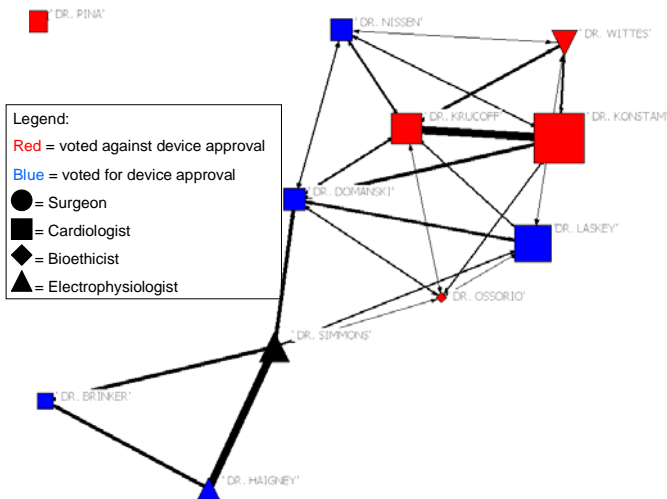


Figure 16 – After-lunch segment of the March 5th, 2002 Circulatory Systems Devices Panel Meeting. This graph shows that, by the second half of the meeting, those who would later vote against device approval had become more strongly linked to those who would later support device approval. This pattern perhaps reflects attempts by the approval voters to convince the non-approval voters to voter differently. Data include utterances 704-1250 of 1250 total utterances.

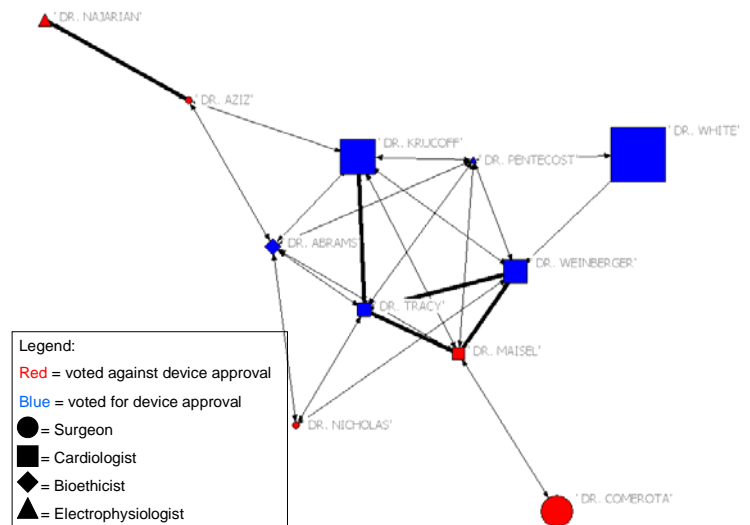


Figure 18 - After-lunch segment of the April 21st, 2004 Circulatory Systems Devices Panel Meeting. This graph shows that the well-defined coalitions of the before-lunch segment have broken down – particularly the anti-device coalition. This may well be due to attempts by members of one coalition to influence the other, leading to cross-coalition dialogue.. Data include utterances 877-1822 of 1822 total utterances.

Finally, we examine a meeting held on April 24, 2004. This meeting was originally divided into four parts. Given that the voting members did not speak during the first two quarters of the meeting (leading to a fully disconnected graph), we present only the last two parts of the meeting, displayed in Figure 17 and Figure 18.

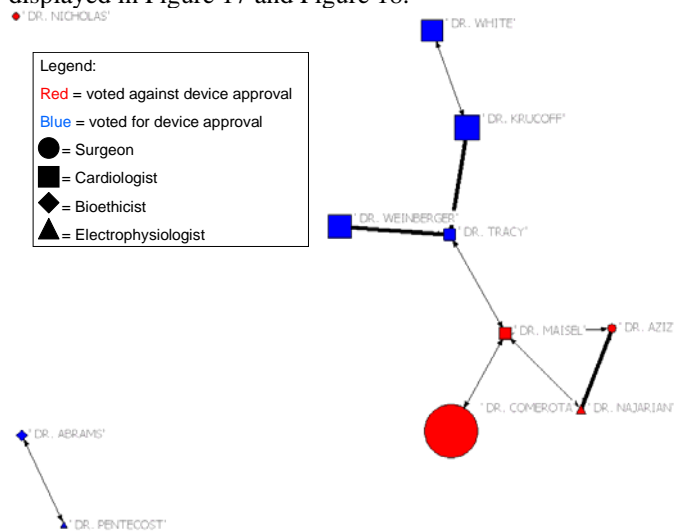


Figure 17 – Before-lunch segment of the April 21st, 2004 Circulatory Systems Devices Panel Meeting. This graph shows well-defined coalitions having been formed relatively early in the meeting. It is interesting that voting patterns seem to largely respect the boundaries of particular medical specialties (i.e., surgeons vs. cardiologists). Data include utterances 399-876 of 1822 total utterances.

The first meeting segment shows the formation of two coalitions that ultimately voted oppositely. It is interesting that the pro-approval coalition is composed largely of cardiologists, whereas the anti-approval coalition is composed largely of non-cardiologists. Furthermore, the bridging members, Drs. Tracy, an electro-physiologist, and Maisel, a cardiologist with a public health background were outliers within their own group. Both Drs. Tracy and Maisel have served as chairs of previous meetings, and are therefore perhaps more driven to achieve consensus among panel members.

The second meeting segment shows the breakdown and fragmentation of the anti-approval coalition and the consolidation of the pro-approval coalition prior to voting.

6 Conclusions –Methodological Limitations and Future Work

The method outlined in this paper relies on meeting transcripts to generate empirical findings regarding committee decision-making. It is therefore sensitive to the limitation that not all committee members might express their views truthfully. Although this is a major concern in the political sciences, it is less likely to be a problem when one considers that technical expert committees are aiming to construct a mutual understanding of a technical device; a situation in which misrepresentation of one’s expertise would likely not be consistent with individual or group goals. Furthermore, even though political actors may strategically misrepresent their ultimate aims, it is very

difficult for individuals to avoid using jargon that they are familiar with.

Perhaps a larger concern is the inability to differentiate between agreement and argument. Two actors are linked if they discuss the same topics. They may do so because they agree on some aspect of the device review, or because they are debating over interpretation of a given element of the debate. This is evident in the figures shown above, where linkage does not always indicate voting similarity. Therefore, a major area for future research is the determination of valence on each of the links in the graph. Determining signs for these graph links will enable a more direct comparison of the voting record to the graphical structure. This, in turn, would enable the analysis of reputation effects – e.g., why don't people vote the way they say they will? Existing work on sentiment classification (e.g., [30]) might potentially be applicable to this problem.

This research is aimed at the development of a quantitative methodology that may be applied to analyze multi-stakeholder decisions made by technical expert. These decisions are of the sort that one might expect to encounter frequently during design processes for large-scale engineered systems. The methodology presented in this paper, although still preliminary, has been used to generate meaningful social networks from transcripts of FDA medical device advisory panel meetings. Future work will focus on applying this method to a larger number of cases with the intention of producing generalizable findings and developing theory.

7 References

- [1] Andy Dong, "The latent semantic approach to studying design team communication," *Design Studies* 26 (2005): 445-461.
- [2] Scott Deerwester et al., "Indexing by Latent Semantic Analysis," *Journal of the American Society for Information Science* 41, no. 6 (1990): 391-407.
- [3] David André Broniatowski et al., "The Influence of Institutional Background on the Approval of Engineered Systems," in (presented at the CSER 2008, Redondo Beach, CA, 2008), 1-11.
- [4] Michal Rosen-Zvi et al., "The author-topic model for authors and documents," in *Proceedings of the 20th conference on Uncertainty in artificial intelligence* (Banff, Canada: AUAI Press, 2004), 487-494, <http://portal.acm.org/citation.cfm?id=1036843.1036902>.
- [5] Annetine C. Gelijns et al., "Evidence, Politics, And Technological Change," *Health Affairs* 24, no. 1 (2005): 29-40.
- [6] Mary Douglas and Aaron Wildavsky, *Risk and Culture* (Berkeley, CA: University of California Press, 1982).

- [7] Marc. P. Freiman, "The Rate of Adoption of New Procedures among Physicians: The Impact of Specialty and Practice Characteristics," *Medical Care* 23, no. 8 (1985): 939-945.
- [8] Deborah A. Savage, "Professional Sovereignty Revisited: The Network Transformation of American Medicine?," *Journal of Health Politics, Policy and Law* 29, no. 4-5 (2004): 611-677.
- [9] Deborah A. Savage and Paul L. Robertson, "The Maintenance of Professional Authority: The Case of Physicians and Hospitals in the United States," in *Authority and Control in Modern Industry* (London: Routledge, 1999), 155-172, <http://sp.uconn.edu/~wwwciom/Maintenance.pdf>.
- [10] James D. Morrow, *Game Theory for Political Scientists* (Princeton University Press, 1994).
- [11] Wulf Gaertner, *A Primer in Social Choice Theory*, Revised edition. (Oxford University Press, 2009).
- [12] Bauke Visser and Otto H. Swank, "On Committees of Experts," *Quarterly Journal of Economics* (2007): 337-372.
- [13] Susan Mohammed and Erika Ringseis, "Cognitive Diversity and Consensus in Group Decision Making: The Role of Inputs, Processes, and Outcomes," *Organizational Behavior and Human Decision Processes* 85, no. 2 (July 2001): 335, 310.
- [14] Richard Klimoski and Susan Mohammed, "Team Mental Model: Construct or Metaphor?," *Journal of Management* 20, no. 2 (April 1, 1994): 403-437, doi:10.1177/014920639402000206.
- [15] Susan Mohammed, "Toward an Understanding of Cognitive Consensus in a Group Decision-Making Context," *Journal of Applied Behavioral Science* 37, no. 4 (December 1, 2001): 408-425, doi:10.1177/0021886301374002.
- [16] J. R. Winqvist and J. R. Larson, "Information Pooling: When It Impacts Group Decision Making," *JOURNAL OF PERSONALITY AND SOCIAL PSYCHOLOGY* 74 (1998): 371-377.
- [17] Tatsuya Kameda, Yohsuke Ohtsubo, and Masanori Takezawa, "Centrality in Sociocognitive Networks and Social Influence: An Illustration in a Group Decision-Making Context," *Journal of Personality and Social Psychology* 73, no. 2 (1997): 309, 296.
- [18] Mary Douglas, *How Institutions Think* (Syracuse, New York: Syracuse University Press, 1986).
- [19] Richard R. Nelson, "On the uneven evolution of human know-how," *Research Policy* 32, no. 6 (June 2003): 909-922, doi:10.1016/S0048-7333(02)00093-8.
- [20] M. Mulkay, T. Pinch, and M. Ashmore, "Colonizing the Mind: Dilemmas in the Application of Social Science," *Social Studies of Science* 17, no. 2 (1987): 231.
- [21] R. Rapp, "Extra chromosomes and blue tulips: medico-familial interpretations," *Living and Working with the New Medical Technologies: Intersections of Inquiry* (2000).

- [22] L. Winner, "On not hitting the tar-baby," *The Whale and the Reactor: A Search for Limits in an Age of High Technology*. Chicago 111 (1986): 138-154.
- [23] Harry Collins and Robert Evans, *Rethinking Expertise* (University Of Chicago Press, 2009).
- [24] Carol Cohn, "Sex and Death in the Rational World of Defense Intellectuals," *Signs* 12, no. 4 (1987): 718, 687.
- [25] Roger W. Cobb and Charles D. Elder, *Participation in American Politics: The Dynamics of Agenda-Building* (Baltimore and London: The Johns Hopkins University Press, 1983).
- [26] T. L. Griffiths, M. Steyvers, and J. B. Tenenbaum, "Topics in Semantic Representation," *PSYCHOLOGICAL REVIEW-NEW YORK* 114, no. 2 (2007): 211.
- [27] M. F. Porter, *Snowball: A language for stemming algorithms*. October 2001.
- [28] Thomas L. Griffiths and Mark Steyvers, "Finding scientific topics," *Proceedings of the National Academy of Sciences of the United States of America* 101, no. Suppl 1 (April 6, 2004): 5228-5235, doi:10.1073/pnas.0307752101.
- [29] Stanley Wasserman and Katherine Faust, *Social Network Analysis: Methods and Applications*, 1st ed. (Cambridge University Press, 1994).
- [30] Matt Thomas, Bo Pang, and Lillian Lee, "Get out the vote: Determining support or opposition from Congressional floor-debate transcripts," in (Sydney, Australia: Association for Computational Linguistics, 2006), 327-335, <http://www.aclweb.org/anthology/W/W06/W06-1639>.