

Detecting Long Distance Conditional Correlations Between Anatomical Regions Using Gaussian Graphical Models

Stéphanie Allassonnière, Pierre Jolivet and Christophe Giraud *

CMAP Ecole Polytechnique, Route de Saclay, 91128 Palaiseau cedex, France
<http://www.cmap.polytechnique.fr>

Abstract. The conditional correlation patterns of an anatomical shape may provide some important information on the structure of this shape. We propose to investigate these patterns by Gaussian Graphical Modelling. We design a model which takes into account both local and long-distance dependencies. We provide an algorithm which estimates sparse long-distance conditional correlations, highlighting the most significant ones. The selection procedure is based on a criterion which quantifies the quality of the conditional correlation graph in terms of prediction. The preliminary results on AD versus control population show noticeable differences.

Keywords: Gaussian Graphical Model, long-distance conditional correlation, sparse dependencies, deformation of Alzheimer's disease

1 Introduction

It is quite natural to think that any region of the brain depends on all the other regions at least via some detours. This is equivalent to say that these regions are all correlated to each other. Several studies [10, 12, 5] have tried to design correlation patterns by computing the correlation matrix given by correlations between any two regions (using a PCA decomposition for example). Then, they highlight the most significant or stable ones by thresholding. However, correlations describe the global statistical dependencies between variables, corresponding to both direct and indirect interactions. The *direct* relations between two of these regions are less numerous and harder to capture but they carry some interesting information as well. Focusing on these direct dependencies, called *conditional correlations*, enables to see which areas directly affect the behaviour of a given region and avoids the burden given by the indirect interactions.

We want to study the geometric conditional dependencies of a response given by a real valued signal carried by a discrete grid. A natural approach to estimate these conditional correlations is given by Gaussian Graphical Modelling (GGM)

* The authors thank Johns Hopkins University for the data and Dominique for the cartoon. Support for the JHU group was provided in part by P41-RR015241 (JHU) and U24 RR021382 (MGH) to BIRN, National Center for Research Resources (NIH), P01-AG0568 and P01-AG03991 (WU), and the NSF via DMS-0456253.

[11, 15, 7]. The response is modelled as a Gaussian random vector. The conditional correlations of the response variables are depicted by a graph. Each node represents a variable and an edge in the graph is set between two variables if they are conditionally dependent given all the remaining ones. This corresponds to a non-zero entry of the inverse covariance matrix called the precision matrix.

In the context of medical data, we face two problems with Gaussian Graphical Modelling. The first one is the high dimension of the data compared to the low sample size. The second problem is that in this framework, the neighbour points are very likely to be conditionally correlated since the medical phenomenon is usually continuous and these local dependencies can be predominant in the estimated graph possibly hiding any other links. We propose here an algorithm and a selection criterion which address both questions.

The high-dimension-low-sample-size paradigm is a common issue in medical imaging. This usually leads to non accurate and non stable statistical analysis. In our graphical framework, this is equivalent to a high dimensional graph and small number of measures of the random signal. To face this situation, sparse representation is well known to be a very powerful tool in Computer Vision and Pattern recognition [26]. Although we know that the underlying real graph is not sparse, it is interesting to perform a *sparse estimation of its structure* [13, 14]. Such an estimation provides both a stable estimate of the conditional dependencies and a selection of the edges with the highest conditional correlations. The procedure proposed by [15, 13] is based on the LASSO algorithm [21]. In addition, the elastic net algorithm [27] slightly relaxes this sparse constrain providing more accurate and robust-to-noise sparsity patterns [9]. This sparsity constraint is crucial as it prunes the graph keeping a small number of edges.

As noticed above, local dependencies of medical images are predominant while estimating the conditional correlations. We want to go beyond that and focus on long-distance dependencies. To show only long distance dependencies, the authors in [10] post-process their graph merging some nodes and edges when the nodes are closer than a fixed threshold. However, we want to introduce this knowledge as a prior to the estimation. We have brought in the algorithm a neighbourhood prior: knowing that the neighbours are linked, we only look for the other conditional dependencies which are modelled as sparse. This enables to highlight the long distance conditional correlations and to gain in term of statistical accuracy since the estimation of long-distance relations is less affected by the noise and the neighbours. The prior knowledge is modelled as a neighbouring graph provided by the user. Because the neighbouring graph is given by the user, it can take into account geographical closeness or some anatomical proximity given for example by fibres.

The estimation procedure depends on some free parameters. In order to find suitable values, one can either use a cross-validation [13] or a theoretical analysis based on a restrictive condition on the data [14]. Unfortunately these choices are not satisfactory in practice and the final choice is usually done by hand looking at the estimations. To avoid this subjective choice, Giraud [7] proposes, for the usual sparse GGM framework, a criterion which optimises the choice of the graph

depending on its capacity to predict one random variable given the others. We generalise this criterion to the case of neighbourhood prior.

In Computational Anatomy, learning what characterises the population in terms of dependencies as in [19] for Alzheimer's disease (AD) is an important aspect of the analysis of the pathology. Different but complementary methods have been used to analyse populations for instance Deformation Based-Morphometry (DBM) [3], Tensor Based-Morphometry [25] or Feature Based-Morphometry [22]. The general idea is to compute some statistics from a sample of template-to-subject deformations. Statistical analysis can also be done directly on the deformations [3, 6, 2] or by modelling the generation of images [18, 1] showing the admissible deformations in a population. Although the global deformations give an interesting interpretation, we may want to know what parts of the shape are deformed jointly. Our proposed algorithm based on LASSO or elastic net has been tested on a training set of high dimension data representing some deformations of a template hippocampus towards 101 patients' hippocampi. The results show some correlated regions when dealing with deformations which suggest that this structure has regions which move together. Moreover, the population is divided into two groups, controls and patients suffering from AD. The comparison of the estimated graphs of the two groups show some noticeable differences which may be a new characterisation of AD.

2 Gaussian Graphical Models

Let us consider p points on a given shape that will compose the nodes of the graph. On these points, we observe n random responses such as a discretisation (with p points) of a quantification of template-to-subject deformations. The p nodes of the graph are thus identified to p random variables denoted (X_1, \dots, X_p) , this vector is assumed to be distributed as a multivariate Gaussian $\mathcal{N}_p(0, \Sigma)$ (data are re-centred if necessary). The graph G_Σ of conditional dependencies is defined as follows: there exists an edge between nodes a and b if and only if the variables X_a and X_b are dependent given all the remaining variables. This will be denoted $a \stackrel{G_\Sigma}{\sim} b$. To illustrate the notion of conditional dependency, let us give a toy example illustrated in Fig. 1. The traffic jam intensity and the number of snowmen in town are correlated due to snowstorm. But conditionally on the height of snow, the number of snowmen is independent of the intensity of traffic jams. This corresponds to the graph in Fig 1 with only two edges. In particular, there is no edge between the traffic jam and snowmen variables.

2.1 Estimation processes

The estimation is done by a regression as presented in [15]. Let X_a be the a^{th} node of the graph. The goal is to estimate the matrix θ such that: $X_a =$

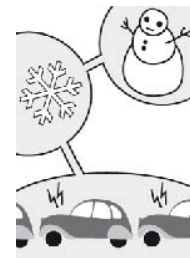


Fig. 1. Conditionally on the height of snow, the number of snowmen is independent of the intensity of traffic jams. This is represented by a two edges graph.

$\sum_{b \neq a} \theta_{a,b} X_b + \varepsilon_a$ where ε_a is assumed to follow a normal distribution with zero mean and variance $1/(\Sigma^{-1})_{a,a}$. An important property of this modelling is that $\theta_{a,b} = -\frac{K_{a,b}}{K_{a,a}}$ where $K = \Sigma^{-1}$ is the precision matrix. Therefore $\theta_{a,b} = 0$ is equivalent to $a \not\stackrel{G_\Sigma}{\sim} b$. Let $\theta_{\cdot,b}$ be the p -dimensional vector of $(\theta_{a,b})_{1 \leq a \leq p}$ and X_{-a} the $(p-1)$ -dimensional random vector $(X_1, \dots, X_{a-1}, X_{a+1}, \dots, X_p)$. The estimation of θ is done by minimising an energy (which is actually the negative penalised log likelihood of the model) given by two terms. The first one is the l^2 distance between the observations and their estimation (coming from the Gaussian distribution of the noise). The second term is a prior on the parameters θ to estimate. In the case we consider, the prior is a Laplacian prior with parameter λ . This corresponds to an l^1 penalty on the energy side. The energy to minimise is therefore the following:

$$\forall b \in \{1, \dots, p\}, \hat{\theta}_{\cdot,b} = \operatorname{argmin}_{\theta_{\cdot,b}} \|X_b - X_{-b}\theta_{\cdot,b}\|_2^2 + \lambda \|\theta_{\cdot,b}\|_1. \quad (1)$$

This regression technique is known as the *LASSO* algorithm [21]. Introducing a penalty on the estimated parameter θ enables to estimate it. Indeed, θ may be of large dimension which prevents from using a small sample size n . Reducing the estimation to more likely or expected matrices, enables to handle this typical case in medical imaging.

The minimum argument of this energy is sparse thanks to the l^1 penalty. Besides, it provides a stable estimation in terms of prediction of a variable given the other ones. Including this constraint on the candidate matrices enables to really focus on the most important conditional correlations. Therefore predicting one variable from the others is not much dependent on the training set.

However, one may want to relax slightly this constraint by balancing the effect of this l^1 penalty by an l^2 one which has the opposite behaviour. Instead of looking for sparse matrix with big coefficients, it tends to spread to all elements in the matrix. This is known as the elastic net algorithm [27] which provides more accurate sparsity patterns [9]. The energy is:

$$\forall b \in \{1, \dots, p\}, \hat{\theta}_{\cdot,b} = \operatorname{argmin}_{\theta_{\cdot,b}} \|X_b - X_{-b}\theta_{\cdot,b}\|_2^2 + \lambda \|\theta_{\cdot,b}\|_1 + \gamma \|\theta_{\cdot,b}\|_2^2. \quad (2)$$

These two algorithms take each node independently one after the other. This has a drawback considering that we would like to have a symmetric relation between nodes. Minimising either energy above does not guarantee that $\hat{\theta}_{a,b} \neq 0$ at the same time as $\hat{\theta}_{b,a} \neq 0$. The construction of the graph is therefore not straightforward. Two natural choices appear to enforce the symmetry [15]. On the one hand, we can say that only one of $\hat{\theta}_{a,b}$ or $\hat{\theta}_{b,a}$ needs to be non-zero to set an edge in the graph between a and b . This yields $a \stackrel{\hat{G}_\Sigma}{\sim} b$ if and only if $\hat{\theta}_{a,b} \neq 0$ or $\hat{\theta}_{b,a} \neq 0$. On the other hand, we can prefer a harder decision process by saying that only one of the $\hat{\theta}_{a,b}$ or $\hat{\theta}_{b,a}$ needs to be zero to remove the edge between a and b in the graph. This leads to $a \stackrel{\hat{G}_\Sigma}{\sim} b$ if and only if $\hat{\theta}_{a,b} \neq 0$ and $\hat{\theta}_{b,a} \neq 0$. Both will be tested in the sequel.

2.2 Penalised criterion to choose among a family of estimated graphs

In the two previous energies (1) and (2), there are two parameters which have to be chosen. This may be very difficult and data depending. A classical choice is to use cross-validation. Giraud in [7] suggests instead to set a criterion so that after computing a family of possible graphs with different parameters, it enables to select the best one with respect to this criterion which aims at answering the question of how good the network is to predict one variable from the others.

To define the criterion $\text{crit}(\cdot)$ we need to introduce a few notations. We associate to any graph G in $\hat{\mathcal{G}}$, the $p \times p$ matrix

$$\hat{\theta}^G = \arg \min_{\theta \in \Theta_G} \sum_{1 \leq i \leq n, 1 \leq a \leq p} [X - X\theta]_{i,a}^2,$$

where Θ_G is the set of $p \times p$ matrices θ such that $\theta_{a,b}$ is non-zero if and only if there is an edge between a and b in G . Then, we define the criterion $\text{crit}(G)$ by

$$\text{crit}(G) = \sum_{a=1}^p \left[\|X_a - \sum_b X_b \hat{\theta}_{a,b}^G\|_n^2 \left(1 + \frac{\text{pen}[d_a(G)]}{n - d_a(G)} \right) \right], \quad (3)$$

where the penalty function is defined by

$$\text{pen}(d) = 1.1 \times \frac{n-d}{n-d-1} \text{EDKhi} \left[d+1, n-d-1, \left(\binom{p-1}{d} (d+1)^2 \right)^{-1} \right]. \quad (4)$$

The function $\text{EDKhi}[d, N, \cdot]$ is the inverse of the function

$$\text{DKhi}[d, N, \cdot] : x \mapsto \mathbb{P} \left(F_{d+2, N} \geq \frac{x}{d+2} \right) - \frac{x}{d} \mathbb{P} \left(F_{d, N+2} \geq \frac{N+2}{Nd} x \right),$$

where $F_{d, N}$ denotes a Fisher random variable with d and N degrees of freedom. See [7] and [4] for more details and more explanations on this penalty. This Criterion (3) is implemented in a R library GGMselect [8].

3 Non local Gaussian Graphical Models

3.1 Introduction of G_0

As noticed above, the neighbouring points - neighbour nodes of the graph- are very likely to be conditionally correlated. We want to put more attention onto the other correlations -which will be called long-distance ones in the sequel since they do not affect the neighbours. To this purpose, we have introduced in the estimation a neighbouring graph G_0 which carries the neighbour nodes of all the graph nodes. We assume that there exist correlations between these points but we are not estimating them rather looking for the other ones. There are two reasons for that. On the one hand, we are not interested in the local correlations since they appear to be obvious. The long-distance ones however may reveal

some important behaviour which may be characteristic from the population we are studying. These non obvious relations between regions may show some non random effect on the shape. This will be illustrated in the experiments.

On the other hand, the value of the penalty is directly related to the complexity of the families of possible candidate graphs. Shrinking the families to those of graphs containing G_0 reduces their complexity and therefore the penalty. The effect is that more edges appear and therefore the long-distance ones.

The way we introduce G_0 in the estimation is as follows. The estimation of the conditional correlations is done in the orthogonal space of the neighbouring graph G_0 . This leads to replace the random variable X_a by

$$X_a - X_{m_a}(X_{m_a}^T X_{m_a})^{-1} X_{m_a}^T X_a, \quad (5)$$

where X^T is the transposition and X_{m_a} is the matrix defined as follow: if we denote X the $n \times p$ matrix of all the data and m_a the list of neighbours of node a in G_0 , then $X_{m_a} = X(\cdot, m_a)$ is of dimension $n \times \text{card}(m_a)$.

This orthogonality constraint may have to be relaxed since the projection may lead to ill-conditioned matrices. Moreover, it will also enable to capture some edges with small projection onto the orthogonal of G_0 which may appear stronger since we know that the local dependencies are more likely to be the strongest and summarise the main information. This can be express in terms of introducing a small ridge (driven by a new parameter γ_0 chosen via the criterion). Denoting Id the identity matrix, this yields

$$X_a - X_{m_a}(X_{m_a}^T X_{m_a} + \gamma_0 Id)^{-1} X_{m_a}^T X_a. \quad (6)$$

3.2 Selection of graph: Change in the penalty

When we introduce an a priori graph G_0 , we shall change the selection criterion to take into account this a priori knowledge. To fit with this situation, the Criterion (3) remains unchanged, but the penalty (4) is replaced by

$$\text{pen}_{G_0}(a, d) = 1.1 \times \frac{n-d}{n-d-1} \text{EDKhi} \left[d+1, n-d-1, \left(\binom{p-d_0[a]-1}{d-d_0[a]} (d-d_0[a]+1)^2 \right)^{-1} \right] \quad (7)$$

where $d_0[a]$ is the degree of the node a in the graph G_0 . This correction reflects the change of complexity induced by the graph G_0 . It ensures a control of the prediction error similar to Theorem 1 in [7]. For further details, we refer to the Appendix. We have therefore modified GGMselect in two ways (which is included in the new update of the package). If a prior is given, we change the data so that the new random variable is given by eq. (5) or (6). Then we changed the penalty criterion for eq. (7).

4 Experiments

We have tested this model on meshed surfaces of hippocampi from Johns Hopkins University [16]. The data base contains $n = 101$ vectors of dimension $p = 713$

corresponding to the vertices of the graph. A template surface of a hippocampus was mapped onto each subject surface using Large Diffeomorphic Deformation Metric Mapping [17]. Prior rigid body registration has been done on each subject. In this common coordinate system, the log of the absolute value of the Jacobian determinant of this deformation was discretised onto the template meshed surface. This leads to 101 vectors with real value signal which correspond to scalar deformation fields over the surface. The subjects belong to three sub-populations. The first 57 were control patients. The 32 next were suffering from mild Alzheimer's disease or semantic dementia (denoted mild AD) and the 12 last were in a late stage of the disease (denote late AD in the sequel). Some examples of the training set are presented in Fig. 2.

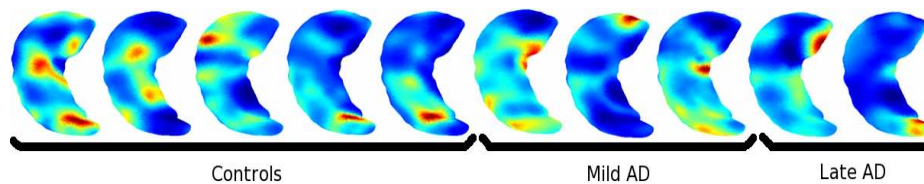


Fig. 2. Examples of the training set. The colour depends on the intensity of the Jacobian of the deformation. Blue means a contraction and red dilatation. The intensity itself is not important but rather its relative value with respect to the others.

4.1 Gaussian Graphical estimation without neighbour prior

The results of the graph estimations without any neighborhood prior are presented in the four graphs on the left hand side of Fig. 3. The four plots correspond respectively to the solution of eq.(1) with both choice of symmetrisation ("or" and "and" resp.) and of eq.(2) again with both symmetrisation processes. For each of these minimisation, we present the result with the lowest criterion. All these graph present some similarities. Indeed, they all find the conditional correlations between close nodes of the graph with respect to the euclidean distance. This is what we expect since the deformations are smooth. Therefore, the deformation at one point is close to the deformation on its neighbours leading to a correlation between all nodes but only local conditional correlations. However, one of the graphs (left one, with a slightly lower criterion) suggests that they should be some conditional correlation not only between neighbour nodes. As we can see, not all the part of the surface are linked through long edges. However, this does not appear in the three other graphs which makes them uncertain.

4.2 Non local GGM estimation

To overcome the problem of the predominance of the conditional correlations between close nodes, we have introduced a neighbouring graph G_0 . Two examples of G_0 s are presented in Fig. 3 (right). The estimated graphs using our proposed

algorithm are presented in Fig. 4 where the top row corresponds to the first G_0 of Fig. 3 and the bottom row to the second one. All the eight estimations are presented: lasso (with the two possible symmetrisations) using the orthogonal projection and the ridge regularisation and next, in the same order, the four estimation with the elastic net energy. The orthogonal projection may appear too strong since it requires to invert the neighbouring matrix (cf eq. (5)). This yields no graph estimated with the lasso minimisation process with a strong symmetrisation constraint on the first line.

We can notice that they almost all carry some long-distance correlations. In particular, four of them show some common patterns pointing the same connected areas (row 1, columns 2,4,5,6). The estimation remains stable when adding or removing some edges in G_0 . This is shown in the bottom row of Fig. 4. The results are very similar. Noticing that the orthogonal of this new G_0 is smaller (the maximum degree of this G_0 is 7 whereas the first one was 3) the resulting graphs look alike with a little less edges. In this case, the estimation of the first graph is possible but presents very few connections. As well as for the first G_0 , the lasso with a relaxed projection constrains provides very interesting graphs and the elastic net does not requires this relaxation. Since some graphs appear very dense, we removed the edges between close points w.r.t. the Euclidean distance. We did this pruning for the marked graph (see below for meaning of this mark) of Fig. 4. The result is presented in Fig. 5. Thanks to this representation, the long-distance conditional correlation appear clearer and can be easier interpreted. Indeed, this suggest that there is a strong conditional correlation between the bottom part of the head of the hippocampus and two regions of the body. This means that the hippocampus is subjected to deformations that are not random. When the bottom area of the head is deformed so are the two local parts of the body. Note that the conditional correlation can be positive or negative which correspond to same and opposite behaviour respectively.

Thanks to the criterion we can point the graph which, among the ones we estimate, has the best power for prediction purposes. This graph is marked with a red star. This graph looks indeed very interesting showing long-distance dependencies that are not trivial. The power of this criterion we propose is that

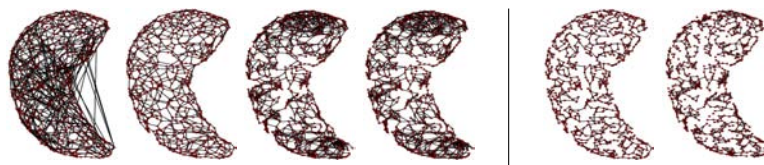


Fig. 3. Left: Estimated graphs using the lasso algorithm and both symmetrisations (called "or" and "and" in the sequel) and the elastic net algorithm with both symmetrisations as well ("or" and "and" resp.). The results show the predominance of the neighbours conditional correlations except for the first estimation where long-distance conditional correlation appear. Right: two examples of possible G_0 s.

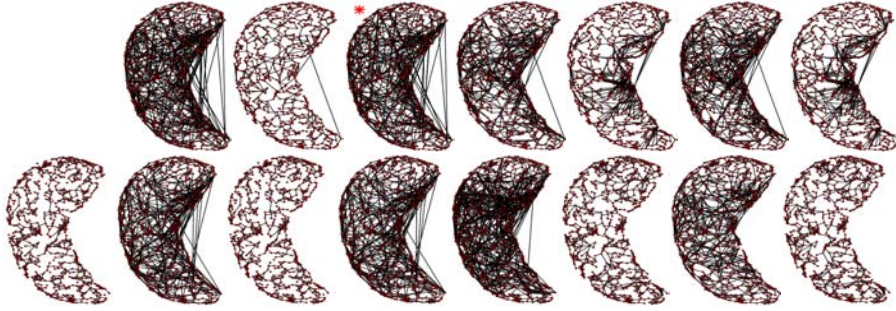


Fig. 4. Estimated graphs with the eight methods using the two neighbouring graphs presented in Fig. 3 on the top and bottom line resp. From left to right, each method with both symmetrisations ("or" and "and" resp.): lasso with the orthogonal projection, lasso with a relaxation on the projection, elastic net with projection, elastic net with relaxed projection. The projection constraint may be too strong preventing from computing the estimation (top left). In almost all graphs, long-distance dependencies appear and are stable with respect to the prior neighbours.

it can quantify the predictive power of any other graph estimated by any other method. We not only provide some estimation technics which take into account a neighbourhood prior but also a way to compare them to other estimated graphs.

4.3 Clustering of the graph

In order to see the different areas which are conditionally correlated, we use the spectral clustering method which highlights these dependencies. Spectral clustering is a technique based on eigen-properties of a similarity matrix Q that partitions data into disjoint clusters. We compute the p eigenvectors of Q , $(V_i)_{1 \leq i \leq p}$. Let the vectors $Y_j = (V_i^j)_{2 \leq i \leq p}$ be the concatenation of the j^{th} coordinates of each eigenvector. The clustering is done using a k -mean algorithm on these vectors. The choice of the similarity matrix is given as follows. If two nodes i and j are connected through the graph but not in the neighbouring graph G_0 then $Q(i, j) = 1/Z$. If this condition is not satisfied then $Q(i, j) = f(\|x_i - x_j\|)/Z$ where Z is a renormalisation constant, x_i is the pose of node i and f is the Gaussian density function with fixed variance (chosen with respect to the spreading of the data, here 50 whereas $\max(x_i - x_j)^2 = 7185$). $Q(i, i)$ is set so that the sum of the i^{th} row is 0. The result using 5 clusters is presented in Fig. 5 where each node has a colour corresponding to its cluster. We can clearly see that the regions that are in the same cluster (same colour) correspond to areas that share a link in the pruned graph (Left of Fig. 5). This shows that several connected parts of the shape (red, black, green and dark blue) seem to have conditionally independent behaviour whereas some nodes (cyan) are clearly related. This result confirms the segmentation in [24].

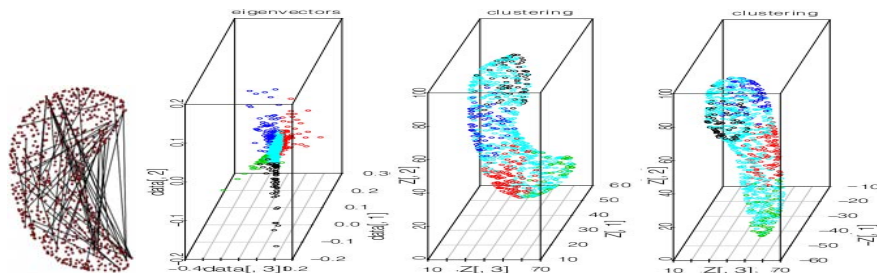


Fig. 5. Best graph analysed. Left: the edges between close nodes have been removed for better visualisation. Right: k-mean clustering.

4.4 Population different

Since we are provided with three sub-groups, it is interesting to see whether they carry differences with respect to conditional correlations. However, since the late AD group is pretty small, we only cluster it into two groups: controls versus AD (both mild and late together). The goal is to see if the disease changes the kind of conditional correlated deformations that are present in a population. This is actually what Fig 6 tends to show. The deformations of the hippocampus in a control population are not completely random since some important long distance conditional correlations appear. This reduces the safe deformations to a subset where these joint behaviour occurs. However, the AD population seems to have less long-distance conditional correlation. This would suggest that the disease affects the hippocampus by removing the direct dependencies of different regions. This effect cannot be seen when no neighbourhood obvious correlations are taken into account (graphs on the left of Fig. 6). However, when introducing the projection onto the orthogonal of a neighbouring graph (the first one presented above), the differences appear and are stable with respect to the estimation process (the criterions are of the same range for these four tests).

5 Conclusion

This paper presents a new way of analysing populations of shapes in terms of random graphs which carry sparse conditional correlations between areas of a shape. We have introduced a neighbourhood prior. It stabilises the estimation and highlights the long-distance conditional correlations which are the non obvious ones. This neighbouring graph is given by the user allowing for non trivial closeness notion as some anatomical ones. The results on the deformed hippocampi reveal some important conditional correlations between particular sub-regions which are stable along the estimation processes as well as with respect to a change in the neighbouring graph. Moreover, it emphasises differences between the control subjects which have more long-distance edges than the AD group. Provided with a larger database would help confirming this first trend. In

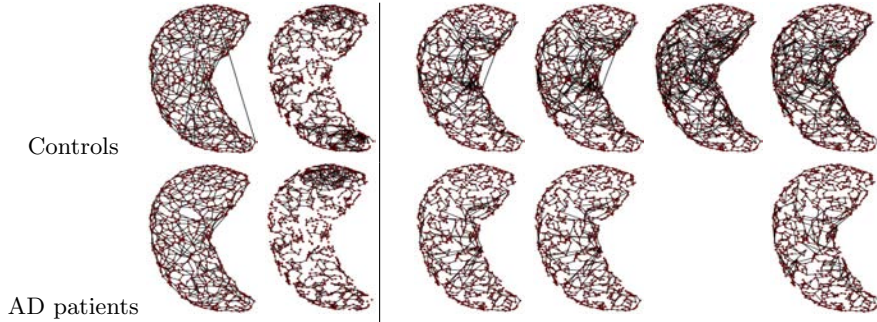


Fig. 6. Comparison of the estimated graphs from the two sub-groups. Left: estimations without a prior on the neighbours (lasso "and" and elastic net "or" resp.). Right: estimations using the first G_0 presented above (lasso "and" with orthogonal projection and the ridge relaxation, elastic net "or" with projection (when computable) and ridge relaxation resp.). The graph without the prior are very similar and does not enable any discrimination. Introducing the prior knowledge enables to catch the differences: the AD group present less long-distance conditional correlations.

addition, we provide the user with a criterion which quantifies the quality of any graph (not only the one estimated with our algorithm). The estimation has been done here for anatomical shapes but is more general and it would be interested to test this on functional data to catch *direct long-distance* dependencies in the brain connectivity network and compare with for eg [23] and [20].

6 Appendix

The penalty pen_{G_0} is a complexity penalty built from the theory of Giraud [7]. For a given integer D and a graph G_0 , we introduce the collection $\mathcal{M}_D(G_0)$ of graphs of degree less than D and containing G_0 . A careful inspection of the proof of Theorem 1 in [7] shows that for each node a the penalty $\text{pen}_{G_0}(a, d)$ must ensures the control

$$\sum_{m_a \in \mathcal{M}_0(a)} (|m_a| + 1) \text{DKhi} \left(|m_a| + 1, n - |m_a| - 1, \frac{(n - |m_a| - 1) \text{pen}(a, m_a)}{K(n - |m_a|)} \right) \leq \log(n),$$

where $\mathcal{M}_0(a) = \{b : b \stackrel{\mathcal{G}}{\sim} a\}$. This control is achieved by eq (7).

References

1. Allasonnière, S., Kuhn, E., Trouvé, A.: Bayesian deformable models building via stochastic approximation algorithm: A convergence study. *Bernoulli J.* 16(3), 641–678 (2010)
2. Allasonnière, S., Younes, L.: A stochastic algorithm for probabilistic independent component analysis. In revision for *Annals of Applied Statistics*
3. Ashburner, J., et al.: Identifying global anatomical differences: deformation-based morphometry. *Human Brain Mapping* 6(5-6), 348–357 (1998)

4. Baraud, Y., Giraud, C., Huet, S.: Gaussian model selection with an unknown variance. *Annals of Statistics* 37(2), 630–672 (2009)
5. Bickel, P., Levina, E.: Covariance regularization by thresholding. *Annals of Statistics* 36(6), 2577–2604 (2008)
6. Durrleman, S., et al.: Inferring brain variability from diffeomorphic deformations of currents: an integrative approach. *Medical Image Analysis* 12(5), 626–637 (2008)
7. Giraud, C.: Estimation of gaussian graphs by model selection. *Electron. J. Stat.* 2, 542–563 (2008)
8. Giraud, C., Huet, S., Verzelen, N.: Graph selection with ggmselect. arXiv:0907.0619
9. Jia, J., Yu, B.: On model selection consistency of the elastic net when $p \gg n$ (2008)
10. Kim, S.G., et al.: Structural connectivity via the tensor-based morphometry. *Bianual Meeting of Korean Society of Human Brain Mapping* (2010)
11. Lauritzen, S.: *Graphical models* 17 (1996)
12. Lee, H., et al.: Discriminative persistent homology of brain networks (2011)
13. Lee, H., et al.: Sparse brain network using penalized linear regression. *SPIE Medical Imaging* (2011)
14. Lee, H.e.a.: Sparse brain network recovery under compressed sensing. *IEEE Transactions on Medical Imaging* (2011)
15. Meinshausen, N., Bühlmann, P.: High-dimensional graphs and variable selection with the lasso. *Ann. Statist.* 34(3), 1436–1462 (2006)
16. Miller, M., et., al.: Morphometry BIRN. collaborative computational anatomy: An MRI morphometry study of the human brain via diffeomorphic metric mapping. *Human Brain Mapping* 30(7), 2132–2141 (2009)
17. Miller, M.I., Trouvé, A., Younes, L.: On the metrics and Euler-Lagrange equations of computational anatomy. *Annual Review of biomedical Engineering* 4 (2002)
18. Sabuncu, M., Balci, S.K., Golland, P.: Discovering modes of an image population through mixture modeling. *MICCAI LNCS(5242)*, 381–389 (2008)
19. Seeley, W., et al.: Neurodegenerative diseases target large-scale human brain networks 62(1), 42–52 (2009)
20. Smith, S., et al.: Network modelling methods for fMRI. *NeuroImage* 54(2), 875–891 (2011)
21. Tibshirani, R.: Regression shrinkage and selection via the lasso. *J. Royal. Statist. Soc B* 58 (1), 267–288 (1996)
22. Toews, M., et al.: Feature-based morphometry: Discovering group-related anatomical patterns. *Neuroimage* 49(3), 2318–2327 (2010)
23. Varoquaux, G., et al.: Brain covariance selection: better individual functional connectivity models using population prior. *NIPS* (2010)
24. Wang, L., et al.: Abnormalities of hippocampal surface structure in very mild dementia of the alzheimer type. *Neuroimage* 30:52-60 (2006)
25. Wang, Y., Chan, T.F., Toga, A.W., Thompson, P.M.: Shape analysis with multi-variate tensor-based morphometry and holomorphic differentials (2009)
26. Wright, J., Ma, Y., Mairal, J., Sapiro, G., Huang, T., Yan, S.: Sparse representation for computer vision and pattern recognition. *Proceedings of the IEEE* (2010)
27. Zou, H., Hastie, T.: Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B* (2005)