

STATISTICS/DEAS (EE)
JOINT
COLLOQUIUM

FRIDAY, APRIL 18, 2008

TALK: 4:00 P.M. - MAXWELL DWORKIN – G125

“Entropic graphs for high dimensional data analysis”

Alfred Hero
Electrical Engineering and Computer Science Department
University of Michigan

ABSTRACT

A minimal spanning tree (MST) spanning random points has total spanning length that converges to the entropy of the underlying density generating the points. This celebrated result was first established by Beardwood, Halton and Hammersley (1958) and has since been extended to other random Euclidean and non-Euclidean graphs, such as the geodesic MST (GMST) and the k-nearest neighbor graph (kNNG) over a random set of point. Using the BHH theory of random graphs one can construct graph-based estimates of topological properties of a high dimensional distribution of a data sample. This leads, for example, to a model-free consistent estimator of intrinsic dimension of a data manifold and a high performance non-parametric anomaly detector. We will illustrate this entropic graph approach for applications including: anomaly detection in Internet traffic; activity detection in a MICA2 wireless network; and intrinsic dimension estimation of image databases.