

The R package **FANet**: sparse Factor Analysis model for high dimensional gene co-expression Networks

Yuna Blum^{1*}, Magalie Houée-Bigot² Sandrine Lagarrigue³ David Causeur²

1. Department of Human Genetics, University of California, Los Angeles CA. USA

2. Agrocampus Ouest- Applied Mathematics Department, Rennes, FRANCE

3. Agrocampus Ouest - UMR PEGASE INRA, Agrocampus Ouest, Rennes, FRANCE.

*Contact author: yuna.blum@gmail.com

Keywords: co-expression networks, factor analysis, high dimension, sparsity, R

Inference on gene regulatory networks from high-throughput expression data turns out to be one of the main current challenges in systems biology. Such interaction networks are very insightful for the deep understanding of biological relationships between genes. In particular, a functional characterization of gene modules of highly interacting genes enables the identification of biological processes underlying complex traits as diseases. Inference on this dependence structure shall account for both the high dimension of the data and the sparsity of the interaction network.

The R package **FANet** provides a powerful method for estimating high dimensional co-expression networks. Extending the idea introduced for differential analysis by Blum et al. [1] and Friguet et al. [2] we suggest to take advantage of a low-dimensional latent linear structure of dependence to improve the stability of correlation estimations. We propose an EM algorithm to fit a sparse factor model for correlations and demonstrate how it helps extracting modules of genes and more generally improves the gene clustering performance. Two functions are available in **FANet** package in order to introduce sparsity in the network estimation. One function is based on a LASSO estimation using a cyclic coordinate descent algorithm. As an alternative, the second function is based on biological knowledge integration as Gene Ontology annotation. Finally, **FANet** results can serve as an input for WGCNA (Langfelder and Horvath [3]) procedure for gene modules detection.

References

- [1] Blum, Y., G. Le Mignon, S. Lagarrigue, and D. Causeur (2010). A factor model to analyze heterogeneity in gene expression. *BMC bioinformatics* 11(1), 368.
- [2] Friguet, C., M. Kloareg, and D. Causeur (2009). A factor model approach to multiple testing under dependence. *Journal of the American Statistical Association* 104(488), 1406–1415.
- [3] Langfelder, P. and S. Horvath (2008). Wgcna: an r package for weighted correlation network analysis. *BMC bioinformatics* 9(1), 559.