A Bayesian Transition Network Topic Model for Inferring Conceptual Networks

Nicholas Beauchamp
Northeastern University
n.beauchamp@northeastern.edu

Abstract

Multiple areas of research have suggested that ideas and beliefs are interconnected in complex networks within the minds of individuals. Inferring and validating those networks remains a challenge, however. Our previous work has suggested that conceptual networks can be inferred via purpose-built experimental designs, but inferring them from naturally-occuring opinion data such as free-form text is considerably more difficult. To address this problem, we develop here a new Bayesian network topic model that can infer networks among topics from relatively short, single-author texts. This Bayesian hierarchical Transition Network Topic Model (TNTM) builds upon a Latent Dirichlet Allocation (LDA) foundation, but also infers topics for individual sentences, the network of connections between topics based on sentence transitions, and temporal and multilevel structures of author-specific networks within a larger pooled network. We test this model using political speeches and experimental data, showing both face validity, and that fundamental network features like centralization vary across individuals, ideology, subject areas and time. The potential to infer individual conceptual networks from text in a fully Bayesian framework with readily interpretable posterior distributions, on both individual and pooled levels, should allow new opportunities for studying political opinion and how conceptual connections affect discourse, deliberation, and persuasion.

Keywords: Bayesian methods, natural language processing, network analysis, topic models, American politics, political psychology, political communication

Draft 7/12/19. Prepared for Polmeth XXXVI – please do not circulate.

1 Introduction

This paper introduces a new Bayesian topic model which, in addition to inferring topics in the tradition of Latent Dirichlet Allocation (Blei, Ng and Jordan, 2003), also infers the transition matrix between topics that governs the sentence-by-sentence topical evolution of a document. On that base model we then add hierarchical structure such as documents clustered by author, topical evolution over time, multi-author multi-subject structure, and the inference of grammar and syntax using a sequence of single-word "documents." We show though a variety of applications that the inferred topics are substantively coherent despite being inferred from as few as one document, and that the transition matrices can be treated as networks that relate to many aspects of the author and their context, such as partisanship, historical period, and personality.

This model was originally motivated by our previous work inferring conceptual networks (Beauchamp, Shugars and Levine, 2019) – the ideas in people's minds and the connections between them – and is introduced below largely in that context. But the model itself has a much broader applicability to any domain where one is interested not just in the topics but in their sequential interrelationships, or when one merely wishes to be able to infer plausible topics from a small number of documents, down to a single one. The hierarchical extensions show that these transition networks can be seen as stable across individuals, subject areas, and time, while simultaneously allowing the inference of parameters at the individual level. Since the model fitting and inference is relatively straightforward, time-consuming estimation strategies are unnecessary even for the hierarchical versions, at least for moderately-sized corpora. We show here that this transition network topic model (TNTM) readily yields substantive insights into a variety of areas of political text, such as speeches and free-text essays, and we suggest that it may do so in many other domains due to its ease of implementation and the readily interpretable networks it produces.

1.1 Previous work

Our work in this area began with a previous project to infer conceptual networks at the individual level and examine the relationship between personality and ideology on the one hand, and individual network structures (such as hierarchy and order) on the other (Beauchamp, Shugars and Levine, 2019). There is a large literature across a wide array of unconnected disciplines that presumes either explicitly or implicitly that ideas and beliefs are connected in people's minds in something like a network of support or implication (Toulmin, 1958; Quillian, 1967; Shavelson, 1974; Collins and Loftus, 1975; Axelrod, 1976; Carley, 1993; Shaffer et al., 2009; Navigli and Ponzetto, 2012; Speer and Havasi, 2012; Yuste, 2015). But inferring these posited network has generally been done only at the collective (cultural) level, either using large corpora of texts and inferring simple correlations between words, or based on crude surveys of associations between terms. Even though individual networks form the basis of most of these theories, inferring networks at the individual level is a challenge. How to even measure such things or validate them? Our previous work designed a number of approaches, including (1) a dynamic graphical interface, (2) a chatbot, and (3) requesting short free-text essays from individuals from which networks were inferred using natural language processing methods. We found that (1) and (3) in particular were able to infer individual networks whose fundamental structures (such as their centralization statistic) varied systematically with personality traits, moral foundation scores, and ideology.

However, while these individual networks were varied, substantively interesting, and statistically related to other individual traits, they were in a sense too particular: the terms (nodes) are too idiosyncratic to each individual, with no easily discerned collective structure, and no way to draw power from the collective in inferring the noisy individual network. In addition, many network statistics vary with network size as well as with more fundamental structural elements, so it would be very useful to measure individual networks with a shared set of nodes to allow us to focus exclusively on the ways individuals differ in the connections between those nodes and the overall structure of their networks. The free-text approaches in (3) have a number of advantages over the interactive methods, mainly that they can be utilized in a wide variety of settings without the need for elaborate survey mechanisms like dynamic graphical interfaces or chatbots. (3) encompasses a set of related approaches, all of which have precedent in the semantic network discipline where such networks are generally inferred at a single, cultural level. In our approach 3a, we simply extract all nouns and connect them if they are within some window of adjacency; this produces robust, interpretable, and highly varied networks with statistics that relate closely to personality, but these network are very difficult to reconcile across individuals since the array of possible nouns even within a relatively constrained domain such as abortion can be quite large.

In approach 3b we ran a standard topic model, assigning each noun to one of n topics, and connecting those topics by adjacency as before. This allows for a unified set of shared nodes, but these topics are noisy due to being inferred crudely at the sentence level, and the individual networks are both highly noisy and depend on this arbitrary and theoretically under-movitated adjacency window. Figure 1 shows an example network for an individual on the topic of abortion, with many of these drawbacks evident. Thus there is a strong need here for a model that includes the power to infer a shared set of larger-scale topical nodes, but whose edges are more theoretically meaningful than just adjacency or similarity, and which generates a collective network both as a substantive interest in its own right, and in order to borrow power from the collective for the individually noisy networks.

2 Model

The model developed here has considerably broader applicability than our particular application to conceptual networks. The model emerges from the Latent Dirichlet Allocation (LDA) family of generative models first devised by Blei, Ng and Jordan (2003), where the simplest version consists of the following: each document is considered as an unordered "bag of words" and each "topic" is a distribution over words, and each document is generated

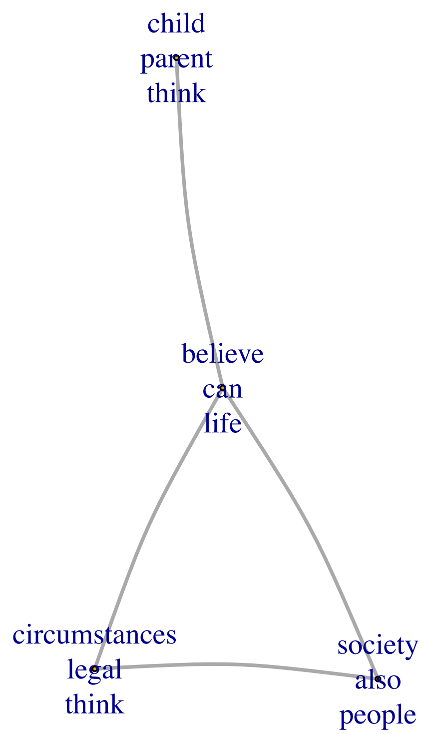

Figure 1: An individual network inferred via Model 3b from Beauchamp, Shugars and Levine (2019) applied to a set of short essays on abortion.

by first drawing a topic, and then drawing words from the distribution associated with that topic until a number of words equal to the length of the document has been generated. Subsequent variants generalize this basic model by positing topic distributions (instead of single topics) for each document; hierarchical authorship structures among documents (Grimmer, 2010); correlations among topics (Blei and Lafferty, 2007); temporal evolution of topics (Blei and Lafferty, 2006); and various others.

The model developed here builds upon that LDA foundation. We seek a model that can take relatively short pieces of text and infer not just topics, but the connections between them. The "Correlated Topic Model" (CTM) of Blei and Lafferty (2007) would seem ideal here, but for two things: first, it requires large quantities of documents in order to infer both the topics and their correlations. And second, these correlations tend to merely reflect correlations or co-occurences among topics. By contrast, here we desire something closer to the traditional "semantic network" (Quillian, 1967; Collins and Loftus, 1975; Navigli and

Ponzetto, 2012; Speer and Havasi, 2012) where speech reflects something like a walk through this network; connections are not just related topics or ideas, but specifically links that bridge ideas in a manner reflective of the thought process. A network of correlated topics would merely generate a sequence of words that stuck within the same general region of synonymy or gradually drifted from one related meaning to another. Instead, we seek something that captures the sorts of topical jumps characteristic of an individual thinking through relatively high-level ideas: something directed, temporally rapid, and reflective of associative or logical connections between ideas, and not just topical similarities.

At this more temporally granular level, a more appropriate predecessor might be the Hidden Topic Markov Model (HTMM, Gruber, Weiss and Rosen-Zvi (2007), which builds upon Mulbregt et al. (1998), Yamron et al. (1998) and Blei and Moreno (2001)). In this model, each sentence is treated as a document and is assigned a single topic, but each topic/sentence draw is not independent. However, contrary to the name, the HTMM augments independent topical draws only by positing a "stickiness" from sentence to sentence, where the speaker is likely to stick with the same topic from sentence to sentence, but when she shifts topics (with some latent probability) a new topic is drawn at random, with no inference of a Markov transition matrix among topics.

The present model, therefore, builds upon these two predecessors by adopting the sentence-level topic approach to the HTMM, but also attempting to infer the entire Markov transition matrix among topics, somewhat like the CTM. This transition matrix is then interpreted as the adjacency matrix of the conceptual network we wish to infer. Whether that interpretation is psychologically correct or not is the project of our associated validation paper; here, we focus on developing the model and applying it to a few test domains to validate the fundamental approach.

Our base model begins with a single document from which both the topics and the transition matrix is inferred. We show that, perhaps contrary to expectations, we are able to infer both sensible topics and plausible connections between them even from a single document.

We then augment this base model in a number of ways. First, and most importantly, we are fundamentally interested in inferring individual conceptual networks that also fit within a larger collectively pooled network. This is both psychologically plausible, and allows use to draw on the power of the pooled parameters in the usual manner of hierarchical models in order to infer individual networks even from very sparse textual data. Second, we describe a variant where a single transition network itself evolves, eg in the case where we have numerous documents from the same author over a long period of time. And third, reflective of the data structure in our previous experimentally-elicited data, we present a hierarchical structure where there are known *a priori* multiple unrelated subject areas with the same users writing on multiple different areas. These three extensions – authorship hierarchy, temporal drift, and subject-area clustering – reflect many of the extensions previously devised for the base LDA model, but also infer the temporal transition networks that are crucial for our application, and potentially many others.

2.1 Formal description

Formally, the base Model 1 (for a single document) is as follows:

- Each sentence s has a single associated topic t_s.
- The topic of the next sentence t_{s+1} is drawn from a transition matrix $\mathbf{T}$, where $T_{i,j}$ is the probability of transitioning to topic t_i from previous topic t_j.
- The n_s words in each sentence s are drawn from a multinomial where β_{t_s,w_k} is the probability of drawing word k given sentence topic t_s.

The generative process is:

$$
\begin{aligned}
p(\mathbf{X}_{s,\mathbf{w}}|\boldsymbol{\beta}_{t_s,\mathbf{w}}) &\sim \text{Multinomial}(n_s, \boldsymbol{\beta}_{t_s,\mathbf{w}}) \\
p(t_s|t_{s-1}) &\sim \text{Multinomial}(1, \mathbf{T}_{\cdot,t_{s-1}}) \\
p(\mathbf{T}_{\cdot,t}) &\sim \text{Dirichlet}(\boldsymbol{\alpha}_1) \\
p(\boldsymbol{\beta}_{t,\mathbf{w}}) &\sim \text{Dirichlet}(\boldsymbol{\alpha}_2)
\end{aligned}
$$

where

Data matrix $\mathbf{X}$ = Number of sentences x Number of words

Topic-word matrix $\boldsymbol{\beta}$ = Number of topics x Number of words

Topic transition matrix $\mathbf{T}$ = Number of topics x Number of topics

The model is fit using standard MCMC methods.[1] While there are many parameters to infer (on the order of 5000 for the base model, and 10 times more for the hierarchical variants), there are only two user-defined free parameters: the hyperparameters $\boldsymbol{\alpha}$. The Dirichlet prior essentially just serves to constrain the rows of $\mathbf{T}$ or $\boldsymbol{\beta}$ to sum to 1, ie constraining them to probability distributions. However, while low $\boldsymbol{\alpha}$ values generally serve as uninformative priors, values at and below 1 function like a Jeffries or Haldane prior vs the Bayes-Laplace prior in the Beta distribution: lower values tend to push the probability distributions into the corners of the simplex, essentially putting more of the mass on a few of the topics or words, while higher values do the opposite. Like the transition parameter in the HTMM, this parameter also affects the "stickiness" of topics over time. For most of the models presented here, $\boldsymbol{\alpha}$ values are set at 1 or 0.1, and while the effects of these variations are modest, there is some subjectivity in selecting the values that elicit the most substantively plausible results.[2]

[1]Models are estimated using JAGS in the R environment, with a single chain usually of 10,000 iterations due to the difficulty of establishing mixing between substantively identical chains with permuted topics. The stationarity of the single chain is verified using the Gelman-Rubin diagnostic test.

[2]The $\boldsymbol{\alpha}$ can itself of course be parameterized; when this is done, the model tends to converge on a quite low value, on the order of 0.01. However, while this may be the best Bayesian fit for the model, slightly higher values appear to produce more substantively interpretable results, though none of the results presented here differ very significantly with different alpha priors.

For the extensions of the model, we have:

2. Individual transition networks with pooled global network.

   $\mathbf{T}_i$ = Number of topics x Number of topics, per individual i

   $p(\mathbf{T}_{\cdot,i,t}) \quad \sim \text{Dirichlet}(\alpha \boldsymbol{T}_{\cdot,t})$

3. Temporal autocorrelation among transition networks.

   $\mathbf{T}_\tau$ = Number of topics x Number of topics, per time step τ

   $p(\mathbf{T}_{\cdot,\tau,t}) \quad \sim \text{Dirichlet}(\alpha \boldsymbol{T}_{\cdot,\tau-1,t})$

4. Multiple super-topics (subject areas) with individual transition networks.

   $\mathbf{T}_{1,s}$ = Number of topics x Number of topics, per individual i and subject-area s

   $p(\mathbf{T}_{\cdot,s,i,t}) \quad \sim \text{Dirichlet}(\alpha \boldsymbol{T}_{\cdot,s,t})$

All other aspects of these variant models are as in Model 1.

3 Applications and Validation

We provide four applications here. First, we apply the base model (Model 1) to a single document: the first State of the Union address by Donald Trump (technically, his inaugural address). We show that, even with just a single text where each sentence is treated as a "document," the model produces interpretable word distributions per topic, plausibly sticky topic transitions, and a readily-interpretable "conceptual network" based on the topic transition matrix. Second, we apply the multilevel, multi-author model (Model 2) to a collection of inaugural addresses from 1860 to 2016. We show that in addition to the interpretable topics and global transition matrix, topics vary plausibly with time and ideology, and the centralization scores of individual networks appear to vary with ideology as well. Third, we apply the multi-author, multi-subject model (Model 4) to our experimental data, showing that network structures vary recognizably by subject area.[3] Finally, we apply a variant of Model 1 not at the sentence level, but at the word level. Obviously discerning topics with single-word "documents" would be impossible with standard LDA models, but the latent transition matrix allows us to nevertheless discern "topics" even at the single word-per-"document" level. Notably, these "topics" appear less like the traditional subject areas of standard LDA topics and more like grammatical parts-of-speech, where the transition matrix itself constitutes the "grammar" and the generative output is akin to a familiar markov bot, but operating at the part-of-speech/grammar level. In addition to being interesting in its own right, this suggests that the more semantic sentence-level topics and transition matrix does discern something like the semantic "grammar" of thought as expressed in text.

3.1 Model 1

The first inaugural address by Donald Trump is about 600 sentences long. Table 1 shows the top loading words (highest scoring in $\boldsymbol{\beta}_{t,\mathbf{w}}$), along with the topic proportions in the document in parentheses.[4] In addition to the readily interpretable word clusters, the model gives very little weight to four of the topics (6-9; numbers and topic order are arbitrary). This is notable because a long-outstanding problem in topic modeling is selecting the number of topics. Here 10 was selected mainly because it is a readily human-comprehensible number, but the model apparently finds that only 6 are necessary. This tendency is amplified by higher α parameters, while lower values (such as 0.1 and below) tend to produce topics assignments that fill out all the available slots, but which are somewhat less interpretable. Figure 2 shows the topics assigned to each sentence in the speech, from top to bottom. As can be seen, no sentences are assigned to topics 6-9. There is also discernible both a stickiness to topics, and a slightly visible tendency to transition to adjacent topics, though

[3] We do not present an application of Model 3 here.

[4] To preprocess the corpus, stop-words are removed and the top 500 remaining words are retained; stemming is avoided because, in these contexts, apparently minor differences between words with the same stem can be quite meaningful, such as the important differences between America, American, and Americans.

while real, this visibility is due to sheer chance in the topic ordering.

Table 1: Topic words, Trump inaugural address (topic weights in parentheses).

1. (0.16) north tonight regime years states america united nuclear joined first
2. (0.14) tonight american one kenton joshua herman soldiers isis justin year
3. (0.25) us american america together must every people can america's nation
4. (0.25) new years american jobs united now last congress states trade
5. (0.10) border immigration illegal country southern system barrier areas immigrants
6. (0.01) police hours sergeant longer liberated officers away prices action recent
7. (0.01) longer syria evil reached afghanistan deserve yet barrier police try
8. (0.01) police national evil longer try crimes accountability necessary cities tell
9. (0.01) police elvin train wanted hour crisis please longer lower perhaps
10. (0.07) tax unemployment small lowest americans business reached also corey

Figure 3 shows the network generated by the transition matrix $\mathbf{T}$, with nodes colored by average time in the speech (blue earlier, red later), and only the top 5% most heavily weighted edges (cells in matrix $\mathbf{T}$) displayed.[5] The connections between nodes reflect plausible similarity between topics, such as the generic rhetoric of topics 3 and 4 or the foreign policy orientation of 1 and 5. The nodes are placed according to the Fruchterman-Reingold algorithm (Fruchterman and Reingold, 1991) which, like most placement algorithms, places more centrally-connected nodes in the center. Thus the centrality of 3 and 4 – the most frequent, empty-rhetoric topics – is reflected in the plot and the network itself. Since edges are directed, though, we discern more than simple similarity: nodes 1-4 tend to be sinks, node 10 is a source, and 5 is a transitional topic. These characteristics are of course dependent on the weight thresholding, but in general reflect the mean weighted in-degree and out-degree of the nodes for the full transition matrix. Topic 10 (economics) is one Trump discussed briefly at the beginning of the speech and then rarely thereafter, whereas topic 5 (immigration) is in some ways the transitional lynchpin of the speech, connecting the first, second and third "acts" of the speech together, as evident in Figure 2. More broadly, the network shows not just the connections between similar topics, but the "grammar" of their connections: which lead into which as Trump (or rather, his speechwriter) moves through the main ideas of the address.

[5]The motivation for this choice is explained below.

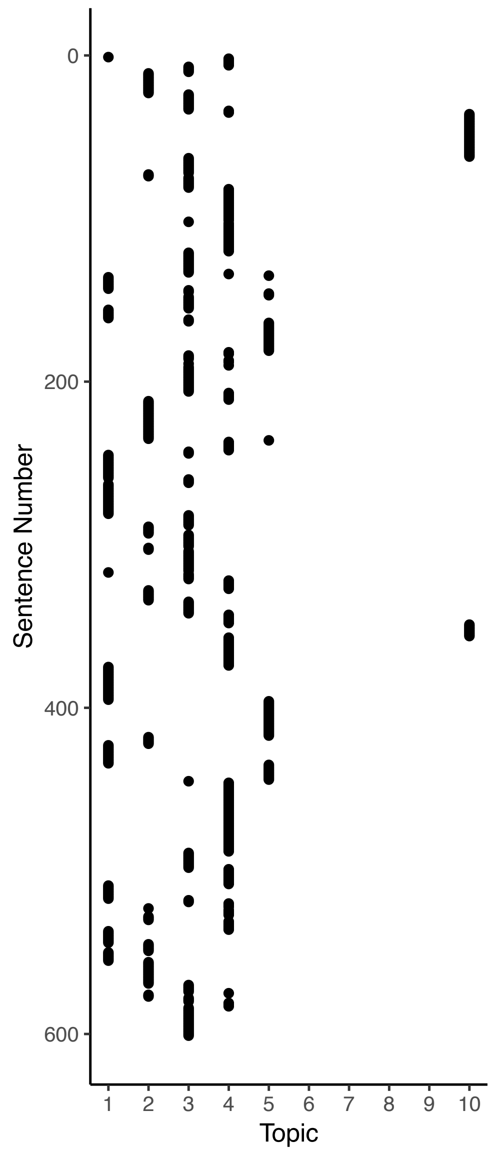

Figure 2: The topics employed sequentially in the first Trump SOTU address.

3.2 Model 2

Although there would appear to be face validity to the topics, timeline, and network inferred via Model 1 from the single Trump document, fitting a model with approximately $K^2 + K * W$ parameters $(10 * 10 + 10 * 500)$ based on a document-term matrix of similar scale (600 * 500 with $\sim$ 4500 non-zero values) must produce a fairly noisy $\mathbf{T}$ (the parameter set of central interest here). Because the $\mathbf{T}$ parameters are constrained by the Dirichlet prior to be $\in [0, 1]$, we cannot simply check whether credible intervals contain 0. Figure 4 shows a histogram of

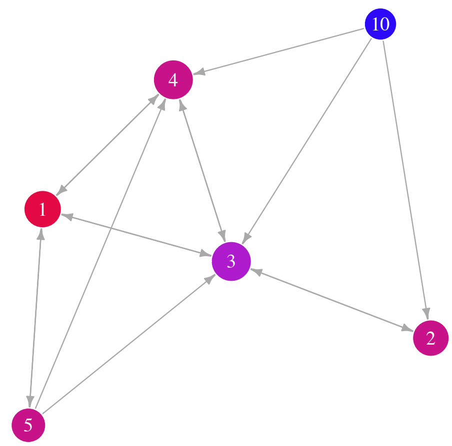

Figure 3: Inferred topic transition network for the first Trump SOTU address.

(logged) mean parameter values in $\mathbf{T}$ for a 40-topic model, which better illustrates the typical parameter distribution than the smaller and noisier 100-parameter/10-topic model. As can be seen, and as appears characteristic of this family of models, there is a large distribution of mean values on the order of 0.01, but there is also a right tail of larger values distinct from this distribution. Akin to the "elbow" test in principal component analysis or the Wigner semi-circle test (Arnold, 1971), we take the values to the right of the noise-like distribution as likely significant values, which often amount to 5-10% of the distribution, which leads to the 5% cutoff heuristic for the network figures. However, while one can quibble about this decision, it mainly affects the visual network representation rathe than something more fundamental, and insofar as we employ the network for downstream statistical analysis, that analysis can be checked for robustness by varying the threshold. More fundamentally, there is good reason to think that even with these relatively long documents, the inference of $\mathbf{T}$ for a single author may be quite noisy, and certainly that is the case for the third application here where the documents are only a few sentences long. Therefore there is a need – the same as that which motivates many hierarchical approaches – to be able, in a multi-author environment, to estimate both individual $\mathbf{T}_i$ matrices as well as a global pooled $\mathbf{T}$ from which each $\mathbf{T}_i$ may draw power.

We begin by estimating Model 2 using a collection of inaugural addresses from 1860 to

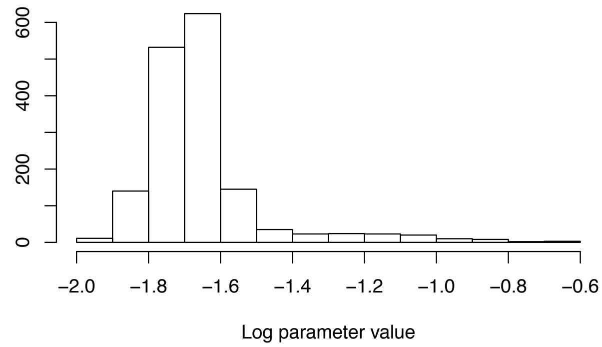

Figure 4: Distribution of (log 10) mean parameter values for 40-topic Trump transition matrix.

2016 (Woolley and Peters, 2008). Table 2 shows the words of a 10-topic model, Figure 5 shows the topic plot across all concatenated sentences, and Figure 6 shows the global pooled network, with nodes again colored by time and sized by topic prevalence. Note here that all the available topic slots are used, and increasing the topic number continues to do so, suggesting that unlike the single-author case, the "natural" number of topics may be larger. In this case, the topics inferred appear strongly correlated with historical time, with topics 4 and 7 early ("government") while 5 ("us, world, america") prevalent later. The network likewise shows the dominance of time, suggesting that perhaps this timespan is too long to infer anything other than shifting topical preferences over the centuries.

Table 2: Topic words, Inaugural Addresses 1860-2016 (topic weights in parentheses).

1. (0.00) address fellow-citizens presence oath office constitution equally problem states
2. (0.02) god bless thank america oath states united president taken office
3. (0.02) president mr vice fellow chief citizens senator friends justice hand
4. (0.13) government business upon congress policy public revenue country tariff every
5. (0.33) us new let world america nation must time people together
6. (0.01) race feeling employees negro interstate south passed sympathy among
7. (0.19) government people constitution states upon laws shall public law can
8. (0.00) helped failure efforts money lies can values perils comes minds
9. (0.11) life things new great upon men must people us leadership
10. (0.17) world peace nations free freedom can must peoples shall human

This is a useful illustration of the limits of combining both within-document sentence sequence and across-document long-term sequences when the latter spans such long periods. To reduce this effect, we next estimate Model 2 using only addresses since 1950, which

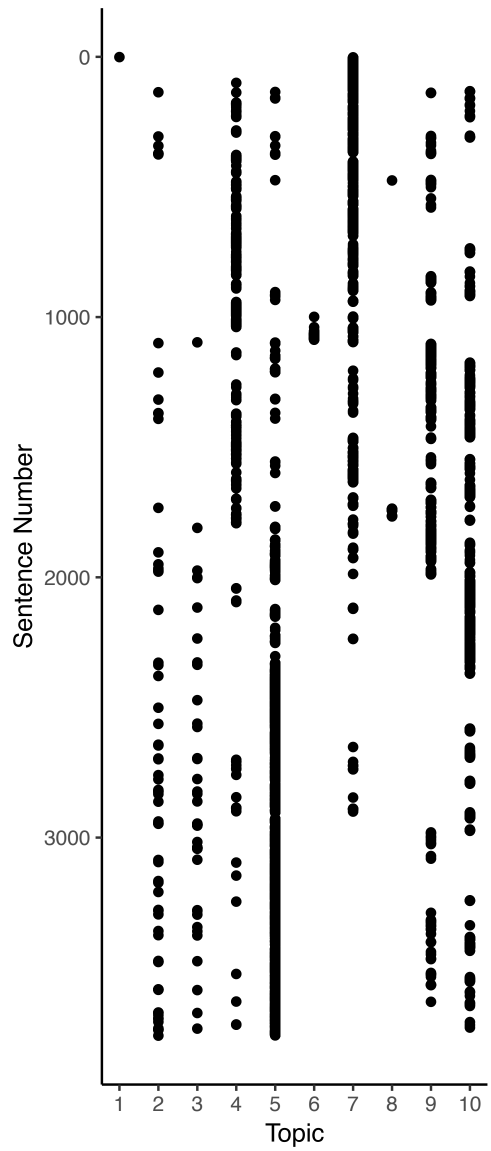

Figure 5: Inaugural Addresses 1860-2016

produces a more diverse and less temporally dominated set of topics, with words shown in Table 3. Figure 7 shows two different node colorings for the global network **T** – colored by time (top left) and by mean partisanship (top right). As can be seen, there is a notable correlation between partisanship and time, with the exception of topic 6, which is both later in the time period and more Democratic. More importantly, while there are slight tendencies for nodes to cluster by time period and partisanship, the network overall is much less dominated by either than is the 1860-2016 network. This is important because is shows

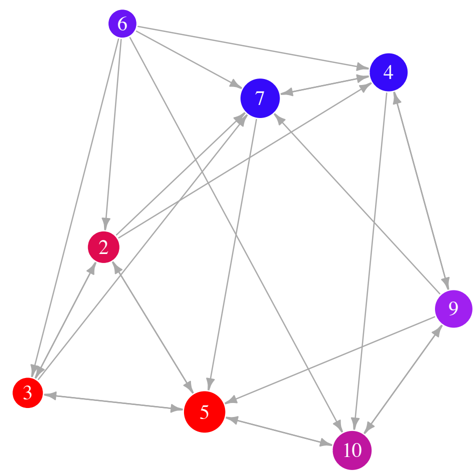

Figure 6: Inaugural Addresses 1860-2016

that, in addition to topical similarity, partisanship, and long-term shifts, the network also infers the sorts of short-term, within-speech transitions that we are most interested in. In particular, the bland rhetoric of topic 6 is the linchpin much as it was in the Trump speech alone, and we can also discern common rhetorical patterns such as the opening verbiage in topic 3 (which appears mainly as a source node in the directed network) and also topic 2, suggesting a general tendency to open with the economy (at least in the modern era) before getting to the redder meat.

Table 3: Topic words, Inaugural Addresses 1950-2016 (topic weights in parentheses).

1. (0.08) faith law arms last years equal country creed free force
2. (0.06) americans loyalty work enterprise products loyal millions heal values taxes
3. (0.07) president mr bless vice god fellow oath chief thank years
4. (0.10) story day america better right laws small generation care land
5. (0.10) weapons america every nuclear freedom rights body mind human america's
6. (0.22) america let nation us promise forward go one spirit government
7. (0.06) let us sides new abroad begin merely things independence away
8. (0.14) nations economic can free security effort world united peace strive
9. (0.10) us act future middle upon dream knowledge freedom afford others
10. (0.08) shall together need states government federal congress problems action days

But the key innovation in Model 2 is to be able to infer a $\mathbf{T}_i$ for each unique speaker that is both particular to them, but also draws on the power of the global $\mathbf{T}$. We theorize, for instance, that individuals largely all draw upon the same pool of ideas, but weight them and connect them differently. Thus Figure 7 shows two individual networks fit within the same model, with nodes positioned the same as the nodes in the global network and colored by partisanship as in the top right figure. The connects are relatively similar to the global network, showing the pooling effect, but also differ in significant ways, as well as in node/topic weights. Obama gives more weight to Democratic topics like 9 ("future") while Trump gives more attention to Republican topics 4 and 5 ("story"; "weapons"). More interestingly, the networks also show that in addition to different topic weights, they are also connected differently, with Obama's network more resembling the hub-and-spoke layout of the global network, while Trump's network places heavier connections on the more Republican nodes with less apparent structure.

Although these descriptive accounts of the network are plausible, they are neither systematic nor statistically grounded. In our previous work (Beauchamp, Shugars and Levine, 2019) we examine in great detail how network structures vary with personality and demographic traits, and find strong correlations between ideology, moral foundations, and network structure as measured by an array of different statistics. Here, we simply examine perhaps the most fundamental network measure: whether it is hierarchical, centralized, and ordered, vs whether is it more homogeneously connected. We take as our basic measure of this the centralization of the network, which is higher when there are core and periphery nodes, and lower when all the nodes are more homogeneously connected.[6] Figure 8 shows the centralization measures for Republicans (blue) and Democrats (red)[7]. While the n is too small for a firm statistical test, as can be seen, by either measure of centralization, the Democratic networks appear to be more centralized than the Republican networks, as we saw with Obama vs Trump.[8] Whether this says something deep about the psychological organization of ideas among Democrats and Republicans is the focus of our related project; here, we mainly note that this new method allows the individual $\mathbf{T}_i$ to retain their signficant individual variation

[6]In our previous work, however, we find that a wide variety of network statistics measuring structure vs homogeneity are all strongly correlated with each other.

[7][Sorry, just noticed the colors were flipped – will fix soon.]

[8]Note that these network statistics presume an unweighted network, which depends on the threshold used to convert the full transition matrix to an adjacency matrix. However, we find that this result is robust to a variety of thresholds from 5% to 25%.

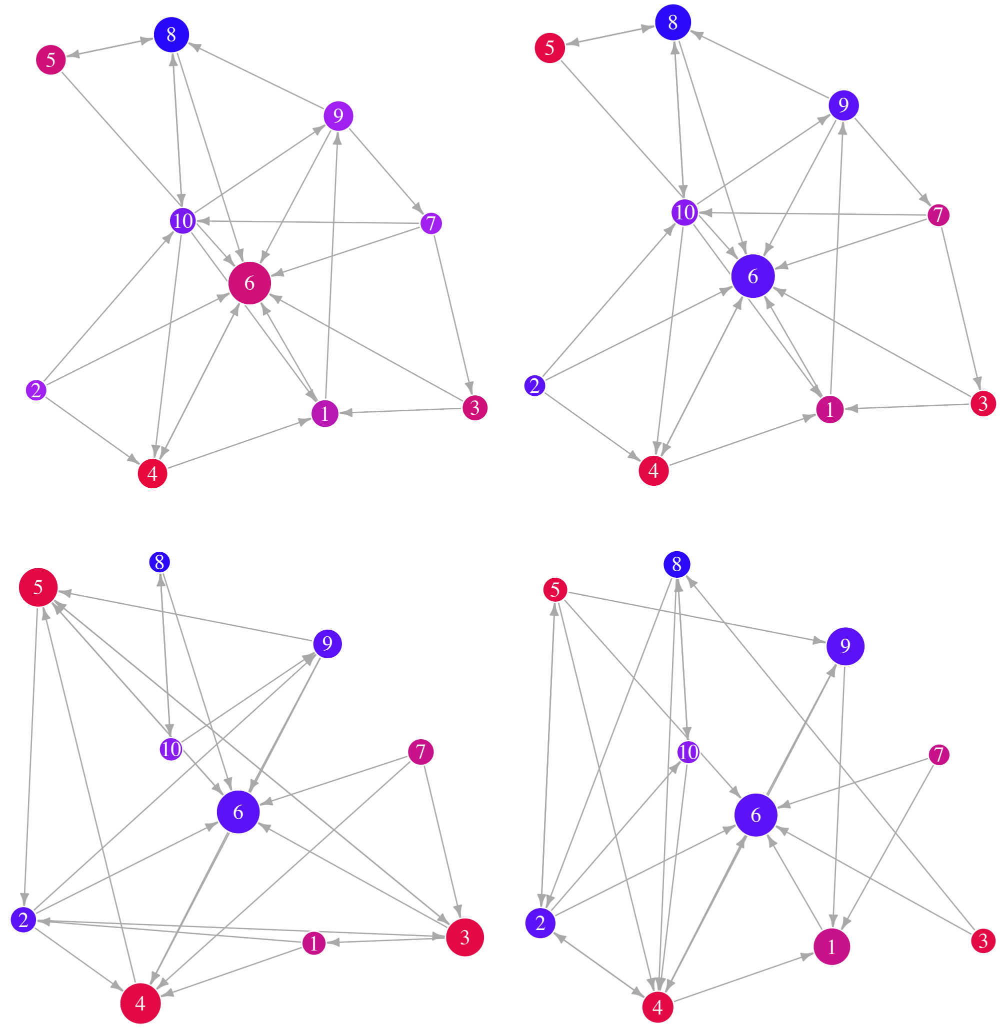

Figure 7: Top: Inaugural Addresses 1950-2016, with nodes colored by time (top left) and party (top right). Bottom: Inaugural addresses for single authors: Trump (bottom left) and Obama (bottom right); node color by party as per top right. Node side by topic prevalence for all figures.

even while drawing on the pooled topics and global network.

3.3 Model 4

As a third example, we apply Model 4 to our experimentally-elicited data.[9] In this experiment, 100 subjects were asked their views on abortion, childrearing philosophy (which is an important predictor of authoritarianism (Feldman and Stenner, 1997)), and healthcare

[9]We do not present an application of Model 3 here; the implementation is straightforward.

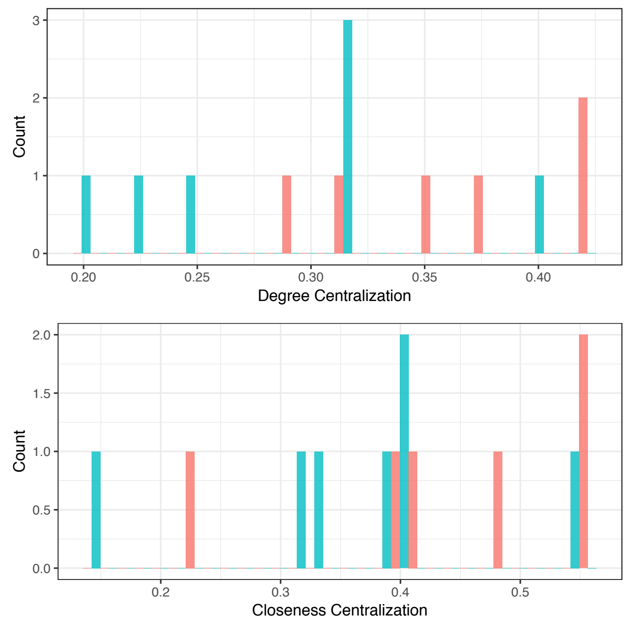

Figure 8: Inaugural Addresses 1950-2016, centralization by party (Democrats red, Republicans blue.)

policy. They were also asked to draw a cognitive network; to participate in a chatbot conversation that elicited keywords, which they then connected together; and to produce a short, 50-word essay on their views on the topic. The drawback of the first two approaches is that they allow for no alignment of individual networks into any global whole, which both limits statistical power, and causes issues if, for instance, we believe that different individuals may interpret an "idea" at different levels of particularity or aggregation. Additionally, if we theorize that individuals have consistent network structures across different subject areas, we would ideally create a model that allows for that pooling within individuals across different areas and texts, as well as across individuals in a global subject-level network.

Table 4 shows the model output with topic words for the three subject areas. Note again that the individual documents are quite short and of only a few sentences. Nevertheless, the model is able to infer sensible topics for all three, with high loadings on the words and issues primed by the subject question (eg, the standard childrearing prompt asks about the tradeoffs between fostering curiosity and independence on the one hand, and obedience and manners on the other). In this case, because the individual documents are so short, they are heavily influenced by the global transition matrix: Figure 9 shows an example of the global network for abortion, as well as examples of networks from a conservative and a liberal subject. The individual networks are much more similar to the global network here than with the much longer inaugural address dataset, yet their few differences show important ideological tendencies: the conservative example emphasizes the more conservative topics (7, 10) which are also more centrally connected, while the liberal example emphasizes 8 in particular ("woman, body, choice"). The liberal example is more like the global network because the subject pool (Mechanical Turk) is generally biased left relative to the US as a whole. Figure 10 shows similar examples for childrearing and healthcare. Childrearing differs little by ideology, though the conservative example emphasizes curiosity in particular much less, consistent with the standard authoritarian model. Healthcare is also less polarized than abortion in the terminology used by subjects, but again we can see how the liberal example differs from the global network in particular via their emphasis on federal responsibility – the topic is not more prevalent in their essay, but is far more central in the network.

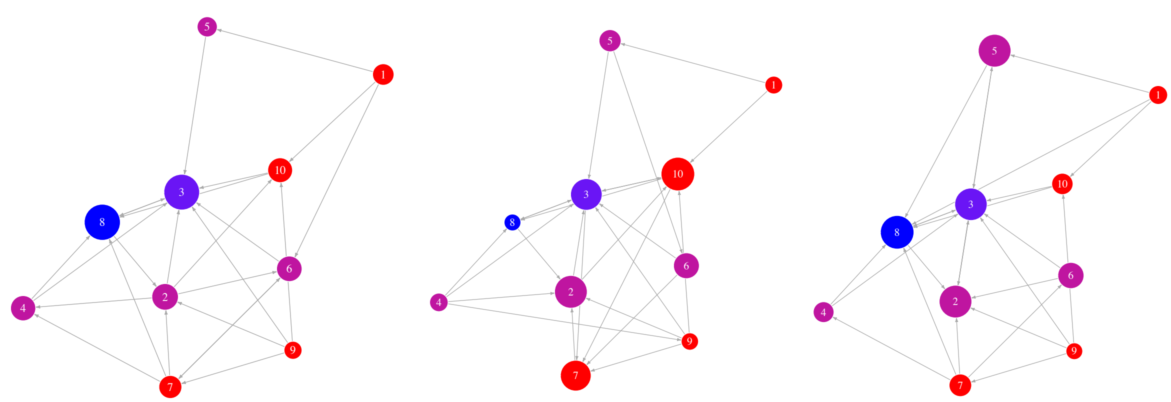

Figure 9: Abortion: Global, Conservative Example, Liberal Example. Nodes colored by mean ideology of all users of that topic.

In our previous work we have found that network statistics differ by personality and ideology, but Figure 11 shows that for the simple centralization measure here, there is no consistent difference by ideology, but there is a difference by subject area, with more centralized (hub and spoke) structures related to healthcare. Why this should be the case we can only speculate, but it may be that concepts on this area tend to be either more sparsely

Table 4: Topic words across three subject areas, all jointly estimated via Model 4.

Abortion

1. (0.06) child children mother mental health
2. (0.09) baby medical mother really parent
3. (0.22) legal circumstances certain abortion
4. (0.08) life child responsibility dont also
5. (0.05) women decision made one illegal
6. (0.08) burden control birth society access
7. (0.06) choice decisions months thing male
8. (0.23) woman body choice pregnant want
9. (0.04) cases extreme rape result incest
10. (0.08) think sex person never moral

Childrearing

1. (0.06) family question change really kids
2. (0.06) told subject even button people
3. (0.09) well need child obedient better
4. (0.16) curiosity important independence obedience manners
5. (0.06) kids learn like much want
6. (0.16) manners good respect obedience
7. (0.04) someone things forward rather make
8. (0.20) children curiosity can allows think
9. (0.03) lead traits child avoid child's
10. (0.13) obedience manners good person

Healthcare

1. (0.04) creating service well idea medicare
2. (0.05) life quality provide healthy individual
3. (0.10) can go take medical without
4. (0.08) rights citizens many can cancer
5. (0.05) states countries united one cost
6. (0.22) americans responsibility federal care government
7. (0.06) one purchase like support interest
8. (0.04) believe just healthy individuals society
9. (0.14) free think healthcare expensive help
10. (0.22) care coverage health can take

connected, or more directly emergent from a few core principles such as healthcare as a right vs cost control.

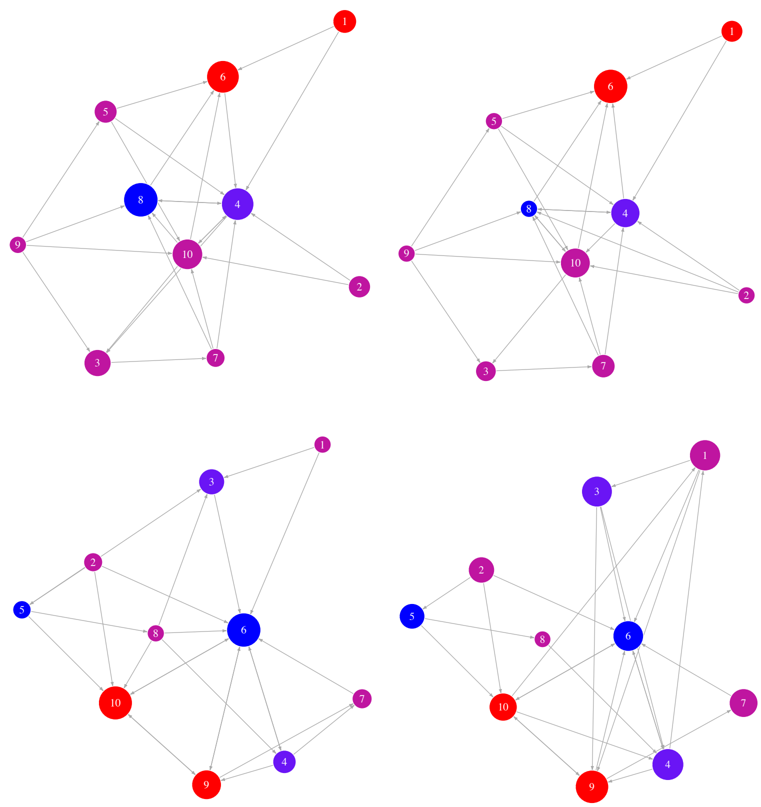

Figure 10: Top: Childrearing. Global (top left); Conservative Example (top right). Bottom: Healthcare. Global (bottom left); Liberal Example (bottom right).

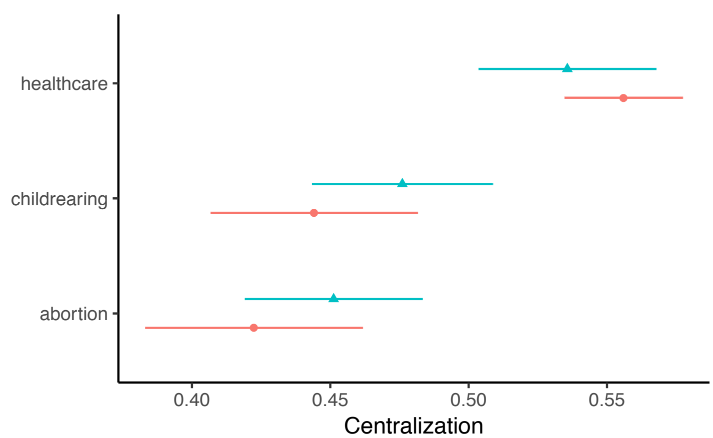

Figure 11: Centralization varies by topic but not by ideology.

3.4 Grammar

We have emphasized our interpretation that these networks represent not just the correlations between topics, but something like the "grammar" of thought: the directed edges are conceptual links that capture the flow of thought and reasoning from one idea to the next, which may bridge entirely different domains, eg as values or facts lead to a set of conclusions or policies. At the relatively high level of topics, this interpretation is fairly hard to validate substantively, even though there is plenty of statistical evidence that the transitions matrix values are significant and not at all symmetric.

To provide some support for the claim that this model does indeed infer something like the "grammar" of semantic thought, we provide one final application that shows that the model can in fact infer literal grammar quite readily. In this example, we take the same Trump address as before, but instead of ~ 600 sentences as our "documents," we employ all ~ 5000 words each as their own document, yielding a 5000x500 document-term matrix which is 0 for all but one value in each row (ie, a "one-hot" encoding in the lingo of natural language processing). The model is exactly the same as in Model 1, but obviously here "topics" cannot be inferred by word co-occurence at all, but instead all clustering information happens via the transition matrix that is simultaneously being inferred. Nevertheless, the resultant "topics" are quite recognizable. Table 5 shows the top words from a 20-topic model, which we have subjectively clustered according to very recognizable patterns, particularly among the first three highest loading words of each topic. The model, based on a single document, appears to infer almost all of the major parts of speech: verbs, conjunctions, adverbs, adjectives, pronouns, prepositions, nouns, and even punctuation. If these are the constituent parts of grammar, the network shown in Figure 12 shows the "rules" of this grammar. Walking through this network produces rudimentary markov-bot-like speech,[10] but without the n-gram-level approach common in those models. Figure 13 shows a similar network for a word-level 40-topic model of Obama's first inaugural, which infers similar grammatical patterns, but produces much more Obama-sounding pseudo-speech if you walk randomly though it.[11]

[10] Eg, for Trump: "we have together of americans. also keep our opportunity in when lowest building american government plans was to building of our almost year - and the better also create ever. and immigrants as our city. victory come: love, we should it during an us. "

[11] Perhaps unsurprisingly, we subsequently discovered that a similar model was proposed in the computational linguistics community (Goldwater and Griffiths, 2007; Christodoulopoulos, Goldwater and Steedman, 2010), oriented towards the unsupervised inference of parts of speech. The model here differs in a number of

These results suggest that when applied to a domain where we know for certain that there is a strong temporal, markov structure (grammar) the model is indeed able to infer both the relevant clusters as well as the "rules" (transition matrix) for combining those clusters. This lends credence to our interpretation that at the semantic level, the model is also inferring something like the rules for thought and the network connecting the constituent elements.

Table 5: Topic words, word-level model, Trump inaugural address (sorted manually).

Verbs, to be	Pronouns
1. (0.05) be been know have asking	13. (0.05) we i they he who
15. (0.03) is - was] has	6. (0.12) the our a their this
5. (0.05) are have will can must	
	Prepositions
Conjunction	19. (0.1) of in for to on
17. (0.04) and but as " if	
12. (0.04) and or including where were	Punctuation/Conjunction
7. (0.03) - this it that tonight	2. (0.08) . , that . of
	3. (0.06) , - has american since
Adverbs	
18. (0.04) to not also never always	Nouns
	4. (0.1) country states people world nation
Adjectives	9. (0.02) congress time going right you
14. (0.05) new american united great same	11. (0.03) one all tonight more members
8. (0.02) last in two after tax	20. (0.04) us america $ freedom home
10. (0.04) more come it work together	
	Time
	16. (0.03) years year ago months decades

4 Conclusion

We have presented here a Bayesian model for inferring topics and the markov transition matrix governing the order of those topics in a document. We have also presented extensions of this model to hierarchical settings with multiple texts and subject areas per author, as well as showing an application that infers grammar from single-word "documents." We interpret the inferred transition matrices as conceptual networks, and show both here and more elaborately in Beauchamp, Shugars and Levine (2019) that these networks vary in their

essential ways: the semantic, sentence-level approach; the MCMC inference with the full posterior parameter distributions; the hierarchical structures; and in a number of other minor details.

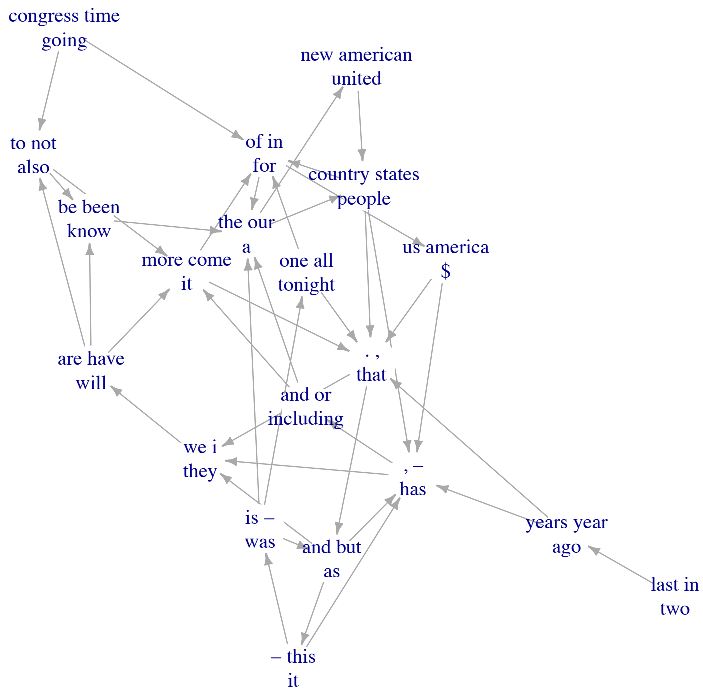

Figure 12: Trump's grammar.

structure by individual. But whether or not one invests in this interpretation of the transition matrix, the model stands on its own and adduces important structural information about the micro-scale sequential ordering tendencies of topics. The extensive literature employing the HTMM shows the need for such temporal structure at the level of single documents (Titov and McDonald, 2008; Jameel and Lam, 2013; Yan et al., 2013; Wang et al., 2017), even though the HTMM itself does not fully provide that. Knowing how topics lead into each other is essential for understanding both the temporal dynamics of documents, and the latent structures among topics that go beyond simple word and topic clustering. Furthermore this transition matrix, as we saw in the grammatical example, allows highly coherent topics to be inferred even from very short documents. The hierarchical structure follows quite easily from the base model, with no need for complex variational approaches as long as the corpus is relatively small – yet even from a small corpus, meaningful topics and matrices can be inferred, even at the level of individual authors who may not have generated much text each individually. Finally, if one does accept our interpretation the word clusters are

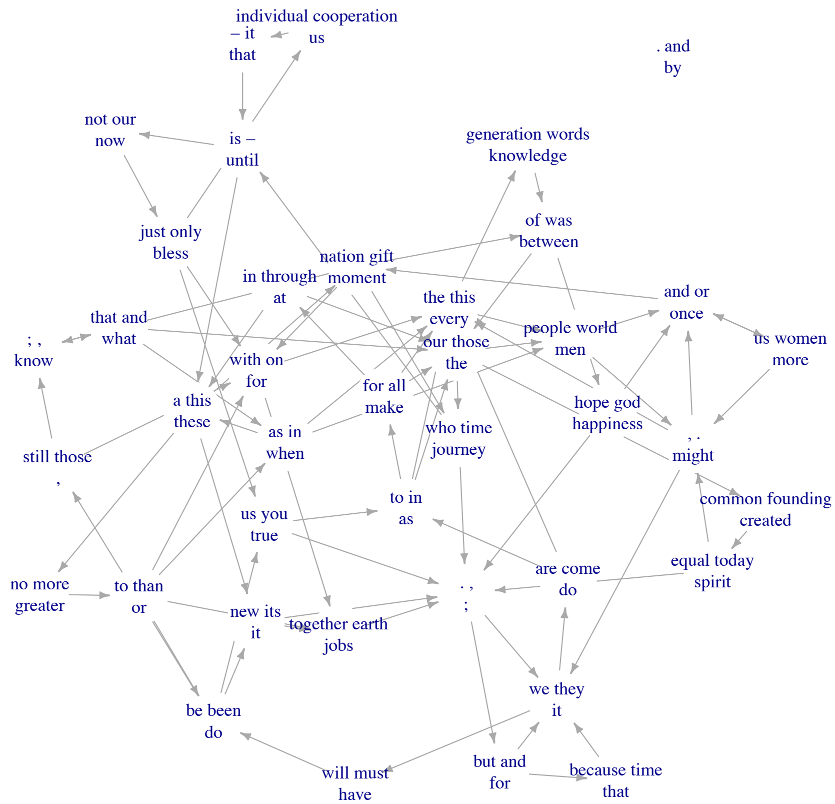

Figure 13: Obama's grammar.

akin to "ideas" at some level of granularity, and that the rules for transitioning between these is akin to a "grammar" of thought, then there remains much important work to be done extending this family of models to map the complex, multilevel network of political thought; to explore how those networks varying with ideology and personality; and to test how network-based conceptual models might inform age-old questions of opinion formation and political behavior.

References

Arnold, Ludwig. 1971. “On Wigner’s semicircle law for the eigenvalues of random matrices.” *Probability Theory and Related Fields* 19(3):191–198.

Axelrod, Robert. 1976. *Structure of decision: The cognitive maps of political elites*. Princeton university press.

Beauchamp, Nicholas, Sarah Shugars and Peter Levine. 2019. “Mapping Conceptual Networks.”.

Blei, David M, Andrew Y Ng and Michael I Jordan. 2003. “Latent dirichlet allocation.” *Journal of machine Learning research* 3(Jan):993–1022.

Blei, David M and John D Lafferty. 2006. Dynamic topic models. In *Proceedings of the 23rd international conference on Machine learning*. ACM pp. 113–120.

Blei, David M and John D Lafferty. 2007. “A correlated topic model of science.” *The Annals of Applied Statistics* pp. 17–35.

Blei, David M and Pedro J Moreno. 2001. Topic segmentation with an aspect hidden Markov model. In *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM pp. 343–348.

Carley, Kathleen. 1993. “Coding Choices for Textual Analysis: A Comparison of Content Analysis and Map Analysis.” *Sociological Methodology* 23:75–126.

Christodoulopoulos, Christos, Sharon Goldwater and Mark Steedman. 2010. Two decades of unsupervised POS induction: How far have we come? In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics pp. 575–584.

Collins, Allan M and Elizabeth F Loftus. 1975. “A spreading-activation theory of semantic processing.” *Psychological review* 82(6):407.

Feldman, Stanley and Karen Stenner. 1997. “Perceived threat and authoritarianism.” *Political Psychology* 18(4):741–770.

Fruchterman, Thomas MJ and Edward M Reingold. 1991. “Graph drawing by force-directed placement.” *Software: Practice and experience* 21(11):1129–1164.

Goldwater, Sharon and Tom Griffiths. 2007. A fully Bayesian approach to unsupervised part-of-speech tagging. In *Proceedings of the 45th annual meeting of the association of computational linguistics*. pp. 744–751.

Grimmer, Justin. 2010. “A Bayesian hierarchical topic model for political texts: Measuring expressed agendas in Senate press releases.” *Political Analysis* 18(1):1–35.

Gruber, Amit, Yair Weiss and Michal Rosen-Zvi. 2007. Hidden topic markov models. In *Artificial intelligence and statistics*. pp. 163–170.

Jameel, Shoaib and Wai Lam. 2013. An unsupervised topic segmentation model incorporating word order. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*. ACM pp. 203–212.

Mulbregt, Paul van, Ira Carp, Lawrence Gillick, Steve Lowe and Jon Yamron. 1998. Text segmentation and topic tracking on broadcast news via a hidden Markov model approach. In *Fifth International Conference on Spoken Language Processing*.

Navigli, Roberto and Simone Paolo Ponzetto. 2012. "BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network." *Artificial Intelligence* 193:217–250.

Quillian, M Ross. 1967. "Word concepts: A theory and simulation of some basic semantic capabilities." *Systems Research and Behavioral Science* 12(5):410–430.

Shaffer, David Williamson, David Hatfield, Gina Navoa Svarovsky, Padraig Nash, Aran Nulty, Elizabeth Bagley, Ken Frank, Andre A Rupp and Robert Mislevy. 2009. "Epistemic network analysis: A prototype for 21st-century assessment of learning." *International Journal of Learning and Media* .

Shavelson, Richard J. 1974. *Journal of Research in Science Teaching* 11(3):231–249.

Speer, Robert and Catherine Havasi. 2012. Representing General Relational Knowledge in ConceptNet 5. In *LREC*. pp. 3679–3686.

Titov, Ivan and Ryan McDonald. 2008. A joint model of text and aspect ratings for sentiment summarization. In *proceedings of ACL-08: HLT*. pp. 308–316.

Toulmin, Stephen E. 1958. *The uses of argument*. Cambridge University Press.

Wang, Lu, Nick Beauchamp, Sarah Shugars and Kechen Qin. 2017. "Winning on the merits: The joint effects of content and style on debate outcomes." *Transactions of the Association for Computational Linguistics* 5:219–232.

Woolley, John T and Gerhard Peters. 2008. "The American presidency project." *Santa Barbara, CA. Available from World Wide Web: http://www. presidency. ucsb. edu/ws* .

Yamron, Jonathan P, Ira Carp, Larry Gillick, Steve Lowe and Paul van Mulbregt. 1998. A hidden Markov model approach to text segmentation and event tracking. In *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'98 (Cat. No. 98CH36181)*. Vol. 1 IEEE pp. 333–336.

Yan, Xiaohui, Jiafeng Guo, Yanyan Lan and Xueqi Cheng. 2013. A biterm topic model for short texts. In *Proceedings of the 22nd international conference on World Wide Web*. ACM pp. 1445–1456.

Yuste, Rafael. 2015. "From the neuron doctrine to neural networks." *Nat Rev Neurosci* 16(8):487–497.