Date: March 3, 2022
Time: 1:30-2:30pm
Title: Topic modeling for understanding multi-domain data from the Vaginal Microbiome
Speaker: Susan Holmes, Professor of Statistics, & Laura Symul, Postdoctoral Scholar, Statistics

Abstract:

Diverse and non-Lactobacillus-dominated vaginal microbial communities are associated with adverse health outcomes such as preterm birth and acquisition of sexually transmitted infections. Despite the importance of recognizing and understanding the key risk-associated features of these communities, their heterogeneous structure and properties remain ill-defined. Clustering approaches have been commonly used to characterize vaginal communities, but they lack sensitivity and robustness in resolving community substructures and revealing transitions between potential sub-communities. We used a more highly resolved approach based on mixed membership topic models with multi-domain longitudinal data from cohorts of pregnant and non-pregnant subjects. We have developed a specific tool, alto, that facilitates the visualization and choice of the number and interpretation of the topics involved.

Overall, our analyses based on mixed membership models revealed new substructures of the vaginal ecosystems which may have potentially important clinical or biological associations.

Suggested Readings:
1. Temporal and spatial variation of the human microbiota during pregnancy
2. Sub-communities of the vaginal ecosystem in pregnant and non-pregnant women
3. Multiscale Analysis of Count Data through Topic Alignment