

## **Title: Making Topic Models more Human(e)**

Jordan Boyd-Graber

Imagine you need to get the gist of what's going on in a large text dataset such as all tweets that mention Obama, all e-mails sent within a company, or all newspaper articles published by the New York Times in the 1990s. Topic models, which automatically discover the themes which permeate a corpus, are a popular tool for discovering what's being discussed. However, topic models aren't perfect; errors hamper adoption of the model, performance in downstream computational tasks, and human understanding of the data. However, humans can easily diagnose and fix these errors. We present a statistically sound model to incorporate hints and suggestions from humans to iteratively refine topic models to better model large datasets.

We also examine how topic models can be used to understand topic control in debates and discussions. We demonstrate a technique that can identify when speakers are "controlling" the topic of a conversation, which can identify events such as when participants in a debate don't answer a question, when pundits steer a conversation toward talking points, or when a moderator exerts her influence on a discourse.

Bio: Jordan Boyd-Graber is an assistant professor in Maryland's iSchool and UMIACS, and a member of the Cloud Computing Center and the Computational Linguistics and Information Processing (CLIP) Lab.

Jordan received his PhD from Princeton University in 2010, advised by Dave Blei, and has bachelors degrees in history and computer science from the California Institute of Technology. He received a best student paper honorable mention at NIPS 2009 and a Computing Innovation Fellowship (declined).