

Collaborative graphical lasso

Alessio Albanese^{1,2}, Wouter Kohlen², and Pariya Behrouzi*¹

¹Mathematical and Statistical Methods (Biometris), Wageningen University and Research

²Cellular and Developmental Biology, Wageningen University and Research

Abstract

In recent years, the availability of multi-omics data has increased substantially. Multi-omics data integration methods mainly aim to leverage different molecular data sets to gain a complete molecular description of biological processes. An attractive integration approach is the reconstruction of multi-omics networks. However, the development of effective multi-omics network reconstruction strategies lags behind. This hinders maximizing the potential of multi-omics data sets. With this study, we advance the frontier of multi-omics network reconstruction by introducing *collaborative graphical lasso* as a novel strategy. Our proposed algorithm synergizes *graphical lasso* with the concept of *collaboration*, effectively harmonizing multi-omics data sets integration, thereby enhancing the accuracy of network inference. Besides, to tackle model selection in this framework, we designed an *ad hoc* procedure based on network stability. We assess the performance of *collaborative graphical lasso* and the corresponding model selection procedure through simulations, and we apply them to publicly available multi-omics data. This demonstrated *collaborative graphical lasso* is able to reconstruct known biological connections and suggest previously unknown and biologically coherent interactions, enabling the generation of novel hypotheses. We implemented *collaborative graphical lasso* as an R package, available on CRAN as *coglasso*.

Keywords: Graphical model; Multi-omics data integration; Lasso; Collaborative graphical lasso; High-dimensional data.

1 Introduction

A successful multi-omics data integration strategy would provide a holistic molecular explanation of any biological phenomenon of interest. Therefore, multi-omics data integration can be considered a paramount topic of biostatistics. An attractive strategy to integrate multi-omics data is to estimate networks of interactions from them, connecting the molecular units under investigation. These estimated interactions hint directly at the regulatory mechanisms that underlie the biological phenomenon during which multi-omics data were recorded. A preferred statistical framework to reconstruct these networks are Gaussian Graphical Models (GGMs) because the interactions they estimate encode simultaneous interactions. These are a form of pairwise relationship that can remarkably separate direct from induced effects (Altenbuchinger et al., 2020). Nevertheless, available GGM estimating strategies are tailored to handle datasets containing only single sets of variables, while multi-omics data are characterized by an intrinsic multi-assay nature. Thus, although other network estimation strategies like mixed graphical models have been adapted for multi-omics data integration (Altenbuchinger et al., 2020), the area of multi-omics data integration through GGM estimation is only marginally explored.

*To whom correspondence should be addressed. Email: pariya.behrouzi@wur.nl

While the field of multi-omics GGMs estimation is still in its infancy, biostatisticians have already developed several prediction techniques based on linear regression to integrate multi-omics data. Some of these techniques have introduced concepts that could strongly benefit current GGM estimation methods. For example, Gross and Tibshirani (2015) presented *collaborative regression*, a regression strategy that handles the multi-assay nature of multi-omics data in a remarkable way. This technique encourages “collaboration” between two datasets by penalizing the difference between the linear predictors obtained from each of the two datasets. In this way, the two datasets comparably contribute to the prediction. Despite this original and elegant way to deal with the multi-assay nature of multi-omics data, the concept of collaboration has not been implemented in the field of GGM estimation yet.

In this study, we further explore the potential of multi-omics GGM estimation by proposing a new algorithm we name *collaborative graphical lasso* (*coglasso*). *Coglasso* incorporates the concept of collaboration as presented by Gross and Tibshirani (2015) into one of the most popular algorithms for GGM estimation, the *graphical lasso* (*glasso*) (Friedman et al., 2008). Our strategy estimates GGMs with a coordinate descent algorithm based on a novel objective function that includes a collaborative term. This term takes advantage of the multi-assay nature of data by ensuring a comparable contribution of the two datasets in estimating the final network structure.

This paper is structured as follows. In Section 2 we briefly provide some background on GGMs, *glasso* and *collaborative regression*. Section 3 illustrates the novelties introduced with *coglasso* and a corresponding new model selection procedure based on the stability criterion. In Section 4 we test the performance of our method in recovering the structure of simulated networks and we compare it with the *glasso* method. Section 5 shows the usefulness of this new method with an application on available multi-omics data. The algorithm is available on CRAN as the R package *coglasso*, while the reproducible code to replicate the analysis performed in Section 5 is available as a vignette.

2 Methodological background

Collaborative graphical lasso (*coglasso*) is a novel algorithm to estimate gaussian graphical models (GGMs) from data with a multi-assay nature, i.e. multi-omics data. To develop this new algorithm, we heavily modified the *glasso* algorithm proposed by Friedman et al. (2008), which is exclusively designed to deal with single datasets. The resulting algorithm can estimate GGMs by leveraging multiple datasets. The approach designed to leverage these different datasets is inspired by *collaborative regression* introduced by Gross and Tibshirani (2015). Similar to how *collaborative regression* integrates multiple datasets by modifying the objective function of the basic linear regression, *coglasso* achieves a similar level of integration by modifying the objective function at the basis of the *glasso* algorithm. The following subsections summarize the essential elements of GGMs, *glasso*, and *collaborative regression* that lay the foundations of *coglasso*.

2.1 Gaussian Graphical Models and *glasso*

Gaussian Graphical Models (GGMs) are a powerful framework to represent variables distributed according to a multivariate normal distribution and their conditional dependence relationships. They can be represented by graphs. A graph $G = (V, E)$ is defined by a set of nodes V and a set of edges E that represent connections between pairs of nodes. In the GGM framework, nodes are the variables of the model, while edges, the connections between such variables, are used to represent the conditional dependence relationships between them. This means that two variables are connected if they are conditionally dependent, while they are not connected if they are conditionally independent. Assume, for example, that we are interested

in estimating a GGM with p variables. We assume that these variables follow a multivariate normal distribution with mean μ and variance-covariance matrix Σ . We call the inverse of a variance-covariance matrix Σ a precision matrix $\Theta = \Sigma^{-1}$. The entries of a precision matrix represent conditional dependencies between variables. In particular, if the entry θ_{ij} is zero it means that the variables i and j are conditionally independent. Therefore, since in the GGM framework connections represent conditional dependence relationships, the set of connections of a GGMs is purely determined by non-zeros in the precision matrix. These connections can be further represented in the adjacency matrix A , where all non-zero elements (the edges of the network) are encoded as 1s and the remaining ones as 0s.

A well-established algorithm to estimate GGMs from a single dataset is *glasso* (Friedman et al., 2008). This algorithm performs an iterative coordinate descent optimizing procedure to find \mathbf{W} , the estimate of the variance-covariance matrix Σ , and $\hat{\Theta}$, the estimate of its inverse. *Glasso*'s strategy allows positive definiteness, hence invertibility, of the matrix \mathbf{W} resulting from the estimation. Moreover, *glasso* guarantees the inverse matrix $\hat{\Theta} = \mathbf{W}^{-1}$ to be sparse, meaning that the final GGM determined by $\hat{\Theta}$ is not too dense with connections. *Glasso* achieves these positive definiteness and sparsity by fragmenting the estimation of \mathbf{W} into several *lasso* regressions that optimize and estimate separately each row and column of the matrix. Every *lasso* regression solves Equation (1).

$$\hat{\beta}_i = \arg \min_{\beta_i} \left\{ \frac{1}{2} \left\| (\mathbf{W}_{\setminus i})^{-1/2} \mathbf{S}_i - (\mathbf{W}_{\setminus i})^{1/2} \beta_i \right\|^2 + \lambda \|\beta_i\|_1 \right\} \quad (1)$$

Here \mathbf{S} is the empirical variance-covariance matrix estimated from the data and λ is the penalty parameter. $\|\cdot\|_1$ and $\|\cdot\|^2$ represent, respectively, the $L1$ - and the $L2$ -norm. \mathbf{S}_i is the i -th column (and row) of \mathbf{S} without the diagonal element, while $\mathbf{W}_{\setminus i}$ is the submatrix of \mathbf{W} without the i -th row and column. $\hat{\beta}_i$ solving equation (1) is stored as the i -th column of the matrix $\hat{\mathbf{B}}$ and used to update the i -th row (and column) of \mathbf{W} , \mathbf{W}_i , as $\mathbf{W}_i = \mathbf{W}_{ii} \beta_i$. \mathbf{W}_{ii} is the i -th diagonal element of the matrix \mathbf{W} , which is initiated to $\mathbf{W}_{ii} = \mathbf{S}_{ii} + \lambda$ at the beginning of the algorithm. The algorithm cycles through $i = 1, 2, \dots, p, 1, 2, \dots, p, \dots$ until convergence, updating β_i in each iteration. For any i , β_i is estimated element by element through a coordinate descent procedure as described in Friedman et al. (2007), iteratively updating the j -th coordinate of the vector β_i , $(\hat{\beta}_i)_j$. Algorithm 1 describes the update rule

Algorithm 1 *glasso* coordinate descent rule

```

for  $j = 1, 2, \dots, p, 1, 2, \dots, p, \dots$ , until convergence criterion is met do
     $r \leftarrow \mathbf{S}_{ji} - \sum_{k \neq j} (\mathbf{W}_{\setminus i})_{jk} (\hat{\beta}_i)_k$ 
    if  $r < -\lambda$  then
         $(\hat{\beta}_i)_j \leftarrow \frac{r+\lambda}{\mathbf{W}_{jj}}$ 
    else if  $r > \lambda$  then
         $(\hat{\beta}_i)_j \leftarrow \frac{r-\lambda}{\mathbf{W}_{jj}}$ 
    else
         $(\hat{\beta}_i)_j \leftarrow 0$ 
    end if
end for

```

to update the j -th element of $\hat{\beta}_i$. It begins with computing an initial value r . The value assigned to r is computed using two terms: the entry \mathbf{S}_{ji} of the empirical variance-covariance matrix and a dot product that is subtracted to \mathbf{S}_{ji} . After r is computed, its value can either fall outside or inside of the penalty interval $(-\lambda, \lambda)$. In the first case, the algorithm assigns to $(\hat{\beta}_i)_j$ the value of r shrunk towards the interval. In the other, $(\hat{\beta}_i)_j$ is fixed to 0. The updates are run multiple times for $j = 1, 2, \dots, p, 1, 2, \dots, p, \dots$ until convergence of the vector $\hat{\beta}_i$.

2.2 Collaborative regression

We aimed at extending *glasso* to develop a novel algorithm that leverages multi-assay data. To do so, we chose as a launching platform the collaborative regression introduced by Gross and Tibshirani (2015). *Collaborative regression* is a strategy to predict a response variable from two datasets \mathbf{X} and \mathbf{Z} that contain two different sets of variables measured over the same set of n samples. The sizes of these two sets of variables are, respectively, p_x and p_z . To achieve collaboration among the two datasets, Gross and Tibshirani (2015) modified the typical objective function of a linear regression leading to Equation (2).

$$\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta}} \left\{ \frac{b_{xy}}{2} \|\mathbf{y} - \mathbf{X}(\boldsymbol{\beta})_X\|^2 + \frac{b_{zy}}{2} \|\mathbf{y} - \mathbf{Z}(\boldsymbol{\beta})_Z\|^2 + \frac{b_{xz}}{2} \|\mathbf{X}(\boldsymbol{\beta})_X - \mathbf{Z}(\boldsymbol{\beta})_Z\|^2 \right\} \quad (2)$$

Here, the coefficients $\boldsymbol{\beta}$ are split into two sets: the set of coefficients related to the p_x predictors from dataset \mathbf{X} , $(\boldsymbol{\beta})_X$ and the one related to the p_z predictors from dataset \mathbf{Z} , $(\boldsymbol{\beta})_Z$. The first two terms in the objective function separately optimize these two sets of coefficients, while the third term, $\|\mathbf{X}(\boldsymbol{\beta})_X - \mathbf{Z}(\boldsymbol{\beta})_Z\|^2$, promotes collaboration between the two datasets. Collaboration sprouts because this last term works as an $L2$ -penalty over the difference between the two different linear predictors, encouraging similarity between them. This stimulates a comparable contribution of the two datasets in predicting the response variable y . The weights of the three terms are, respectively, b_{xy} , b_{zy} and b_{xz} , and they control the importance of the single terms in the optimization.

3 Collaborative graphical lasso

Collaborative graphical lasso (coglasso) is the result of joining the advantages of *collaborative regression* with the well-established method for GGM estimation *glasso*. This is possible by incorporating the former's ability to leverage multiple data sets in the latter. Let us have two data sets, \mathbf{X} and \mathbf{Z} , respectively containing two different kinds of variables of size p_x and p_z measured over the same set of samples. Let $p = p_x + p_z$ be the total number of variables. Like *glasso*, *coglasso* computes an invertible estimate of the variance-covariance matrix, \mathbf{W} , whose inverse matrix is sparse. An issue with using *glasso* with this problem is that this algorithm is designed to handle only single data sets. Therefore, the only way to apply it would be to unify the two data sets as if they were measuring the same kind of variables. In contrast, when estimating a GGM, *coglasso* uses an original strategy to carefully leverage the contributions of each data set. This is possible due to a major rewriting of the *glasso* objective function reported in Equation (1), similar to how *collaborative regression* rewrote the basic regression strategy. First, *coglasso*'s objective function separates the contribution of the data sets \mathbf{X} and \mathbf{Z} into two separate terms. It then adds a third term, the collaborative term, that leverages the contributions of these two different components. Finally, a fourth term allows for a system of two different $L1$ -penalties: one penalizing the interactions *within* the variables belonging to the same data set and one for the interaction *between* different data sets. After these elements are incorporated, *coglasso*'s objective function takes the form of Equation (3).

$$\begin{aligned} \hat{\boldsymbol{\beta}}_i = \arg \min_{\boldsymbol{\beta}_i} & \left\{ \frac{1}{2} \left\| (\mathbf{W}_{\setminus i \setminus i})^{-1/2} \mathbf{S}_i - (\mathbf{W}_{\setminus i \setminus i})_X^{1/2} (\boldsymbol{\beta}_i)_X \right\|^2 + \frac{1}{2} \left\| (\mathbf{W}_{\setminus i \setminus i})^{-1/2} \mathbf{S}_i - (\mathbf{W}_{\setminus i \setminus i})_Z^{1/2} (\boldsymbol{\beta}_i)_Z \right\|^2 \right. \\ & \left. + \frac{c}{2} \left\| (\mathbf{W}_{\setminus i \setminus i})_X^{1/2} (\boldsymbol{\beta}_i)_X - (\mathbf{W}_{\setminus i \setminus i})_Z^{1/2} (\boldsymbol{\beta}_i)_Z \right\|^2 + \|\boldsymbol{\Lambda}_i \odot \boldsymbol{\beta}_i\|_1 \right\} \end{aligned} \quad (3)$$

In Equation (3) we partition the vector of coefficients $\boldsymbol{\beta}_i$ and of the $\mathbf{W}_{\setminus i \setminus i}$ sub-matrix from Equation (1) into two main components, one for each data set. This partitioning is

performed as follows. First, let us partition β_i into two sub-vectors: one containing the first p_x coefficients, $(\beta_i)_X$, and the other containing the remaining p_z coefficients, $(\beta_i)_Z$. These two sub-vectors of coefficients quantify, respectively, the contribution of the data set \mathbf{X} and that of the data set \mathbf{Z} to the optimization. As a second partitioning, let the square matrix $\mathbf{W}_{\setminus i \setminus i}$, of dimension $(p - 1) * (p - 1)$, be split into two rectangular vertical matrixes, such that $\mathbf{W}_{\setminus i \setminus i} = ((\mathbf{W}_{\setminus i \setminus i})_X, (\mathbf{W}_{\setminus i \setminus i})_Z)$. Each one of these two vertical sections can have two different dimensions, depending on the position of the index i . The vertical section $(\mathbf{W}_{\setminus i \setminus i})_X$ has dimensions $(p - 1) * p_x$ when the index i belongs to the set of coefficients p_z , while it has dimensions $(p - 1) * (p_x - 1)$ if the index i belongs to the set of coefficients p_x . Comparably, $(\mathbf{W}_{\setminus i \setminus i})_Z$ can either have dimensions $(p - 1) * (p_z - 1)$ if i belongs to p_z and $(p - 1) * p_z$ if it does not.

The first two terms in the objective function (3) separately optimize the value of the two sub-vectors $(\beta_i)_X$ and $(\beta_i)_Z$ in predicting the object $(\mathbf{W}_{\setminus i \setminus i})^{-1/2} \mathbf{S}_i$. Simultaneously, the third term penalizes the distance between the two predictors, harmonizing the contributions of the two data sets. This is the *collaborative term*. Here, c is the collaboration value that determines the relative importance of the collaborative term with respect to the first two terms. The fourth term is the penalty term. In this last term, the \odot symbolizes the Hadamard product, and Λ is the matrix of penalty parameters that allows to differently penalize *within* data set and *between* data set interactions. It is a $p * p$ matrix with two square diagonal blocks of dimensions $p_x * p_x$ and $p_z * p_z$, respectively, and two off-diagonal blocks of dimensions $p_x * p_z$ and $p_z * p_x$. The diagonal blocks contain the penalty parameter regulating the within-data set interactions, λ_w , while the off-diagonal blocks contain the parameter penalizing the between-data set interactions, λ_b . In practice, the former controls the density of connections between nodes belonging to the same datasets, while the latter controls the density of the connections between nodes belonging to different datasets. Finally, we fixed to 1 the value of the hyperparameters responsible for weighting the first two terms, corresponding to the parameters b_{xy} and b_{zy} of *collaborative regression* (see Equation (2)). This allows to search through the hyperparameter space focusing only on the hyperparameters we consider essential to be explored, namely c , λ_w and λ_b .

We solve Equation (3) iteratively to obtain $\hat{\beta}_i$ for $i = 1, 2, \dots, p, 1, 2, \dots, p, \dots$ until convergence. At every iteration this vector is used to update \mathbf{W} as described in Section 2.1 ($\mathbf{W}_i = \mathbf{W}_{\setminus i \setminus i} \beta_i$), then it is stored as the i -th column of a matrix $\hat{\mathbf{B}}$. To estimate the j -th element of β_i , *coglasso* adopts a coordinate descent procedure. We derived the coordinate update rule from Equation (3), and we present it in Algorithm 2. For the details about the derivation, we refer the reader to the Supplementary Materials. To keep the update rule more general, let j belong to the generic set of indexes A , which could index, for instance, variables of the data set \mathbf{X} .

Algorithm 2 *coglasso* coordinate descent rule

```

for  $j = 1, 2, \dots, p, 1, 2, \dots, p, \dots$ , until convergence criterion is met do
     $r \leftarrow \alpha S_{ji} + c\alpha \sum_{k \notin A} (\mathbf{W}_{\setminus i \setminus i})_{jk} (\hat{\beta}_i)_k - \sum_{k \in A, k \neq j} (\mathbf{W}_{\setminus i \setminus i})_{jk} (\hat{\beta}_i)_k$ 
    if  $r < -\alpha \Lambda_{ji}$  then
         $(\hat{\beta}_i)_j \leftarrow \frac{r + \alpha \Lambda_{ji}}{\mathbf{W}_{jj}}$ 
    else if  $r > \alpha \Lambda_{ji}$  then
         $(\hat{\beta}_i)_j \leftarrow \frac{r - \alpha \Lambda_{ji}}{\mathbf{W}_{jj}}$ 
    else
         $(\hat{\beta}_i)_j \leftarrow 0$ 
    end if
end for

```

Here, $\alpha = \frac{1}{1+c}$. There are remarkable differences with *glasso*'s update rule seen in Algorithm 1 to update $(\hat{\beta}_i)_j$. First, here entry S_{ji} is weighted by α . Second, the dot product that *glasso* subtracts to S_{ji} here is divided into two components with two different roles. The second component of the two is subtracted to αS_{ji} , corresponding to the elements with indexes contained in the set A , the same set where index j belongs. In contrast, the first component, added after being weighted by $c\alpha$, corresponds to all the other elements that have indexes not contained in A . This added component allows *coglasso* to harmonize the role of the elements contained in the set A in computing r with all the other ones not contained in it. A final difference is that in *coglasso* the penalty interval is now defined as $(-\alpha\Lambda_{ji}, \alpha\Lambda_{ji})$. Then, similarly to *glasso*, if r falls out of it, we assign to $(\hat{\beta}_i)_j$ a value closer than r to the interval, while if r falls inside it we assign 0 to it.

We implemented *coglasso* algorithm in C++, and we developed an R package as an interface to it, naming it *coglasso*. *Coglasso* is currently distributed through CRAN, and, for the moment, it only accepts two data sets. For the C++ implementation, we were inspired by Zhao et al. (2012).

Remark 3.1 - Even if the objective function and the current implementation of *coglasso* are tailored over a group of two data sets of interest, a generalization to a higher number of data sets is straightforward. This can be seen in Algorithm 2, where it is not specified to which data set the A index set belongs. In principle, A could indicate the indexes of one of many possible data sets. In this case, the third element of the formula to compute r would directly take into account the contribution of *all the other* data sets.

3.1 Stability selection for collaborative graphical lasso

Coglasso requires the setting of three hyperparameters: λ_w , λ_b and c . The optimal choice of hyperparameters can be a challenging task that is usually tackled with model selection procedures. An attractive model selection procedure developed to choose the best λ for *glasso* is *StARS* (Liu et al., 2010). *StARS* chooses the value of λ between a descending array of values for which the network is as sparse as possible, yet stable. Nevertheless, to select the optimal *coglasso* network the hyperparameter space to be explored spans three dimensions. Of these three dimensions, λ_w and λ_b directly influence the sparsity of a network, hence its stability, while c exerts its main effect on the global model performance. We designed an alternative version of *StARS* able to explore the three-dimensional hyperparameter space of *coglasso*. Algorithm 3 describes the outline of *StARS* for *coglasso*, with c , λ_b , and λ_w being the arrays of explored hyperparameters. Here, the function *StARS*(λ) returns the λ in an array λ yielding the most stable *coglasso* network, after fixing c and one of the two penalty parameters to a given value.

4 Simulations

We simulated networks to assess *coglasso*'s network-reconstructing performance. In this evaluation, *coglasso* was compared with the well-established original *glasso* algorithm from Friedman et al. (2008). The designed simulation scheme was the following.

The *HUGE* R package (Zhao et al., 2012) was used to generate six sets of 200 networks each, arranged in two triplets. Of these, a triplet of sets belonged to the “scale-free” network family, while the other belonged to the “random” family. “scale-free” networks are generated with the Barabási–Albert algorithm (Barabási and Albert, 1999), and “random” networks by establishing a connection between every pair of nodes with probability $P(\text{edge}) = 3/p$. Each triplet was composed of a set of networks of 25 nodes, one of 100 nodes, and one of 200 nodes. With this design, we aimed to explore scenarios of increasing p to n ratios. For each

Algorithm 3 *StARS* for *coglasso*

```
MAX_ITER  $\leftarrow 10$  (as a default value)
for  $c_i$  in  $\mathbf{c}$  do
    converged  $\leftarrow FALSE$ 
     $\hat{\lambda}_b \leftarrow Max(\boldsymbol{\lambda}_b)$ 
     $\hat{\lambda}_w \leftarrow -1$ 
    iter  $\leftarrow 0$ 
    while NOT converged AND iter  $< MAX\_ITER$  do
         $\lambda_w^{temp} \leftarrow StARS(\boldsymbol{\lambda}_w)$ , with  $c \leftarrow c_i$  and  $\lambda_b \leftarrow \hat{\lambda}_b$ 
        if  $\lambda_w^{temp} = \hat{\lambda}_w$  then
            converged  $\leftarrow TRUE$ 
        else
             $\hat{\lambda}_w \leftarrow \lambda_w^{temp}$ 
             $\lambda_b^{temp} \leftarrow StARS(\boldsymbol{\lambda}_b)$ , with  $c \leftarrow c_i$  and  $\lambda_w \leftarrow \hat{\lambda}_w$ 
            if  $\lambda_b^{temp} = \hat{\lambda}_b$  then
                converged  $\leftarrow TRUE$ 
            else
                 $\hat{\lambda}_b \leftarrow \lambda_b^{temp}$ 
            end if
        end if
        iter  $\leftarrow iter + 1$ 
    end while
end for
```

Select the $(\hat{c}_i, \hat{\lambda}_b, \hat{\lambda}_w)$ combination that yielded the most stable *coglasso* network

network in these six sets, we extracted the corresponding precision matrix. This is computed as $\boldsymbol{\Theta} = v\mathbf{A} + (\|e\|_1 + 0.1 + u)\mathbf{I}$, where A is the adjacency matrix encoding the connections of the generated network. Here, we chose the default values for $v = 0.3$ and $u = 0.1$, while e is the smallest eigenvalue of the matrix $v\mathbf{A}$. For each one of the extracted precision matrices, we simulated a data set of sample size 100 distributed according to the associated multivariate normal distribution. These simulated data sets were used as input data to the compared algorithms *coglasso* and to *glasso*. Since *coglasso* expects to receive two “collaborating” data sets as input, we fictitiously split each of the generated data sets into two halved data sets without altering the data themselves. Remarkably, this situation is not ideal for *coglasso*. The reason is that our algorithm assumes a more intertwined role of the two data sets when it incorporates the collaborative term. Instead, here we used only a single data set, sampled from a single multivariate normal distribution, and divided it into two halves as if we had two different data sets.

Among the three possible hyperparameters of *coglasso*: λ_w , λ_b and c , we were mostly interested in exploring its performance under different collaboration amounts (i.e. different values for c). c is the collaboration value, the weight given to the collaborative term in the objective function, and it fixes the relative importance of such term with respect to the first two terms in equation 3. We hence explored four values: $c = 0.2$ (1/5 the importance of the other terms), $c = 1$ (same importance), $c = 2$ (twice the importance) and $c = 5$ (5 times the importance). Moreover, we decided to set $\lambda = \lambda_w = \lambda_b$ since we wanted to compare *coglasso* with *glasso*, and the latter has a single penalty parameter λ controlling the general sparsity uniformly throughout the network. We performed network reconstruction with the same 30 λ values for both *coglasso* and *glasso*. Hence, for each simulated data set and lambda value, we performed four network reconstructions with *coglasso* (one for each c value) and one with *glasso*. For each reconstruction, the recovered connections were compared to the ones of the

reference simulated network, and a confusion matrix of the reconstructed connections was built. For each reference network and collaboration value, the computed confusion matrixes were used to compute the Area Under the Receiver Operator Curve (*AUC*) over the 30 λ values for performance comparison with *glasso*. Each *AUC* value was scaled by the maximum *FPR* value of the curve multiplied by the maximum *TPR* value to obtain comparable values on a scale from 0 to 1. The results of these simulation analyses are displayed in Figure 1.

Each plot in Figure 1 illustrates scaled *AUCs* computed over 200 generated networks and 30 λ values, for the four collaboration values and for *glasso*. Each one represents a different combination of network family and network size, namely “scale-free” and 25 nodes, “scale-free” and 100 nodes, “scale-free” and 200 nodes, “random” and 25 nodes, “random” and 100 nodes, “random” and 200 nodes. On the whole, *glasso* and *coglasso* did not have remarkably different performances. This result was unexpected since the simulated data are not ideal for *coglasso*, while they represent a typical use case for *glasso*. Nevertheless, variation in performance for different collaboration values suggests that a parameter selection procedure may be required to reconstruct the most fine-tuned network when running *coglasso*.

4.1 Model selection performance

We employed the same set of simulated networks used to compare *coglasso* with *glasso* to test the performance of *StARS* for *coglasso*. In particular, we used “scale-free” and “random” networks of 25 and 200 nodes. We chose to measure the F_1 scores of the selected networks across the 200 replicates for each category to measure model selection performance. We compare the F_1 scores of *StARS*-selected models with the F_1 scores of “oracle” models. For every replicate, the “oracle” model is the one with the hyperparameters combination yielding the best performance *coglasso* could achieve. In Table 1 we compare the average F_1 scores of the networks selected via *StARS* for *coglasso* with the ones of “oracle” networks. The table shows how on average *StARS* for *coglasso* selects models that are close in performance from the average oracle model that *coglasso* can achieve.

4.2 Timing comparison

Figure 2 shows the average time required for *coglasso* and for *glasso* to estimate a single network from an input empirical variance-covariance matrix. The average time is taken over the same range of 30 λ values used to evaluate network reconstruction performance. The explored c values are 0.2, 1 and 5, and the explored network sizes are 25, 100 and 200 nodes. The two algorithms exhibited comparable computing times even though *glasso* performed moderately better. For each one of the three explored sizes, the median duration of network estimation by *coglasso* across all the explored c values was, respectively, 31.2%, 42.0% and 57.3% higher than the median duration of *glasso*. The reason behind this lies in the different update rules of the two algorithms. To ensure collaboration between the two datasets, it is unavoidable to make a distinction between them in the update rule, which causes the algorithm to perform more operations per iteration. Nevertheless, we see that different collaboration values do not remarkably impact the computing times of *coglasso*.

5 Real data example

We illustrate the proposed method using a multi-omics data set from Diessler et al. (2018) to study the molecular biology behind sleep deprivation (SD). This study compares transcriptomic and metabolomic measurements in sleep-deprived mice with non-sleep-deprived mice. The data set we selected consists of a sample of 30 diverse mouse lines, $p_x = 14896$ genes

measured in the cortex and $p_z = 124$ blood circulating metabolites, measured in the SD condition. The original study also performs differential expression analysis, providing a list of differentially expressed genes and differentially expressed metabolites. Moreover, the authors mention a list of 78 genes known to be associated with SD from Mongrain et al. (2010). We decided to use this information to reduce the dimensionality of the two omics data sets. From the transcriptomic data set, we extracted the union of the 78 known genes and the top 100 differentially expressed genes, resulting in a list of 162 genes. Similarly, from the metabolomic data set we extracted the 76 differentially expressed metabolites. We used *coglasso* to estimate the network from the two aggregated data sets. Then, we applied the adapted *StARS* model selection to select the optimal combination of λ_b , λ_w and c . We run our method over 30 possible λ_b and λ_w values, and over 9 c values, leading to a total of 8100 combinations of hyperparameters. Figure 3 presents the inferred network by setting the selected combination of hyperparameters ($\lambda_b = 0.37$, $\lambda_w = 0.70$ and $c = 100$). Blue nodes represent transcripts, while pink nodes represent metabolites. The figure shows how the two groups of variables are highly blended, representing a large amount of connections shared among them.

To explore the biological meaningfulness of a network generated with *coglasso*, we followed two strategies: the first one knowledge-based and the second one data-driven. In both instances, we were able to show how *coglasso* could retrieve connections that have previously been experimentally validated.

For the knowledge-based driven approach, we assessed the available literature on the molecular biology behind sleep. Hoekstra et al. (2019) investigated the role of Cold-Induced RNA Binding Protein (*Cirbp*) as a molecular regulator of SD response. Figure 4, on the left, shows the subnetwork of *Cirbp* and its neighbouring nodes. Among the targets Hoekstra et al. (2019) investigated, two genes encoding for heat-shock proteins: *Hspa5* and *Hsp90b1*, had an acutely increased induction upon SD in *cirbp* knock-out mutant mice. This indicates that these two genes act in response to SD under the regulation of *Cirbp*. In the selected *coglasso* network, *Cirbp* was connected to both these genes, meaning that *coglasso* was able to reconstruct connections that have previously been validated. These two genes are involved in unfolded protein response (UPR), a process known to be activated during SD (Ji et al., 2011; Sun et al., 2019; Mackiewicz et al., 2007). Interestingly, eight additional genes in the neighbourhood of *Cirbp* are associated with UPR or, in general, with protein folding (Soldà et al., 2006; Vekich et al., 2012; Kern et al., 2021; Lin et al., 2021; Mizobuchi et al., 2007; Genereux et al., 2015; Tao and Sha, 2011; Yamagishi et al., 2011). Hence, *coglasso* estimated biologically coherent connections between *Cirbp* and other UPR-involved genes. This suggests that *Cirbp* acts as a regulator of the UPR upon SD, a hypothesis that could be experimentally tested. Additionally, the connection between *Cirbp* and tryptophan (Trp) suggests its regulatory role in response to SD, since Trp is a precursor to serotonin, one of the key hormones in sleep regulation (Minet-Ringuet et al., 2004).

We then inspected the generated network by a data-driven approach, performing community discovery with the algorithm described by Clauset et al. (2004). We focused on the second-largest community that the algorithm identified (shown in Figure 4, on the right, the community is highlighted in the global network in Figure S1). Among other metabolites, the second community contained the majority of the amino acids recorded in the data set: alanine, arginine, asparagine, histidine, isoleucine, leucine, lysine, methionine, ornithine, phenylalanine, serine, threonine, tyrosine and valine. Among the transcripts, it included all the genes belonging to the *Fos/Jun* and the *Egr* transcription factor families that were present in our network: *Fos*, *Fosb*, *Fosl2*, *Junb*, *Egr1*, *Egr2*, and *Egr3*. Their nodes are those highlighted in the figure. Several previous studies reported that various members of the two families participate in the amino acid starvation response (AAR, (Kilberg et al., 2012)). For example, transcriptomic studies in mouse (Deval et al., 2009) and human cells (Shan et al., 2010) showed induction of members of both families. In particular, in Shan et al. (2010), *Junb* exhibited the

greatest induction upon histidine deprivation, while *Fos* was the most activated member upon leucine starvation in Deval et al. (2009). *Coglasso* recovered both these connections in the second community. Previously, Pohjanpelto and H\"olt\"a (1990) separately studied the induction of targeted members of the *Fos/Jun* family to methionine deprivation in hamster cells, with *Junb* showing again the strongest response. This connection is present in the community too. Other targeted studies verified the role during AAR of *Egr1* (Shan et al., 2014, 2019) and *Fos* (Shan et al., 2015) in human cells. Since the two transcription factor families belong to the group of immediate-early response genes and take part in AAR, their expression is expected to fluctuate with high sensitivity accordingly with variations in amino acid concentrations (Healy et al., 2013). Altogether, these data suggest that *coglasso* was able to reconstruct several to-be-expected or previously validated connections between transcripts and amino acids, grouping the respective nodes in a densely interconnected, biologically relevant community.

6 Discussion

In this study, we addressed the issue of estimating a network from multiple high-dimensional data sets of variables that are different in nature recorded on the same individuals (*e.g.* multi-omics data sets). In particular, we focused on GGM estimation in a multi-assay setting. To tackle this question, we introduced *coglasso*, a new algorithm that is able to leverage the contribution of diverse data sets in the process of GGM estimation. *Coglasso* is based on major changes in *glasso* (Friedman et al., 2008), a well-known GGM estimation algorithm. The cornerstone of this alteration was the introduction of a collaborative term in the objective function, similar to what Gross and Tibshirani (2015) described with *collaborative regression*. After designing and implementing *coglasso*, we then compared network reconstruction performances of *coglasso* and *glasso* by simulation studies. The comparisons showed that *coglasso* and *glasso* perform similarly well in a *glasso*-favourable scenario. Regardless of the network size or the network family, the two methods estimated networks of comparable quality and in comparable amounts of time. This similar performance was in part encouraging since simulated data sets represent a typical use-case for *glasso*, while they do not represent the ideal scenario *coglasso* is designed for. The real-world multi-omics data sets are expected to have a remarkably intertwined behaviour, which is our algorithm's ideal, but to our knowledge, currently not reproducible by simulations. Across our simulations, we observed instances where *coglasso* struggled to reach convergence, even though this was restricted to regions of the hyperparameter space characterized by low c value and that lead to high network density. Even though this phenomenon does not originate from the region of the hyperparameter space generating sparse networks, we intend to investigate it further in the future. Simulation studies also showed that different weights given to the collaborative term influence network reconstruction performance. This indicates that collaboration has a role in determining the best network structure. This role of the c value, together with the regulatory role of the other two hyperparameters (*i.e.* λ_w and λ_b) made it necessary to develop an appropriate model selection strategy for *coglasso*. One of the best approaches to address this for *glasso*, based on the network stability criterion, is *StARS* (Liu et al., 2010). Nevertheless, this strategy was designed to handle a single hyperparameter. Therefore, we had to extend *StARS* to cope with the additional hyperparameters. Simulations show that the extended version of *StARS* for *coglasso* tends to select a model that is not far from the best one our algorithm can achieve, making it a suitable model selection procedure.

We then illustrated *coglasso*'s ability to reconstruct networks from real-world multi-omics data. We chose a multi-omics data set studying transcriptomics and metabolomics of sleep deprivation in mice (Hoekstra et al., 2019). *Coglasso*, together with our novel model selection procedure, was able to retrieve connections that had previously been validated, both

surrounding the gene *Cirbp* and in the community of the network most enriched with amino acids. Moreover, the first example allowed the formulation of additional hypotheses on the regulatory behaviour of the gene *Cirbp* in sleep deprivation, and we encourage the field of sleep studies to assess these on biological relevance.

Our results show that *coglasso* represents a novel relevant alternative in the quest for multi-omics network reconstruction, as it uniquely leverages the multi-assay nature of multi-omics data. Nevertheless, like many other GGM estimation algorithms, *coglasso* relies on the normality assumption, which is rarely true for multi-omics data. In recent years biostatisticians have attempted to solve this major issue by resorting to copula-based techniques (Behrouzi and Wit, 2019; Cougoul et al., 2019; Vinciotti et al., 2022). Therefore we expect that flanking our algorithm with a copula-based approach could further increase the effectiveness of *coglasso*, allowing the transition of multi-omics data to the normal realm. This would create a novel strategy to reconstruct networks from any combination of omics data. Additionally, the current implementation of *coglasso* as an R package accepts two data sets. However, multi-omics studies could generate more than two types of data. Therefore, future implementations of *coglasso* should be able to handle even more data sets.

Finally, even though we designed *coglasso* as an effective strategy to integrate multi-omics data, it remains a general method. This implies that *coglasso* has a wide margin of applicability in any scientific field that could be interested in reconstructing networks leveraging data sets of different natures. A relevant example could be in psychological science, for instance in the fields of personality research or psychopathology. In both these fields, the main GGM estimation strategy employed is *glasso*, despite variables of different kinds arising routinely. For example, personality research could aim to build a model linking personality traits and motivational goals, while psychopathology could be interested in connecting observable signs and subjective symptoms (Borsboom et al., 2021). This suggests that both these fields and, in general, psychological networks, could benefit from the multi-assay strategy provided by *coglasso*.

Acknowledgments

We thank Prof. Fred van Eeuwijk and Dr. Gwenaël G.R. Leday (Wageningen University and Research) for precious discussions during the development of *coglasso*. We thank Dr. Michael Schon (Wageningen University and Research) for the valuable intellectual discussions on the biological application. This work was partially funded by the Plant Science Group, and by the departments of Mathematical and Statistical Methods (Biometris) and of Cellular and Developmental Biology at Wageningen University and Research (project number 3183500140), and by the NWO VIDI grant with number VI.Vidi.193.119.

Conflict of Interest: None declared.

References

- Michael Altenbuchinger, Antoine Weihs, John Quackenbush, Hans Jörgen Grabe, and Helena U. Zacharias. Gaussian and Mixed Graphical Models as (multi-)omics data analysis tools. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1863(6):194418, June 2020. ISSN 1874-9399. doi: 10.1016/j.bbagr.2019.194418. URL <https://www.sciencedirect.com/science/article/pii/S187493991930224X>.
- Albert-László Barabási and Réka Albert. Emergence of Scaling in Random Networks. *Science*, 286(5439):509–512, October 1999. doi: 10.1126/science.286.5439.509. URL <https://www.science.org/doi/full/10.1126/science.286.5439.509>.

Pariya Behrouzi and Ernst C. Wit. Detecting epistatic selection with partially observed genotype data by using copula graphical models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 68(1):141–160, 2019. ISSN 1467-9876. doi: 10.1111/rssc.12287. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/rssc.12287>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/rssc.12287>.

Denny Borsboom, Marie K. Deserno, Mijke Rhemtulla, Sacha Epskamp, Eiko I. Fried, Richard J. McNally, Donald J. Robinaugh, Marco Perugini, Jonas Dalege, Giulio Costantini, Adela-Maria Isvoranu, Anna C. Wysocki, Claudia D. van Borkulo, Riet van Bork, and Lourens J. Waldorp. Network analysis of multivariate data in psychological science. *Nature Reviews Methods Primers*, 1(1):1–18, August 2021. ISSN 2662-8449. doi: 10.1038/s43586-021-00055-w. URL <https://www.nature.com/articles/s43586-021-00055-w>.

Aaron Clauset, M. E. J. Newman, and Christopher Moore. Finding community structure in very large networks. *Physical Review E*, 70(6):066111, December 2004. doi: 10.1103/PhysRevE.70.066111. URL <https://link.aps.org/doi/10.1103/PhysRevE.70.066111>.

Arnaud Cougoul, Xavier Bailly, and Ernst C. Wit. MAGMA: inference of sparse microbial association networks, February 2019. URL <https://www.biorxiv.org/content/10.1101/538579v1>. Pages: 538579 Section: New Results.

Christiane Deval, Cédric Chaveroux, Anne-Catherine Maurin, Yoan Cherasse, Laurent Parry, Valérie Carraro, Dragan Milenkovic, Marc Ferrara, Alain Bruhat, Céline Jousse, and Pierre Fafournoux. Amino acid limitation regulates the expression of genes involved in several specific biological processes through GCN2-dependent and GCN2-independent pathways. *The FEBS Journal*, 276(3):707–718, 2009. ISSN 1742-4658. doi: 10.1111/j.1742-4658.2008.06818.x. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1742-4658.2008.06818.x>.

Shanaz Diessler, Maxime Jan, Yann Emmenegger, Nicolas Guex, Benita Middleton, Debra J. Skene, Mark Ibberson, Frederic Burdet, Lou Götz, Marco Pagni, Martial Sankar, Robin Liechti, Charlotte N. Hor, Ioannis Xenarios, and Paul Franken. A systems genetics resource and analysis of sleep regulation in the mouse. *PLOS Biology*, 16(8):e2005750, August 2018. ISSN 1545-7885. doi: 10.1371/journal.pbio.2005750. URL <https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.2005750>. Publisher: Public Library of Science.

Jerome Friedman, Trevor Hastie, Holger Höfling, and Robert Tibshirani. Pathwise coordinate optimization. *The Annals of Applied Statistics*, 1(2):302–332, December 2007. ISSN 1932-6157, 1941-7330. doi: 10.1214/07-AOAS131. URL <https://projecteuclid.org/journals/annals-of-applied-statistics/volume-1/issue-2/Pathwise-coordinate-optimization/10.1214/07-AOAS131.full>.

Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, July 2008. ISSN 1465-4644. doi: 10.1093/biostatistics/kxm045. URL <https://doi.org/10.1093/biostatistics/kxm045>.

Joseph C Genereux, Song Qu, Minghai Zhou, Lisa M Ryno, Shiyu Wang, Matthew D Shoulters, Randal J Kaufman, Corinne I Lasmézas, Jeffery W Kelly, and R Luke Wiseman. Unfolded protein response-induced ERdj3 secretion links ER stress to extracellular proteostasis. *The EMBO Journal*, 34(1):4–19, January 2015. ISSN 0261-4189. doi: 10.15252/embj.201488896. URL <https://www.embopress.org/doi/full/10.15252/embj.201488896>.

Samuel M. Gross and Robert Tibshirani. Collaborative regression. *Biostatistics*, 16(2):326–338, April 2015. ISSN 1465-4644. doi: 10.1093/biostatistics/kxu047. URL <https://doi.org/10.1093/biostatistics/kxu047>.

Shannon Healy, Protiti Khan, and James R. Davie. Immediate early response genes and cell transformation. *Pharmacology & Therapeutics*, 137(1):64–77, January 2013. ISSN 0163-7258. doi: 10.1016/j.pharmthera.2012.09.001. URL <https://www.sciencedirect.com/science/article/pii/S0163725812002033>.

Marieke MB Hoekstra, Yann Emmenegger, Jeffrey Hubbard, and Paul Franken. Cold-inducible RNA-binding protein (CIRBP) adjusts clock-gene expression and REM-sleep recovery following sleep deprivation. *eLife*, 8:e43400, February 2019. ISSN 2050-084X. doi: 10.7554/eLife.43400. URL <https://doi.org/10.7554/eLife.43400>.

Cheng Ji, Neil Kaplowitz, Mo Yin Lau, Eddy Kao, Lydia M. Petrovic, and Amy S. Lee. Liver-specific loss of GRP78 perturbs the global unfolded protein response and exacerbates a spectrum of acute and chronic liver diseases. *Hepatology (Baltimore, Md.)*, 54(1):229–239, July 2011. ISSN 0270-9139. doi: 10.1002/hep.24368. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3125405/>.

Paul Kern, Nora R. Balzer, Nelli Blank, Cornelia Cygon, Klaus Wunderling, Franziska Bender, Alex Frolov, Jan-Peter Sowa, Lorenzo Bonaguro, Thomas Ulas, Mirka Homrich, Eva Kiermaier, Christoph Thiele, Joachim L. Schultze, Ali Canbay, Reinhart Bauer, and Elvira Mass. Creld2 function during unfolded protein response is essential for liver metabolism homeostasis. *The FASEB Journal*, 35(10):e21939, 2021. ISSN 1530-6860. doi: 10.1096/fj.202002713RR. URL <https://onlinelibrary.wiley.com/doi/abs/10.1096/fj.202002713RR>.

Michael S. Kilberg, Mukundh Balasubramanian, Lingchen Fu, and Jixiu Shan. The Transcription Factor Network Associated With the Amino Acid Response in Mammalian Cells. *Advances in Nutrition*, 3(3):295–306, May 2012. ISSN 2161-8313. doi: 10.3945/an.112.001891. URL <https://www.sciencedirect.com/science/article/pii/S216183132200998X>.

Lilian Tsai-Wei Lin, Abdul Razzaq, Sonja E. Di Gregorio, Soojie Hong, Brendan Charles, Marilene H. Lopes, Flavio Beraldo, Vania F. Prado, Marco A. M. Prado, and Martin L. Duennwald. Hsp90 and its co-chaperone Sti1 control TDP-43 misfolding and toxicity. *The FASEB Journal*, 35(5):e21594, 2021. ISSN 1530-6860. doi: 10.1096/fj.202002645R. URL <https://onlinelibrary.wiley.com/doi/abs/10.1096/fj.202002645R>.

Han Liu, Kathryn Roeder, and Larry Wasserman. Stability Approach to Regularization Selection (StARS) for High Dimensional Graphical Models, June 2010. URL <http://arxiv.org/abs/1006.3316>. arXiv:1006.3316 [stat].

Miroslaw Mackiewicz, Keith R. Shockley, Micah A. Romer, Raymond J. Galante, John E. Zimmerman, Nirinjini Naidoo, Donald A. Baldwin, Shane T. Jensen, Gary A. Churchill, and Allan I. Pack. Macromolecule biosynthesis: a key function of sleep. *Physiological Genomics*, 31(3):441–457, November 2007. ISSN 1531-2267. doi: 10.1152/physiolgenomics.00275.2006.

J Minet-Ringuet, P. M Le Ruyet, D Tomé, and P. C Even. A tryptophan-rich protein diet efficiently restores sleep after food deprivation in the rat. *Behavioural Brain Research*, 152(2):335–340, July 2004. ISSN 0166-4328. doi: 10.1016/j.bbr.2003.10.018. URL <https://www.sciencedirect.com/science/article/pii/S0166432803003814>.

Naomi Mizobuchi, Jun Hoseki, Hiroshi Kubota, Shinya Toyokuni, Jun-ichi Nozaki, Motoko Naitoh, Akio Koizumi, and Kazuhiro Nagata. ARMET is a soluble ER protein induced by

the unfolded protein response via ERSE-II element. *Cell Structure and Function*, 32(1): 41–50, 2007. ISSN 1347-3700. doi: 10.1247/csf.07001.

Valérie Mongrain, Susana A. Hernandez, Sylvain Pradervand, Stéphane Dorsaz, Thomas Curie, Grace Hagiwara, Phung Gip, H . Craig Heller, and Paul Franken. Separating the Contribution of Glucocorticoids and Wakefulness to the Molecular and Electrophysiological Correlates of Sleep Homeostasis. *Sleep*, 33(9):1147–1157, September 2010. ISSN 0161-8105. doi: 10.1093/sleep/33.9.1147. URL <https://doi.org/10.1093/sleep/33.9.1147>.

Pirkko Pohjanpelto and Erkki Hölttä. Deprivation of a Single Amino Acid Induces Protein SynthesisDependent Increases in c-jun, c-myc, and Ornithine Decarboxylase mRNAs in Chinese Hamster Ovary Cells. *Molecular and Cellular Biology*, 10(11):5814–5821, November 1990. ISSN null. doi: 10.1128/mcb.10.11.5814-5821.1990. URL <https://doi.org/10.1128/mcb.10.11.5814-5821.1990>.

Jixiu Shan, Maria-Cecilia Lopez, Henry V. Baker, and Michael S. Kilberg. Expression profiling after activation of amino acid deprivation response in HepG2 human hepatoma cells. *Physiological Genomics*, 41(3):315–327, May 2010. ISSN 1094-8341. doi: 10.1152/physiolgenomics.00217.2009. URL <https://journals.physiology.org/doi/full/10.1152/physiolgenomics.00217.2009>.

Jixiu Shan, Mukundh N. Balasubramanian, William Donelan, Lingchen Fu, Jaclyn Hayner, Maria-Cecilia Lopez, Henry V. Baker, and Michael S. Kilberg. A Mitogen-activated Protein Kinase/Extracellular Signal-regulated Kinase Kinase (MEK)-dependent Transcriptional Program Controls Activation of the Early Growth Response 1 (EGR1) Gene during Amino Acid Limitation *. *Journal of Biological Chemistry*, 289(35):24665–24679, August 2014. ISSN 0021-9258, 1083-351X. doi: 10.1074/jbc.M114.565028. URL [https://www.jbc.org/article/S0021-9258\(20\)31997-9/abstract](https://www.jbc.org/article/S0021-9258(20)31997-9/abstract).

Jixiu Shan, William Donelan, Jaclyn N. Hayner, Fan Zhang, Elizabeth E. Dudenhausen, and Michael S. Kilberg. MAPK signaling triggers transcriptional induction of cFOS during amino acid limitation of HepG2 cells. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1853(3):539–548, March 2015. ISSN 0167-4889. doi: 10.1016/j.bbamcr.2014.12.013. URL <https://www.sciencedirect.com/science/article/pii/S0167488914004418>.

Jixiu Shan, Elizabeth Dudenhausen, and Michael S. Kilberg. Induction of early growth response gene 1 (EGR1) by endoplasmic reticulum stress is mediated by the extracellular regulated kinase (ERK) arm of the MAPK pathways. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1866(3):371–381, March 2019. ISSN 0167-4889. doi: 10.1016/j.bbamcr.2018.09.009. URL <https://www.sciencedirect.com/science/article/pii/S0167488918304130>.

Tatiana Soldà, Natalio Garbi, Günter J. Hä默ling, and Maurizio Molinari. Consequences of ERp57 Deletion on Oxidative Folding of Obligate and Facultative Clients of the Calnexin Cycle*. *Journal of Biological Chemistry*, 281(10):6219–6226, March 2006. ISSN 0021-9258. doi: 10.1074/jbc.M513595200. URL <https://www.sciencedirect.com/science/article/pii/S0021925819579765>.

Ming Sun, Judy L. M. Kotler, Shanshan Liu, and Timothy O. Street. The endoplasmic reticulum (ER) chaperones BiP and Grp94 selectively associate when BiP is in the ADP conformation. *Journal of Biological Chemistry*, 294(16):6387–6396, April 2019. ISSN 0021-9258. doi: 10.1074/jbc.RA118.007050. URL <https://www.sciencedirect.com/science/article/pii/S0021925820362967>.

Jiahui Tao and Bingdong Sha. Chapter fifteen - Structural Insight into the Protective Role of P58(IPK) during Unfolded Protein Response. In P. Michael Conn, editor, *Methods in Enzymology*, volume 490 of *The Unfolded Protein Response and Cellular Stress, Part B*, pages 259–270. Academic Press, January 2011. doi: 10.1016/B978-0-12-385114-7.00015-5. URL <https://www.sciencedirect.com/science/article/pii/B9780123851147000155>.

John A. Vekich, Peter J. Belmont, Donna J. Thuerauf, and Christopher C. Glembotski. Protein disulfide isomerase-associated 6 is an ATF6-inducible ER stress response protein that protects cardiac myocytes from ischemia/reperfusion-mediated cell death. *Journal of Molecular and Cellular Cardiology*, 53(2):259–267, August 2012. ISSN 0022-2828. doi: 10.1016/j.yjmcc.2012.05.005. URL <https://www.sciencedirect.com/science/article/pii/S0022282812001824>.

Veronica Vinciotti, Pariya Behrouzi, and Reza Mohammadi. Bayesian structural learning of microbiota systems from count metagenomic data, March 2022. URL <http://arxiv.org/abs/2203.10118>. arXiv:2203.10118 [stat].

Nobuyuki Yamagishi, Masayasu Yokota, Kunihiko Yasuda, Youhei Saito, Kazuhiro Nagata, and Takumi Hatayama. Characterization of stress sensitivity and chaperone activity of Hsp105 in mammalian cells. *Biochemical and Biophysical Research Communications*, 409(1):90–95, May 2011. ISSN 0006-291X. doi: 10.1016/j.bbrc.2011.04.114. URL <https://www.sciencedirect.com/science/article/pii/S0006291X11007121>.

Tuo Zhao, Han Liu, Kathryn Roeder, John Lafferty, and Larry Wasserman. The huge Package for High-dimensional Undirected Graph Estimation in R. *Journal of Machine Learning Research*, 13(37):1059–1062, 2012. ISSN 1533-7928. URL <http://jmlr.org/papers/v13/zha012a.html>.

Table 1: Evaluation of model selection with *StARS* for *coglasso*. Values represent the average of the F_1 scores of 200 replicates of network reconstruction, with their relative standard error in brackets. In all instances, the sample size is 100.

Method	scale-free, 25 nodes	scale-free, 200 nodes
<i>StARS</i>	0.63 (0.14)	0.20 (0.04)
Oracle	0.75 (0.07)	0.28 (0.05)
Method	random, 25 nodes	random, 200 nodes
<i>StARS</i>	0.61 (0.11)	0.42 (0.02)
Oracle	0.75 (0.06)	0.53 (0.04)

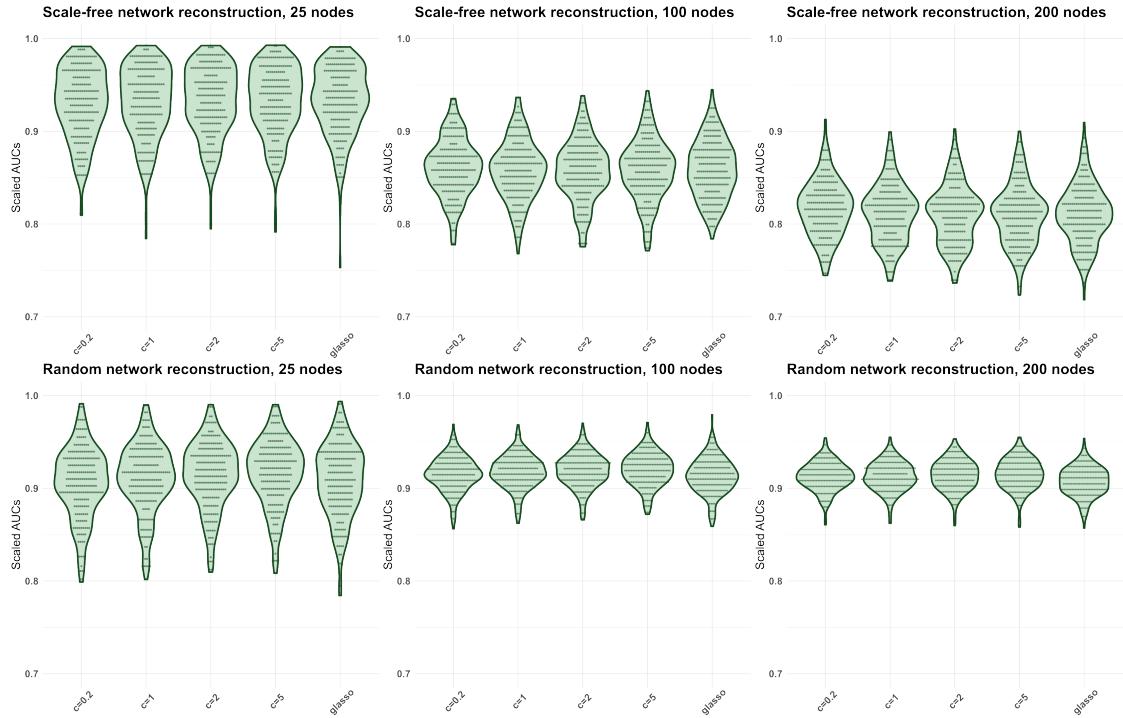


Figure 1: Results of simulated networks reconstruction. Each plot represents scaled AUC values of a different combination of network family (“scale-free” and “random”) and network size (25, 100 and 200 nodes), measured for *coglasso* over four c values (0.2, 1, 2 and 5) and for *glasso*. The sample size of the data sets generated from each network is 100.

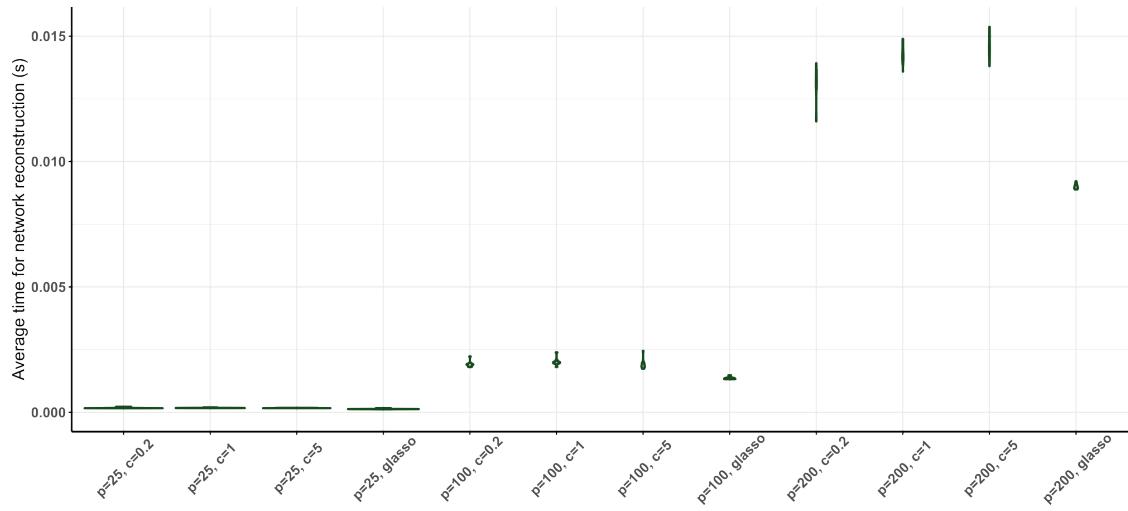


Figure 2: Time for simulated network reconstruction, averaged over 30 λ parameters. Networks of three different sizes (25, 100 and 200 nodes) were reconstructed by *glasso* and by *coglasso* over three c values (0.2, 1 and 5). The sample size of the data sets generated from each network is 100.

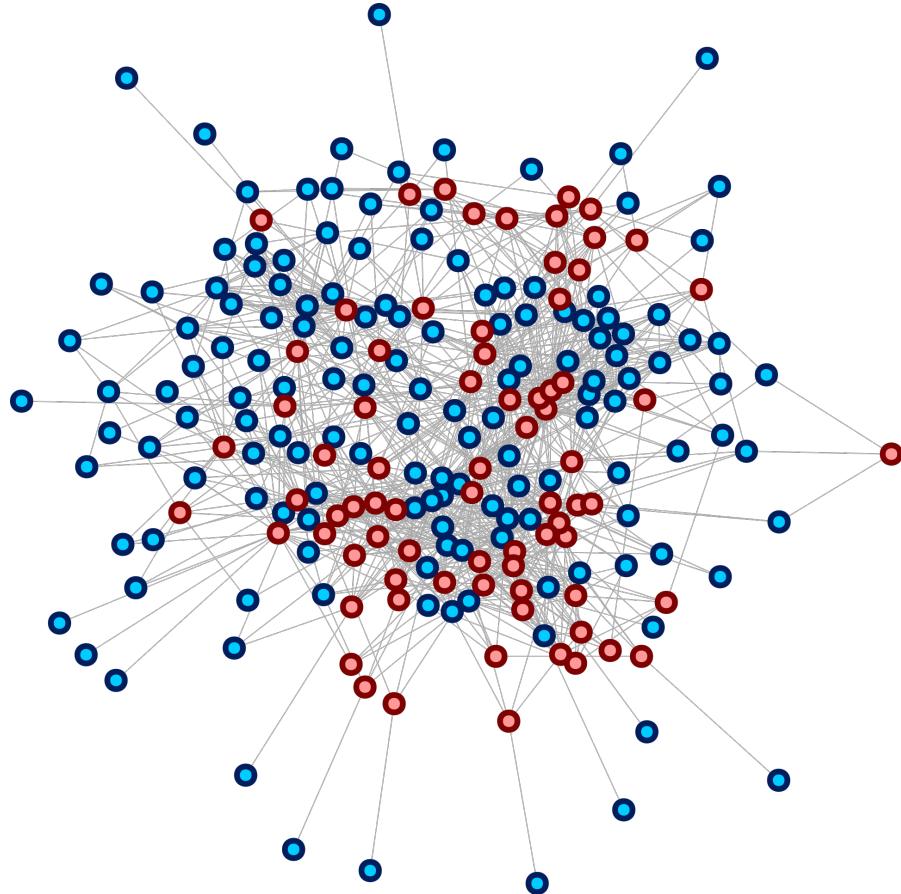


Figure 3: Multi-omics network inferred using *coglasso*, selected by the *coglasso*-adapted *StARS* procedure. Blue nodes represent transcripts, while pink nodes represent metabolites.

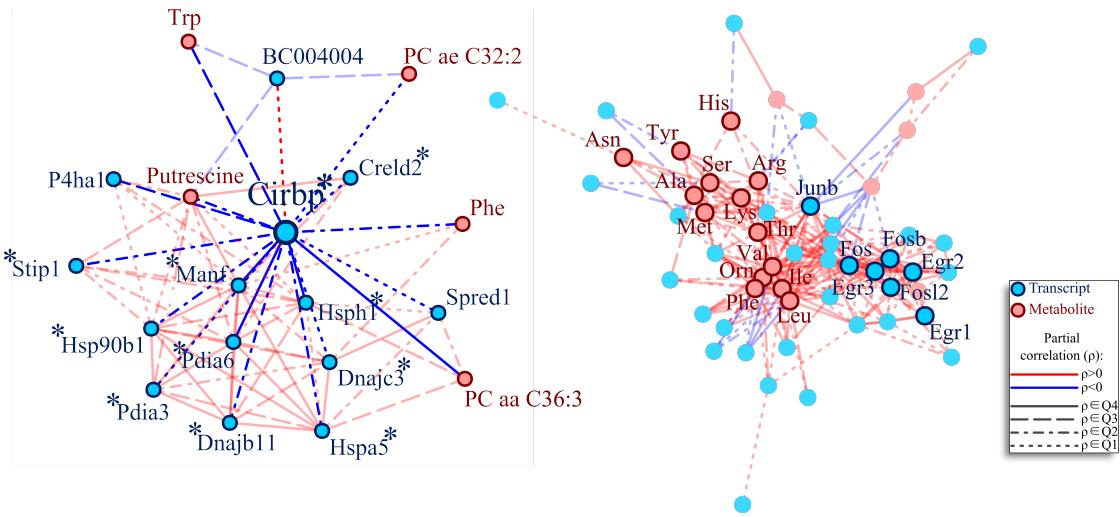


Figure 4: Subnetwork of *Cirbp* and its neighbouring nodes (left) and second largest community (right) from the *coglasso* network shown in Figure 3. Blue nodes represent transcripts, while pink nodes represent metabolites. Blue edges represent negative partial correlations, while red edges stand for positive partial correlations. There are four line intensities, representing the strength of the edges. Dotted lines represent the first quartile of edge strengths of the network, while full lines represent the last quartile. On the left, node labels with asterisks belong to genes known to be involved with unfolded protein response or protein folding in general. *Cirbp* shows a negative relation to most of the genes of its neighbourhood.

Collaborative graphical lasso supplementary material

Alessio Albanese^{1,2}, Wouter Kohlen², and Pariya Behrouzi*¹

¹Mathematical and Statistical Methods (Biometris), Wageningen University and Research

²Cellular and Developmental Biology, Wageningen University and Research

Appendix A: Derivation of *coglasso* coordinate descent update rule

In this appendix, we report the original *coglasso* objective function, we develop it and we minimize it with respect to the j th component of the vector β_i . Here is the objective function:

$$F_{\Lambda, \mathbf{W}_{\setminus i \setminus i}}(\beta_i) = \frac{1}{2} \left\| (\mathbf{W}_{\setminus i \setminus i})^{-1/2} \mathbf{S}_i - (\mathbf{W}_{\setminus i \setminus i})_X^{1/2}(\beta_i)_X \right\|^2 + \frac{1}{2} \left\| (\mathbf{W}_{\setminus i \setminus i})^{-1/2} \mathbf{S}_i - (\mathbf{W}_{\setminus i \setminus i})_Z^{1/2}(\beta_i)_Z \right\|^2 \\ + \frac{c}{2} \left\| (\mathbf{W}_{\setminus i \setminus i})_X^{1/2}(\beta_i)_X - (\mathbf{W}_{\setminus i \setminus i})_Z^{1/2}(\beta_i)_Z \right\|^2 + \|\Lambda_i \odot \beta_i\|_1$$

For an explanation of the elements of the objective function, we refer the reader to Sections 2 and 3 of the manuscript. Now, developing the elements of the objective function:

$$F_{\Lambda, \mathbf{W}_{\setminus i \setminus i}}(\beta_i) = \mathbf{S}_i^T (\mathbf{W}_{\setminus i \setminus i}^{-1/2})^T \mathbf{W}_{\setminus i \setminus i}^{-1/2} \mathbf{S}_i + \frac{1+c}{2} (\beta_i)_X^T (\mathbf{W}_{\setminus i \setminus i}^{1/2})_X^T (\mathbf{W}_{\setminus i \setminus i}^{1/2})_X (\beta_i)_X \\ + \frac{1+c}{2} (\beta_i)_Z^T (\mathbf{W}_{\setminus i \setminus i}^{1/2})_Z^T (\mathbf{W}_{\setminus i \setminus i}^{1/2})_Z (\beta_i)_Z - \frac{c}{2} (\beta_i)_X^T (\mathbf{W}_{\setminus i \setminus i}^{1/2})_X^T (\mathbf{W}_{\setminus i \setminus i}^{1/2})_Z (\beta_i)_Z \\ - \frac{c}{2} (\beta_i)_Z^T (\mathbf{W}_{\setminus i \setminus i}^{1/2})_Z^T (\mathbf{W}_{\setminus i \setminus i}^{1/2})_X (\beta_i)_X - (\beta_i)_X^T (\mathbf{W}_{\setminus i \setminus i}^{1/2})_X^T \mathbf{W}_{\setminus i \setminus i}^{-1/2} \mathbf{S}_i \\ - (\beta_i)_Z^T (\mathbf{W}_{\setminus i \setminus i}^{1/2})_Z^T \mathbf{W}_{\setminus i \setminus i}^{-1/2} \mathbf{S}_i + \|\Lambda_i \odot \beta_i\|_1$$

Considering that the matrix $\mathbf{W}_{\setminus i \setminus i}$ can be divided in the following block matrix:

$$\mathbf{W}_{\setminus i \setminus i} = \begin{pmatrix} (\mathbf{W}_{\setminus i \setminus i})_{XX} & (\mathbf{W}_{\setminus i \setminus i})_{XZ} \\ (\mathbf{W}_{\setminus i \setminus i})_{XZ}^T & (\mathbf{W}_{\setminus i \setminus i})_{ZZ} \end{pmatrix}$$

and taking into account the equivalences shown in Appendix B, it's easy to see what follows:

- $(\mathbf{W}_{\setminus i \setminus i}^{-1/2})^T \mathbf{W}_{\setminus i \setminus i}^{-1/2} = \Theta_{\setminus i \setminus i}$
- $(\mathbf{W}_{\setminus i \setminus i}^{1/2})_X^T (\mathbf{W}_{\setminus i \setminus i}^{1/2})_X = (\mathbf{W}_{\setminus i \setminus i})_{XX}$ and $(\mathbf{W}_{\setminus i \setminus i}^{1/2})_Z^T (\mathbf{W}_{\setminus i \setminus i}^{1/2})_Z = (\mathbf{W}_{\setminus i \setminus i})_{ZZ}$
- $(\mathbf{W}_{\setminus i \setminus i}^{1/2})_X^T (\mathbf{W}_{\setminus i \setminus i}^{1/2})_Z = (\mathbf{W}_{\setminus i \setminus i})_{XZ}$ and $(\mathbf{W}_{\setminus i \setminus i}^{1/2})_Z^T (\mathbf{W}_{\setminus i \setminus i}^{1/2})_X = \mathbf{W}_{\setminus i \setminus i}^{1/2} \mathbf{W}_{\setminus i \setminus i}^T$
- $(\mathbf{W}_{\setminus i \setminus i}^{1/2})_X^T \mathbf{W}_{\setminus i \setminus i}^{-1/2} = (\mathbf{I}_{XX}, \mathbf{O}_{XZ})$ and $(\mathbf{W}_{\setminus i \setminus i}^{1/2})_Z^T \mathbf{W}_{\setminus i \setminus i}^{-1/2} = (\mathbf{O}_{XZ}^T, \mathbf{I}_{ZZ})$

*To whom correspondence should be addressed. Email: pariya.behrouzi@wur.nl

Considering the previous rewritings, the objective function can be simplified to the following form:

$$\begin{aligned} F_{\Lambda, \mathbf{W}_{\setminus i \setminus i}}(\boldsymbol{\beta}_i) &= \mathbf{S}_i^T \boldsymbol{\Theta}_{\setminus i \setminus i} \mathbf{S}_i + \frac{1+c}{2} (\boldsymbol{\beta}_i)_X^T (\mathbf{W}_{\setminus i \setminus i})_{XX} (\boldsymbol{\beta}_i)_X + \frac{1+c}{2} (\boldsymbol{\beta}_i)_Z^T (\mathbf{W}_{\setminus i \setminus i})_{ZZ} (\boldsymbol{\beta}_i)_Z \\ &\quad - \frac{c}{2} (\boldsymbol{\beta}_i)_X^T (\mathbf{W}_{\setminus i \setminus i})_{XZ} (\boldsymbol{\beta}_i)_Z - \frac{c}{2} (\boldsymbol{\beta}_i)_Z^T (\mathbf{W}_{\setminus i \setminus i})_{XZ}^T (\boldsymbol{\beta}_i)_X - (\boldsymbol{\beta}_i)_X^T (\mathbf{I}_{XX}, \mathbf{O}_{XZ}) \mathbf{S}_i \\ &\quad - (\boldsymbol{\beta}_i)_Z^T (\mathbf{O}_{XZ}^T, \mathbf{I}_{ZZ}) \mathbf{S}_i + \|\boldsymbol{\Lambda}_i \odot \boldsymbol{\beta}_i\|_1 \end{aligned}$$

We now define the matrix of off-diagonal blocks obtained from $\mathbf{W}_{\setminus i \setminus i}$ as:

$$\mathbf{W}_{\setminus i \setminus i}^{ODB} = \begin{pmatrix} \mathbf{0}_{XX} & (\mathbf{W}_{\setminus i \setminus i})_{XZ} \\ (\mathbf{W}_{\setminus i \setminus i})_{XZ}^T & \mathbf{0}_{ZZ} \end{pmatrix}$$

It's now straightforward to see that

$$(\boldsymbol{\beta}_i)_X^T (\mathbf{W}_{\setminus i \setminus i})_{XZ} (\boldsymbol{\beta}_i)_Z + (\boldsymbol{\beta}_i)_Z^T (\mathbf{W}_{\setminus i \setminus i})_{XZ}^T (\boldsymbol{\beta}_i)_X = \boldsymbol{\beta}_i^T \mathbf{W}_{\setminus i \setminus i}^{ODB} \boldsymbol{\beta}_i,$$

and that

$$\begin{aligned} &(\boldsymbol{\beta}_i)_X^T (\mathbf{W}_{\setminus i \setminus i})_{XX} (\boldsymbol{\beta}_i)_X + (\boldsymbol{\beta}_i)_Z^T (\mathbf{W}_{\setminus i \setminus i})_{ZZ} (\boldsymbol{\beta}_i)_Z \\ &= \boldsymbol{\beta}_i^T \mathbf{W}_{\setminus i \setminus i} \boldsymbol{\beta}_i - (\boldsymbol{\beta}_i)_X^T (\mathbf{W}_{\setminus i \setminus i})_{XZ} (\boldsymbol{\beta}_i)_Z - (\boldsymbol{\beta}_i)_Z^T (\mathbf{W}_{\setminus i \setminus i})_{XZ}^T (\boldsymbol{\beta}_i)_X \\ &= \boldsymbol{\beta}_i^T \mathbf{W}_{\setminus i \setminus i} \boldsymbol{\beta}_i - \boldsymbol{\beta}_i^T \mathbf{W}_{\setminus i \setminus i}^{ODB} \boldsymbol{\beta}_i. \end{aligned}$$

Moreover, considering that

$$(\boldsymbol{\beta}_i)_X^T (\mathbf{I}_{XX}, \mathbf{O}_{XZ}) \mathbf{S}_i + (\boldsymbol{\beta}_i)_Z^T (\mathbf{O}_{XZ}^T, \mathbf{I}_{ZZ}) \mathbf{S}_i = \boldsymbol{\beta}_i^T \mathbf{S}_i,$$

the objective function becomes:

$$F_{\Lambda, \mathbf{W}_{\setminus i \setminus i}}(\boldsymbol{\beta}_i) = \mathbf{S}_i^T \boldsymbol{\Theta}_{\setminus i \setminus i} \mathbf{S}_i + \frac{1+c}{2} \boldsymbol{\beta}_i^T \mathbf{W}_{\setminus i \setminus i} \boldsymbol{\beta}_i - \frac{1+2c}{2} \boldsymbol{\beta}_i^T \mathbf{W}_{\setminus i \setminus i}^{ODB} \boldsymbol{\beta}_i - \boldsymbol{\beta}_i^T \mathbf{S}_i + \|\boldsymbol{\Lambda}_i \odot \boldsymbol{\beta}_i\|_1$$

Now we minimize with respect to the j th element of $\boldsymbol{\beta}_i$, symbolized here by β_j . Of the dot products in the previous equation, we display in the next only the terms dependent on β_j . Moreover, we remark that all the penalty coefficients that are part of $\boldsymbol{\Lambda}$ are defined to be positive.

$$\begin{aligned} \frac{\partial F_{\Lambda, \mathbf{W}_{\setminus i \setminus i}}(\boldsymbol{\beta}_i)}{\partial \beta_j} &= \frac{\partial}{\partial \beta_j} \left\{ \frac{1+c}{2} \left(\beta_j^2 (\mathbf{W}_{\setminus i \setminus i})_{jj} + 2\beta_j \sum_{k \neq j} (\mathbf{W}_{\setminus i \setminus i})_{jk} \beta_k \right) \right. \\ &\quad \left. - \frac{1+2c}{2} \left(\beta_j^2 (\mathbf{W}_{\setminus i \setminus i})_{jj}^{ODB} + 2\beta_j \sum_{k \neq j} (\mathbf{W}_{\setminus i \setminus i})_{jk}^{ODB} \beta_k \right) - \beta_j S_{ji} + \|\boldsymbol{\Lambda}_i \odot \boldsymbol{\beta}_i\|_1 \right\} \\ &= (1+c)\beta_j (\mathbf{W}_{\setminus i \setminus i})_{jj} + (1+c) \sum_{k \neq j} (\mathbf{W}_{\setminus i \setminus i})_{jk} \beta_k - (1+2c) \sum_{k \neq j} (\mathbf{W}_{\setminus i \setminus i})_{jk}^{ODB} \beta_k \\ &\quad - S_{ji} + \Lambda_{ji} \frac{\partial |\beta_j|}{\partial \beta_j} \end{aligned}$$

Notice how the term containing $\mathbf{W}_{\setminus i \setminus i}^{ODB}_{jj}$ disappears, as all diagonal elements of $\mathbf{W}_{\setminus i \setminus i}^{ODB}$ are zero. In particular, with j being part of the generic dataset A , that could either be X or Z :

$$(1+c) \sum_{k \neq j} (\mathbf{W}_{\setminus i \setminus i})_{jk} \beta_k = (1+c) \sum_{k \notin A} (\mathbf{W}_{\setminus i \setminus i})_{jk} \beta_k + (1+c) \sum_{\substack{k \in A \\ k \neq j}} (\mathbf{W}_{\setminus i \setminus i})_{jk} \beta_k$$

$$(1+2c) \sum_{k \neq j} (\mathbf{W}_{\setminus i \setminus i})_{jk}^{ODB} \beta_k = (1+2c) \sum_{k \notin A} (\mathbf{W}_{\setminus i \setminus i})_{jk}^{ODB} \beta_k = (1+2c) \sum_{\substack{k \in A \\ k \neq j}} (\mathbf{W}_{\setminus i \setminus i})_{jk} \beta_k$$

Again, here all the elements $(\mathbf{W}_{\setminus i \setminus i})_{jk}^{ODB}$ are zero for every $k \in A$, since the diagonal blocks of the matrix are zero, and all the other elements coincide with the corresponding elements in $\mathbf{W}_{\setminus i \setminus i}$. Hence, setting $\frac{\partial F_{\Lambda, \mathbf{W}_{\setminus i \setminus i}}(\boldsymbol{\beta}_i)}{\partial \beta_j} = 0$, we can rewrite the equation as:

$$\beta_j = \frac{1}{(1+c)(\mathbf{W}_{\setminus i \setminus i})_{jj}} \left\{ S_{ji} - (1+c) \sum_{k \notin A} (\mathbf{W}_{\setminus i \setminus i})_{jk} \beta_k - (1+c) \sum_{\substack{k \in A \\ k \neq j}} (\mathbf{W}_{\setminus i \setminus i})_{jk} \beta_k + (1+2c) \sum_{k \notin A} (\mathbf{W}_{\setminus i \setminus i})_{jk} \beta_k - \Lambda_{ji} \frac{\partial |\beta_j|}{\partial \beta_j} \right\},$$

which, for $\alpha = \frac{1}{1+c}$ and simplifying, can be written as:

$$\beta_j = \frac{1}{(\mathbf{W}_{\setminus i \setminus i})_{jj}} \left\{ \alpha S_{ji} - \sum_{\substack{k \in A \\ k \neq j}} (\mathbf{W}_{\setminus i \setminus i})_{jk} \beta_k + c\alpha \sum_{k \notin A} (\mathbf{W}_{\setminus i \setminus i})_{jk} \beta_k - \alpha \Lambda_{ji} \frac{\partial |\beta_j|}{\partial \beta_j} \right\},$$

From here, by applying the soft-threshold operator as seen in Friedman et al. (2007), one can straightforwardly obtain Algorithm 1 as shown in Section 3.

Appendix B: Equivalences involving the estimated variance-covariance matrix

It can be proven that the squared root of a symmetric positive semidefinite matrix is a symmetric positive semidefinite matrix itself. Nevertheless, such proof is out of the scope of this manuscript. Given this statement, since both $\mathbf{W}_{\setminus i \setminus i}$ and $\Theta_{\setminus i \setminus i}$ are symmetric positive semidefinite, so are their squared roots. This allows us to prove the following three sets of equivalences:

$$\begin{aligned} & \begin{pmatrix} (\mathbf{W}_{\setminus i \setminus i})_{XX} & (\mathbf{W}_{\setminus i \setminus i})_{XZ} \\ (\mathbf{W}_{\setminus i \setminus i})_{XZ}^T & (\mathbf{W}_{\setminus i \setminus i})_{ZZ} \end{pmatrix} = \mathbf{W}_{\setminus i \setminus i} = \mathbf{W}_{\setminus i \setminus i}^{1/2} \mathbf{W}_{\setminus i \setminus i}^{1/2} = (\mathbf{W}_{\setminus i \setminus i}^{1/2})^T \mathbf{W}_{\setminus i \setminus i}^{1/2} \\ &= \begin{pmatrix} (\mathbf{W}_{\setminus i \setminus i}^{1/2})_X^T \\ (\mathbf{W}_{\setminus i \setminus i}^{1/2})_Z^T \end{pmatrix} \begin{pmatrix} (\mathbf{W}_{\setminus i \setminus i}^{1/2})_X & (\mathbf{W}_{\setminus i \setminus i}^{1/2})_Z \end{pmatrix} = \begin{pmatrix} (\mathbf{W}_{\setminus i \setminus i}^{1/2})_X^T (\mathbf{W}_{\setminus i \setminus i}^{1/2})_X & (\mathbf{W}_{\setminus i \setminus i}^{1/2})_X^T (\mathbf{W}_{\setminus i \setminus i}^{1/2})_Z \\ (\mathbf{W}_{\setminus i \setminus i}^{1/2})_Z^T (\mathbf{W}_{\setminus i \setminus i}^{1/2})_X & (\mathbf{W}_{\setminus i \setminus i}^{1/2})_Z^T (\mathbf{W}_{\setminus i \setminus i}^{1/2})_Z \end{pmatrix} \\ & \begin{pmatrix} \mathbf{I}_{XX} & \mathbf{O}_{XZ} \\ \mathbf{O}_{XZ}^T & \mathbf{I}_{ZZ} \end{pmatrix} = \mathbf{W}_{\setminus i \setminus i}^{1/2} \mathbf{W}_{\setminus i \setminus i}^{-1/2} = (\mathbf{W}_{\setminus i \setminus i}^{1/2})^T \mathbf{W}_{\setminus i \setminus i}^{-1/2} = \begin{pmatrix} (\mathbf{W}_{\setminus i \setminus i}^{1/2})_X^T \\ (\mathbf{W}_{\setminus i \setminus i}^{1/2})_Z^T \end{pmatrix} \mathbf{W}_{\setminus i \setminus i}^{-1/2} = \begin{pmatrix} (\mathbf{W}_{\setminus i \setminus i}^{1/2})_X^T \mathbf{W}_{\setminus i \setminus i}^{-1/2} \\ (\mathbf{W}_{\setminus i \setminus i}^{1/2})_Z^T \mathbf{W}_{\setminus i \setminus i}^{-1/2} \end{pmatrix} \\ & \Theta_{\setminus i \setminus i} = \mathbf{W}_{\setminus i \setminus i}^{-1} = \mathbf{W}_{\setminus i \setminus i}^{-1/2} \mathbf{W}_{\setminus i \setminus i}^{-1/2} = (\mathbf{W}_{\setminus i \setminus i}^{-1/2})^T \mathbf{W}_{\setminus i \setminus i}^{-1/2} \end{aligned}$$

References

- Jerome Friedman, Trevor Hastie, Holger Höfling, and Robert Tibshirani. Pathwise coordinate optimization. *The Annals of Applied Statistics*, 1(2):302–332, December 2007. ISSN 1932-6157, 1941-7330. doi: 10.1214/07-AOAS131. URL <https://projecteuclid.org/journals/annals-of-applied-statistics/volume-1-issue-2/Pathwise-coordinate-optimization/10.1214/07-AOAS131.full>.

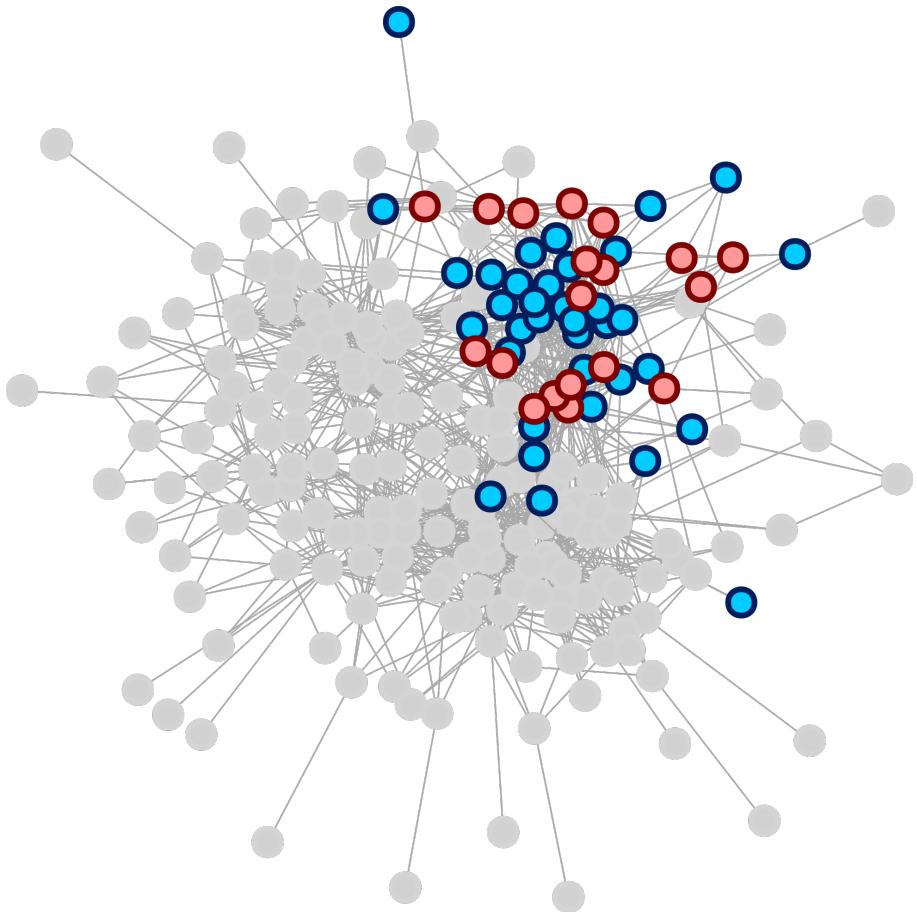


Fig. S1: Multi-omics network inferred using *coglasso*. The nodes belonging to community two are highlighted here.