Statistical Learning for multilayer biological networks
MIA Paris-Saclay, INRAE-AgroParisTech

October 4, 2021

- **Team**: Team SOLsTIS, MIA-Paris-Saclay, Unité Mixte de recherche INRAE - AgroParisTech

  - **Localisation**: Paris until may 2021 and then Palaiseau.

  - **Contact**:
    * pierre.barbillon@agroparistech.fr,
    * julien.chiquet@inrae.fr,
    * sophie.donnet@inrae.fr

**Context**

A gene (or genetic) regulatory network (GRN) is a collection of molecular regulators that interact with each other. They are in general obtained from mRNA levels of expression by thresholding the covariances. Another popular solution is sparse Gaussian graphical models (Banerjee et al., 2008), which select strong partial correlations in the inverse covariance matrix (aka the precision matrix). The inferred network is then studied to get its organization and link it to the gene functions. The Stochastic block models (Holland et al., 1983) provide a set of tools for this purpose.

Ambroise et al. (2009) proposed to perform the two tasks (namely inference and structure characterization) jointly by introducing a prior distribution on the network such as Stochastic Block Models, which is more or less equivalent to optimizing the likelihood with an adapted penalty term.

The last years have witnessed the need to adopt a more integrative view, and not consider only genes but also the metabolites and / or proteins networks, resulting into a multi-omics integration. The objective of that post-doctoral contract is to work on the global network including genes, proteins and metabolites. The network has to be inferred from heterogeneous and structured data,
including nodes with various types (genes, proteins, ...). Stochastic Block Models adapted to multi-type nodes must be considered (such as multilayer, multiplex or nested SBM, see Han et al., 2015; Barbillon et al., 2015; Peixoto, 2019). The network reconstruction will rely on structured versions of the penalized likelihood approach, obtained by modifying the Graphical-Lasso algorithm (Friedman et al., 2008) so that it can deal with hierarchical or tree-like penalties (Bach et al., 2012). Moreover, the high number of nodes (≈ 20000 genes, ≈ 2000 metabolites) will make the problem computationally challenging.

In a nutshell, the objective of the contract is to propose a new method to jointly infer and cluster biological networks involving nodes of various nature and from heterogeneous data. The work will be completed by a diffusable tool such as an R or Python package and applied to real datasets.

Scientific environment

The MIA Paris-Saclay UMR is a research unit in statistics and computer science oriented towards applications in life sciences and environment. Its main objective is to develop and disseminate generic methods in the disciplines of origin of its members (statistics, mathematics and computer science), motivated by applications in life sciences and the environment. The unit is composed of two teams, EkiNocs (Expert Knowledge, InTernactive modellIng and learnIng for unDerstanding and decisiOn making In dyNamic complex systeMs) and SOLsTIS. SOLsTIS is a team dedicated to statistical modelling and learning for environment and life sciences, composed of around 20 permanent researchers and 10 PhD and post-doctoral students. The post-doctoral candidate will belong to that team. The team has long-term collaborations with researchers in applied genomics.

At the Paris Saclay, the candidate will benefit from a great scientific environment, and will rely in particular on the work groups “Variational Inference”, “Le séminaire Palaisien en Stat/Math/Info” or Institute such as DATA-IA.

Expected skills

Beyond the classical skills expected (curiosity, motivation, rigor, ability to read and understand English, autonomy and taste for teamwork), and particular motivations for the analysis of natural systems and the treatment of environmental issues, the candidates must hold a PhD in applied statistics or machine learning with strong skills in algorithmic tools and optimization.

References


