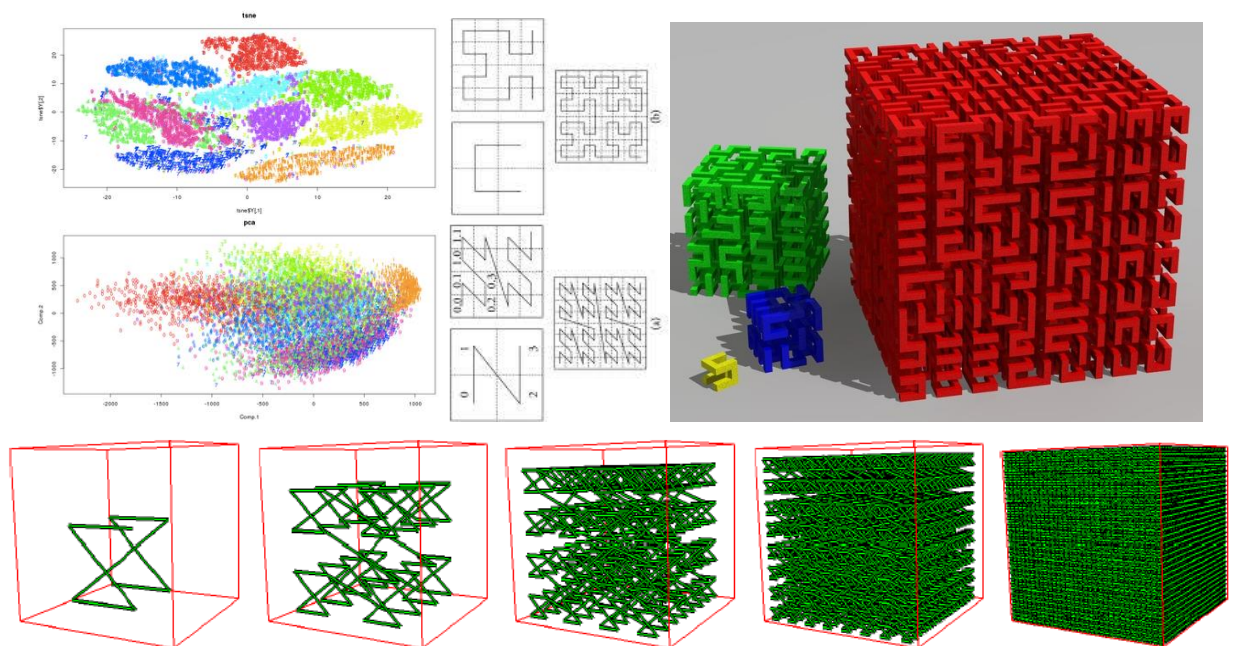


Space-Filling-Curve-based Dimension Reduction for Big Data Visual Analytics

Project Description: Big data come usually as a large set of multidimensional instances. Visualizing this kind of data, such as actigraphy data produced by wearable sensors (e.g. Fitbit), is crucial for instance to discover typical behaviors that would distinguish diabetic from non-diabetic patients, a major disease in Qatar.

But visualizing these large and high dimensional data is very challenging. Approaches like Principal Component Analysis generate many artefacts due to their linear assumption, while t-SNE, a well-known nonlinear technique does not scale well to large datasets. We want to explore how we could use multidimensional space filling curves as a way to represent multidimensional data as a 2 dimensional multiscale scatterplot.



Left shows tSNE (top) and PCA (bottom) 2 dimensional projection of originally multidimensional data. Center shows a Hilbert space filling curve (top) and Z-order one (bottom) in 2D space. Right shows Hilbert in 3D space. Bottom row shows Z-order in 3D space. Hilbert and Z-order can be generalized to multidimensional spaces as well. We want to encode multidimensional data on such curves and use this indexing to display data in 2D.

Duties/Activities: You will implement Hilbert and Z-order indexing and projection in Python

Required Skills: Python

Preferred Intern Academic Level: BSc

Learning Opportunities: Python programming language

Expected Team Size: 1 or 2 students

Mentors

Michael Aupetit email: maupetit@hbku.edu.qa

Abdelkader Baggag email : abaggag@hbku.edu.qa