Structure Learning via Hadamard Product of Correlation and Partial Correlation Matrices

Karina Ashurbekova
GIPSA-lab, Inria
Univ. Grenoble Alpes
Grenoble, France
karina.ashurbekova@gipsa-lab.fr

Sophie Achard
CNRS, Grenoble INP, GIPSA-lab
Univ. Grenoble Alpes
Grenoble, France
sophie.achard@gipsa-lab.fr

Florence Forbes
Inria, CNRS, Grenoble INP, LJK
Univ. Grenoble Alpes
Grenoble, France
florence.forbes@inria.fr

Abstract—**Structure learning is an active topic nowadays in different application areas, i.e. genetics, neuroscience. Classical conditional independences or marginal independences may not be sufficient to express complex relationships. This paper is introducing a new structure learning procedure where an edge in the graph corresponds to a non zero value of both correlation and partial correlation. Based on this new paradigm, we define an estimator and derive its theoretical properties. The asymptotic convergence of the proposed graph estimator and its rate are derived. Illustrations on a synthetic example and application to brain connectivity are displayed.**

Index Terms—**structure learning, conditional independence, marginal independence, Hadamard product, soft-thresholding, brain connectivity**

I. Introduction

Finding relationships among a large number of random variables is a difficult problem that has generated a substantial interest in recent years. This is due to the large number of application areas where relationships are important features to extract. For example, in neuroscience, brain networks providing relationships between brain regions, have brought new understandings on the brain mechanisms under a pathological condition. Other examples can be found in genetics, medicine, finance, social networks. Graphical representations of probabilistic relationships have been extensively studied in machine learning and statistics. In particular, the so-called Graphical Models [11] are a way to represent probability density functions (pdf) that makes explicit factorization properties of the pdf and as a consequence helps in designing efficient inference algorithms. Given a set of p variables $Y = (Y_1, Y_2, \cdots Y_p)$, one important issue in practice, is therefore how to infer such a graph structure from a n-sample of $(Y_1, Y_2, \cdots Y_p)$. The appropriate procedure may of course depend on additional assumptions on the joint pdf of the Y_i's and on the type of relationships that is seek. Among the various dependence measures, conditional dependence is the most frequent as it is often seen as a more suitable concept to catch direct link than standard dependence, referred to as marginal dependence. Conditional independence of two variables Y_i and Y_j is defined as the independence of these two variables after regressing out the other variables. It is denoted by $Y_i \perp Y_j | Y_{-(i,j)}$ where $Y_{-(i,j)}$ is the set of variables Y after removing Y_i and Y_j.

A graph $G = (V, E)$ can then be constructed. The set of nodes V corresponds to the p variables in Y, and the edge set $E \subset V \times V$ is defined as follows

$$Y_i \perp Y_j | Y_{-(i,j)} \Leftrightarrow (i,j) \notin E. \qquad (1)$$

We will refer to this graph as a *partial correlation graph*.

Similarly, marginal independence of two variables Y_i and Y_j is the standard probabilistic independence denoted by $Y_i \perp Y_j$. A graph can also be defined as,

$$Y_i \perp Y_j \Leftrightarrow (i,j) \notin E. \qquad (2)$$

We will refer to this graph as a *correlation graph*. The two dependence measures and therefore the two graphs are not equivalent nor one is included in the other (eg. [8]). This is illustrated in Figure 1. Figure 1 (a) illustrates the so-called *common effect* phenomenon or v-structure. Using the Bayesian network representation, this graph corresponds to the joint distribution $p(x_1, x_2, x_3) = p(x_3|x_2, x_1) \; p(x_1) \; p(x_2)$ for which it is easy to check that X_2 and X_1 are independent but not conditionally independent. The correlation graph is the one to use in Figure 1 (a) to find the v-structure, while a partial correlation graph would give a triangle for this pdf. Conversely for the pdf represented in Figure 1 (b), $p(x_1, x_3, x_4) = p(x_3|x_1) \; p(x_4|x_1) \; p(x_1)$, X_3 and X_4 are marginally dependent but not conditionally independent. The partial correlation graph is the one to use while the correlation graph would be a triangle. More generally, for more complex graphs that would mix the two phenomenons, neither covariance graphs nor precision graphs would express our understanding of the variables relationships in a satisfying manner (see Figure 1 (c)). However, the intersection of the two graphs seems promising in structure learning.

In this work, we therefore propose to consider simultaneously both conditional and marginal dependencies and to address the problem of estimating them from observed data.

If the joint pdf of $(Y_1, Y_2, \cdots Y_p)$ is Gaussian, i.e. $Y \sim N_p(\mu, \Sigma)$, then conditional independence can be read directly from the precision matrix $\Theta = \Sigma^{-1}$ of the pdf. This provides a relatively straightforward way to assess conditional independencies from a data sample, by estimating the empirical precision matrix. Similarly, marginal independence can be read on the covariance matrix Σ which can also be estimated

(a) common effect   (b) common cause   (c) mixing model
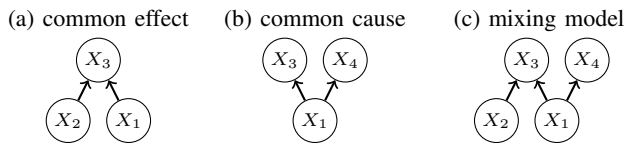

Fig. 1: Illustration of the common cause, common effect phenomena and the combination of common cause and common effect phenomena where neither correlation nor partial correlation can retrieve the real structure of the graph. This problem is a standard issue encountered in directed graphs exploration.

empirically. Most popular approaches resort to penalized likelihood estimation with the idea of favoring sparse matrices. Reviews of recent results including methods adapted for high-dimensional setting and heavy-tailed distributions can be found in [6], [8]. Starting with *graphical lasso (glasso)* [10], penalized likelihood approaches based on L_1-norm of the precision matrix Θ have been increasingly popular [1], [23]. Several estimators have been studied based on the use of Dantzig selector [22], CLIME algorithm [5] or thresholding operator [19]. In contrast to most other proposals, the estimator proposed in [19] has a closed-form solution as long as it is invertible. Similarly, [3] uses a *glasso* approach for the covariance matrix while recent progress on this problem has been made in [18] with a new optimization method based on coordinate descent. Other approaches [2], [14] consider generalized thresholding of the covariance matrix to get a sparse estimate. These methods have good theoretical properties and are computationally efficient. Finally, recent developments based on graph signal processing have brought new perspectives on this problem [7], [13]. Solutions are based on the study of eigenvalues and eigenvectors of Laplacian matrix using either correlation or partial correlation as inputs.

However, to our knowledge the existing methods use either Θ (conditional dependencies) or Σ (marginal dependencies) to infer the graph. The purpose of the present work is to develop a method to construct a graph where the absence of an edge between Y_i and Y_j corresponds to the absence of one of the properties: partial correlation or correlation. The graph is then defined by:

$$Y_i \perp Y_j | Y_{-(i,j)} \text{ or } Y_i \perp Y_j \Leftrightarrow (i,j) \notin E. \tag{3}$$

This paper is organized as follows. We briefly give an overview of the related work. A new definition of the structure of the graph using Hadamard product is introduced in section III. Theoretical properties of the proposed estimator are derived in section IV and an illustration is presented in section V. Finally, a validation on real data is described in section VI.

II. Prior work

Despite the very large number of existing papers dealing with conditional independence (i.e. classical Gaussian graphical models), there are few papers studying the *common cause - common effect* problem.

Using specific conditions on sparsity, [9] showed that positions of zeros in partial correlation and correlation matrices are equal. As a consequence for these conditions, both dependence measures, partial correlation or correlation, are estimating the same graphs. However, hypotheses on sparsity are hard to check, and we argue that they are too restrictive in many applications.

In the framework of Bayesian networks, heuristic approaches have been developed to iteratively decompose the graph and to find the v-structures. In [21] the authors propose to infer the dependence structure after identifying subset of nodes with a possible v-structure. This approach may be very costly as it is not providing an analytic solution. Alternative approaches based on graph signal processing tend to propose an optimization framework independent of the input matrices correlation or partial correlation. In [16], the authors derive a criterion based on eigenvectors decomposition of covariance matrix. Its minimization under the constraint of sparsity is achieved corresponding to either correlation or partial correlation. Using this method it is possible to retrieve the two examples given in Fig. 1 (a) or Fig. 1 (b). However for Fig. 1 (c), this method will not give the correct solution because of the combination of the two phenomena.

III. Hadamard structure learning

In this section, we describe a novel method for defining graphical models in the way that the absence of edge between two nodes means that either the correlation of these variables equal to zero or the partial correlation of these variables given the rest equal to zero. The idea is to incorporate both correlation matrix $R = (r_{ij})$ and partial correlation matrix $\Gamma = (\gamma_{ij})$ in one product. We focus on Hadamard product for its tractability and ability to reveal the absence of at least one of the properties among partial correlation or marginal correlation: $\mathcal{H} = R \circ \Gamma$, $\mathcal{H} = (h_{ij})$. Therefore, h_{ij} have zero on the place ij if at least one of r_{ij} and γ_{ij} equals to zero. The graph (3) is then defined by zeros of matrix $\mathcal{H}$ and we call this graph a *Hadamard graph*. The edges of this graph correspond to non-zeros values for both partial and marginal correlations.

As an example of the application of this new definition on real data, we used brain data using functional Magnetic Resonance Imaging (fMRI). The details of the preprocessing of the data is explained in [17]. Figure 2 is displaying the obtained correlation graph (a), partial correlation graph (b) and Hadamard graph (c). Each graph is obtained by defining 40 edges corresponding to the first 40th highest absolute values of respectively R (Figure 2 (a)), Γ (Figure 2 (b)) and $R \circ \Gamma$ (Figure 2 (c)). Different positioning of the edges on the structure of the brain are obtained. Correlation graph is highlighting more short-range connections (clustering of regions close in space), whereas partial correlation is capturing long-range connections (connection of regions that are far from each other). These two features can be related to the cost-efficiency trade-off of the brain [4]. It is assumed that the brain networks are constructed in order to both optimize the efficiency (facility to connect
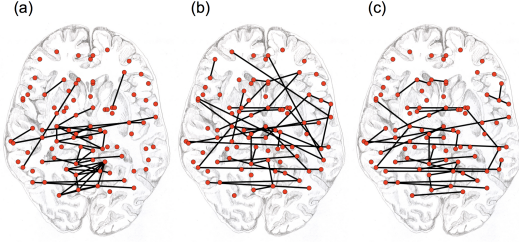
Fig. 2: Structure learning on brain graphs using fMRI data: (a) correlation graph; (b) partial correlation graph; (c) Hadamard graph. Short-range connections are prevalent in (a) whereas edges in (b) are mostly long-range connections. (c) is a combination of (a) and (b).

any two regions of the brain), and the cost (the cost to build structural connections). The Hadamard graph can be seen as an illustration of this trade-off where both short-range and long-range connections are present in the brain.

IV. Estimation of Hadamard graphs

For now, for simplicity, we assume that the sample correlation matrix $\widehat{R}$ is invertible. To recover the Hadamard product parameter $\mathcal{H}$ we propose the following estimator which allows us to regularize the number of connections defined as the non-zero elements in $\mathcal{H}$:

$$\arg\min_{\mathcal{H}} \ \|\mathcal{H}\|_{1,\text{off}}$$
$$\text{s.t.:} \ |\mathcal{H} - \widehat{\mathcal{H}}|_{\infty,\text{off}} \leqslant \lambda, \tag{4}$$

where $\widehat{\mathcal{H}} = \widehat{R} \circ \widehat{\Gamma}$ and $\widehat{\Gamma}$ is the estimator of the partial correlation matrix obtained by normalizing $\widehat{R}^{-1}$, for matrix A: $|A|_{\infty,\text{off}} = \max_{i \neq j} |A_{ij}|$, $\|A\|_{1,\text{off}} = \sum_{i \neq j} |A_{ij}|$. Note that similar approaches were considered in [19] and [20] but for the estimation of the precision and covariance matrices respectively. Optimization problem (4) is decomposable into independent element-wise sub-problems, where each sub-problem corresponds to soft-thresholding of $\mathcal{H}$ [20]. Thus, the estimator (4) is available in closed-form: $\widehat{\mathcal{H}}_\lambda = T_\lambda(\widehat{\mathcal{H}})$, where $[T_\lambda(u)]_{ij} = sign(u_{ij})max(|u_{ij}|-\lambda, 0)$. Now we derive theoretical properties for the thresholding of the Hadamard product.

We consider the following uniformity class of "approximately sparse" correlation matrices invariant under permutations presented in papers [2] and [14]:

$$\mathcal{U}_\tau(q_0, c_o(p)) = \big\{R, \sum_{j=1}^{p} |r_{ij}|^{q_0} \leq c_0(p) \ \forall i\big\}$$

for $0 \leq q_0 < 1$. When $q_0 = 0$, $\sum_{j=1}^{p} |r_{ij}|^{q_0} = \sum_{j=1}^{p} \mathbb{1}(r_{ij} \neq 0)$ which corresponds to the class of sparse matrices ($\mathbb{1}$ means the indicator function). The dependence of $c_0(p)$ on the dimension p allows to be less restricted on the sparsity [14]. We assume that the partial correlation matrix also belongs to

the class of "approximately sparse" matrices invariant under permutations:

$$\mathcal{U}_\tau(q_1, c_1(p)) = \big\{\Gamma, \sum_{j=1}^{p} |\gamma_{ij}|^{q_1} \leq c_1(p) \ \forall i\big\}$$

for $0 \leq q_1 < 1$. Then the Hadamard product of the correlation and partial correlation matrices is also in the class of "approximately sparse" matrices:

Proposition 1. *The Hadamard product $\mathcal{H}$ of correlation matrix R in the class $\mathcal{U}_\tau(q_0, c_o(p))$ and partial correlation matrix Γ in the class $\mathcal{U}_\tau(q_1, c_1(p))$ also belongs to the class of "approximately sparse" matrices $\mathcal{U}_\tau(q_2, c_2(p))$ for $0 \leq q_2 < 1$.*

Proof. Let us assume $q_2 = \frac{1}{2}min(q_0, q_1)$ and $c_2(p) = max(c_0(p), c_1(p))$. Then for i, $1 \leq i \leq p$,

$$\sum_{j=1}^{p} |h_{ij}|^{q_2} = \sum_{j=1}^{p} |r_{ij}\gamma_{ij}|^{q_2} = \sum_{j=1}^{p} |r_{ij}|^{q_2}|\gamma_{ij}|^{q_2}$$

$$\leq \sqrt{\sum_{j=1}^{p} |r_{ij}|^{2q_2}}\sqrt{\sum_{j=1}^{p} |\gamma_{ij}|^{2q_2}} \leq \sqrt{\sum_{j=1}^{p} |r_{ij}|^{q_0}}\sqrt{\sum_{j=1}^{p} |\gamma_{ij}|^{q_1}}$$

$$\leq (c_0(p))^{1/2}(c_1(p))^{1/2} \leq max(c_o(p), c_1(p))$$

This means that the Hadamard product belongs to

$$\mathcal{U}_\tau(q_2, c_2(p)) = \big\{H, \sum_{j=1}^{p} |h_{ij}|^{q_2} \leq c_2(p) \ \forall i\big\}$$

for some $0 \leq q_2 < 1$. ☐

The following theorem provides the convergence rate for $T_\lambda(\widehat{\mathcal{H}})$ in the matrix operator norm, for matrix A: $\|A\|^2 = \lambda_{\max}(AA^T)$, where $\lambda_{\max}(AA^T)$ represents the largest eigenvalue of matrix AA^T.

Theorem 1. *Suppose the distribution of the data is Gaussian with Hadamard product $\mathcal{H} = R \circ \Gamma$. Then, uniformly on $\mathcal{U}_\tau(q_2, c_2(p))$, for sufficiently large M', if $\lambda = M'\sqrt{\frac{\log p}{n}}$ and $\frac{\log p}{n} = o(1)$, then:*

$$P\Big[\|T_\lambda(\widehat{\mathcal{H}}) - \mathcal{H}\| < c_2(p)\Big(\frac{\log p}{n}\Big)^{(1-q_2)/2}\Big] \geq 1 - \epsilon_{n,p}, \tag{5}$$

where $\epsilon_{n,p}$ is a deterministic sequence that decreases to zero as $n, p \to \infty$.

Proof. By examining the proof of Theorem 1 in [14], the only property we need to prove is the following:

$$P\Big[\max_{ij} |h_{ij} - \widehat{h}_{ij}| < \frac{\log p}{n}\Big] \geq 1 - \epsilon_{n,p} \tag{6}$$

$$\max_{ij} |h_{ij} - \widehat{h}_{ij}| = \max_{ij} |r_{ij}\gamma_{ij} - \widehat{r}_{ij}\widehat{\gamma}_{ij}| \leq$$

$$\leq \max_{ij} |-\widehat{r}_{ij}\widehat{\gamma}_{ij} + \widehat{r}_{ij}\gamma_{ij}| + \max_{ij} |-\widehat{r}_{ij}\gamma_{ij} + r_{ij}\gamma_{ij}| =$$

$$= \max_{ij} |\widehat{r}_{ij}(\gamma_{ij} - \widehat{\gamma}_{ij})| + \max_{ij} |\gamma_{ij}(r_{ij} - \widehat{r}_{ij})| \leq$$

$$\leq \max_{ij} |\widehat{r}_{ij}| \max_{ij} |\gamma_{ij} - \widehat{\gamma}_{ij}| + \max_{ij} |\gamma_{ij}| \max_{ij} |r_{ij} - \widehat{r}_{ij}|$$

Applying theorem 1 [14] to each of the two terms of the right-hand side of the above inequality finishes the proof. □

Remark 1. *Sample covariance (correlation) matrices are not invertible unless the sample size n is larger than the number of variables p and we cannot compute a partial correlation matrix using the sample correlation matrix R. This issue can be solved using the thresholding operator of sample correlation matrix, i.e. Γ can be derived from $[T_\nu(\widehat{R})]^{-1}$. This idea was implemented in [20] and under additional assumptions on precision matrix it is shown that this estimator is well-defined and well-behaved.*

V. ILLUSTRATION ON SIMULATED DATA

We illustrate the behaviour of soft-thresholding operator of Hadamard product of estimated correlation matrix and partial correlation matrix in terms of the recovering of the sparsity pattern. In order to confirm the usefulness of our framework, we construct an example where neither correlation graph nor partial correlation graph could reconstruct the correct structure. First, we simulate two positive definite matrices R_1 and R_2 of size 100 each using models 1 and 2 respectively. **Model 1**: "Triangular" correlation, $r_{ij} = (1 - ((|i - j|)/k))_+$, for $k = [p/4]$. **Model 2**: inverse of "Triangular" correlation.
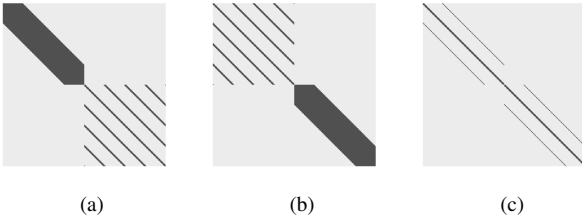

(a)　　　　　　(b)　　　　　　(c)

Fig. 3: Adjacency matrix of (a) correlation, (b) partial correlation and (c) Hadamard matrix.

Then we take the true correlation R matrix of size 200 as a block-diagonal matrix of R_1 and R_2. Model 1 is a standard test case in the literature. The adjacency matrices of the true correlation matrix R, partial correlation matrix and their Hadamard product are presented in Figure 3. We generated $n = 300$ independent and identically distributed p-variate normal random vectors with mean 0 and correlation matrix equal to R. The ability to recover sparsity patterns of the true Hadamard product was evaluated via ROC curves.

We compare two methods: soft-thresholding of Hadamard product $\widehat{\mathcal{H}}$, i.e. $T_\lambda(\widehat{\mathcal{H}}) = T_\lambda(\widehat{R} \circ \widehat{\Gamma})$, where $\widehat{R}$ is the sample correlation matrix and $\widehat{\Gamma}$ is the estimator of the partial correlation matrix obtained by normalizing the matrix $\widehat{R}$; Hadamard product of soft-thresholding of correlation matrix $\widehat{R}$ and soft-thresholding of partial correlation matrix $\widehat{\Gamma}$, i.e. $T_\lambda(\widehat{\mathcal{R}}) \circ T_\lambda(\widehat{\Gamma})$. The whole process was run 50 times. The corresponding ROC in Figure 4 shows that result for $T_\lambda(\widehat{\mathcal{H}})$ outperforms the result for $T_\lambda(\widehat{\mathcal{R}}) \circ T_\lambda(\widehat{\Gamma})$ for the FPR greater then 6%.

VI. VALIDATION ON REAL DATA

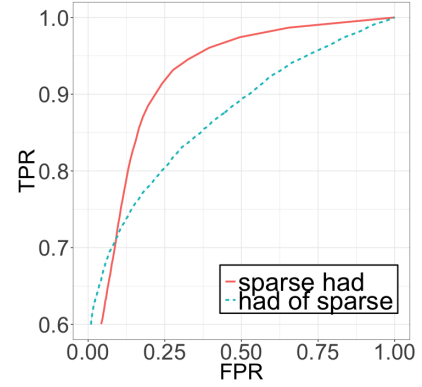In order to validate our approach on real data, we tested our graph structure learning approach on brain signals using

Fig. 4: ROC curve representing the performance of $T_\lambda(\widehat{\mathcal{R}}) \circ T_\lambda(\widehat{\Gamma})$ and $T_\lambda(\widehat{\mathcal{H}})$. The ROC curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings λ. For any estimated matrix $M_\lambda = (m(\lambda)_{ij})$ and the true Hadamard matrix $H = (h_{ij})$ TPR and FPR are defined as: TPR $= \frac{\#\{(i,j):m(\lambda)_{ij} \neq 0 \text{ and } h_{ij} \neq 0\}}{\#\{(i,j):h_{ij} \neq 0\}}$ and FPR $= \frac{\#\{(i,j):m(\lambda)_{ij} \neq 0 \text{ and } h_{ij} = 0\}}{\#\{(i,j):h_{ij} = 0\}}$.

fMRI. This provides brain connectivity networks that are used in the identification of cognitive impairments for example. As ground truth for evaluating brain connectivity networks is not available, we used test-retest datasets (see [17] for details) to quantify the performances of the different approaches. 100 subjects were scanned twice in two different sessions. The reliability of each structure learning approach is evaluated by measuring the difference of graph properties for the two sessions obtained for the same subject. For each datasets, we compute a graph using four different approaches: $\widehat{R}$, $\widehat{\Gamma}$, $\widehat{\mathcal{H}} = \widehat{R} \circ \widehat{\Gamma}$ and partial correlation matrix obtained from *glasso* algorithm; the tuning parameter λ in *glasso* algorithm for each subject in each session was obtained by cross-validation with $k = 5$ folds. In order to produce a fixed number of edges, we apply the soft-thresholding for each of the matrix. Thus we obtain an adjacency matrix that defines an unweighted graph for each subject in each session. We compute a graph metric called "global efficiency" which is related to the communication efficiency of a node i with all other nodes (detailed information can be found in [15]). For each graph we obtain one value of this parameter. Accordingly, there are 200 values of global efficiency (one per each subject in each session) for fixed value of percentage of edges in the graphs. To evaluate the reliability of each method we compute intraclass correlation coefficient (ICC) which is based on the comparison of the within-subject and between-subject variability. The Figure 5 represents the scheme of computing the ICC for fixed number of edges. ICC is close to 0 when the reliability is low, and close to 1 when the reliability is high. Note that ICC may take negative values when the variance within subjects is larger than between subjects. This is due to statistical errors given a particular dataset and should be considered as a non reliable estimation.

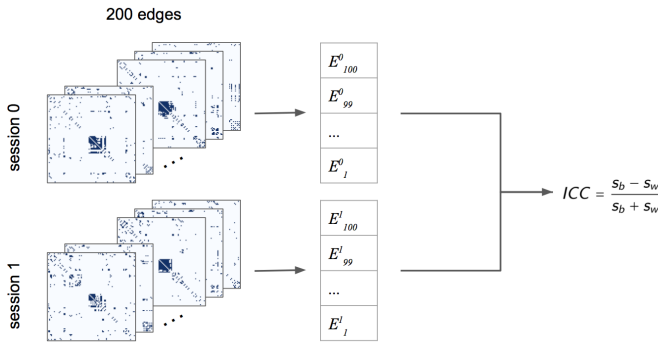The results are shown in Figure 6. It is clear that methods

Fig. 5: The computation of ICC for global efficiency for fixed number of edges equal to 200. $E_1^0, \ldots E_{100}^0, E_1^1, \ldots, E_{100}^1$ are values of global efficiency for each subject in both sessions; s_b is the variance between subjects, s_w is the variance within subjects in 2 sessions; $ICC = (s_b - s_w)/(s_b + s_w)$ (see [17] for more details about computing s_b and s_w).

based on partial correlation are the least reliable. Using *glasso* improves the reliability however correlation approaches are still better. For this datasets, Hadamard approach gives similar results as sample correlation, which means that brain connectivity networks seem to be better characterized by marginal dependence. The highest value of ICC is obtained when using Hadamard graphs for 30% of edges present in the graphs. When there are no hypotheses on the dependence structure of the graphs, these results show that Hadamard graph is a good trade off for structure learning.
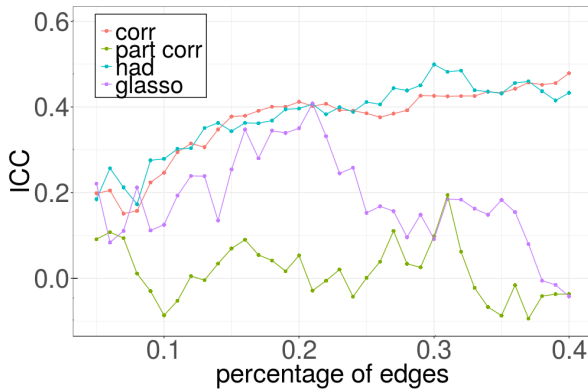

Fig. 6: Reliability measures using ICC for global efficiency for the sample correlation matrix $\widehat{R}$, the partial correlation matrix $\widehat{\Gamma} = \widehat{R}^{-1}$, Hadamard product $\widehat{\mathcal{H}}$ and partial correlation matrix obtained from *glasso* algorithm

VII. CONCLUSION

We introduce a new graph construction method taking into account both conditional and marginal independences. We derive a convergence bound for the proposed estimator and show on synthetic data the ability of the estimator to recover the sparsity patterns of the true matrix. Our approach was applied on brain signals and showed better results than

classical correlations or partial correlations. A nice advantage of our method is its simplicity where we only need to choose the estimation procedure for correlation and partial correlation matrices. In perspective, the estimation procedure for these matrices can be chosen to improve robustness to non-Gaussian data, for example with Kendall tau. Also, it is possible to use estimation based on shrinkaging of covariance matrix for high-dimensional case [12].

REFERENCES

[1] O. Banerjee, L. El Ghaoui, and A. d'Aspremont. Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. *J. Mach. Learn. Res.*, 9(Mar):485–516, 2008.

[2] P. J Bickel, E. Levina, et al. Covariance regularization by thresholding. *The Annals of Statistics*, 36(6):2577–2604, 2008.

[3] J. Bien and R. Tibshirani. Sparse estimation of a covariance matrix. *Biometrika*, 98(4):807–820, 2011.

[4] E. Bullmore and O. Sporns. The economy of brain network organization. *Nature Reviews Neuroscience*, 13(5):336, 2012.

[5] T. Cai, W. Liu, and Xi Luo. A constrained 1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, 106(494):594–607, 2011.

[6] T. Cai, Z. Ren, and H. Zhou. Estimating structured high-dimensional covariance and precision matrices: Optimal rates and adaptive estimation. *Electron. J. Stat.*, 10(1):1–59, 2016.

[7] X. Dong, D. Thanou, P. Frossard, and P. Vandergheynst. Learning laplacian matrix in smooth graph signal representations. *IEEE Transactions on Signal Processing*, 64(23):6160–6173, 2016.

[8] J. Fan, Y. Liao, and H. Liu. An overview of the estimation of large covariance and precision matrices. *Econom. J.*, 19(1), 2016.

[9] S. Fattahi and S. Sojoudi. Graphical lasso and thresholding: Equivalence and closed-form solutions. *arXiv preprint arXiv:1708.09479*, 2017.

[10] J. Friedman, T. Hastie, and R. Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.

[11] N. Friedman and D. Koller. Probabilistic graphical models: Principles and techniques, 2009.

[12] O. Ledoit and M. Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis*, 88(2):365–411, 2004.

[13] J. Mei and J. Moura. Signal processing on graphs: Estimating the structure of a graph. In *International Conference on Acoustics, Speech and Signal Processing*, pages 5495–5499. IEEE, 2015.

[14] A. J. Rothman, E. Levina, and Ji Zhu. Generalized thresholding of large covariance matrices. *Journal of the American Statistical Association*, 104(485):177–186, 2009.

[15] M. Rubinov and O. Sporns. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, 52(3):1059–1069, 2010.

[16] S. Segarra, A. G. Marques, G. Mateos, and A. Ribeiro. Network topology inference from spectral templates. *IEEE Transactions on Signal and Information Processing over Networks*, 3(3):467–483, 2017.

[17] M. Termenon, C. Delon-Martin, A. Jaillard, and S. Achard. Reliability of graph analysis of resting state fMRI using test-retest dataset from the human connectome project. *Neuroimage*, 142(15):172–187, 2016.

[18] H. Wang. Coordinate descent algorithm for covariance graphical lasso. *Statistics and Computing*, 24(4):521–529, 2014.

[19] E. Yang, A. C. Lozano, and P. Ravikumar. Elementary estimators for graphical models. In *Advances in neural information processing systems*, pages 2159–2167, 2014.

[20] E. Yang, A. C. Lozano, and P. Ravikumar. Elementary estimators for sparse covariance matrices and other structured moments. In *International conference on machine learning*, pages 397–405, 2014.

[21] R. Yehezkel and B. Lerner. Bayesian network structure learning by recursive autonomy identification. *Journal of Machine Learning Research*, 10(Jul):1527–1570, 2009.

[22] M. Yuan. High dimensional inverse covariance matrix estimation via linear programming. *Journal of Machine Learning Research*, 11(Aug):2261–2286, 2010.

[23] M. Yuan and Y. Lin. Model selection and estimation in the Gaussian graphical model. *Biometrika*, 94(1):19–35, 2007.