BIOCHEMICAL CHARACTERIZATION OF TISSUE VIA SPECTRAL CLUSTERING METHODS AND IMAGING MASS SPECTROMETRY

This topic is meant for 1 student.
The location (Research unit) is ESAT - SISTA - BIOI

Thesis topic

BACKGROUND
Imaging mass spectrometry is a technology that facilitates the detection of biomolecules such as proteins, peptides, and metabolites directly from organic tissue while keeping the spatial information intact. The generated data enables us to unravel the mechanisms underlying disease and to identify possible targets for drug delivery. A typical IMS experiment consists of a grid of measurement locations or pixels covering the tissue section, with an individual mass spectrum attached to each pixel. The resulting data structure can be considered as a three-dimensional array with two spatial dimensions \((x,y)\) and one mass-over-charge \((m/z)\) dimension.

Spectral Clustering
Clustering methods correspond to a family of statistics and machine learning techniques in which the objective is to group “similar” patterns, or structures, within a dataset without using external information. Classical algorithms for clustering, for instance k-means, only work in high dimensional settings unless pre-processing and model reduction is first performed. On the contrary, recent spectral clustering methods exploit these high dimensional spaces with the use of (non)-linear kernel functions. Moreover no assumptions are made in the modeling and one can incorporate prior knowledge, such as region constraints, to improve differentiation between tissue regions starting from the mass profiles.

CONTENT
The first step is to construct a simulated imaging mass spectrometry dataset to benchmark the performance of spectral clustering algorithms on this type of data. The second step will be to apply spectral clustering algorithms to a real IMS dataset. The third step would be to adapt the kernel metric by, for instance, including the spatial information of the mass profiles. A final step consists of incorporating prior knowledge by imposing a cluster membership to a set of tissue grid points.

Profile
The thesis is situated in the booming field of bioinformatics, a multidisciplinary field with great applications for mathematical engineering techniques.

Prerequisites:
Basic Matlab knowledge.
Strong basis in algebra is a plus.
Machine learning courses are a plus.

Literature (35%), Programming (35%), Writing (%30)

Daily supervisors
Nico Verbeeck, nico.verbeeck@esat.kuleuven.be, +3216328647, Kasteelpark Arenberg 10 - bus 2446, 3001 Heverlee
Yousef El Aalamat, yousef.elaalamat@esat.kuleuven.be, +3216328647, Kasteelpark Arenberg 10 - bus 2446, 3001 Heverlee

Promotors
Bart De Moor