Topics and Trends in Cognitive Science (2000-2017)

Anselm Rothe[1,*] (anselm@nyu.edu)
Alexander S. Rich[1,*] (asr443@nyu.edu)
Zhi-Wei Li[2,*] (zhiwei.li@nyu.edu)
[1]Department of Psychology, [2]Center for Neural Science
New York University

Abstract

What are the major topics of the Cognitive Science Society conference? How have they changed over the years? To answer these questions, we applied an unsupervised learning algorithm known as dynamic topic modeling (Blei & Lafferty, 2006) to the 2000–2017 Proceedings of the Cognitive Science Society. Unlike traditional topic models, a dynamic topic model is sensitive to the temporal context of documents and can characterize the evolution of each topic across years. Using this model, we identify historical trends in the popularity of topics over time, and shifts in word use within topics indicative of changing focuses within the field. We also measure the correlation across topics, and use the model to highlight the topic structure of particular papers and labs. We believe dynamic topic models present an important tool towards understanding Cognitive Science as it continues to grow and evolve over time.

Keywords: topic models; trends; scientometrics; cognitive science

From August 13th to 16th in 1979 the first conference of the Cognitive Science Society took place. The conference program listed talks by 42 researchers, grouped into five categories in a single track: cognitive science and education, psychology of categorization, human development, language processing, and belief systems. In comparison, last year's conference in 2017 had 255 talks grouped into 54 categories that ran in 11 parallel tracks. The 11-fold increase in categories and tracks is modest evidence for the increasing complexity of the field of cognitive science.

To shed more light on the evolution of topics within the field, we used a dynamic topic model to analyze the raw text of the annual Proceedings of the Cognitive Science Society.

Dynamic Topic Models

Topic modeling is an approach to unsupervised text understanding in which documents are posited to be generated by a set of underlying word distributions known as topics. Topic models have been used successfully to capture structure in large text corpora and make these corpora more human-understandable (Blei et al., 2003). However, traditional topic models do not capture the temporal ordering of documents, thus not directly modeling how topics change over time. Dynamic topic models address this issue by allowing the distribution of words in topics to change as a function of time (Blei & Lafferty, 2006). For example, a topic devoted to articles about communication could give high probability to words related to fax machines 20 years ago, but shift to include more words related to the Internet today.

[*]All authors contributed equally to this work.

Related work

Topic models have been used to understand the structure of academic publishing in the past. The archives of the journal Science have been used as a test dataset for dynamic topic models (Blei & Lafferty, 2006), as well as other topic model extensions (Blei & Lafferty, 2007). Cohen Priva & Austerweil (2015) applied topic modeling to the field of cognitive psychology, using the archives of the journal Cognition. However, they used a static topic model, and visualized the change of topics over time in an ad-hoc manner using the empirical frequency of words within each topic. Thus, the present project represents the first attempt to directly model changes in the field of cognitive science over time.

Problem definition and algorithm

In standard topic modeling, also known as Latent Dirichlet Allocation (LDA; Blei et al., 2003), the goal is to infer a fixed set of latent topics underlying a corpus of documents. Let $\beta_{1:K}$ be K topics, each of which is a multinomial distribution over a fixed vocabulary, and let α be a hyperparameter that governs the topic-distribution of documents. For each document, we assume that topic proportions θ are drawn from $Dirichlet(\alpha)$. We then assume that each word in the document is drawn with topic assignment $z \sim Mult(\theta)$ and identity $w \sim Mult(\beta_z)$.

In LDA, the time at which a document was published would have no effect on word distributions of its underlying topics. In the dynamic topic model (DTM), we relax the assumption that $\beta_{1:K}$ are fixed over all time points. Instead, we replace β_k with $\beta_{t,k}$, denoting the word distribution of topic k at time t. This means that DTM allows for the words within a given topic to change over time.

To model the drifting of β over time, we represent β in an unconstrained space described by the natural parameters of the multinomial. We assume that the terms of β drift over time according to Gaussian noise,

$$\beta_{t,k}|\beta_{t-1,k} \sim \mathcal{N}(\beta_{t-1,k}, \sigma^2 I)$$

The function π then maps β back to a standard representation of a multinomial,

$$\pi(\beta_{k,t})_w = \frac{\exp(\beta_{k,t,w})}{\sum_w \exp(\beta_{k,t,w})}$$

This means that the complete generative process assumed by the DTM is:

1. Draw topics $\beta_t|\beta_{t-1} \sim \mathcal{N}(\beta t-1, \sigma^2 I)$.
2. For each document, choose topic proportions θ from $Dirichlet(\alpha)$.
3. For each word in each document:
   (a) Choose a topic assignment $Z \sim Mult(\theta)$.
   (b) Choose a word $W \sim Mult(\pi(\beta_{t,z}))$.

Once the generative model has been defined, variational inference is used to infer β, as well as θ for each document.

We note that in a more complete model the prior over topic proportions α could vary across topics and over time, and could be inferred from the data along with β. However, in the existing implementation of variational inference for the DTM, α is fixed, presumably to decrease computational complexity. Thus our analyses assume a fixed α across topics and over time.

Method

Data

We downloaded 6920 PDF files from the Cognitive Science Conference archives, representing submissions from 2000 to 2017[1]. In general, each submission is a 6 page paper. We converted each entire PDF to text using an automated pdftotxt utility. We tokenized the text based on whitespace, and removed lines in which few tokens were English words, because these lines tended to contain equations. We also removed words that were less than four characters long and were not in a standard English dictionary, as these were often produced by errors in the PDF parser. We then lemmatized the words to standardize pluralizations and verb tenses. Finally, we removed tokens that occurred in fewer than 36 documents (i.e., 2 documents per year on average), as well as tokens that occurred in more than 50% of documents. Our final vocabulary contained 9710 words.

Model fitting

We used a modified version of the Blei lab's C implementation of the dynamic topic model (https://github.com/blei-lab/dtm) which uses a variational inference algorithm to estimate an approximate model posterior from data. Following Blei & Lafferty's analysis of Science we assumed 20 topics[2], and chose $\alpha = .05$ (using $1/(\text{num. topics})$ as a rule of thumb). The model used for our qualitative results was fit using all 6920 pdf files from all 18 years. To determine optimal topic variance parameter (σ^2), we fit a series of DTM models with $\sigma^2 = \{.0001, .0003, .001, .003, .01\}$ up to 2017, and evaluated them by determining the negative log likelihood assigned to the 2017 documents.

[1]For years before 2000, only large PDF files concatenating all papers from a year were available and not one PDF per paper.

[2]We found that the log-likelihood of held-out data actually improved slightly up through 100 topics, the highest number tested, but remained with a 20 topic model for reasons of computational costs and human interpretability.

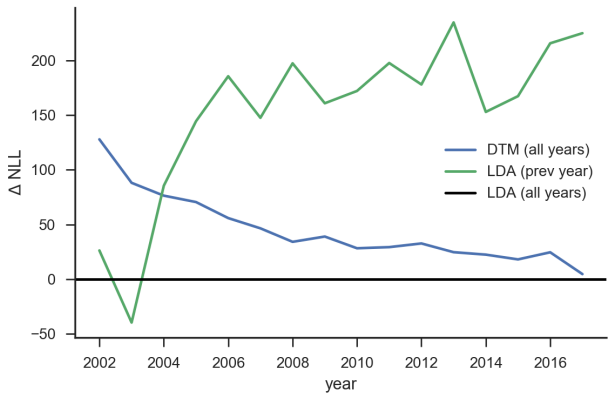

Figure 1: Model performances on the next year's data, after training on previous years. The y-axis shows negative log-likelihoods (i.e., lower score is better) relative to the LDA (all years) model (i.e., the black base line). LDA (all years) and DTM (all years) were trained on all data up to the held-out year. LDA (prev year) was only trained on the year prior to the held-out year.

Model comparison

To evaluate the DTM compared to LDA, we fit a series of models in which we trained the model up to a given year, and then inferred the negative log-likelihood of the data for the given year. We compared the DTM with optimal σ^2 trained on all data up to the held-out year, LDA trained on all data up to the held-out year, and LDA trained only on the last year of data before the held-out year. The DTM for each year was initialized using the inferred topics from the LDA fit on the same training set, to maximize comparability across conditions. We conducted this procedure for each year from 2002 to 2017.

Results

Model evaluation

We found that the optimal value of σ^2 was .001, slightly lower than the value of .005 used by Blei & Lafferty (2006) in their DTM analysis of Science. This optimal value was used for all of our other fits of the DTM model. The results of testing the DTM and LDA models on a held out year are displayed in Figure 1. Because the all-years LDA model tended to perform best over all, the results of the DTM and of the previous-year LDA model are shown in relation to this model. We found that with fewer years, the DTM performs far worse than LDA. However as the number of training years increases, the DTM's relative performance gradually improves, becoming roughly equal to the static LDA model by 2017. LDA trained on only one previous year, in contrast, gradually loses ground as the the last year's data becomes a smaller proportion of the

training set.

Our finding that the DTM performs poorly with few years, and improves with a temporally broader training set, comports with the idea that the DTM is a more complex model that can capture temporal variation but can also over-fit training sets with little temporal structure. However, this result differs from that of Blei & Lafferty (2006), who found that the DTM outperformed LDA strongly when trained on only a few early years of Science, and that LDA's relative performance improved over time. It will be interesting to see if the performance of the DTM relative to LDA continues to improve as the training set of Cognitive Science Proceedings increases in future years.

Topics and trends

Topics in DTM are generated in an unsupervised way and thus do not naturally come with a meaning, but it is useful to name these topics in order to talk about them. There are at least two ways to interpret a topic: by looking at the terms with highest weights in the topic-term distribution (β), and the papers with highest weights on this topic (θ). The interpretations from these two perspectives should agree with each other.

Our procedure of deciding topic labels proceeded as follows. First, we looked into the most frequent words in each topic cluster. For example, the most frequent term for topic 17 is "probability", followed by "distribution", "parameter" and "prior"—all are typical in a probabilistic modeling research. Second, we looked into the most typical papers in each topic, formally defined as the paper with the highest proportion on the given topic among all papers in the same year. We checked their titles and keywords to confirm our intuition. Through this method we manually labeled all the 20 topics, and these labels are used throughout.

Trends in CogSci Using the DTM's inferred topic-year-word distributions (β) and topic-document distributions (θ), we created several visualizations to understand how the field of Cognitive Science has changed over the last two decades.

Figure 2 shows the overall proportion of each of the 20 topics over the last 18 years. Since we assumed a flat prior on topic proportions (α), these proportions were estimated empirically by averaging the topic proportions in all documents in a given year, and then smoothing the curve using a LOESS regression. Some topics have remained fairly stable over the years. Others have become much more or less popular. The topic *Probabilistic modeling*, for example, has more than doubled in popularity to become the most popular topic. *Decision making* has become much more popular as well. The topic *Neural network*, in contrast, has decreased in popularity from its earlier heyday, but is starting to show a resurgence.

Trends in the topic *Neural network* Here we provide a detailed study of one specific topic, the *Neural network* topic, which seems to contain subtopics with different trends across

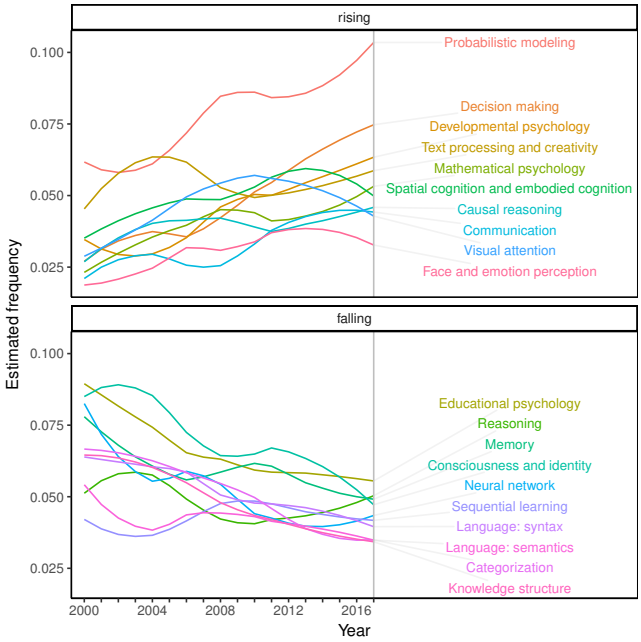

Figure 2: Rising and falling popularity of each topic over the last 18 years. The upper panel shows topics with a rising trend, defined by a higher estimated frequency in 2017 than in 2000, and the lower panel shows all remaining topics.

years. This can be demonstrated by the changing theme in typical papers across years. For each year, the paper that had the largest topic proportion (θ) for this topic, was selected as the most typical. When inspecting these papers, there seems to be a shift over the years in the predominant theme of the topic. From 2000 to 2005, most typical papers are shown in the upper half of Table 1. The overarching theme seems to be connectionist neural models depicting cognitive processes. This differs from the typical papers in the last five years, shown in the lower half of Table 1 where the models become more and more biologically focused and related to neuroscience studies on neurons and circuits.

The shift within the *Neural network* topic is also noticeable in the trends of particular words relevant for the topic. This can be shown with the terms whose weights increased or decreased most over the years (see Figure 3).

A similar analysis also applies to other topics. For example, we found in the *Probabilistic modeling* topic, words such as "Bayesian," "fit," "prior," and "sample" show the most increasing trends, which may indicate that Bayesian methods have become more prominent over time.

Similarities of topics Figure 4 shows a visualization of the similarity structure of the topics. The similarity between two topics is obtained by correlating their document vectors (of θ values), where a higher correlation indicates a more similar scoring pattern across documents. The complete topic-by-topic correlation matrix was projected into two dimensions using the R package `qgraph` (Epskamp et al., 2012).

Table 1: Each year's most typical paper for the topic *Neural network*. Up to 2005 the theme is more tuned towards artificial intelligence (e.g., connectionism), after 2012 more towards neuroscience (e.g., neural circuits).

Year	Title
2000	Representing Categories in Artificial Neural Networks Using Perceptual Derived Feature Networks
2001	Neural Synchrony Through Controlled Tracking
2002	Preventing Catastrophic Interference in Multiple-Sequence Learning Using Coupled Reverberating Elman Networks
2003	A Split Model to Deal with Semantic Anomalies in the Task of Word Prediction
2004	A Neural Model of Episodic and Semantic Spatiotemporal Memory
2005	A Connectionist Implementation of Identical Elements
2012	How many Neurons for your Grandmother Three Arguments for Localised Representations
2013	Simultaneous unsupervised and supervised learning of cognitive functions in biologically plausible spiking neural networks
2014	Learning and Variability in Spiking Neural Networks
2015	Lateral Inhibition Overcomes Limits of Temporal Difference Learning
2016	Improving with Practice A Neural Model of Mathematical Development
2017	A Plausible Micro Neural Circuit for Decision-Making

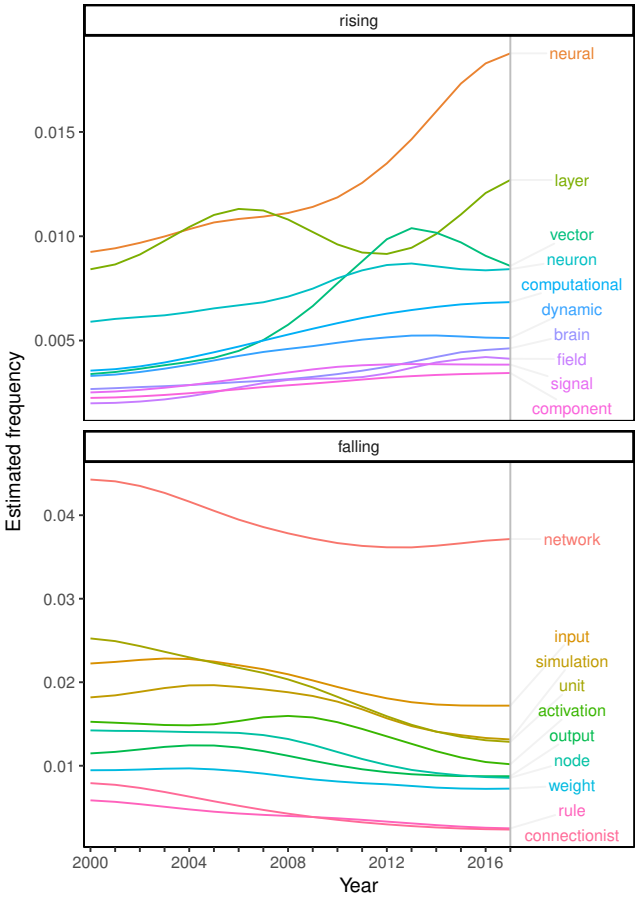

Figure 3: Trends within the *Neural network* topic. The upper panel shows the ten words with largest increase in frequency from 2000 to 2017, as estimated by the DTM. The lower panel shows the ten words with the largest decrease. We can see that the words becoming less popular in this topic are more closely related to connectionist neural networks (e.g., connectionist, nodes, input, output) while the more rising words are more related to biologically relevant neural models (e.g., brain, neuron), and maybe also to deep learning models (e.g., layer, vector).

Due to the fact that all θs for a given document must add to 1 and are thus in competition, the topics were naturally slightly anti-correlated at -0.05, as determined by shuffling topic-assignments within documents. We set this value to be the baseline and rescaled all correlation coefficients, resulting in the line widths in Figure 4. Interestingly, of all topics, *Educational psychology* had both the strongest similarity with another topic, namely *Text processing and creativity*, as well as the strongest dissimilarity, with *Probabilistic modeling*.

Characterizing a lab or author in the topic space DTM gives us a method not only for identifying trends in the whole field but also a reference frame to characterize a subset of documents of interest. Here we provide an example of this kind of analysis, namely locating a lab in the topic space by averaging the topic proportions of all the documents produced by this lab. We chose the Computation and Cognition Lab at New York University (our database had 29 out of 32 CogSci publications listed on http://smash.psych.nyu.edu/papers.php), and the Stanford Language and Cognition Lab (with 58 out of 59 publications listed on http://langcog.stanford.edu/). The averaged topic proportion across all papers in each lab are shown in Figure 5, which agrees with our intuition for these labs' themes. Worth noting is that this analysis is specific to publications in CogSci only. It may well be a lab has other directions of study that are not published in CogSci and are therefore not visible in this analysis.

Recommending interesting papers We can also use the model to identify papers that are similar to a target paper, based on the cosine angle between their topic vectors. Since our model is fitted to all years at once, the similarity search is not constrained to papers that were published in the same year as the target paper. For demonstration[3], we took a recent CogSci publication from one author of the present paper: Rothe et al. (2016). Table 2 shows the most similar papers, which come from the same, earlier and later years then the pa-

[3]See `https://anselmrothe.github.io/dtm/` for a web-based, interactive version of our recommendation system.

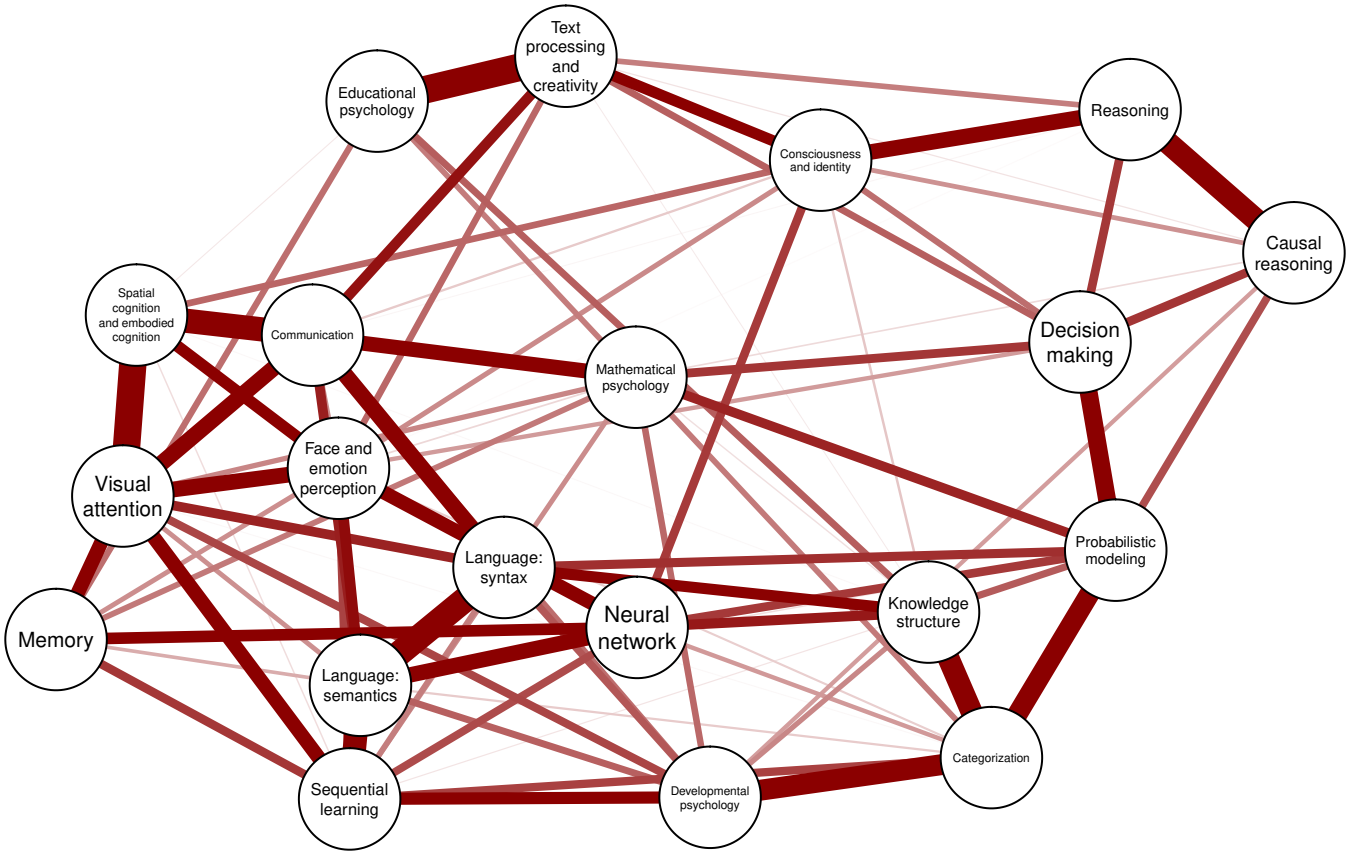

Figure 4: Similarity structure of the topics. Stronger links between two nodes indicates a stronger similarity of the topics.

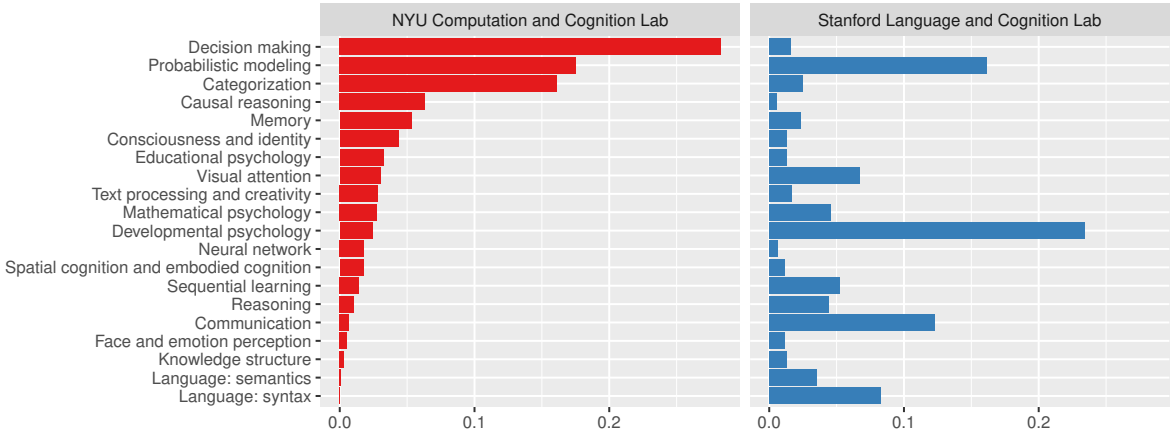

Figure 5: Dominant topics in two selected labs, the Stanford Language and Cognition Lab and the NYU Computation and Cognition Lab. For ease of comparison, the topics were ordered by their score for the NYU lab.

Table 2: Papers similar to *Asking and evaluating natural language questions* (Rothe et al., 2016).

Cosine	Year	Title
0.986	2016	The distorting effect of deciding to stop sampling
0.983	2013	Non-parametric estimation of the individuals utility map
0.980	2016	Searching large hypothesis spaces by asking questions
0.978	2017	A computational model for decision tree search
0.977	2010	Cognitive Models and the Wisdom of Crowds A Case Study Using the Bandit Problem

per itself, and which have cosine angles close to 1, indicating a strong similarity. Interestingly, on a first glance the match of the titles is not striking at all. For example, words from the titles such as "sampling", "utility", or "computational models" do not seem very close to "natural language questions" from the target paper's title. However, knowing the content of the target paper, these words fit neatly to the account on question asking pursued in this article.

Discussion

We applied a dynamic topic model (DTM) to the last 18 years of Cognitive Science Society Proceedings. The model inferred 20 distinct topics, which appear to reasonably correspond to different subfields of cognitive science. We applied the DTM to two directions of analysis. First, we identified historical trends in the whole field as well as the detailed evolution of a specific topic. Second, we characterized individual documents in the topic space, which enabled comparing different labs' topic orientations as well as recommending papers of similar topic composition.

It is worth noting that the interpretations we offer are only one of many ways to look at the field of cognitive science and simplify its complexity. Specifically, the topics we identified are dependent on stochasticity in the dataset and limitations in data preprocessing and model fitting and specification. Future work can extend our analyses to yield new insight in a number of ways.

DTM improvements Our model assumes a fixed number of 20 topics. Future models could be more flexible and estimate the best fitting number of topics, perhaps better capturing human intuitions about the subdivisions of the field (see, e.g., Griffiths & Steyvers, 2004). Also, our model assumed a uniform prior over topic frequencies. In future work, a prior should be estimated from the data that can reflect the distribution of topic frequencies (i.e., not assuming not all topics are equally likely), and that can evolve over time (Blei & Lafferty, 2006). This would allow one to measure the changing popularity of topics in a more Bayesian manner. Finally, simply allowing more time to pass (and more data, in the from of CogSci proceedings, to be collected) may continue to improve the performance of the DTM relative to LDA.

LDA extensions Adding temporal drift is one way to extend traditional topic models. But it is not the only extension that might capture interesting patterns in the field of cognitive science. Correlated topic models (Blei & Lafferty, 2007) can explicitly account for topics that tend to occur together in the same article, providing a formal avenue to understanding the kinds of connections between topics visualized in Figure 4. The document influence model (Gerrish & Blei, 2010) is an extension of the DTM that integrates the idea that some articles have a greater influence than others on future articles' word composition. Using this approach, one could identify influential papers in an automated way, and investigate whether this measure of influence maps well to citation count or tends to be associated with mixtures of previously less-connected topics. The approach could also be combined with that of Leydesdorff & Goldstone (2014) to determine how document influence relates to the citation of journals outside the field.

Acknowledgments We thank everyone in the NYU Computation and Cognition Lab for helpful discussions.

References

Blei, D. M., & Lafferty, J. D. (2006). Dynamic topic models. *Proceedings of the 23rd international conference on Machine learning - ICML '06*, 113–120.

Blei, D. M., & Lafferty, J. D. (2007). A correlated topic model of science. *The Annals of Applied Statistics*, *1*(1), 17–35.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022.

Cohen Priva, U., & Austerweil, J. L. (2015). Analyzing the history of cognition using topic models. *Cognition*, *135*, 4–9.

Epskamp, S., Cramer, A., Waldorp, L., Schmittmann, V., & Borsboom, D. (2012). qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software, Articles*, *48*(4), 1–18.

Gerrish, S., & Blei, D. M. (2010). A language-based approach to measuring scholarly impact. In *Icml* (Vol. 10, pp. 375–382).

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*, *101*, 5228–5235.

Leydesdorff, L., & Goldstone, R. L. (2014). Interdisciplinarity at the journal and specialty level: The changing knowledge bases of the journal cognitive science. *Journal of the Association for Information Science and Technology*, *65*(1), 164–177.

Rothe, A., Lake, B. M., & Gureckis, T. M. (2016). Asking and evaluating natural language questions. In A. Papafragou, D. Grodner, D. Mirman, & J. Trueswell (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society.* Austin, TX.