Use of sparse principal component analysis approaches to build gene co-expression networks in a systems biology context

A lot of information is encoded in the relationships between biological features (gene expression, protein concentrations, metabolites concentrations...) levels, rather than in their levels alone. The multiple regulatory layers that underlie the functioning of a living system organize the response to a perturbation and/or modification of the system, causing the biological components, such as metabolites, to change in a coordinate way upon these modifications. The associations and the interconnections between the building blocks of the individual metabolic phenotype can be described through biological networks which, in turn, can be compared across different pathophysiological conditions or phenotypes. This can help the identification of those biological processes that are perturbed and pinpoint the key regulatory mechanisms of disease pathophysiology.

Many methods have been developed to infer gene expression networks: in this project we are interested in investigating if a recently developed tool, the Group-wise Principal Component Analysis (GPCA)(Camacho, Rodríguez-Gómez, Saccenti 2016) can be used to successfully detect clusters of functionally related genes in the context of co-expression gene network analysis.
GPCA is an extension of PCA where sparsity is imposed. The structure to impose sparsity is defined in terms of groups of correlated variables found in correlation matrices or maps that can be computed with minimum noise level using a Missing-Data for Exploratory Data Analysis (MEDA) approach.

You will become acquainted with Matlab and with both theory and practice of principal component analysis and (basic) theory of network inference in a systems biology context by making use of real and simulated data.

References:
Camacho, J., R. A. Rodríguez-Gómez, E. Saccenti (2016). *Group-wise Principal Component Analysis for Exploratory Data Analysis*. Journal of Computational and Graphical Statistics, 0-0 doi:10.1080/10618600.2016.1265527

Contact person: Edoardo Saccenti (edoardo.saccenti@wur.nl) Laboratory of Systems and Synthetic Biology.