Eigenvectors, Heat Kernels, and Low Dimensional Representation of Data Sets

Peter Jones, Yale University

For several years Diffusion Geometry has been used for low dimensional representations of data sets. We will start by giving results of PJ, M. Maggioni, R. Schul on why theory predicts these representations (via eigenvectors) discover various geometric structures in data. We then discuss a new approach that uses heat kernels instead. The underlying idea for this goes back several decades and is due to S. Varadhan. Theory explains why this procedure is much more stable than choosing a fixed number of eigenvectors. In addition, there are canonical methods for the representations. Unlike the case for Diffusion Geometry, these methods require no human intervention. For very large, high dimensional data sets one needs fast linear algorithms for initial processing of the data. We will discuss numerical methods introduced by PJ, A. Osipov, and V. Rokhlin that are used for this step.