A Geometry-Driven Approach To Longitudinal Topic Modeling Of Nuclear-Scientific Literature

Presenter: Conrad Hougen

Advisor: Alfred O. Hero III, University of Michigan

Collaborator: Karl Pazdernik, Pacific Northwest National Laboratory

Email: chougen@umich.edu

Abstract:

Recent work in topic modeling has demonstrated the potential for combining latent topic models with methods from computational geometry to link together topics within time-evolving document corpora. However, while the aforementioned approach has been utilized to visualize topic paths in conversational text sourced from Twitter, this work is the first to extend that framework for dynamic topic modeling of scientific literature. In particular, in nuclear non-proliferation, it is essential to understand the evolution of topics present in research manuscripts from the field of nuclear science. Given a time-evolving set of nuclear science manuscript abstracts from the OSTI dataset, we extract dynamic topic paths from these text data, primarily for visualization in a low-dimensional embedding space. This work explains the first steps toward development of a novel framework for anomaly detection based on text data from scientific publication literature and collaboration network data between associated authors. To demonstrate the efficacy of the method, we compare extracted topics from the model to self-labeled topics from nuclear science manuscripts in the OSTI dataset.