

Inference and analysis of molecular multilevel networks in plant

MIA Paris-Saclay in collaboration with IPS2

FMJH Grant Application

Abstract

The ambition of this project is to decipher the effect of increasing atmospheric CO₂ on plant by building multilevel networks to compare the response of transcripts, proteins and metabolites to an increase in temperature between both CO₂ conditions and identify subnets that will be characterized by the biologists of the consortium.

1 Context

A hallmark of the Anthropocene is the increase in atmospheric CO₂ concentration with a value of 415 ppm in 2021 and an expected level of 700-1000 ppm in 2100 if no action is taken. A consequence of elevated CO₂ (eCO₂) will be increased periods of a more elevated mean temperature (eT). Such changes will adversely affect plant development, stress response and crop yield. However, crop yields must increase to sustain growing human food demands. The impact of eCO₂ on plant physiology is not fully understood. On the one hand, there is a large body of evidence supporting a positive impact of eCO₂ on thermotolerance but on the other hand the increase in C3 plant yield in response to long-term eCO₂ is 50% lower than predicted. Therefore it becomes urgent to determine the underlying mechanisms leading to the impact of eCO₂.

The challenge is biologically complex because eCO₂ and eT affect cell physiology at multiple levels (transcripts/proteins/metabolites) and combined stresses are not the sum of the individual stresses. Moreover, because molecular biology teams are often specialised in a specific process, there are few genomic studies focused on the interconnections between molecular levels. **We propose an original project based on computational approaches to construct a holistic understanding of how plant response is altered by elevated atmospheric CO₂.**

2 Dataset

Thanks to past INRAE funding, we generated a dataset to investigate the impact of eCO₂. Arabidopsis thaliana plants were grown under either 400 ppm or 1000 ppm CO₂ at 20°C during 3 weeks. Plants were then either transferred to an eT of 30°C during 30 h or they were kept at 20°C. Rosette leaves from the 4 conditions (400 ppm, 20°C; 400 ppm, 30°C; 1000 ppm, 20°C and 1000 ppm, 30°C) were collected and -omics measurements (transcripts/proteins/metabolites) were carried out. Furthermore photosynthetic parameters were measured at the same time and 5 days after the plants were finally returned to 20°C. **Aware of the lack of power of high-throughput technologies, we anticipated the challenges they could introduce to data integration by generating 22 biological replicates.** Preliminary analyses have highlighted the quality of the dataset, and its biological relevance.

3 Objective

The first objective is to construct a multilevel network associated with each of the 4 conditions (Fig 1). To do this, we propose to use state-of-the-art techniques relying on partial correlation matrices (Graphical-LASSO, Genie3, Poisson Log-Normal) to infer the network at each level. Then levels will be interconnected using biological knowledge from databases related to gene annotation (TAIR10 and Araport11), metabolite pathways (Kegg, AraCyc, MapMan), transcription factor binding sites (JASPAR, PlantTFDB, ReMap). This step will be performed in close collaboration with biologists

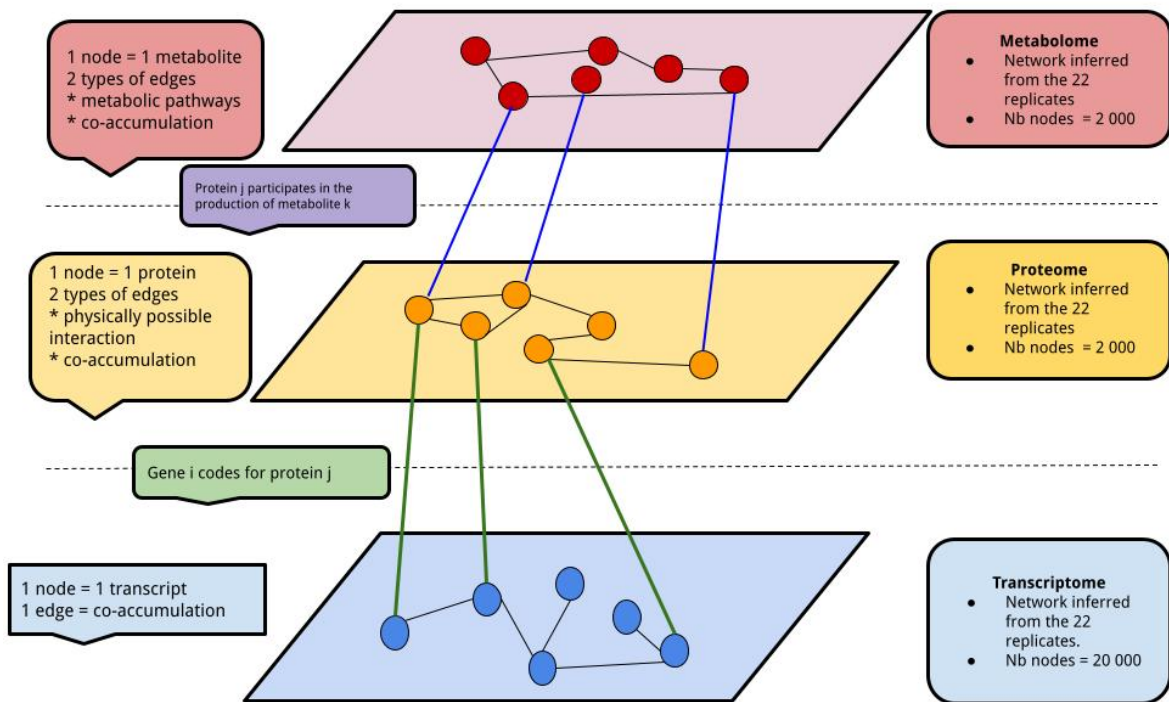


Figure 1: Schematic view of a multilevel network

of IPS2. Structural and functional annotations will be added to the biomolecules by a bioinformatics characterization. Finally the structure of each multilevel network will be identify by adapting existing work on multilayer networks (Bar-Hen et al., 2022; Barbillon et al., 2017; Chabert-Liddell et al., 2021). It will identify biomolecule groups with notable patterns.

The second objective is to identify pathways impacted by $e\text{CO}_2$ by comparing the 4 multilevel networks. To start we propose to study their structural differences to highlight subnets impacted by $e\text{CO}_2$. This could be achieved by fitting a model for collections of networks and then providing a clustering of networks on this basis. An adaptation of Chabert-Liddell et al. (2024) to multilevel network would be necessary.

Although the lines of the analysis are defined, the challenges lies in dealing with the huge sizes of molecular networks and in adapting the existing methods. These methods are indeed not directly suited to the specificity of the multilevel layers encountered in molecular networks.

4 Profile required

PhD in applied mathematics, with an interest in interface work and applications in the life sciences; proficiency in a programming language essential; scientific rigour, intellectual curiosity, good communication skills.

5 Practical arrangements and supervision

The position will be based at AgroParisTech in the MIA Paris-Saclay unit. This project will require numerous discussions with bioinformaticians and biologists of IPS2, located near AgroParisTech. These discussions will be organized at the beginning by Marie-Laure Martin who is affiliated to both units and then the applicant will manage these discussions.

Contacts of the supervisors

Marie-Laure Martin: marie-laure.martin@inrae.fr

Sophie Donnet: sophie.donnet@inrae.fr

Pierre Barbillon: pierre.barbillon@agroparistech.fr

References

Bar-Hen, A., P. Barbillon, and S. Donnet (2022). Block models for generalized multipartite networks: Applications in ecology and ethnobiology. *Statistical modelling* 22(4), 273–296.

Barbillon, P., S. Donnet, E. Lazega, and A. Bar-Hen (2017). Stochastic block models for multiplex networks: an application to a multilevel network of researchers. *Journal of the Royal Statistical Society Series A: Statistics in Society* 180(1), 295–314.

Chabert-Liddell, S.-C., P. Barbillon, and S. Donnet (2024). Learning common structures in a collection of networks. an application to food webs. *The Annals of Applied Statistics* 18(2), 1213–1235.

Chabert-Liddell, S.-C., P. Barbillon, S. Donnet, and E. Lazega (2021). A stochastic block model approach for the analysis of multilevel networks: An application to the sociology of organizations. *Computational Statistics & Data Analysis* 158, 107179.