

## Data Science Seminar

Hosted by the Department of Mathematics and Statistics

- Date: Tuesday, November 15, 2022
- Time: 12:00pm - 1:00pm
- Room: Whitney Hall 100E
- Speaker: Dr. Rachael Hageman Blair (University at Buffalo)
- Title: Harnessing stability estimation for module detection, clustering, and ensemble clustering.

### **Abstract**

Clustering data is a challenging problem in unsupervised learning where no gold standard exists. The selection of a clustering method, measures of dissimilarity, parameters, and the determination of the number of reliable groupings, are often viewed as subjective processes. Stability has become a valuable surrogate to performance and robustness that can guide an investigator in selecting and prioritizing clusters. This talk presents a framework for stability measurements based on resampling and out-of-bag estimation. Bootstrapping methods for cluster stability can be prone to overfitting in a setting analogous to poor delineation of test and training sets in supervised learning. Out-of-bag stability, which overcomes this issue, is observed to be consistently more conservative than traditional measures and is uniquely not conditional on a reference clustering. Furthermore, out-of-bag stability estimates can be estimated at different levels: item level, cluster level, and as an overall summary, which has good interpretive value for the investigator. This framework is extended to develop stability estimates for determining the number of clusters (model selection) through contrasts with simulated reference data with no signal. Finally, new out-of-bag stability estimates are developed to address the problems of ensemble clustering and multi-modal clustering. Applications in the Biomedical Sciences are presented. Stability estimation can be implemented using the &quot;bootcluster &quot; package on the Comprehensive R Archive Network (CRAN).

Biography of the speaker: Dr. Hageman Blair is an Associate Professor in Biostatistics at The University at Buffalo. She received her PhD in 2007 in Applied Mathematics from Case Western Reserve University and trained as a post-doc in statistical genetics at The Jackson Laboratory in Bar Harbor, Maine. Her methodological research interests include Computational Biology, Mathematical Biology, Network Theory, Cluster Analysis and Stability. She maintains several collaborations across the Biomedical Sciences and School of Engineering. She serves as the Associate Director of Education in UB's new Institute of Artificial Intelligence and Data Science, which is home to a PhD program and four interdisciplinary masters programs.

From:

<https://www2.math.binghamton.edu/> - **Department of Mathematics and Statistics, Binghamton University**

Permanent link:

<https://www2.math.binghamton.edu/p/seminars/datasci/111522>

Last update: **2022/11/09 20:37**

