Título/Title:
Unsupervised detection of galaxy clusters in future large photometric surveys

Orientador/Supervisor:
António da Silva (IA-Lisboa), Alberto Krone-Martins (CENTRA-SIM) ajosilva@fc.ul.pt, algol@sim.ul.pt

Descrição/Description:
Galaxy clusters are the largest bound structures in the universe. They are formed by gravitational accretion of infalling material at the intersection of filamentary structures and contain hundreds to several thousand galaxies in dynamical equilibrium within the gravitational potential. Clusters are therefore essential objects to understand the evolution of galaxies and to test structure formation and cosmology models featuring unknown forms of dark matter and dark energy. Near future galaxy surveys, such as that by the ESA/Euclid satellite and LSST, will provide large catalogues with photometric information for billions of galaxies over the whole sky. The identification of galaxy clusters in these datasets require a correct assessment of their member galaxies (membership identification problem). This a non-trivial task when individual galaxy distances/redshifts are unknown or only poorly determined. Moreover, good quality distance information is usually only available for a small fraction of objects in these datasets, and traditional methods of cluster identification often assume model dependent information or human supervision in the identification process.
The objective of this project is to study the problem of identification of galaxy clusters in large galaxy photometric surveys using a fully data driven (i.e. unsupervised) way that does not necessarily rely on distance estimations. The proposed work involves the study and modification of the UPMASK (Unsupervised Photometric Membership Assignment in Stellar Clusters) method and associated software package. This method, originally developed to study star clusters without or poor knowledge about distances, uses heuristics and statistical analysis to separate clusters from the field, without any basis on theoretical models, and consequently without strong a priori statements, of what a cluster is. The student will first review the problem of galaxy clustering and membership identification in clusters. She/he will then study the UPMASK algorithm and investigate ways to improve the code's performance for the identification of clusters in large galaxy surveys. Afterwards he/she will validate code modifications with well-known observed datasets and with simulations of the forthcoming ESA/Euclid space mission. Depending on progress, the modified method will be applied to existing data and Euclid simulations to estimate cluster survey yields and assess model constraints.

Requisitos/Requirements:
Previous knowledge of programming in any language is required. Preference will be given to candidates with previous knowledge of the R language and contact with the problem of cluster identification using unsupervised techniques.