

# MAD Style: Multivalent Authorship Detection (MAD) Topic Models

David Dohan, Charles Marsh, Shubhro Saha, Max Simchowitz  
Princeton University, Department of Computer Science

## Goals

- Classify author writing style in a wide range of media.
- Extract compact representation of stylistic tendency.
- Determine which features are most indicative of writing style.

## Introduction

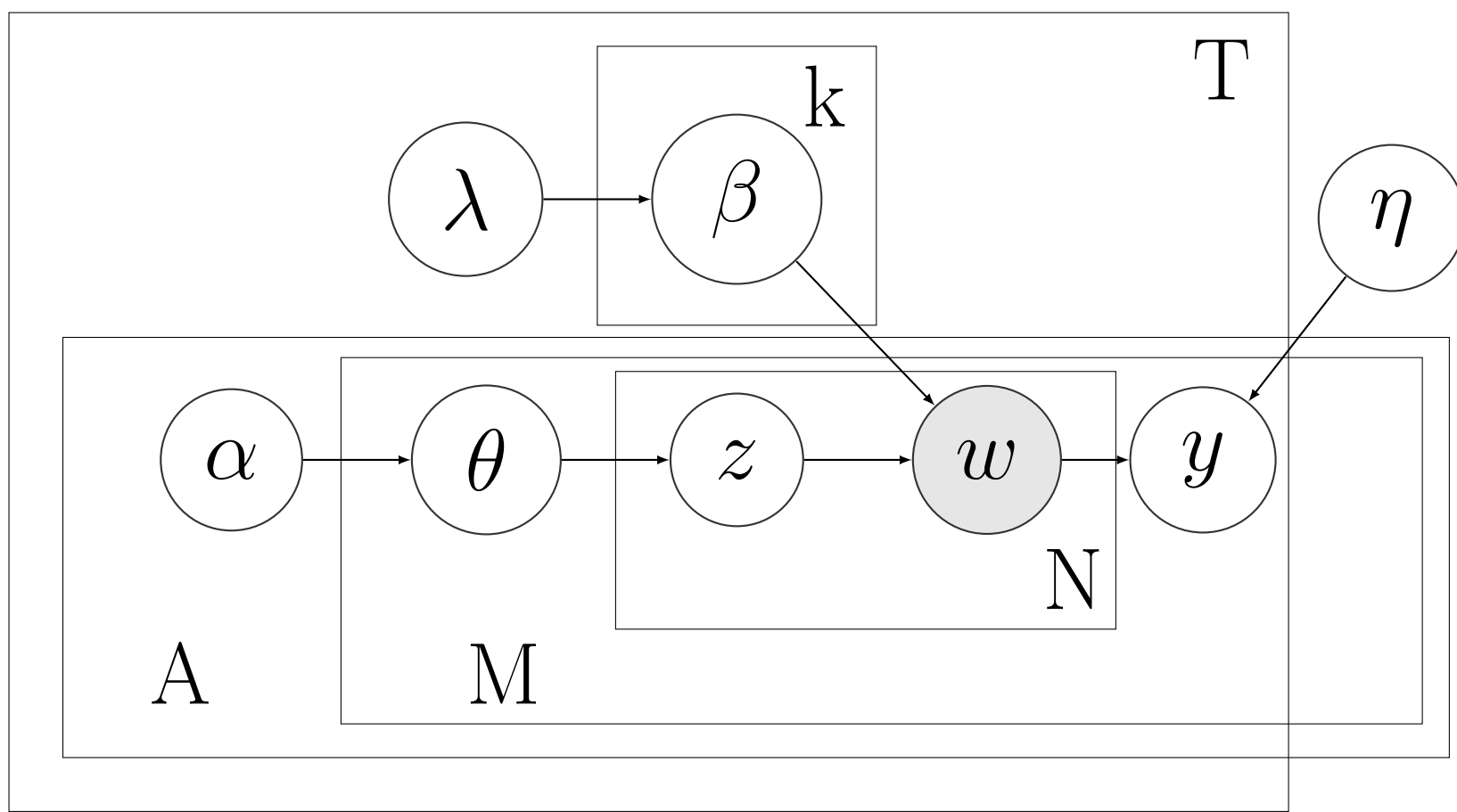
In the *authorship detection* problem, one is given:

- A set of documents labeled (by author) on which to train.
- A set of anonymized documents to classify.

Methods for authorship detection traditionally depended on careful feature extraction and rather black-box methods. Hence, they rely on extensive domain specific knowledge, and can be difficult to decipher. Here, we present the *MAD Topic Model*, which uses syntactic and stylometric  $n$ -gram features (e.g., part-of-speech tags, meter). MAD fits separate topic models to each of these  $n$ -gram vocabularies, and then combines the models with a multiclass logistic regression classifier. After fitting the topic model parameters, new data can be classified using the multiclass component.

INSERT GRAPHIC

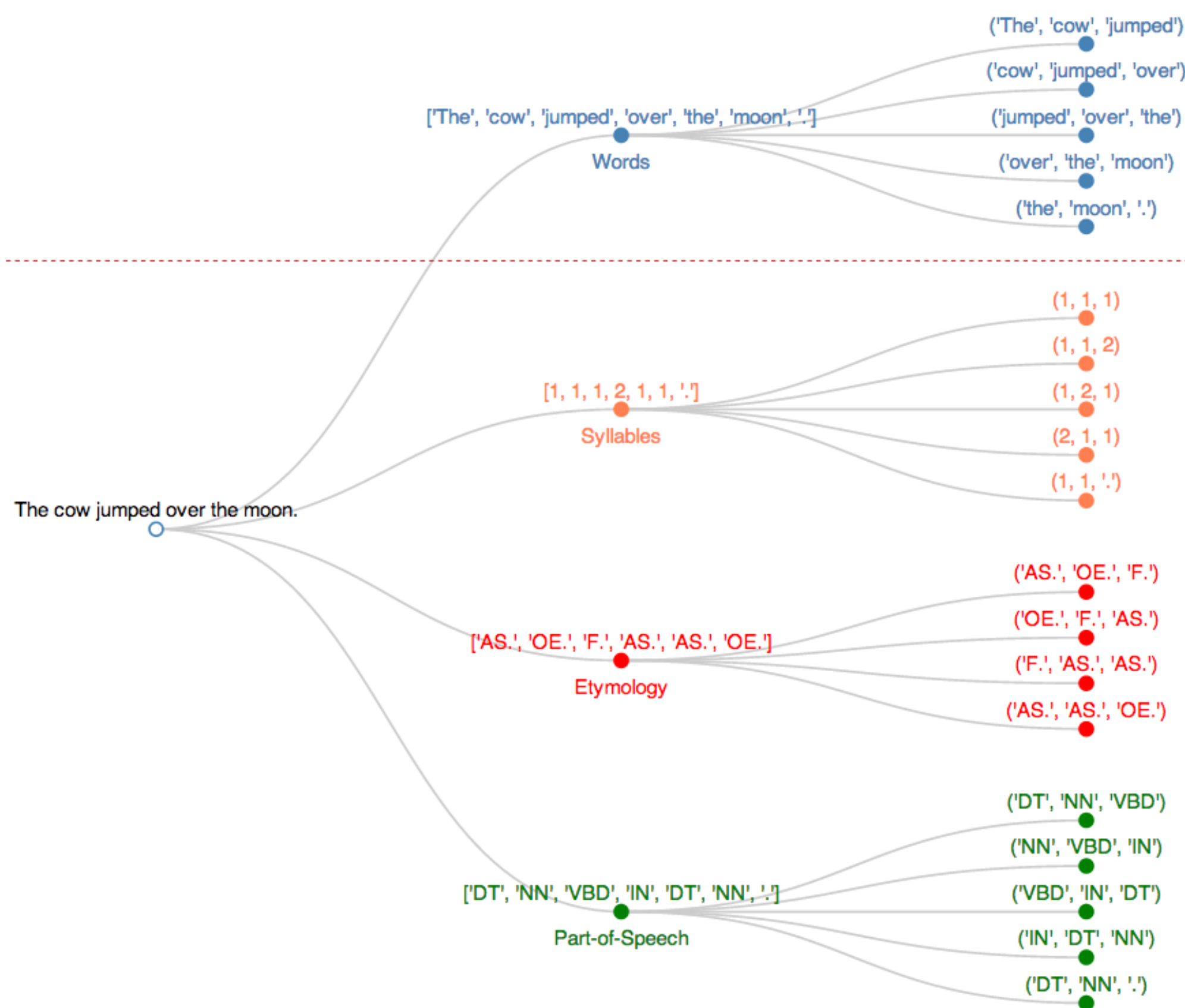
## Model



Graphical Model for the MAD Topic Model

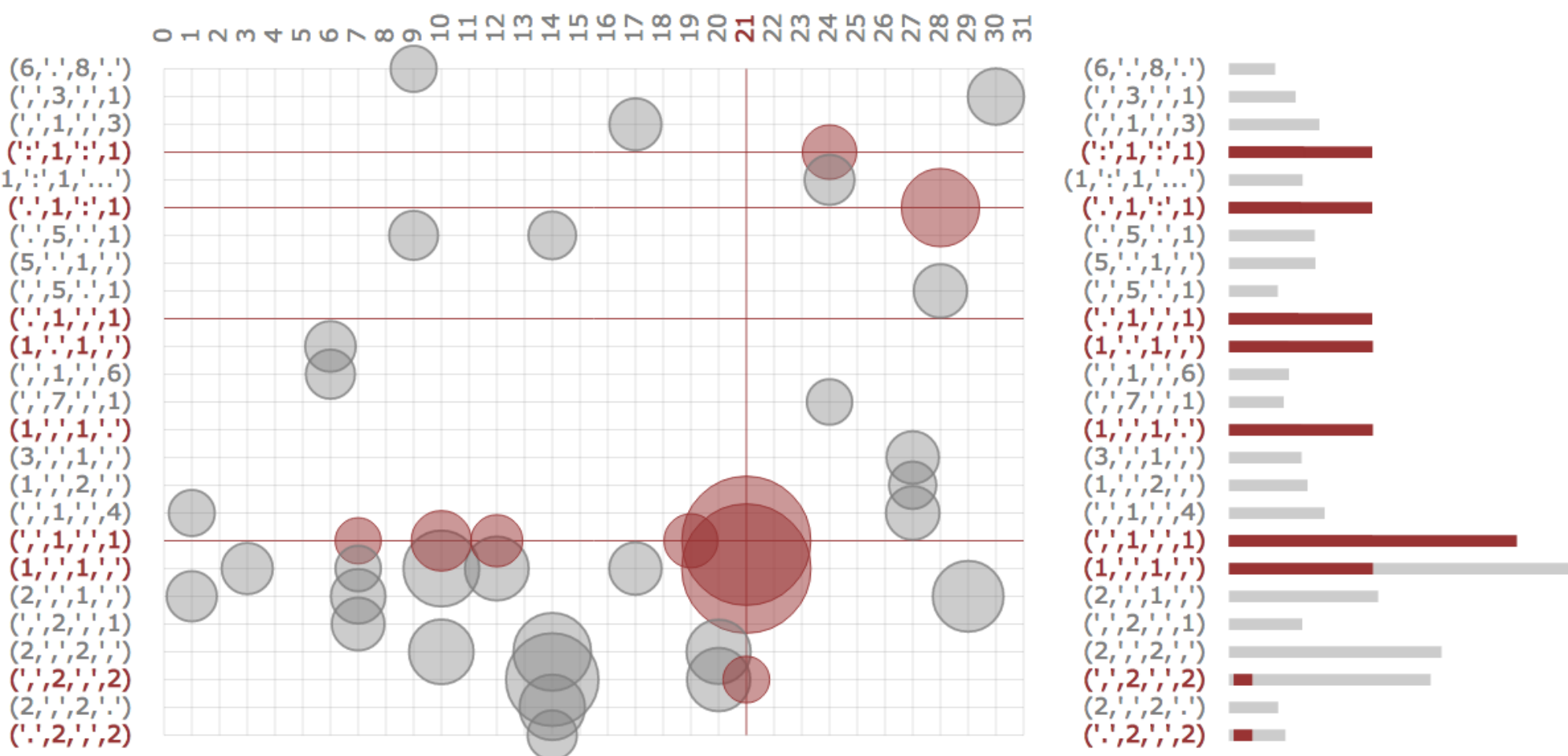
The MAD topic model combines the SLDA algorithm presented in [4] with the Author Topic Model in [3], and extending both to account for multiple word types. The model is variational inference, following coordinate ascent updates in [4]. Stochastic variational inference was also tested, but proved impractical for these rather small data sets.

## Features



Feature extraction for the MAD Topic Model. Word and syllable counts (between punctuation) were also included.

## Visualization



Topic model proportions for the word count  $n$ -grams. Topic 21 represents short, staccato sentences.

## Conclusion

Our (short) conclusion.

## References

- David M Blei and Jon D McAuliffe. Supervised Topic Models. In *NIPS*, volume 7, pages 121–128, 2007.
- David M Blei, Andrew Y Ng, and Michael I Jordan. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.
- Michal Rosen-Zvi, Thomas Griffiths, Mark Steyvers, and Padhraic Smyth. The Author-Topic Model for Authors and Documents. In *Proceedings of the 20th conference on Uncertainty in artificial intelligence*, pages 487–494. AUAI Press, 2004.
- Chong Wang, David Blei, and Fei-Fei Li. Simultaneous Image Classification and Annotation. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1903–1910. IEEE, 2009.



## Data

To collect data for training and testing, we wrote Python scrapers for Project Gutenberg, Nassau Weekly, and Quora.

Datasets collected for training and testing

Source	Authors	Docs/Author
Project Gutenberg	5	50
Nassau Weekly	550	200
Quora	1600	100

Project Gutenberg contains excerpts from fictional books. Nassau Weekly features narrative & editorial articles from the campus publication. Quora captures responses from top users on the question-answer site. The diversity in topic, language, and length challenges our model to detect consistent for

## Results

Unfortunately, preliminary results show that which MAD fares far worse as using the same features with another classification scheme. This is consistent with [?], which suggests that a Pitman-Yor process better captures power law frequencies in language use than Dirichlet methods. Nevertheless, MAD’s topic models over the  $n$ -gram stylistic features can be used to extract compact representations of stylistic tendency and discern which features are most indicative of individual writing style.