



MIT Open Access Articles

Analysis of Social Dynamics on FDA Panels Using Social Networks Extracted from Meeting Transcripts

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation	Broniatowski, David A., and Christopher L. Magee. "Analysis of Social Dynamics on FDA Panels Using Social Networks Extracted from Meeting Transcripts." IEEE, 2010. 329–334. © 2010 IEEE
As Published	http://dx.doi.org/10.1109/SocialCom.2010.54
Publisher	Institute of Electrical and Electronics Engineers (IEEE)
Version	Final published version
Citable link	http://hdl.handle.net/1721.1/81980
Terms of Use	Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.

Analysis of Social Dynamics on FDA Panels using Social Networks Extracted from Meeting Transcripts

David A. Broniatowski and Christopher L. Magee

Massachusetts Institute of Technology, Engineering Systems Division

77 Massachusetts Ave., E38-450, Cambridge, MA 02139

{david, cmagee}@mit.edu

Abstract

The social elements of technical decision-making are not well understood, particular among expert committees. This is largely due to a lack of methodology for directly studying such interactions in real-world situations. This paper presents a method for the analysis of transcripts of expert committee meetings, with an eye towards understanding the process by which information is communicated in order to reach a decision. In particular, we focus on medical device advisory panels in the US Food and Drug Administration. The method is based upon natural language processing tools, and is designed to extract social networks from these transcripts, which are representative of the flow of information and communication on the panel. Application of this method to a set of 37 meetings from the FDA's Circulatory Systems Devices Panel shows the presence of numerous effects. Prominent among these is the propensity for panel members from similar medical specialties to use similar language. Furthermore, panel members who use similar language have the propensity to vote similarly. We find that these propensities are correlated – i.e., as panel members' language converges by medical specialty, panel members' votes also tend to converge. This suggests that voting behavior is mediated by membership in a medical specialty and supports the notion that voting outcome is, to some extent, dependent on an interpretation of the data associated with training.

Key Words: Quantitative Content Analysis, Group Decision-Making

1. Introduction

In the committees that concern us in this paper, information must be aggregated from multiple expert specialists. Evaluating committee decision processes requires a means of understanding the interaction between the social and technical specifics of the system in question. The decision of what information is important and how it should be interpreted is the subject of exchange up until the time that each committee member casts a vote. That different experts hold different perspectives and values makes it more likely that

additional aspects of a problem will come under consideration. Nevertheless, this does not guarantee consensus on the interpretation of data.

Different experts, having been trained in different areas or components, will tend to pay attention to those elements of the system that they find consistent with their professional training – i.e., cognitively salient [1]. The mechanisms by which this training is achieved include acculturation within specific professional specialties, and require learning that professional institution's language and jargon. By institution, we mean a set of social norms to which a particular community adheres. This leads to a situation wherein individual experts develop different views of the system.

Understanding how best to structure committees such as those described above requires a method of empirically examining communication within a real-world setting. This paper presents an empirical method aimed at extracting communication patterns and social dynamics through a computational analysis of committee meeting transcripts. A computational approach is used for its consistency and reliability across meetings. Furthermore, an algorithmic approach enables any potential biases that might be present in the analysis to be minimal and transparent. In particular, we use a modification of the Author-Topic Model [2], a Bayesian inference tool used in the field of machine learning, to discover linguistic affinity between committee members. We find that the resulting output may be used to construct social networks representing patterns of communication among panel members. Analyses of these networks are then performed.

2. Literature Review

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 [1]. Among technical experts, this is reflected in the fact that each specialty possesses its own unique language and jargon, which carries with it an implicit scheme for categorizing perceived phenomena [3]. 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

The authors would like to acknowledge the MIT-Portugal Program for its generous support of this research.

given interpretation of a problem that is consistent with that expert’s training, while simultaneously directing that attention away from other possible interpretations.

3. Case Study: FDA Advisory Panels

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 committee meetings 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.

4. Methodological Approach

A major challenge to the use of linguistic data for the analysis of social behavior on expert committees stems from the strong assumption that such dynamics are entirely reflected in language, and that differences in language necessarily indicate differences in perception. Another similar concern is absence of data that might result if a particular voting member of the committee remains silent or says little. Neither can strategic attempts by actors to hide preferences and thereby avoid revealing personal information be explicitly captured in this representation. Indeed, work by Pentland [4] has shown that much social signaling occurs through body language and vocal dynamics that are not able to be captured in a transcript. It should therefore be clarified that this paper does not claim that all social dynamics are manifest in language – rather, word-choice provides one source of insight into a complex, multi-modal process. The extent and severity of this challenge is mitigated somewhat by the literature cited above. Differential use of language due, for example, to assigned roles, may reflect a salient role-based difference between decision-makers that is worth studying on its own merits.

4.2. Construction of a Word-Document Matrix

Our analysis begins with a standard “bag of words” representation for natural language processing applications. For the analyses reported in this paper, a word-document matrix, \mathbf{X} , 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 algorithm.

4.2. AT Model Structure and Implementation

The Author-Topic model provides a structured analysis of \mathbf{X} . In particular, each author (in this case, a speaker in the meeting) is modelled as a distribution over topics, where 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 by 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. Details of the MCMC algorithm implementation are given in [2]. The AT model was implemented in MATLAB using the Topic Modelling Toolbox algorithm [5].

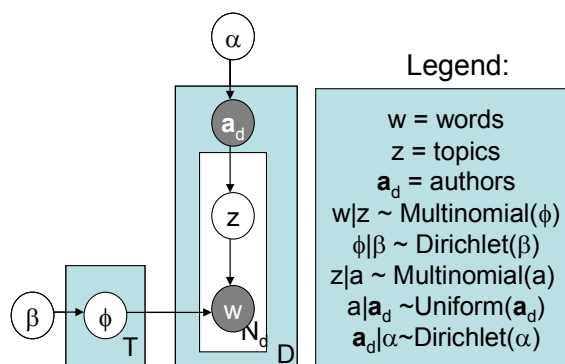


Figure 1. A plate notation representation of the Author-Topic model from [2]. Authors are represented by a multinomial distribution over topics, which are in turn represented by a multinomial distribution over all words in the corpus.

4.3. Hyperparameter Selection

Each author’s topic distribution is modeled as having been drawn from a symmetric Dirichlet distribution, with parameter α . 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, based upon the values suggested by Griffiths and Steyvers [5]. For the numbers of topics considered in these analyses (generally less than 30), this corresponds to a mild smoothing across authors. 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 [5], and empirical testing performed by the author, we set the value of

$\beta = 200 / (\# \text{ words})$. It is interesting that there has been relatively little work within the topic modeling community on the appropriate choice of hyperparameters. Exceptions include algorithms, such as those designed by Wallach [6], which overfit hyperparameters for the purposes of this analysis. To the author’s knowledge, there has been no analysis of the co-selection of topics and hyperparameters.

4.4. Selection of Number of Topics

Given hyperparameter values, as defined above, we may use perplexity as a metric for choosing T . Non-parametric methods may also be used (e.g., [7]) but at a greater computational cost than required for our purposes, especially given our use of fitted priors. The method presented here chooses T so as to be as small as possible (i.e., maximum dimensionality reduction) while still constituting a good model fit. The number of topics is chosen independently for each transcript as follows: 35 AT models are fit to the transcript for $t = 1 \dots 35$ topics (a value determined empirically). As the number of topics increases, model cross-entropy becomes asymptotically smaller. Griffiths and Steyvers [5] report a unique minimum for fitted values of α although they tested topics in increments of 100. In principle, given a sufficiently large number of topics, the perplexity would begin to increase at a relatively mild slope as the model starts overfitting. Lacking such a unique minimum here, we choose the minimum number of topics such that the cross-entropy values are statistically indistinguishable from larger numbers of topics. Thus, for each model, 20 independent samples are generated from one randomly initialized Markov chain after a burn-in of 1000 iterations. Sample independence is guaranteed by introducing a lag of 50 iterations between each sample (lags of 100 iterations were tested, yielding qualitatively similar results). We find the smallest value, t_0 , such that the 95th percentile of all samples for all larger values of t is greater than the 5th percentile of t_0 . Given fitted priors of the sort recommended by Griffiths and Steyvers [5], the asymptotic behavior displayed in Figure 2 is typical of AT Model fits. We set the value of $T = t_0 + 1$ so as to ensure that the model chosen is well beyond the knee in the curve, and therefore in the neighborhood of the minimum perplexity.

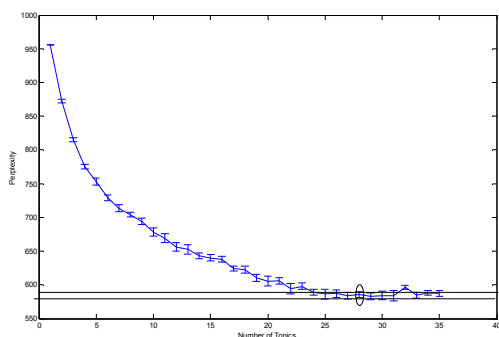


Figure 2. Perplexity vs. number of topics for the meeting of the FDA Circulatory Systems Devices Panel held on July 9, 2001. T , the number of topics, is equal to 28, using the procedure described above. Horizontal lines indicate the 5th and 95th percentiles for perplexity for a 27 topic model fit.

Once the number of topics has been chosen, a T -topic AT Model is again fit to the transcript. Ten samples are taken from 20 randomly initialized Markov chains, such that there are 200 samples in total. These form the basis for all subsequent analysis.

4.4.1. Committee Filtering. Our analysis focuses primarily on the voting members on an advisory panel. It is precisely these members whose evaluations will determine the panel recommendations. Other panel members, such as non-voting guests and consultants, are also included in the analysis because, like the voting members, they play the role of resident experts. Panel members such as the executive secretary, and consumer, patient and industry representatives are not included as part of the committee in the following analyses because they play a relatively small role in panel discussion in the meetings examined. Inclusion of these members is straightforward, and examination of their roles is left to future research.

It is often difficult to differentiate between panel members, especially since the majority of the speech during an FDA panel meeting is occupied by presentations from the sponsor and the FDA. A given voting member might speak relatively rarely. Furthermore, panel members share certain language in common including procedural words and domain-specific words that are sufficiently frequent as to prevent good topic identification. As a result, a large proportion of the words spoken by each committee member may be assigned to the same topic, preventing the AT model from identifying important differences between speakers. In a variant of a technique suggested in [8]¹ 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 language to be identified. In the limiting case where all committee members’ language is common, half of all words would be assigned to “committee” and the other half would be assigned at random to the individual speakers in such a way as to preserve the initial distribution of that author’s words over topics.

¹ The author paper would like to thank Dr. Mark Dredze for suggesting this approach

4.5. AT Model Output

When applied to a transcript, we treat each utterance as a document. Thus, the meeting transcript may be viewed as a corpus. Words within each utterance are grouped into topics with probability proportional to the number of times that word has been previously used in that topic, and the number of times that word’s “author” (i.e., speaker) has previously used that topic.

4.6. Network Construction

We would like to develop a principled way to determine what constitutes a link within a given model iteration. As noted above, we would like to link together speakers who commonly use the same topics of discourse. In particular, we examine each author-pair’s joint probability of speaking about the same topic.

$$P(X_1 \cap X_2) = \sum_i^T P(Z = z_i | X_1) * P(Z = z_i | X_2)$$

We would like to be able to construct an Author-Author matrix, Δ , with entries equal to 1 for each linked author pair, and entries equal to 0 otherwise.

4.6.1. Author-Author Matrix Determination The AT model outputs an Author-Topic matrix, \mathbf{A} , that gives the total number of words assigned to each topic for each author. This information must be reduced to the Δ matrix identified above. The form of the author-topic model makes an explicit assumption regarding an author’s prior distribution over topics. This value is expressed by the hyperparameter α . Given the number of topics fit to a particular model, we may use the value of α to generate a set of *a priori* author-specific topic distributions. These, in turn, can be input into the equation above in order to generate a prior distribution for any given author-pair’s link probability. Such a distribution is shown in Figure 3.

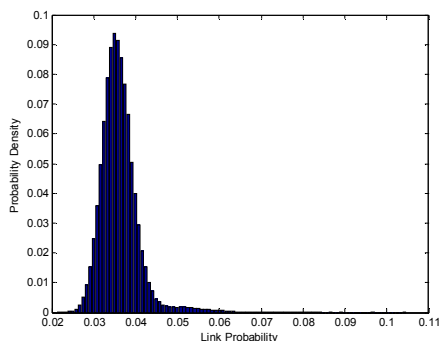


Figure 3. Prior probability distribution for links between speakers in the April 21, 2004 meeting with 28 topics. The median of this distribution is 0.0356; whereas $1/28 = 0.0357$.

In practice, the median value of this distribution becomes arbitrarily close to $1/(\# \text{ topics})$. Therefore, within one iteration we assign a link if the observed probability

that a given author pair discusses the same topic is linked exceeds $1/(\# \text{ topics})$. In other words, it is more likely than not that the author-pair is linked. If there are 10 topics, we would expect every author-pair to have a 10% probability of being linked, *a priori*. This scheme allows network construction to adapt to changing numbers of topics.

As before, we average over multiple MCMC iterations to enable a social network to be created with weighted links, where the weight of each link is proportional to its frequency of occurrence among iterations. Nevertheless, the variability among draws from the MCMC algorithm suggests that links should not be weighted. Histograms of the distribution of these link frequency values tend to show a bimodal structure (see Figure 4) suggesting that a description of author pairs as either connected or not connected is appropriate.

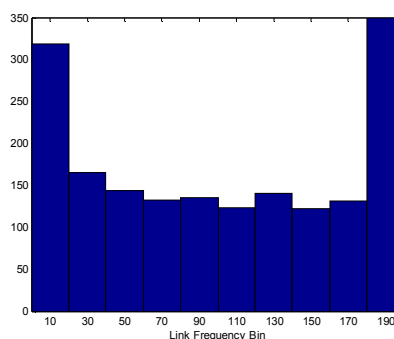


Figure 4. Sample histogram of linkage frequency for an FDA Advisory Panel meeting of April 21, 2004. The horizontal axis is the link weight (i.e., the frequency with which author-pairs are connected over 200 samples from the AT model). The vertical axis is the link frequency of links with the weight specified by the abscissa (i.e., the number of author-pairs that are connected with the frequency specified by the abscissa). Note the existence of two modes located at the extremes of the distribution.

The final challenge in constructing a network is determining where to establish the cutoff beyond which we accept that a pair of speakers is linked.

4.6.1. Bonferroni Cutoff Criterion Two authors are considered to be linked in a network if they are more likely to be connected by an edge in a given sample iteration than not. Since there are 200 samples from which a link might be inferred, we would like to establish a cutoff value that is consistent across networks. The largest committee in our sample of 37 FDA advisory panel meetings possesses 15 potential voting members (not including the committee chair). Therefore, the largest network has $15*14/2 = 105$ potential links among voting members. Each potential link must be tested in order to determine if it occurs more frequently than would be expected by chance. Lacking any prior information on link probabilities, we assume that a given speaker has no predisposition towards either linking or not linking. Therefore, we would expect that a randomly chosen pair

of speakers would be linked 100 times out of 200. We would like to know if a given pair's link frequency is higher than what we would expect under a uniform distribution across conditions of linkage and no linkage. The binomial test may be used for precisely this sort of analysis. Furthermore, given that we are testing up to 105 different independent potential links, the p-value for this test should be subject to a Bonferroni correction. Using a binomial test, and a family-wise error rate of $p=0.05$, a given author pair must be linked at least 125 times out of 200 samples to be considered more frequently linked than we would expect by chance. This is the criterion that we use for the results presented.

5. Results

Consider a graph, Δ , generated by the method outlined in chapter 4. One such graph may be generated for each of the 37 meetings that we analyze. We would like to be able to determine, on a given graph, how likely members of the same medical specialty are to be linked to one another. Suppose that graph Δ has n edges, m of which connect a pair of speakers who have the same medical specialty. We may therefore define *specialty cohesion* as m/n – the proportion of edges in graph Δ connecting members of the same medical specialty. A high specialty cohesion might indicate that members of the same medical specialty are more likely to link than are members of different medical specialties – on the other hand, it might just indicate that the meeting is homogenous – if there is very little diversity on a panel, then we might expect cohesion to be high by definition. We would therefore prefer to compare the observed specialty cohesion to the cohesion of graphs that have similar properties to Δ . We can do this by examining *specialty cohesion percentile*: For each graph, Δ , representing a meeting, 1000 random graphs (a number that has been empirically found to converge to a stable value) are generated having a number of nodes, and a graph density, equal to those found in Δ . Each node is similarly assigned a medical specialty as in Δ . Specialty cohesion is calculated for each of these random graphs, generating a meeting-specific distribution. Specialty cohesion percentile is defined as the proportion of the resultant graphs that have lower specialty cohesion than Δ .

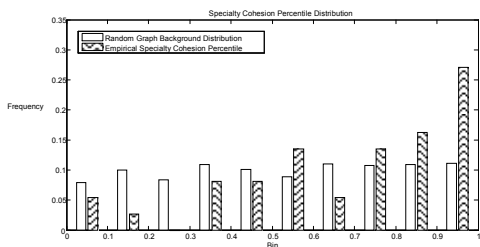


Figure 5. Histogram of Specialty Cohesion Percentiles for the 37 meetings in our sample.

Figure 5 shows the empirical distribution of specialty cohesion percentiles for the 37 meetings analyzed (textured). This is contrasted with the specialty cohesion percentile distribution for 1000 random graphs – a uniform distribution. We may see, by inspection, that the empirical specialty cohesion percentile distribution has a right skew – i.e., probability mass is concentrated near 1 and away from 0. This suggests that specialties are more likely to group together than we might expect under conditions of chance. A Kolmogorov-Smirnov test for equality of distributions finds that the empirical cumulative distribution function (CDF) is significantly less than the uniform background CDF ($p=0.0045$). These results provide support for the notion that members of the same medical specialty tend to preferentially link to one another, but not in a way that totally precludes links to other specialties.

Our experience also suggests a relation between voting behavior and linkage patterns. If people who vote the same way also share linguistic attributes, then this suggests that their attention may be directed towards something that drives their decision outcome. This further suggests the possibility of agreement on a relatively small number of reasons for either approval or non-approval. On the other hand, the absence of links between members who vote the same way suggests that there may be a high diversity of reasons for why individuals vote a certain way. In a similar manner to how we define specialty cohesion, we define *vote cohesion* as the proportion of edges in a graph that connect two panel members who vote the same way. *Vote cohesion percentile* is the proportion of random graphs, out of 1000 samples, that have lower vote cohesion than a graph representing a given meeting. There are 11 meetings in which there is a voting minority that has at least two people in it. These are used to generate a second meeting-specific distribution found (textured) in Figure 6. This is contrasted against the vote cohesion percentile distribution for 1000 random graphs – a uniform distribution.

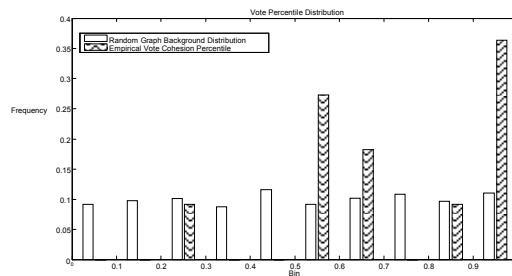


Figure 6. Histogram of Vote Cohesion Percentiles for the 11 meetings with a minority of size 2 or greater.

We may see, by inspection, that the empirical vote cohesion percentile distribution has a right skew – i.e., probability mass is concentrated near 1 and away from 0.

This suggests that people who vote alike are more likely to group together than we might expect under conditions of chance. A Kolmogorov-Smirnov test for equality of distributions finds that the empirical cumulative distribution function (CDF) is significantly less than the uniform background CDF ($p=0.015$). These results provide support for the notion that panel members who vote similarly tend to be linked.

A scatter plot of specialty cohesion percentile vs. vote cohesion percentile for the 11 meetings analyzed shows that the two quantities are correlated (Spearman rho = 0.79, $p=0.0061$; see Figure 7).

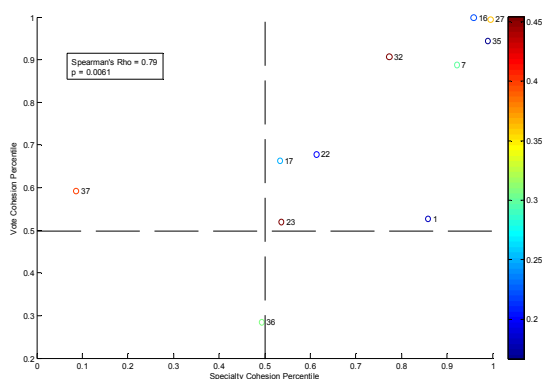


Figure 7. Scatter plot of Vote Cohesion percentile vs. Specialty Cohesion percentile for 11 meetings in which there was a minority of two or more. Datapoints are color-coded by the proportional size of the minority in each meeting, suggesting that this effect holds independent of proportional minority size.

This is a relatively tight correlation, suggesting that as specialty and voting cohesion increase together. In other words, meetings in which individuals' language links them by specialty are also meetings in which individuals' language links them by vote. Of the 11 meetings observed, 9 are located in the upper right quadrant, with high er-than-average specialty cohesion and vote cohesion, suggesting that these factors are dominant on this particular panel. The two outliers were meetings held on June 22 & 23, 2005 – both experimental cardiac devices with large committees and high data ambiguity.

It is likely that within each voting group, a relatively small number of device features might attract the attention of a number of panel members, causing them to vote a certain way for that reason. Common language could suggest a common direction of attention and perhaps common preferences. In cases of mild ambiguity, where a small number of potential interpretations of the data are possible, Douglas [1] notes that institutional membership acts to direct one's attention to a given framing of a situation or problem. This framing mechanism could potentially serve as an antecedent to preference formation. If such is the case, then a correlation between vote

cohesion percentile and specialty cohesion percentile would be expected. In these situations, the data may be difficult to interpret, e.g., due to mixed signals from a device that has a high risk but high potential reward, or sparse or ambiguous data. Under such conditions, many possible interpretations of the data might be possible within each specialty, suggesting that voters could rely on idiosyncratic beliefs. Medical specialties would have a weaker effect on an individual's perception since the data might not match any situation previously encountered. Specialty cohesion would be lower because panel members from the same specialty would have different perceptions of the data. Under these circumstances, individual expertise becomes particularly valuable, although it is unclear whose expertise is most appropriate. Panel members who vote the same way would likely do so for different reasons, thus leading to low vote cohesion.

6. Conclusion

This research is aimed at the development of a quantitative methodology that may be applied to analyze multi-actor decision-making by committees of technical experts. The methodology presented in this paper 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.

12. References

- [1] M. Douglas, *How Institutions Think*, Syracuse, New York: Syracuse University Press, 1986.
- [2] M. Rosen-Zvi, T. Griffiths, M. Steyvers, and P. Smyth, "The author-topic model for authors and documents," *Proceedings of the 20th conference on Uncertainty in artificial intelligence*, Banff, Canada: AUAI Press, 2004, pp. 487-494.
- [3] J.A. Brown, "Professional language: words that succeed," *Radical History Review*, vol. 34, 1986, pp. 33-51.
- [4] A.S. Pentland, *Honest signals: how they shape our world*, The MIT Press, 2008.
- [5] T.L. Griffiths and M. Steyvers, "Finding scientific topics," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, Apr. 2004, pp. 5228-5235.
- [6] H.M. Wallach, "Structured Topic Models for Language," Doctoral Dissertation, University of Cambridge, 2008.
- [7] Y.W. Teh, M.I. Jordan, M.J. Beal, and D.M. Blei, "Hierarchical Dirichlet Processes," *Journal of the American Statistical Association*, vol. 101, 2006, p. 1566.
- [8] M. Rosen-Zvi, T. Griffiths, P. Smyth, and M. Steyvers, "Learning Author Topic Models from Text Corpora," Nov. 2005.