# Stanford University
## Department of Statistics

## Departmental Seminar

4:30pm, Tuesday, March 13, 2018
Sloan Mathematics Center Room 380Y

Refreshments served at 4pm in Sequoia Lounge.

**Speaker:**   Genevera I. Allen
*Departments of Statistics, CS, & ECE,*
*Rice University*

**Title:   Inference, Computation, and Visualization
for Convex Clustering and Biclustering**

**Abstract:**

Hierarchical clustering enjoys wide popularity because of its fast computation, ease of interpretation, and appealing visualizations via the dendogram and cluster heatmap. Recently, several have proposed and studied convex clustering and biclustering which, similar in spirit to hierarchical clustering, achieve cluster merges via convex fusion penalties. While these techniques enjoy superior statistical performance, they suffer from slower computation and are not generally conducive to representation as a dendogram. In the first part of the talk, we present new convex (bi)clustering methods and fast algorithms that inherit all of the advantages of hierarchical clustering. Specifically, we develop a new fast approximation and variation of the convex (bi)clustering solution path that can be represented as a dendogram or cluster heatmap. Also, as one tuning parameter indexes the sequence of convex (bi)clustering solutions, we can use these to develop interactive and dynamic visualization strategies that allow one to watch data form groups as the tuning parameter varies. In the second part of this talk, we consider how to conduct inference for convex clustering solutions that addresses questions like: Are there clusters in my data set? Or, should two clusters be merged into one? To achieve this, we develop a new geometric representation of Hotelling's $T^2$-test that allows us to use the selective inference paradigm to test multivariate hypotheses for the first time. We can use this approach to test hypotheses and calculate confidence ellipsoids on the cluster means resulting from convex clustering. We apply these techniques to examples from text mining and cancer genomics.

This is joint work with John Nagorski and Frederick Campbell.