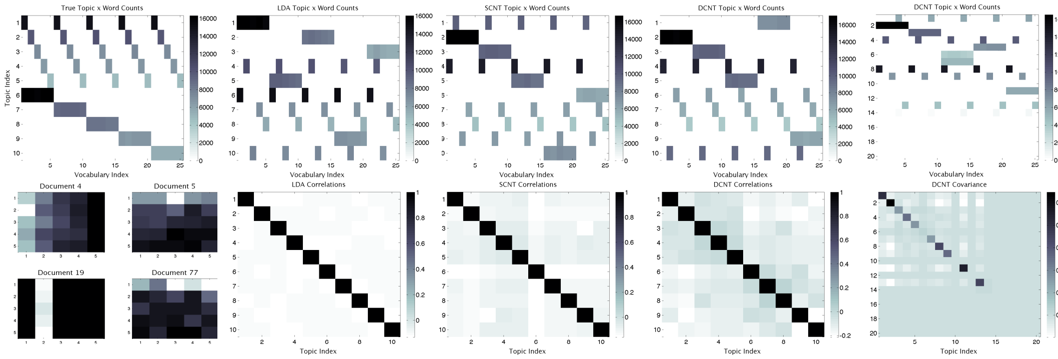

Figure 2: A dataset of correlated toy bars (example document images in bottom left). *Top:* From left to right, the true counts of words generated by each topic, and the recovered counts for LDA ($K = 10$), SCNT ($K = 10$), DCNT ($K = 10$), and DCNT ($K = 20$). Note that the true topic order is not identifiable. *Bottom:* Inferred topic covariance matrices for the four corresponding models. Note that LDA assumes all topics have a slight negative correlation, while the DCNT infers more pronounced positive correlations. With $K = 20$ potential DCNT topics, several are inferred to be unused with high probability, and thus have low variance.

topics recover the correct topics. With $K = 20$ topics, the DCNT recovers the true topics, as well as a redundant copy of one of the bars. This is typical behavior for sampling runs of this length; more extended runs usually merge such redundant bars. The development of more rapidly mixing MCMC methods is an interesting area for future research.

To determine the topic correlations corresponding to a set of learned model parameters, we use a Monte Carlo estimate (details in the supplemental material). To make these matrices easier to visualize, the Hungarian algorithm was used to reorder topic labels for best alignment with the ground truth topic assignments. Note the significant blocks of positive correlations recovered by the DCNT, reflecting the true correlations used to create this toy data.

4.2 NIPS Corpus

The NIPS corpus that we used consisted of publications from previous NIPS conferences 0-12 (1987-1999), including various metadata (year of publication, authors, and section categories). We compared four variants of the DCNT model: a model which ignored metadata, a model with indicator features for the year of publication, a model with indicator features for year of publication and the presence of highly prolific authors (those with more than 10 publications), and a model with features for year of publication and additional authors (those with more than 5 publications). In all cases, the feature matrix ϕ is binary. All models were truncated to use at most $K = 50$ topics, and the sampler initialized as in Sec. 4.1.

4.2.1 Conditioning on Metadata

A learned DCNT model provides predictions for how topic frequencies change given particular metadata associated with a document. In Figure 3, we show how predicted topic frequencies change over time, conditioning also on one of three authors (Michael Jordan, Geoffrey Hinton, or Terrence Sejnowski). For each, words from a relevant topic illustrate how conditioning on a particular author can change the predicted document content. For example, the visualization associated with Michael Jordan shows that the frequency of the topic associated with probabilistic models gradually increases over the years, while the topic associated with neural networks decreases. Conditioning on Geoffrey Hinton puts larger mass on a topic which focuses on models developed by his research group. Finally, conditioning on Terrence Sejnowski dramatically increases the probability of topics related to neuroscience.

4.2.2 Correlations between Topics

The DCNT model can also capture correlations between topics. In Fig. 4, we visualize this using a diagram where the size of a colored grid is proportional to the magnitude of the correlation