

## **RobustMap: A Fast and Robust Algorithm for Dimension Reduction and Clustering**

Databases can be very large due to the number of items and due to the number of attributes (high-dimensionality) associated with each item. Clustering reduces the number of items to their representative clusters and dimension reduction reduces the number of attributes. In addition, visualization of high-dimensional data requires reduction to lower-dimensional views that are often displayed as two or three dimensional plots. Traditional dimension reduction algorithms such as the singular value decomposition based principal components are computationally demanding and can be very slow. As the size of databases continues to grow, so does the demand for faster methods to visualize the data. RobustMap is a new fast and robust dimension reduction method for high-dimensional datasets, which can separate outlying clusters from the main body of the data while computing a low-dimensional representation. It relies on stochastic concepts and on statistical distance distributions. The algorithm considers distance distributions from random and from extreme points to determine projection axes and clusters for dimension reduction. In determining the clusters, RobustMap focuses on the largest cluster, excluding outlying clusters. The visualization applications of this algorithm may be implemented in a range of disciplines, which include medical databases, images, time series, music, data mining, etc.

Student's name:

**Lionel F. Lovett, II**

School student attends:

Jackson State University

Name(s) of mentor(s):

**Dr. George Ostrouchov and Dr. Houssain Kettani**

Division:

Computer Science and Mathematics

Program:

Research Alliance in Math and Science (RAMS)