

Content-Conditioned Hierarchical Latent Space Models for Text-Valued Networks

Matthew Denny *, James ben Aaron *, Hanna Wallach * †, and Bruce Desmarais *,

*University of Massachusetts Amherst, and †Microsoft Research NYC

Statistical topic models are widely used to represent content in textual corpora. When documents in a corpus are attributed with senders and recipients, the corpus constitutes text-valued relational or network data. We develop a class of models for the topical content and relational structure of communications in which the rate of communication between actors is governed by covariate-conditioned latent space embeddings of the actors, and each embedding is associated with a cluster of topics. We present Markov Chain Monte Carlo methods for inference within the Bayesian framework. This model holds the potential for extensive application to real-world social and organizational relational communication data. To illustrate the model and evaluate its performance, we present an application to internal e-mail communications among managers in North Carolina county governments. We illustrate the use of covariates in the model with gender and find that patterns of gender mixing vary with the topical content of communication in a way that is consistent with the content-conditional gendered patterns that have been found to characterize social and organizational networks

Introduction

Results

