



THE UNIVERSITY OF
CHICAGO

COMPUTATIONAL AND APPLIED MATHEMATICS COLLOQUIUM

ANNA LITTLE

Department of Mathematics
University of Utah

**Clustering High-dimensional Data with Path Metrics: A
Balance of Density and Geometry**

THURSDAY, April 22, 2021 at 4:15pm (Central)

Via ZOOM

This talk discusses multiple methods for clustering high-dimensional data, and explores the delicate balance between utilizing data density and data geometry. I will first present path-based spectral clustering, a novel approach which combines a density-based metric with graph-based clustering. This density-based path metric allows for fast algorithms and strong theoretical guarantees when clusters concentrate around low-dimensional sets. However, the method suffers from a loss of geometric information, information which is preserved by simple linear dimension reduction methods such as classic multidimensional scaling (CMDS). The second part of the talk will explore when CMDS followed by a simple clustering algorithm can exactly recover all cluster labels with high probability. However, scaling conditions become increasingly restrictive as the ambient dimension increases, and the method will fail for irregularly shaped clusters. Finally, I will discuss how a more general family of path metrics, when combined with CMDS, give low-dimensional embeddings which respect both data density and data geometry. This new method exhibits promising performance on single cell RNA sequence data and can be computed efficiently by restriction to a sparse graph.

Organizer:

Daniel Sanz-Alonso, Department of Statistics, sanzalonso@uchicago.edu
CAM Colloquium URL: <https://cam.uchicago.edu/events/cam-colloquium/>

For further information and inquiries about building access for persons with disabilities, please contact Zellencia Harris, zellenciah@uchicago.edu. If you wish to subscribe to our email list, please visit the following website: https://lists.uchicago.edu/web/subscribe/cam_colloquium/.