

# ACMS Statistics Seminar

**Martin Barron**  
**University of Notre Dame**  
**Tues, September 26**  
**154 Hurley Hall**  
**3:30– 4:30 PM**



## **Sparse Differential Clustering: A Sparse Clustering Algorithm For Tracking Changing Populations**

Two-condition data arise from a wide range of experimental conditions, e.g. cells before and after a treatment, experimental groups on different diets, and populations of people from different countries. It is often of interest to researchers to identify clusters of similar samples among the different populations. However, for data from two conditions this provides a range of challenges which must be accounted for, in particular, linking clusters across conditions is a struggle as even though samples are of the same “type” they may have changed slightly due to the condition change. The vast majority of current clustering algorithms are wholly unsuitable for this type of analysis as they are designed for single condition data and cannot account for the changes between subpopulations of samples across conditions.

To solve this problem, SparseDC, a new clustering algorithm designed for sparse differential clustering will be presented. This clustering algorithm was designed to solve the challenge of simultaneously clustering two populations of cells from different conditions. SparseDC also generates a sparse solution, that is, only a subset of the features are used in the clustering solution. This feature in particular makes SparseDC scalable to high-dimensional data.

The application of SparseDC will be demonstrated using scRNA-Seq data. Here, it will be shown how SparseDC is able to accurately cluster groups of cells, link the clusters across the populations and identify a set of marker genes for each group.

The Department of Applied and Computational  
Mathematics and Statistics

Please visit [acms.nd.edu](http://acms.nd.edu) to view the full list of speakers.