Author Topic Manifold Summarization for Interpreting Co-Author Networks

Conrad Hougen
Advisor: Prof. Alfred O. Hero and Dr. Karl Pazdernik
University of Michigan
chougen@umich.edu

Abstract:

Text-based scientific corpora are rich sources of multimodal text and network data. In this work, we use computational tools to model non-obvious author collaborative behaviors within the nuclear science research community. Traditionally, analyses of text-based data focus on modeling latent topics as word-frequency distributions, estimated by statistical inference algorithms. Here, we extend Latent Dirichlet Allocation and the related Author-Topic model to build a model of interpretable co-authorship. We propose a scalable and extensible framework for the analysis of co-authored text corpora, which we call Author Topic Manifold Summarization (AToMS). The AToMS framework fuses together probabilistic topic models with topology-based link prediction algorithms, providing both interpretive and predictive capabilities. We leverage a Hierarchical Random Graph model to provide a probabilistic interpretation of paired connections in the observed or predicted co-author network. Prior work in information-geometric topic modeling inspires our design — relationships between authors and documents are approximated by statistical manifolds of their associated topic vectors. Through approximate estimation of these topic manifolds, AToMS captures intuitive visual and probabilistic relationships between documents, authors, and higher-order groupings of these entities.