Group finding in multi-dimensional data-sets: method and application to the identification of substructures in the stellar halo  Sanjib Sharma & Kathryn V Johnston (Columbia University)

- A multi-dimensional group finding algorithm to efficiently and accurately analyze *large and complex datasets* with arbitrary number and type of dimensions.
  - Can be used to analyze numerical simulations of structure formation, astronomical surveys or as a general data mining tool.

- **What is the optimum distance metric?**

- **Method:** uses the information theoretic idea of *Shannon entropy* to calculate a *locally adaptive distance metric* for each data point which helps to extract maximum amount of information from the data.

- **Application:** identify tidal streams produced by accretion of satellite galaxies in *simulated stellar halos* and in the *2MASS data*. 