International Journal of Automation and Computing 14(4), August 2017, 371-385
DOI: 10.1007/s11633-017-1084-9

Inferring Functional Connectivity in fMRI Using Minimum Partial Correlation

Lei Nie[1] Xian Yang[1] Paul M. Matthews[2] Zhi-Wei Xu[3] Yi-Ke Guo[1,4]

[1]Department of Computing, Imperial College London, London SW7 2AZ, UK
[2]Department of Medicine, Imperial College London, London SW7 2AZ, UK
[3]Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China
[4]School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China

Abstract: Functional connectivity has emerged as a promising approach to study the functional organisation of the brain and to define features for prediction of brain state. The most widely used method for inferring functional connectivity is Pearson′s correlation, but it cannot differentiate direct and indirect effects. This disadvantage is often avoided by computing the partial correlation between two regions controlling all other regions, but this method suffers from Berkson′s paradox. Some advanced methods, such as regularised inverse covariance, have been applied. However, these methods usually depend on some parameters. Here we propose use of minimum partial correlation as a parameter-free measure for the skeleton of functional connectivity in functional magnetic resonance imaging (fMRI). The minimum partial correlation between two regions is the minimum of absolute values of partial correlations by controlling all possible subsets of other regions. Theoretically, there is a direct effect between two regions if and only if their minimum partial correlation is non-zero under faithfulness and Gaussian assumptions. The elastic PC-algorithm is designed to efficiently approximate minimum partial correlation within a computational time budget. The simulation study shows that the proposed method outperforms others in most cases and its application is illustrated using a resting-state fMRI dataset from the human connectome project.

Keywords: Functional connectivity, functional magnetic resonance imaging (fMRI), network modelling, partial correlation, PC-algorithm, resting-state networks.

1 Introduction

Functional connectivity is defined as measurable temporal dependences among spatially segregated brain regions of interest (ROIs)[1]. It is a statistical summarization of brain signals rather than a generative model of brain functions. Although functional connectivity cannot give a comprehensive interpretation about how brain works, it can show important features of brain activities at a macroscale level. For example, functional connectivity can inspire hypotheses of brain organisation[2] and contribute to predictive models[3], e.g., to characterise pathological[4] or cognitive states[5].

Implicit in interpretations is the hypothesis that functional connectivity describes true connections (i.e., direct causal effects) between nodes (i.e., distinguishable functional anatomic regions of interest). The accuracy of functional connectivity is typically evaluated in three stages: in terms of the "skeleton" or organisation, strength and, in some instances, directionality. These stages can be considered to be conceptually sequential. This paper focuses on inferring the skeleton, which is to determine whether a connection between two ROIs exists or not. Smith et al.[6] systematically investigated the accuracy of 12 methods for skeleton inference using 28 synthetic networks and their corresponding functional magnetic resonance imaging (fMRI) blood-oxygen-level dependent (BOLD) signals generated from the dynamic causal model (DCM)[7]. The results show that the top methods are inverse covariance (ICOV), fully partial correlation, Bayes net and full correlation.

Among the above four methods, full correlation is the most widely used method. Application of full correlation methods to fMRI datasets enables inferences regarding brain structures[8] and provides powerful features for brain "decoding"[9]. However, while full correlation is robust and parameter-free, there are well-known intrinsic limitations to differentiation of direct and indirect effects. Fig. 1 shows an example of differentiating direct and indirect effects in the network where Node 1 → Node 2 → Node 3. The full correlation between Node 1 and Node 3 is as high as 1.0, but there is no real connection between the two nodes.

Fully partial correlation[10] often can successfully identify the direct effects in Fig. 1 by regressing out the intermediate node. The fully partial correlation of two nodes is their correlation when all other nodes in the network are controlled. "Fully partial correlation" was called "partial correlation" in previous studies about functional connectivity. Mathematically, partial correlation is used to express the correlation when any subset of nodes is controlled. To avoid ambiguity, fully partial correlation and partial correlation are explicitly distinguished in this paper. It is difficult to estimate the fully partial correlation when the number of data points is less than the number of nodes. Methods based on a sparse

Research Article
Manuscript received August 20, 2016; accepted February 21, 2017; published online June 13, 2017
Recommended by Associate Editor Hong Qiao
© The Author(s)

linear regression model with L_1-norm regularisation[11] and on both L_1-norm and L_2-norm regularisation[12] were developed. In [6], it is shown that fully partial correlation can perform well. Recent studies show that fully partial correlation of fMRI signals was correlated with lifestyle, demographic and psychometric measures[13, 14]. However, this strategy sometimes causes two independent nodes to become conditionally dependent, which is known as Berkson's paradox[15]. As illustrated by Fig. 2, there is no connection between Node 1 and Node 2, but the fully partial correlation between these two nodes is very close to 1.0.

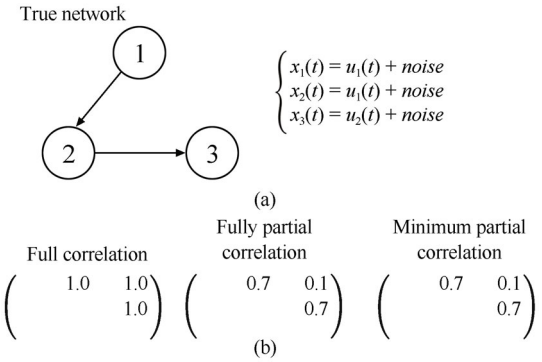


Fig. 1 An example of differentiating direct and indirect effects. (a) The left is the structure of a 3-node network, the right is the dynamics of each node. The input signal $u_1(t)$ is the BOLD signal of a voxel from a real fMRI image. The noise is randomly sampled from a Gaussian distribution with mean 0 and standard deviation 0.01. (b) The absolute values of full correlation, fully partial correlation and minimum partial correlation, respectively. The direct effects (1–2 and 2–3) and the indirect effect (1–3) have the same values of full correlation.

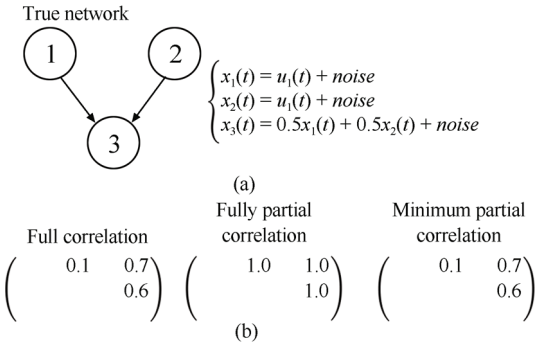


Fig. 2 An example of the Berkson's paradox. (a) The left is the structure of a 3-node network, the right is the dynamics of each node. The input signals $u_1(t)$ and $u_2(t)$ are the BOLD signals of two voxels from a real fMRI image. The correlation between $u_1(t)$ and $u_2(t)$ is -0.035. The noise is randomly sampled from a Gaussian distribution with mean 0 and standard deviation 0.01. (b) The absolute values of full correlation, fully partial correlation and minimum partial correlation, respectively. The fully partial correlation of Node 1 and Node 2 is as high as 1.0, although there is no effect between the two nodes.

ICOV is a regularised version of fully partial correlation, which tends to reduce the number of non-zero elements of the precision matrix[16, 17]. This method was extended to incorporate both functional and diffusion imaging data[18]. ICOV may correct Berkson's paradox if the regularisation parameter is set to an appropriate value. The main drawback of ICOV is that its accuracy highly depends on the regularisation parameter. As is observed in Fig. 3 (a), the absolute value of ICOV between Node 1 and Node 3 is significantly smaller than those of other node pairs when the value of the regularisation parameter is smaller than 1.0, which means the direct and indirect effects can be correctly distinguished. However, the absolute value of ICOV between Node 1 and Node 3 is closed to the values of other node pairs if the value of the regularisation parameter is larger than 1.0. Thus, the direct and indirect effects cannot be identified.

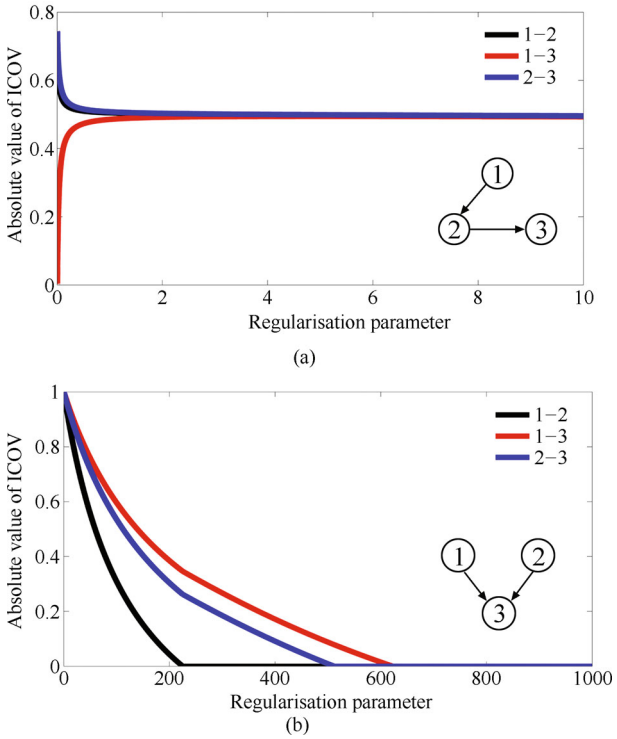


Fig. 3 ICOV with various values of the regularisation parameter. (a) Absolute values of ICOV for the network in Fig. 1 with the regularisation parameter varying from 0 to 10. (b) Absolute values of ICOV for the network in Fig. 2 with the regularisation parameter varying from 0 to 1 000.

One possible way to overcome this drawback is to use cross-validation for automatically adjusting the regularisation parameter. For real datasets, the cross-validation scores were usually calculated based on the accuracies of predicted BOLD signals rather than inferred connections[17], because the real connections are unknown. However, the validity of tuning the regularisation parameter in this way is still doubtful for two reasons. One reason is that a value of the parameter that results in accurate prediction (e.g., fMRI BOLD signals) is not necessarily same as the one that is likely to recover the true model (e.g., real connections)[19]. The other reason is that functional connectivity provides a data summary rather than

a generative model[1], so it may not be reasonable to use functional connectivity to replicate fMRI BOLD signals for cross-validation. Several other approaches were also proposed for overcoming this drawback. Utilizing the monotone property of the lasso path of ICOV, Huang et al.[20] used the largest regularization parameter value that preserves the existence of a connection as a quasi-measure for the strength of this connection. Recently, a statistic also based on the lasso path was proposed to test whether a regularised inverse covariance matrix represents all the real connections[21, 22]. The statements of their method are asymptotic and based on strong mathematical assumptions. Even worse, the proper value of the regularisation parameter for a sub-network may not be suitable for other sub-networks. The existence of an appropriate value for a large-scale network is still unclear.

Here we propose use of minimum partial correlation as a parameter-free measure for the skeleton of functional connectivity in fMRI. The minimum partial correlation between two nodes is the minimum of absolute values of partial correlations by controlling on all possible subsets of other nodes. Fully partial correlation, which controls all other nodes rather than a proper subset of other nodes, can be regarded as an approximation of minimum partial correlation. Theoretically, there is a direct effect between two nodes if and only if their minimum partial correlation is non-zero under the faithfulness assumption[23] and Gaussian assumption. However, it is time-consuming to calculate minimum partial correlation, since the number of all possible controlling subsets grows exponentially with the number of nodes. To address this, we have modified the PC-algorithm to approximate the minimum partial correlation for an "elastic PC-algorithm". Unlike the PC-algorithm and its other variations, the elastic PC-algorithm is not based on a specific significance threshold. Within a given time budget, the elastic PC-algorithm can efficiently approach the minimum partial correlation as possible. We have evaluated the elastic PC-algorithm using the NetSim dataset[6] and illustrate its application with a resting-state fMRI dataset from the human connectome project[24]. This paper is an extension of its conference version[25]. More details of the algorithm and applications are provided in this journal version.

2 Methods

2.1 Preliminaries

In this paper, a graph is denoted as $G = (V, E)$, which consists of a node set $V = \{r_1, r_2, \cdots, r_N\}$ and an edge set E. A directed graph contains only directed edges, an undirected graph contains only undirected edges. A graph is usually represented by an adjacency matrix M_G. For a directed graph, there is an edge from the node r_i to the node r_j if and only if the element m_{ji} in the j-th row and i-th column is non-zero. For an undirected graph, the adjacency matrix is symmetric. A non-zero element in the adjacency matrix represents an undirected edge. Given a directed graph, its skeleton is an undirected graph by stripping away all arrowheads from its edges and getting rid of redundant undirected edges. If there is a directed edge from the node r_i to the node r_j, then the node r_i is a parent of the node r_j and the node r_j is a child of the node r_i. If there is a directed path from the node r_i to the node r_j, then the node r_i is an ancestor of the node r_j and the node r_j is a descendant of the node r_i.

Given a directed acyclic graph (DAG) $G = (V, E)$, each node denotes a random variable and each directed edge represents a causal effect from one random variable to another one. A causal model represented by G generates a joint probability distribution over the nodes, which is denoted as $P(V)$.

Definition 1. For a DAG $G = (V, E)$ and a distribution $P(V)$ generated by the causal model represented by G, G and P satisfy the causal Markov Condition if and only if for every node $r \in V$, r is independent of $V \setminus (Descendants(r) \bigcup Parents(r))$ given $Parents(r)$ [23].

When the causal Markov condition is satisfied, the conditional independence relationships in the generated distribution P are entailed in the DAG G using the concept called d-separation. For a 3-node path $r_1 \rightarrow r_2 \leftarrow r_3$, the middle node r_2 is called collider.

Definition 2. A path is said to be d-separated by a set Z if and only if at least one non-collider is in Z or at least one collider and all its descendants are not in Z. Two nodes are said to be d-separated by a set Z if and only if all paths between the two nodes are d-separated by Z[23].

If two nodes r_i and r_j in DAG G are d-separated by a node set Z, then r_i and r_j are conditionally independent in distribution P given Z. There is no edge between nodes r_i and r_j if and only if they can be d-separated by some node set. This means that the nodes r_i and r_j are conditionally independent in the distribution P given some node set, if there is no edge in DAG G between the two nodes. However, the reverse statement is not true. For some DAG G and distribution P that satisfy the causal Markov condition, the conditional independence between two nodes r_i and r_j does not necessarily imply the absence of an edge between r_i and r_j. In order to exclude these cases, two equivalent concepts, stability[26] and faithfulness[23], are introduced.

Definition 3. A distribution P is faithful to a DAG G if all and only the conditionally independence relations in P are entailed by G, i.e., under the faithfulness assumption, there is no edge between two nodes r_i and r_j in G if and only if r_i and r_j are conditional independent in P given some node set[23].

For the details of Bayesian networks and causal Bayesian networks, please refer to the two seminal texts[23,26]. For references to applications of Bayesian networks in neuroscience, refer to the two recent reviews[27, 28].

2.2 Minimum partial correlation

In this paper, we assume the mechanism that generates BOLD signals can be represented by a DAG $G = (V, E)$, where $V = \{r_1, r_2, \cdots, r_N\}$. The node r_i denotes the ran-

dom variable of the i-th region of interest (ROI). The random variable r_i generates T samples of the BOLD signals of the i-th ROI, which are denoted as $X_i = \{x_i^1, x_i^2, \cdots, x_i^T\}$. All random variables generate the same number of samples. The Pearson′s correlation coefficient $\rho_{i,j}$ between r_i and r_j can be estimated from samples as follows:

$$\hat{\rho}_{i,j} = \frac{\sum_{t=1}^{T}(x_i^t - \bar{x}_i)(x_j^t - \bar{x}_j)}{\sqrt{\sum_{t=1}^{T}(x_i^t - \bar{x}_i)^2 \sum_{t=1}^{T}(x_j^t - \bar{x}_j)^2}} \tag{1}$$

where $\bar{x}_i$ and $\bar{x}_j$ are sample means of random variables r_i and r_j. The Pearson′s correlation coefficient is also called full correlation.

Full correlation counts both direct and indirect effects between two random variables. To remove indirect effects, we can calculate the second order correlation between r_i and r_j by removing the effects of a node set $Z \subseteq (V - \{r_i, r_j\})$. This second order correlation is called the partial correlation $\rho_{i,j\cdot Z}$ by controlling the node set Z. Similarly, partial correlation can be estimated from samples:

$$\hat{\rho}_{i,j\cdot Z} = \frac{\sum_{t=1}^{T}(\varepsilon_{i\cdot Z}^t - \bar{\varepsilon}_{i\cdot Z})(\varepsilon_{j\cdot Z}^t - \bar{\varepsilon}_{j\cdot Z})}{\sqrt{\sum_{t=1}^{T}(\varepsilon_{i\cdot Z}^t - \bar{\varepsilon}_{i\cdot Z})^2 \sum_{t=1}^{T}(\varepsilon_{j\cdot Z}^t - \bar{\varepsilon}_{j\cdot Z})^2}} \tag{2}$$

where $\{\varepsilon_{i\cdot Z}^t\}$ is a set of the residuals of the linear regression with r_i as the response variable and Z as the predictor variable set. So is $\{\varepsilon_{j\cdot Z}^t\}$. The sample partial correlation is the sample full correlation on the residuals when some controlling factors are regressed out.

The partial correlation between two nodes varies with the controlling set. For a node set V, the minimum partial correlation (MPC) between two nodes is the minimum of absolute values of partial correlations by controlling all possible subsets of other nodes, which is formally defined as follows.

Definition 4. Given two nodes r_i and r_j in a node set V, the minimum partial correlation $\omega_{i,j}$ between the two nodes is

$$\omega_{i,j} = \min\{|\rho_{i,j\cdot Z}||Z \in 2^{(V-\{r_i,r_j\})}\} \tag{3}$$

where $\rho_{i,j\cdot Z}$ is the partial correlation between r_i and r_j by controlling the node set Z. $2^{(V-\{r_i,r_j\})}$ represents all subsets of the set $V - \{r_i, r_j\}$. A controlling set Z that satisfies $\rho_{i,j\cdot Z} = \omega_{i,j}$ is called a minimum controlling set for nodes r_i and r_j.

Next, we will discuss an important property of minimum partial correlation that implies that it is a good measure of the existence of functional connectivity.

Proposition 1. Assuming the distribution P of the nodes V in a DAG G is multivariate Gaussian and faithful to G, there is an edge between the two nodes r_i and r_j in G if and only if the minimum partial correlation $\omega_{i,j} \neq 0$.

According to [29] , the partial correlation $\rho_{i,j\cdot z} = 0$ if and only if r_i and r_j are conditionally independent given Z under Gaussian assumption. Based on this result and Definition 3, Proposition 1 can be easily proved.

According to Proposition 1, minimum partial correlation is a measure of the existence of functional connectivity (i.e., the links in functional networks). Because minimum partial correlation is symmetrical, it cannot infer the directions of functional connectivity. According to Definition 1, minimum partial correlation can be calculated only based on partial correlation. Thus, minimum partial correlation can be estimated from samples, which is denoted as $\hat{\omega}_{i,j}$.

We denote the cardinality of the controlling node set Z as $|Z|$. According to [30], the distribution of the estimated partial correlation $\hat{\rho}_{i,j\cdot Z}$ is the same as the full correlation estimated from $T - |Z|$ samples. Under the hypothesis that $\rho_{i,j\cdot Z} = 0$, the z-score $\tilde{\rho}_{i,j\cdot Z}$ of the estimated partial correlation $\hat{\rho}_{i,j\cdot Z}$ can be calculated as follows:

$$\tilde{\rho}_{i,j\cdot Z} = \frac{1}{2}\log\left(\frac{1+\hat{\rho}_{i,j\cdot Z}}{1-\hat{\rho}_{i,j\cdot Z}}\right)\sqrt{T - |Z| - 3}. \tag{4}$$

However, the distribution of estimated minimum partial correlation is unknown. In this paper, we approximate the absolute value of z-score $\tilde{\omega}_{i,j\cdot Z}$ of the estimated minimum partial correlation $\hat{\omega}_{i,j\cdot Z}$ as

$$\tilde{\omega}_{i,j\cdot Z} = \min\{|\tilde{\rho}_{i,j\cdot Z}||Z \in 2^{(V-\{r_i,r_j\})}\}. \tag{5}$$

Given the BOLD signal samples from all ROIs of a brain, we can calculate the estimated value or the absolute value of its z-score of the minimum partial correlation for each pair of ROIs. These estimated values or absolute values of z-scores form a symmetric matrix Ω. The matrix Ω represents the skeleton of functional connectivity in a fractional way. Given a directed graph G, its skeleton is an undirected graph by stripping away all arrowheads from its links and removing redundant undirected links. A pair of ROIs with higher estimated minimum partial correlation value or absolute value of the z-score is more likely to have a link between them. The matrix Ω is therefore a parameter-free measure of functional connectivity, which can be used for decoding cognitive states, evaluating brain disease, etc.

There are three advantages of parameter-free methods for inferring functional connectivity. The most obvious one is that parameter-free methods are user-friendly, prior knowledge and experiences are not required. A second advantage is that parameter-free methods increase the comparability of results from different studies. If there are some tuning parameters in a method, the results usually vary according to the values of the parameters. Even though all experimental conditions are identical in two studies, it is not valid to compare the two results if different values of parameters are used in the two studies. A third advantage is that interpretability of parameter-free methods is usually better than other methods. Taking ICOV as an example, the physiological meaning of results showing that if a connection exists under some values of the regularisation parameter, but not with others is hard to explain.

2.3 Elastic PC-algorithm

It is in polynomial time to estimate partial correlation by controlling a given node set. However, we need to enumerate all possible controlling sets in order to estimate minimum partial correlation, the number of all possible controlling sets is exponential. Several algorithms for causal inference can be easily modified to calculate minimum partial correlation, such as SGS-algorithm[23], PC-algorithm[23], TPDA[31], MMHC[32] and TPMB[33].

This paper focuses on extending PC-algorithm to approximately calculate minimal partial correlation. The PC-algorithm is composed of two main steps: skeleton graph inference and edge direction determination. This paper aims at generating functional connectivity without orientations of edges. Thus we concentrate on the first step of the PC-algorithm.

The PC-algorithm uses an important feature of Bayesian networks to achieve fast computation. That is, if two nodes r_i and r_j can be d-separated, then there is a subset of their neighbours which d-separates them. Therefore, there is no need to enumerate all possible controlling sets. Instead the PC-algorithm only considers subsets of the neighbour sets. Because the structure of the graph is not known in advance, the PC-algorithm implements an iterative strategy: One step in the PC-algorithm is based on the graph reduced by the previous steps. The edges of the graph are removed according to a statistical test and a significance threshold. The more edges are removed, the smaller the search space is. However, if an edge is deleted incorrectly, this search space will be improperly reduced. These mistakes will propagate to the following steps.

The edge deletion is highly depended on the choice of the significance threshold α. The null hypothesis of the statistical test in the PC-algorithm is that there is no edge between the selected two nodes. With a lower significance threshold α, it is less likely that the null hypothesis is rejected. In other words, edges are more likely to be removed, which will increase type II error. To increase the accuracy of edge removal at each step, a larger α is required. However, the increase of α will increase the calculation time as more edges are remained. Therefore, the significance threshold α balances the accuracy of result and complexity of computation.

It is difficult to optimally determine α, because the accuracy and execution time of a given α cannot be predicted in advance. A common choice of α is 0.05. We can first set α to 0.05 and gradually increase its value to obtain more results with higher accuracy. This can avoid a situation in which the PC-algorithm runs for an undesirably long time for evaluation of a large choice of α. For example, we would like to choose α from the set $\{0.05, 0.1, 0.15, 0.2, 0.25\}$. We first carry out the PC-algorithm with $\alpha = 0.05$ and then with larger values. If the algorithm keeps running for a long time when $\alpha = 0.2$, we can just record the results for $\alpha = 0.15$, as it is the most accurate result that can be obtained within an acceptable execution time.

Although it has been demonstrated that the above approach achieved a satisfied accuracy for inferring functional connectivity in simulated fMRI datasets[34], this approach brings in extra computational loads, as the results with lower significance threshold are not used in the calculation with higher significance thresholds. To save computational effort, this paper proposes the elastic PC-algorithm, taking the concept of elasticity in cloud computing[35]. Our method automatically increases the significance threshold within a given time budget to maximally approach the minimum partial correlation. More specifically, an execution of the elastic PC-algorithm includes several iterations. The execution starts from an iteration with a low significance threshold and continues on the iterations with higher thresholds. The execution stops if the given time budge is used up. An iteration with a higher significance threshold can exploit the results of the iterations with lower significance thresholds in order to avoid repeated calculation of same partial correlation in different iterations. An iteration of elastic PC-algorithm is shown in Algorithm 1 and the execution process is shown in Algorithm 2.

Algorithm 1. An iteration of elastic PC-algorithm
Input: signal samples $\{x_i^t \| t \in [1:T], i \in [1:N]\}$, significance threshold α, previous significance threshold β, previous s-cube D for β
Output: s-cube C for α
1) **for each** ordered node pair (r_i, r_j) **do**
2) $C(i, j, 0 : (N-2)) \leftarrow |\tilde{\rho}_{i,j}|$
3) **end for**
4) **for** $k = 1 : (N-2)$ **do**
5) initialize the referenced skeletons S_α and S_β as complete undirected graphs
6) **for each** unordered node pair (r_i, r_j) **do**
7) **if** $C(i, j, k-1) \leq N^{-1}(1 - \frac{\alpha}{2})$ **then**
8) delete the undirected edge (r_i, r_j) from the referenced skeleton S_α
9) **end if**
10) **if** $D(i, j, k-1) \leq N^{-1}(1 - \frac{\beta}{2})$ **then**
11) delete the undirected edge (r_i, r_j) from the referenced skeleton S_β
12) **end if**
13) **end for**
14) $C(:, :, k) \leftarrow D(:, :, k)$
15) **for each** ordered node pair (r_i, r_j) that satisfies $|adj(S_\alpha, r_i) \backslash \{r_j\}| \geq k$ **do**
16) **for each** controlling node set $Z \subseteq adj(S_\alpha, r_i) \backslash \{r_j\}$ that satisfies $|Z| = k$ and $Z \bigcup \{r_j\} \not\subseteq adj(S_\beta, r_i)$ **do**
17) **if** $C(i, j, k) > |\tilde{\rho}_{i,j \cdot Z}|$ **then**
18) $C(i, j, k : (N-2)) \leftarrow |\tilde{\rho}_{i,j \cdot Z}|$
19) $C(j, i, k : (N-2)) \leftarrow |\tilde{\rho}_{i,j \cdot Z}|$
20) **end if**
21) **end for**
22) **end for**
23) **end for**

Algorithm 2. Elastic PC-algorithm (EPC)
Input: signal samples $\{x_i^t \| t \in [1:T], i \in [1:N]\}$, step δ
Output: s-cube C
1) **for each** ordered node pair (r_i, r_j) **do**

2) $C(i,j,0:(N-2)) \leftarrow |\tilde{\rho}_{i,j}|$
3) **end for**
4) $\beta = 0$
5) $\alpha = \beta + \delta$
6) **while** remaining time budget > 0 **do**
7) $C \leftarrow$ Call Algorithm 1 $\left(\{x_i^t\}, \alpha, \beta, C\right)$
8) **end while**

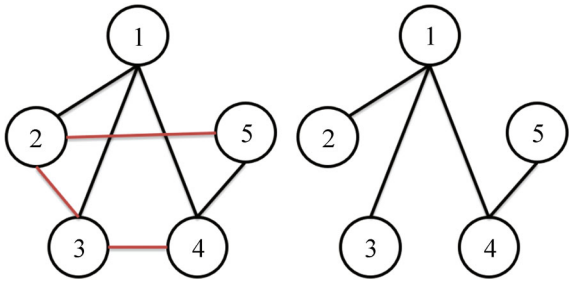
The elastic PC-algorithm is based on PC-stable algorithm[36], which is a recent variant of PC-algorithm. The symbol $N^{-1}(\cdot)$ represents the inverse of the standard normal cumulative distribution function (CDF). The z-score of full correlation $\tilde{\rho}_{i,j}$ and the z-score of partial correlation $\tilde{\rho}_{i,j\cdot Z}$ can be calculated according to (4). This symbol $adj(G, r_i)$ indicates the neighbour set of node r_i in graph G.

In the elastic PC-algorithm, we introduce a 3D matrix that contains the minimum partial correlation values with different steps. We call this matrix as s-cube for short. In the algorithm, there are two s-cubes C and D, which are for the significance thresholds α and β respectively. As there are N nodes in the graph G and k should be smaller than $|adj(G, r_i) \setminus \{r_j\}|$ for any pair of nodes (r_i, r_j), the maximum value of k is $N-2$. Thus, k can vary from 0 to $N-2$, which indicates the number of neighbours. The dimension of s-cube is $N \times N \times (N-1)$, where the k-th slice $C(:,:,k)$ contains minimum partial correlation calculated in Step k.

The inputs of the elastic PC-algorithm include signal samples, significance threshold α, previous significance threshold β and its corresponding s-cube D. At beginning, the s-cube C is initialized using the full correlation between any node pairs. In the k-th iteration, the reference skeletons S_α and S_β are derived using the $(k-1)$-th slice of s-cube C and D respectively. A reference skeleton is used as an approximation of the skeleton of the graph for calculating neighbour sets of nodes. Afterwards, the k-th slice of C is calculated. For an ordered pair of (r_i, r_j) that is adjacent in S_α but not adjacent in S_β, the minimum partial correlation should be calculated as it is not included in D. For an ordered pair of (r_i, r_j) that is adjacent in both S_α and S_β, some steps of minimum partial correlation calculation have been done in D, which can be directly reused. That is, if a subset of neighbours Z for (r_i, r_j) in S_α is also a subset of neighbours for (r_i, r_j) in S_β, there is no need to calculate the partial correlation $\tilde{\rho}_{i,j\cdot Z}$. In this way, computational effort can be saved by avoiding repeated calculations. The estimated z-score matrix $\tilde{\Omega}$ of minimum partial correlation for significance threshold α is $C(j, i, N-2)$. As shown in Algorithm 2, Algorithm 1 is repeatedly executed with increasing significance threshold within the given time budget. An example of the elastic PC-algorithm is illustrated in Fig. 4.

3 Results

In this section, the elastic PC-algorithm was evaluated using a synthetic dataset: NetSim[6]. Its application is illustrated using a resting-state fMRI dataset from the human connectome project[24]. Results with the elastic PC-algorithm were compared with full correlation, fully partial correlation, regularized inverse covariance (ICOV)[16, 17], network deconvolution (ND)[37] and global silencing (GS)[38]. A brief description of these methods is as follows:



#	Partial correlation	Need calculation?	#	Partial correlation	Need calculation?
1	$\bar{\rho}_{1,2\cdot\{3\}}$	No	11	$\bar{\rho}_{2,3\cdot\{5\}}$	Yes
2	$\bar{\rho}_{1,2\cdot\{4\}}$	No	12	$\bar{\rho}_{2,5\cdot\{1\}}$	Yes
3	$\bar{\rho}_{1,2\cdot\{5\}}$	Yes	13	$\bar{\rho}_{2,5\cdot\{3\}}$	Yes
4	$\bar{\rho}_{1,3\cdot\{2\}}$	No	14	$\bar{\rho}_{2,5\cdot\{4\}}$	Yes
5	$\bar{\rho}_{1,3\cdot\{4\}}$	No	15	$\bar{\rho}_{3,4\cdot\{1\}}$	Yes
6	$\bar{\rho}_{1,4\cdot\{2\}}$	No	16	$\bar{\rho}_{3,4\cdot\{2\}}$	Yes
7	$\bar{\rho}_{1,4\cdot\{3\}}$	No	17	$\bar{\rho}_{3,4\cdot\{5\}}$	Yes
8	$\bar{\rho}_{1,4\cdot\{5\}}$	No	18	$\bar{\rho}_{4,5\cdot\{1\}}$	No
9	$\bar{\rho}_{2,3\cdot\{4\}}$	Yes	19	$\bar{\rho}_{4,5\cdot\{2\}}$	Yes
10	$\bar{\rho}_{2,3\cdot\{4\}}$	Yes	20	$\bar{\rho}_{4,5\cdot\{3\}}$	Yes

(c)

Fig. 4 An example of the elastic PC-algorithm. (a) and (b) display the reference skeletons S_α and S_β in Algorithm 1, satisfying $\alpha > \beta$. The table in (c) shows calculations required at Step 1 with the significance threshold α. The column "need calculation?" tells us whether the corresponding partial correlation needs to be calculated or can be directly obtained from D. For node pairs (r_2, r_3), (r_2, r_5) and (r_3, r_4), whose edges are in red in (a), all their partial correlations need to be calculated as S_β in (b) does not contain the edges between them. For the node pair (r_1, r_2), partial correlation on set $\{r_5\}$ needs to be calculated. This is because in S_β, $\{r_5\}$ is not a subset of neighbours for (r_1, r_2).

1) Full correlation (Full): The full correlation of two ROIs r_i and r_j can be calculated using (1). 2) Fully partial correlation (FP): The fully partial correlation of two ROIs r_i and r_j can be calculated using (2) by setting $Z = V \backslash \{r_i, r_j\}$, where V is the set of all ROIs. 3) ICOV: Let $\mathbf{\Sigma}_{\text{sample}}$ be the sample covariance matrix calculated according to (1), the ICOV method finds an optimal matrix Λ according to: $\Lambda = \text{argmin}_{\Lambda \succ 0} \text{tr}\left(\Lambda \Sigma_{\text{sample}}\right) - \text{logdet}\Lambda + \lambda \|\Lambda\|_{l_1}$, where $\|\cdot\|_{l_1}$ is the element-wise l_1 norm of elements in the matrix Λ. ICOV is a parameter dependent method and the value of the regularization parameter λ greatly influences the results. For the simulation dataset, the regularization parameter λ was set to 5 and 100, which were suggested by the designers of this dataset[6]. For the real dataset, a cross-validation method was used to tune the regularization parameter λ as is described in [17], although we do not

think this cross-validation method is appropriate. 4) ND: Let Σ_{sample} be the sample covariance matrix calculated according to (1). In ND, Σ_{sample} is modelled as the sum of direct dependence Λ and indirect dependence due to indirect paths. Λ is derived from: $\Lambda=\frac{\Sigma_{\text{sample}}}{(I+\Sigma_{\text{sample}})}$, where I is a identity (ID) matrix. 5) GS: In the GS method, the relationship between Λ and Σ_{sample} is approximately modelled as $\Lambda=\frac{\Sigma_{\text{sample}}-I+\varphi((\Sigma_{\text{sample}}-I)\Sigma_{\text{sample}})}{\Sigma_{\text{sample}}}$, where $\varphi(\cdot)$ sets the off-diagonal terms of matrix to zero.

3.1 Simulation evaluation

The NetSim dataset[6] designed by Smith et al. was used to evaluate the accuracy of our method. Because this paper focuses on skeletons, we did not evaluate the direction inference. The BOLD signals were generated using DCM with nonlinear balloon model for the "vascular dynamics"[7]. The dynamics of the signals in the neural network was defined as follows: $\dot{s}=\sigma\Phi s+\Psi u$, where s is the neural time series, $\dot{s}$ is its rate of change, u is the external inputs, Ψ is the weights controlling how the external inputs feed into the network, Φ defines the network connections between nodes with the diagonal elements set to -1, σ controls both the within-node temporal inertia and the neural lag between nodes. The neural signals were then fed into the nonlinear balloon model for vascular dynamics to generate the BOLD signals. The values of balloon model parameters were almost identical to the prior means used in [7].

All network topologies were constructed based on a same building block: a 5-node ring. The ring was a variation of a chain, the head of which is pointed to its tail. The networks with more than 5 nodes were built by linking several 5-node rings together. 28 networks were designed for different purposes of evaluation. Please refer to [6] for the details of these networks. For each network, 50 simulated subjects were generated by slightly changing the values of parameters.

Following [6], we used c-sensitivity of skeleton inference to measure the accuracy. Given a directed graph G, its skeleton is an undirected graph by stripping away all arrowheads from its edges and removing redundant undirected edges. Suppose that each element of the symmetric matrix $\hat{A}$ represents the estimated "confidence" about whether there is a connection between the corresponding nodes. All elements of matrix $\hat{A}$ are assumed to be nonnegative. The distribution of "true positive" (TP) connections includes all the elements in the matrix $\hat{A}$, the corresponding connections of which are found in the skeleton of graph G. Similarly, the distribution of "false positive" (FP) connections includes all the elements in the matrix $\hat{A}$, the corresponding connections of which do not exist in the skeleton of graph G. The c-sensitivity is defined as the fraction of true positives that have higher value in the matrix $\hat{A}$ than the 95th percentile of the false positive distribution. The c-sensitivity shows the ability of the matrix $\hat{A}$ for differentiating true connections and spurious connections. If there is no overlap between the distributions of TP and FP, the c-sensitivity is 100%. It is worth noting that diagonal elements are ignored for calculating c-sensitivity. Fig. 5 shows an example of calculating c-sensitivity.

We investigated the performance of different methods by estimating the resulting c-sensitivity values. The results are summarized in Table 1, where each value is the mean value of 50 simulations. ICOV-5 and ICOV-100 represent the ICOV method with the regularization parameter being set to 5 and 100, respectively. These settings can give good results of ICOV according to [6].

The figures highlighted by yellow in Table 1 are the highest c-sensitivity scores in their corresponding simulations. Generally speaking, the elastic PC-algorithm (EPC) performed best. The elastic PC-algorithm gained highest c-sensitivity in 24 out of 28 simulations. Full correlation (Full) and ICOV-100 only worked best for 1 simulation, while fully partial correlation (FP) achieved the maximal c-sensitivity values for 2 simulations. ICOV-5 and network deconvolution (ND) gained the maximal values for 3 simulations, while global silencing (GS) worked best for 4 simulations. For the results obtained from elastic PC-algorithm, 19 simulations had c-sensitivity values larger than 85%. The results from all methods for the 11th simulation were less than 20%. This is because the synthetic signals for this simulation were generated in a way that ROIs signals greatly influence all others. Some values of c-sensitivities in Table 1 are slightly different from the previous results in [6]. These differences must be caused by minor variations of

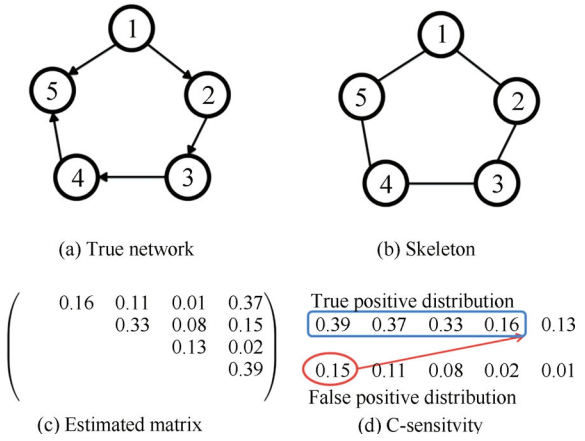


Fig. 5 An example of calculating c-sensitivity. (a) The true network that generates the observations. (b) The skeleton of the true network, which is constructed by stripping away all arrowheads from links of the true network and removing redundant undirected links. (c) The estimated matrix, each element of which represents the "confidence" about whether there is a link between the corresponding nodes. (d) True positive distribution and false positive distribution that are calculated using both the estimated matrix and the skeleton. For example, the element 0.16 in Row 1 and Column 2 belongs to the true positive distribution because Node 1 and Node 2 are linked together in the skeleton. The element 0.11 in Row 1 and Column 3 belongs to the false positive distribution since there is no connection between Node 1 and Node 3 in the skeleton. The 95th percentile of the false distribution is 0.15 that is marked using a red circle. Four elements in the true distribution are larger than 0.15, one element is smaller than 0.15. Thus, the c-sensitivity is 0.8.

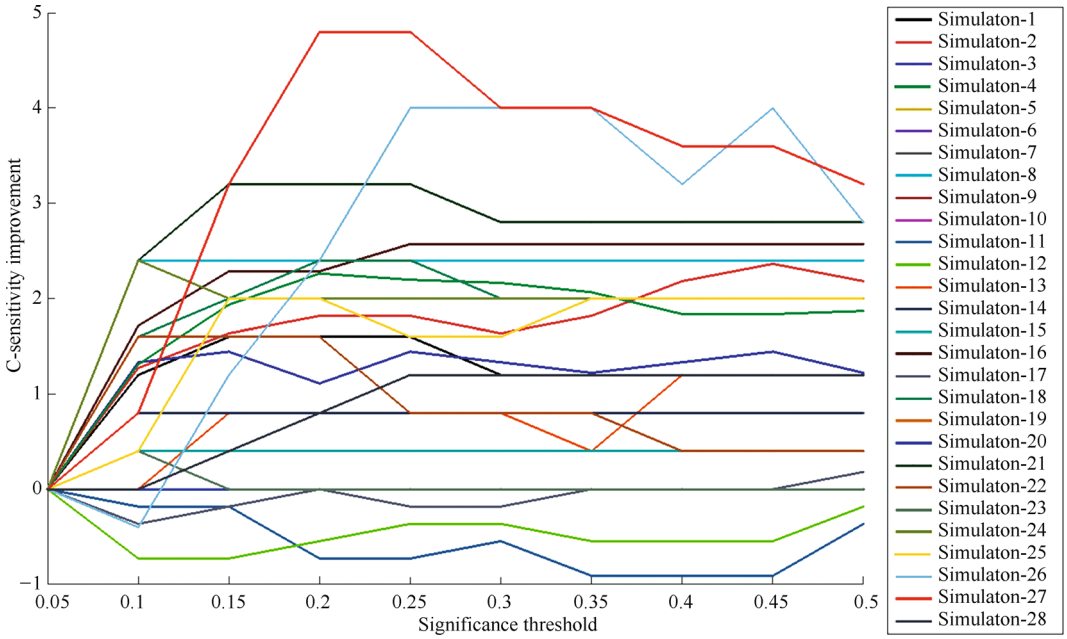


Fig. 6 Relationship between c-sensitivity improvements and significance thresholds in the elastic PC-algorithm

Table 1 Summary of c-sensitivity results over simulations and methods

ID	Full	FP	ICOV-5	ICOV-100	ND	GS	EPC
1	84.00	92.40	93.60	89.60	92.80	92.00	95.60
2	80.00	86.73	88.36	88.18	91.45	88.18	93.82
3	80.67	82.56	84.78	88.00	88.00	71.22	89.44
4	91.51	77.02	82.95	91.48	91.34	45.97	90.16
5	97.20	99.60	100.00	99.60	100.00	100.00	100.00
6	94.00	99.64	99.82	99.27	99.82	99.64	100.00
7	98.80	100.00	100.00	100.00	100.00	100.00	100.00
8	47.20	65.20	65.20	58.40	62.40	60.40	67.20
9	57.60	81.20	81.20	69.20	74.80	76.00	80.00
10	80.00	96.80	96.80	95.20	95.20	90.80	97.20
11	14.91	11.45	12.91	13.64	13.82	15.27	12.18
12	76.00	83.64	84.55	86.00	87.64	84.73	88.55
13	61.20	61.20	62.00	61.20	64.00	63.20	65.20
14	81.20	94.00	94.40	90.00	93.20	94.80	94.40
15	59.20	89.20	90.00	84.80	80.00	78.00	95.20
16	69.14	85.71	86.57	80.57	86.86	83.14	86.86
17	87.64	92.18	93.27	96.73	96.73	80.18	97.45
18	81.60	91.60	91.60	89.20	92.40	91.60	94.40
19	86.00	94.00	94.40	94.80	95.20	94.80	96.40
20	87.20	95.60	97.20	95.20	98.00	96.00	98.80
21	82.40	89.60	90.40	87.60	90.80	89.20	92.80
22	61.20	74.00	73.60	68.40	72.80	66.80	76.80
23	46.40	73.20	78.00	63.20	70.40	69.20	80.40
24	32.00	41.20	41.60	24.80	38.40	42.40	45.60
25	65.60	68.00	67.20	69.60	72.00	71.60	73.60
26	51.20	53.60	54.00	57.60	57.60	52.80	60.00
27	65.20	68.00	69.20	71.20	74.00	61.60	75.60
28	74.40	83.20	84.00	84.00	84.40	83.20	87.60

c-sensitivity in our experiment.

We further investigated the relationship between c-sensitivity and significance threshold for the elastic PC-algorithm. Fig. 6 shows the c-sensitivity improvements with increasing significance thresholds. Negative values in Fig. 6 mean the c-sensitivity decreased with higher significance thresholds. Theoretically, these cases should not happen, but fMRI data do not strictly follow faithfulness and Gaussian assumptions. Although the c-sensitivity improvements in Fig. 6 had some local fluctuations, the overall increasing trend can still be clearly observed. An obvious increase for a large fraction of networks occurred at the first two steps, which are from 0.05 to 0.10 and from 0.10 to 0.15.

The elastic PC-algorithm can save computational effort by avoiding repeated calculations. Some partial correlations at a higher significance threshold can be directly obtained from the results at the previous significance threshold. To assess this, we investigated computational effort saved at each significance threshold by representing the saved computational effort as the proportion of times that the last condition in Line 16 of Algorithm 1 was not satisfied.

Fig. 7 shows the relationship between saved computational effort and significance thresholds for different simulations. It can be seen that the maximum proportion approached to 100%. For some simulations, the proportion values became 100% at some significance thresholds. It means that the elastic PC-algorithm returned stable results when significance threshold was above a certain value. Hence all results at higher threshold were directly obtained from previous results. The computational effort was 100% saved in this case.

The proportion of saved efforts varied from 23% to 100% with the mean equal to 84.3%. When the significance threshold became higher than 0.15, 27 out of 28 networks had the proportion values larger than 50%. The largest

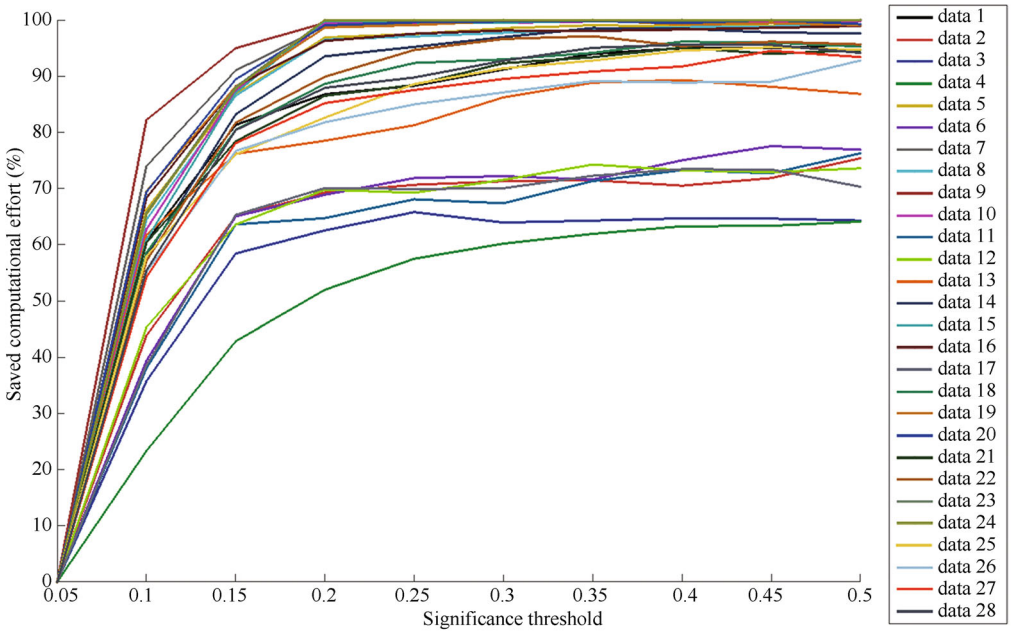


Fig. 7 Relationship between saved computational effort and significance thresholds in the elastic PC-algorithm

increase of proportion value for most simulations occurred when significance threshold increased from 0.05 to 0.15. As has been pointed out, there was a significant increase of c-sensitivity in Fig. 6 occurring at this place. These observations demonstrate that the elastic PC-algorithm can save computational efforts while maintaining accuracy.

3.2 Empirical illustration

A real resting-state fMRI dataset from the human connectome project[24] was used to illustrate the elastic PC-algorithm and compare it with other methods as we listed above. In contrast to the simulation dataset, there was no "ground truth" in the real dataset. Thus accuracy measures of the results, such as c-sensitivity, therefore cannot be calculated. Here we only presented and discussed the properties of the results.

The real dataset included resting-state fMRI data of 10 unrelated subjects. All were healthy (6 women, 4 men). There were two subjects in the 22–25 age range, three subjects in the 26–30 age range, and five subjects in the 31–35 age range. For each subject, the functional data were acquired in four approximately-15-minute runs, which were carried out in two separate imaging sessions. Within each session, the oblique axial acquisition of one run was phase encoded in a right-to-left direction, it was alternated to phase encoding in a left-to-right direction for the other run. All functional images were acquired using a gradient-echo echo-planar imaging (EPI) sequence with the following parameters: repetition time (TR) = 720 ms, echo time (TE) = 33.1 ms, flip angle = 52°, FOV = 208×180 mm, slice thickness = 2.0 mm, 72 slices, frames per run = 1 200, run duration = 14:33. The data was processed using FMRIB software library (FSL)[39], FreeSurfer[40] and Connectome Workbench image analysis suites[41]. The data was de-noised using an independent component analysis based method[42, 43]. The details of the pipelines are presented in [44].

The automated anatomical labeling (AAL) atlas[45] was used to parcel a whole brain into 116 ROIs, which are 90 regions in cerebrum and 26 regions in cerebellum. The time series of a ROI were defined to be the average time series of all voxels in this ROI. Although this paper only aims at estimating the connections between nodes, we have also faced the problem of parcelling a whole brain into nodes in this study. This is because that there is yet no widely-accepted parcellation for functional connectivity inference[46, 47]. Although it has been recommended to use data-driven approaches, such as high-dimensional independent component analysis (ICA), rather than gross structural atlas based parcellation that may not finely correspond to real functional boundaries in the data[48], we chose the AAL atlas (which shows greater stability than other atlases tested[49]), because of its simplicity and interpretability.

For the elastic PC-algorithm, the initial significance threshold was 0.05 and the step size was 0.05. Ten steps were executed for single subject analysis. For multiple subject analysis, to save computational time only three steps were executed. Thus the significance threshold ranged from 0.05 to 0.15. For ICOV, the regularisation parameter was set to 6.5, which was obtained from a cross-validation method as is described in [17], although the validity of this method is still unclear.

Single subject analysis:

We first arbitrarily selected a subject to visually compare the elastic PC-algorithm with other methods. The results of each method formed a matrix, representing the functional connectivity skeleton. All matrices were normalized in the same way that all elements range from 0 to

1. An element with higher score in the matrix indicated a higher confidence on the existence of a connection. Fig. 8 gave us a straightforward overview of different results, the 10 images in the first two columns show results of elastic PC-algorithm over different significance thresholds, 5 images in the last column display results of 5 other methods. Each figure presents a matrix, the dimension of which is 116×116. The colour indicates the values of elements of these matrices, which are from 0 to 1. A figure with more white area means that there are more zero elements in the corresponding matrices.

More elements in the results of the elastic PC-algorithm approached zero with an increasing significance threshold. The resulting matrix with the significance threshold of 0.50 was sparsest. There were many elements consistently large in all results of the elastic PC-algorithm, especially the elements adjacent to the diagonal, suggesting that neighbouring labelled ROIs in AAL atlas are more likely to have connections. Because two homotypic regions in the left and right hemispheres were assigned adjacent labels in the AAL atlas (e.g., lingual regions 47 and 48 in the left and right hemispheres), this observation is consistent with the observation that two homotypic regions are more likely to highly correlate with each other.

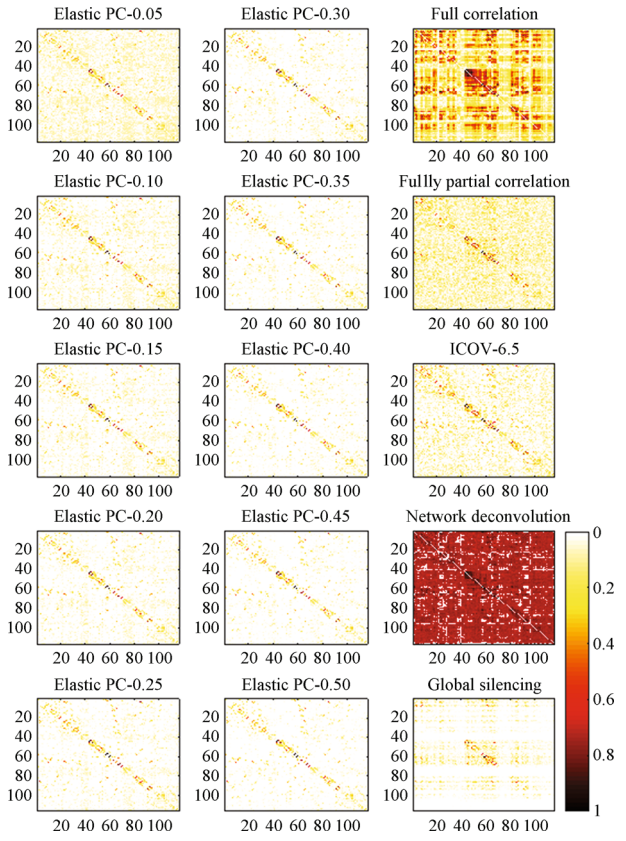


Fig. 8 Single subject results of different methods. The x-y axes indicate the labels of the 116 ROIs

Compared with other methods including full correlation, fully partial correlation, ICOV with regularisation parameter of 6.5, ND and GS, results from elastic PC-algorithm with significance thresholds larger than 0.10 were sparser than results from all other methods. The ND method performed poorly in regards of sparsity, while GS method generates results with patterns different from other methods that appears unreliable. Fully partial correlation and ICOV methods returned results with similar patterns to the elastic PC-algorithm but less sparse. All these observations suggest that results from the elastic PC-algorithm may be more plausible and reliable than results from other methods.

The relationship between the saved computational efforts and significance thresholds was studied. The computational effort saved at each significance threshold was represented as the proportion of times that the last condition in Line 16 of Algorithm 1 was not satisfied. The proportion of saved computational effort varies from 9.2% to 72%. When the significance threshold increased from 0.05 to 0.1, only 9.2% computational effort was saved. It means that the reference skeleton of threshold 0.1 changed greatly from the reference skeleton of threshold 0.05. When the significance threshold further increased above 0.10, the proportion of saved effort was always maintained at a level larger than 50%. Thus, the results of the elastic PC-algorithm reached a relatively stable state, where the number of re-calculations is limited.

The patterns of the results with the elastic PC-algorithm with different significance thresholds were similar, but became sparser with higher significance threshold (Fig. 8). Moreover, the changes between results with two adjacent significance thresholds were calculated. The amounts of changes between two neighbouring significance thresholds were measured using the root mean square difference of elements. The changes at threshold 0.10 was as high as 0.17. When the significance threshold increased above 0.20, the changes remained at a constant level lower than 0.02. It means that the results became relatively stable with significance threshold larger than 0.20. This is consistent with the observation that the proportion of saved computational effort maintained high when the significance threshold was larger than 0.2.

Multi-subject analysis:

After the single subject analysis, we applied different methods to infer functional connectivity for 10 subjects. Similar to single subject analysis, functional connectivity was fractionally represented in matrices, whose elements were normalized between 0 and 1. The matrix for each subject was inferred using different methods, then the average matrix was obtained from each method across 10 subjects. This generated 6 matrices, one for each of the 6 different methods including full correlation (Full), fully partial correlation (FP), ICOV, network deconvolution (ND), global silencing (GS) and the elastic PC-algorithm (EPC). To compare these 6 different results, we selected the top 20 largest elements of each matrix and use their corresponding connections to form a functional network for each method, which was used to estimate the 20 most confident connections. There were 40 connections selected by at least one method. Many connections were detected by at least 2 different methods.

Table 2 shows the details of these connections. There

were three categories, classified according to their patterns of connections between pairs of: 1) homotypic regions in the left and right hemispheres (Type I connection), 2) the heterotypic regions in the same hemisphere (Type II connection), 3) heterotypic regions in different hemispheres (Type III connection). If a connection was found by a certain method, then the corresponding value in the table was set to 1, otherwise it was set to 0. There were 16 Type I connections, 22 Type II connections and 2 Type III connections. Connections discovered by the elastic PC-algorithm were similar with those detected by fully partial correlation and ICOV. However, the results of full correlation were very different from those of other methods, consistent with the expectation that full correlation method detects indirect, as well as direct, effects.

Table 2 Top 20 connections detected by different methods

Type	ROI label	Hemisphere	ROI label	Hemi	Full	FP	ICOV	ND	GS	EPC
I	Frontal_Mid	Left	Frontal_Mid	Right	0	0	0	0	1	0
I	Frontal_Sup_Medial	Left	Frontal_Sup_Medial	Right	0	1	1	0	0	1
I	Frontal_Med_Orb	Left	Frontal_Med_Orb	Right	0	0	0	0	0	1
I	Calcarine	Left	Calcarine	Right	1	1	1	1	1	1
I	Cuneus	Left	Cuneus	Right	1	1	1	1	1	1
I	Lingual	Left	Lingual	Right	1	0	0	0	0	0
I	Occipital_Sup	Left	Occipital_Sup	Right	1	0	0	0	0	0
I	Occipital_Mid	Left	Occipital_Mid	Right	1	1	1	1	1	1
I	Occipital_Inf	Left	Occipital_Inf	Right	0	1	1	1	0	1
I	Postcentral	Left	Postcentral	Right	1	1	1	1	1	1
I	Parietal_Sup	Left	Parietal_Sup	Right	0	1	1	1	1	1
I	SupraMarginal	Left	SupraMarginal	Right	1	1	1	1	1	1
I	Angular	Left	Angular	Right	0	1	1	0	1	0
I	Precuneus	Left	Precuneus	Right	1	1	1	1	1	1
I	Paracentral_Lobule	Left	Paracentral_Lobule	Right	0	1	1	1	0	1
I	Temporal_Sup	Left	Temporal_Sup	Right	0	1	1	1	0	1
II	Frontal_Sup	Left	Frontal_Mid	Left	0	1	1	1	1	1
II	Frontal_Inf_Tri	Left	Frontal_Inf_Orb	Left	0	0	1	0	0	1
II	Calcarine	Left	Cuneus	Left	1	0	0	0	0	0
II	Calcarine	Left	Lingual	Left	1	0	0	0	0	0
II	Cuneus	Left	Lingual	Left	1	0	0	0	0	0
II	Cuneus	Left	Occipital_Sup	Left	1	1	0	1	1	0
II	Occipital_Sup	Left	Occipital_Mid	Left	0	0	0	0	1	0
II	Occipital_Mid	Left	Occipital_Inf	Left	0	0	0	1	0	0
II	Parietal_Sup	Left	Parietal_Inf	Left	0	1	1	1	1	1
II	Angular	Left	Temporal_Mid	Left	0	0	0	0	1	0
II	Cerebelum_Crus1	Left	Cerebelum_Crus2	Left	0	1	1	1	0	1
II	Precentral	Right	Postcentral	Right	1	1	1	1	1	1
II	Frontal_Sup	Right	Frontal_Mid	Right	0	1	1	1	1	1
II	Frontal_Mid	Right	Parietal_Inf	Right	1	0	0	1	0	1
II	Calcarine	Right	Lingual	Right	1	0	0	0	0	0
II	Cuneus	Right	Lingual	Right	1	0	0	0	0	0
II	Cuneus	Right	Occipital_Sup	Right	1	0	0	0	1	0
II	Occipital_Sup	Right	Occipital_Mid	Right	1	0	0	0	1	0
II	Occipital_Mid	Right	Parietal_Sup	Right	0	0	0	0	1	0
II	Parietal_Inf	Right	Angular	Right	0	1	1	1	0	0
II	Angular	Right	Temporal_Mid	Right	0	0	0	0	1	0
II	Cerebelum_Crus1	Right	Cerebelum_Crus2	Right	0	1	1	1	0	1
III	Calcarine	Left	Lingual	Right	1	0	0	0	0	0
III	Cuneus	Right	Occipital_Sup	Left	1	0	0	0	0	0

There were 6 Type I connections that were coherently discovered by all 6 different methods. These were found in the following hemispherically homotypic ROIs: Calcarine, Cuneus, Occipital_Mid, Postcentral, Supramarginal and Precuneus. Compared to the other types, the fraction of Type I connections that were consistently detected by different methods was relatively high. Only 1 out of 22 Type II connections was discovered by all 6 methods (the connection between precentral and postcentral gyri in the right hemisphere). Ten Type II connections were found by one method.

There were only two Type III connections (between left calcarine and right lingual regions, and right cuneus and left superior occipital regions), and these were identified only by the full correlation method. As is shown in Fig. 9, it was quite likely that these Type III connections were enhanced by indirect effects.

We also investigated the distribution of top 1% connections that were found by the elastic PC-algorithm. Totally, 67 connections were selected with the corresponding elements in the matrix ranging from 0.20 to 0.75. These connections were plotted using BrainNet Viewer[50] (Fig. 10). There were 26 Type I connections, 40 Type II connections and only 1 Type III connection. It is interesting to observe that we only obtained a quite limited number of connections between two heterotypic regions in different hemispheres. It suggests that in resting states, the functional connectivity might be composed primarily of Type I and Type II connections.

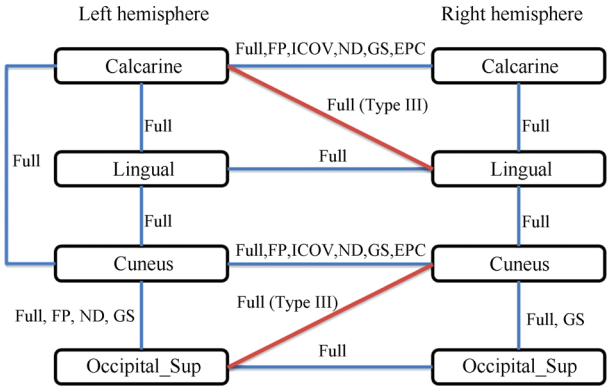


Fig. 9 Connections detected by various methods between the following ROIs: calcarine, lingual, cuneus and superior occipital regions

4 Discussion

We proposed use of minimum partial correlation to infer the skeleton of functional connectivity in fMRI data. We designed an algorithm, called elastic PC-algorithm, to efficiently approximate the minimum partial correlation within a computational time budget. A Matlab implementation of the proposed algorithm is publicly available at https://github.com/LNie/ElasticPC. The simulation study demonstrated the capability of the proposed method to improve the accuracy of skeleton inference. The empirical study showed that direct functional effects between two heterotypic regions in different hemispheres might be rare during resting state.

There are limitations to our approach. One drawback of the proposed method is that minimum partial correlation is always zero if the number of points in fMRI time series is less than or equal to the number of ROIs. This situation is not common in areal level studies, because the number of time points is normally more than 600 and the number of ROI is usually equal to or less than 300. However, minimum partial correlation cannot be directly applied in voxel level studies. This is because the number of voxels is much larger than the number of time points. ICOV[16, 17] and recently proposed L_0-norm based methods[51] can avoid this drawback via proper regularization. Future work is needed to explore the possibility of combining minimum partial correlation and L_0-norm or L_1-norm based approaches.

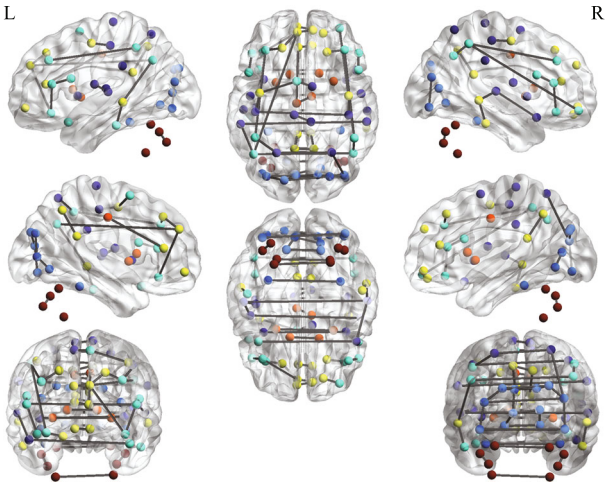


Fig. 10 Top 1% connections selected by the elastic PC-algorithm. The nodes outside the cerebrum are in cerebellum. The figure is generated by BrainNet Viewer[50] to show all sides of the brain, including the lateral and medial sides of each hemisphere, the dorsal and ventral sides, and the anterior and posterior sides of the brain.

Another drawback of our study is that the theoretical analysis of minimum partial correlation is based on the faithfulness, Gaussian and DAG assumptions. As these are approximations for fMRI data, potential confounds that may arise as a consequence need to be considered in the application of the proposed method. Although, the performance of the proposed method was demonstrated on a simulation dataset generated without these assumptions, the simulation dataset was artificially designed and the dimensions of the network structures were highly simplified. Future work is needed to investigate the theoretical bounds on the performances under more realistic assumptions.

The methods for network modelling from fMRI data are based on various assumptions, thus the model complexity of these methods are different. A simpler model, e.g., full correlation, is less neurologically meaningful but it can stably and efficiently infer a network with more nodes, in contrast,

a more complex model, e.g., DCM, is more neurologically meaningful but it may be more sensitive to its assumptions and can only analyse smaller networks[48]. In terms of model complexity, the proposed minimum partial correlation is simpler than DCM and more complex than full correlation, partial correlation and ICOV. The minimum partial correlation is a statistical model rather than a generation model and with few neurophysiological assumptions. Therefore, it may be better to be used to discover network structures as an alternative to full correlation, partial correlation and etc.

A recent study shows that the functional connectivity can be used as "fingerprinting" to identify individuals with a high accuracy[52]. Further work is needed to investigate the performance of minimum partial correlation to preserve subject-specific patterns via identification experiments. It will be of interest to study the heritability of subject-specific patterns on functional connectivity. Future work also may use minimum partial correlation to explore the relationship between brain functions and behavior, genetics or disease related phenotypes[3]. This could be important particularly in using minimum partial correlation as biomarkers[53] to better classify psychiatric diseases.

Acknowledgement

Data were provided by the human connectome project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil, 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research, and by the McDonnell Center for Systems Neuroscience at Washington University. Paul M. Matthews gratefully acknowledges support from the Imperial College NIHR Biomedical Research Centre and personal support from the Edmond Safra Foundation and Lily Safra.

Open Access

This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

[1] K. J. Friston. Functional and effective connectivity: A review. *Brain Connectivity*, vol. 1, no. 1, pp. 13–36, 2011.

[2] R. L. Buckner, F. M. Krienen, B. T. T. Yeo. Opportunities and limitations of intrinsic functional connectivity MRI. *Nature Neuroscience*, vol. 16, no. 7, pp. 832–837, 2013.

[3] R. C. Craddock, S. Jbabdi, C. G. Yan, J. T. Vogelstein, F. X. Castellanos, A. D. Martino, C. Kelly, K. Heberlein, S. Colcombe, M. P. Milham. Imaging human connectomes at the macroscale. *Nature Methods*, vol. 10, no. 6, pp. 524–539, 2013.

[4] D. J. Hawellek, J. F. Hipp, C. M. Lewis, M. Corbetta, A. K. Engel. Increased functional connectivity indicates the severity of cognitive impairment in multiple sclerosis. *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 47, pp. 19066–19071, 2011.

[5] W. R. Shirer, S. Ryali, E. Rykhlevskaia, V. Menon, M. D. Greicius. Decoding subject-driven cognitive states with whole-brain connectivity patterns. *Cerebral Cortex*, vol. 22, no. 1, pp. 158–165, 2012.

[6] S. M. Smith, K. L. Miller, G. Salimi-Khorshidi, M. Webster, C. F. Beckmann, T. E. Nichols, J. D. Ramsey, M. W. Woolrich. Network modelling methods for fMRI. *NeuroImage*, vol. 54, no. 2, pp. 875–891, 2011.

[7] K. J. Friston, L. Harrison, W. Penny. Dynamic causal modelling. *NeuroImage*, vol. 19, no. 4, pp. 1273–1302, 2003.

[8] A. M. Hermundstad, D. S. Bassett, K. S. Brown, E. M. Aminoff, D. Clewett, S. Freeman, A. Frithsen, A. Johnson, C. M. Tipper, M. B. Miller, S. T. Grafton, J. M. Carlson. Structural foundations of resting-state and task-based functional connectivity in the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 15, pp. 6169–6174, 2013.

[9] N. B. Turk-Browne. Functional interactions as big data in the human brain. *Science*, vol. 342, no. 6158, pp. 580–584, 2013.

[10] G. Marrelec, A. Krainik, H. Duffau, M. Pélégrini-Issac, S. Lehéricy, J. Doyon, H. Benali. Partial correlation for functional brain interactivity investigation in functional MRI. *NeuroImage*, vol. 32, no. 1, pp. 228–237, 2006.

[11] H. Lee, D. S. Lee, H. Kang, B. N. Kim, M. K. Chung. Sparse brain network recovery under compressed sensing. *IEEE Transactions on Medical Imaging*, vol. 30, no. 5, pp. 1154–1165, 2011.

[12] S. Ryali, T. W. Chen, K. Supekar, V. Menon. Estimation of functional connectivity in fMRI data using stability selection-based sparse partial correlation with elastic net penalty. *NeuroImage*, vol. 59, no. 4, pp. 3852–3861, 2012.

[13] S. M. Smith, D. Vidaurre, C. F. Beckmann, M. F. Glasser, M. Jenkinson, K. L. Miller, T. E. Nichols, E. C. Robinson, G. Salimi-Khorshidi, M. W. Woolrich, D. M. Barch, K. Uğurbil, D. C. Van Essen. Functional connectomics from resting-state fMRI. *Trends in Cognitive Sciences*, vol. 17, no. 12, pp. 666–682, 2013.

[14] S. M. Smith, T. E. Nichols, D. Vidaurre, A. M. Winkler, T. E. J. Behrens, M. F. Glasser, K. Ugurbil, D. M. Barch, D. C. Van Essen, K. L. Miller. A positive-negative mode of population covariation links brain connectivity, demographics and behavior. *Nature Neuroscience*, vol. 18, no. 11, pp. 1565–1567, 2015.

[15] J. Berkson. Limitations of the application of fourfold table analysis to hospital data. *Biometrics Bulletin*, vol. 2, no. 3, pp. 47–53, 1946.

[16] J. Friedman, T. Hastie, R. Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, vol. 9, no. 3, pp. 432–441, 2008.

[17] G. Varoquaux, A. Gramfort, J. B. Poline, B. Thirion. Brain covariance selection: Better individual functional connectivity models using population prior. In *Proceedings of Neural Information Processing Systems*, NIPS, Vancouver, Canada, pp. 2334–2342, 2010.

[18] M. Hinne, L. Ambrogioni, R. J. Janssen, T. Heskes, M. A. J. van Gerven. Structurally-informed Bayesian functional connectivity analysis. *NeuroImage*, vol. 86, pp. 294–305, 2014.

[19] K. P. Murphy. *Machine Learning: A Probabilistic Perspective*, Cambridge, USA: MIT Press, 2012.

[20] S. Huang, J. Li, L. Sun, J. P. Ye, A. Fleisher, T. Wu, K. W. Chen, E. Reiman. Learning brain connectivity of Alzheimers disease by sparse inverse covariance estimation. *NeuroImage*, vol. 50, no. 3, pp. 935–949, 2010.

[21] M. G. G′Sell, J. Taylor, R. *Tibshirani. Adaptive testing for the graphical lasso*, [Online], Available: https://arxiv.org/abs/1307.4765, 2013.

[22] R. Lockhart, J. Taylor, R. J. Tibshirani, R. Tibshirani. A significance test for the lasso. *The Annals of Statistics*, vol. 42, no. 2, pp. 413–468, 2014.

[23] P. Spirtes, C. Glymour, R. Scheines. *Causation, Prediction, and Search*, 2nd ed., Cambridge, USA: MIT Press, 2000.

[24] S. M. Smith, C. F. Beckmann, J. Andersson, E. J. Auerbach, J. Bijsterbosch, G. Douaud, E. Duff, D. A. Feinberg, L. Griffanti, M. P. Harms, M. Kelly, T. Laumann, K. L. Miller, S. Moeller, S. Petersen, J. Power, G. Salimi-Khorshidi, A. Z. Snyder, A. T. Vu, M. W. Woolrich, J. Q. Xu, E. Yacoub, K. Uğurbil, D. C. Van Essen, M. F. Glasser. Resting-state fMRI in the Human Connectome Project. *NeuroImage*, vol. 80, pp. 144–168, 2013.

[25] L. Nie, X. Yang, P. M. Matthews, Z. X. Xu, Y. K. Guo. Minimum partial correlation: An accurate and parameter-free measure of functional connectivity in fMRI. In *Proceedings of International Conference on Brain Informatics and Health*, Springer, Cham, Switzerland, pp. 125–134, 2015.

[26] J. Pearl. *Causality: Models, Reasoning and Inference*, Cambridge, UK: Cambridge University Press, 2000.

[27] J. A. Mumford, J. D. Ramsey. Bayesian networks for fMRI: A primer. *NeuroImage*, vol. 86, pp. 573–582, 2014.

[28] C. Bielza, P. Larrañaga. Bayesian networks in neuroscience: A survey. *Frontiers in Computational Neuroscience*, vol. 8, Article number 131, 2014.

[29] S. L. Lauritzen. *Graphical Models*, Oxford, UK: Oxford University Press, 1996.

[30] R. A. Fisher. The distribution of the partial correlation coefficient. *Metron*, vol. 3, pp. 329–332, 1924.

[31] J. Cheng, R. Greinera, J. Kelly, D. Bell, W. R. Lius. Learning Bayesian networks from data: An information-theory based approach. *Artificial Intelligence*, vol. 137, no. 1–2, pp. 43–90, 2002.

[32] I. Tsamardinos, L. E. Brown, C. F. Aliferis. The max-min hill-climbing Bayesian network structure learning algorithm. *Machine Learning*, vol. 65, no. 1, pp. 31–78, 2006.

[33] Z. X. Wang, L. W. Chan. Learning Bayesian networks from Markov random fields: An efficient algorithm for linear models. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 6, no. 3, Article number 10, 2012.

[34] S. P. Iyer, I. Shafran, D. Grayson, K. Gates, J. T. Nigg, D. A. Fair. Inferring functional connectivity in MRI using Bayesian network structure learning with a modified PC-algorithm. *NeuroImage*, vol. 75, no. 4, pp. 165–175, 2013.

[35] R. Han, L. Nie, M. M. Ghanem, Y. K. Guo. Elastic algorithms for guaranteeing quality monotonicity in big data mining. In *Proceedings of IEEE International Conference on Big Data*, IEEE, Silicon Valley, USA, pp. 45–50, 2013.

[36] D. Colombo, M. H. Maathuis. Order-independent constraint-based causal structure learning. *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 3741–3782, 2014.

[37] S. Feizi, D. Marbach, M. Médard, M. Kellis. Network deconvolution as a general method to distinguish direct dependencies in networks. *Nature Biotechnology*, vol. 31, no. 8, pp. 726–733, 2013.

[38] B. Barzel, A. L. Barabási. Network link prediction by global silencing of indirect correlations. *Nature Biotechnology*, vol. 31, no. 8, pp. 720–725, 2013.

[39] M. Jenkinson, C. F. Beckmann, T. E. J. Behrens, M. W. Woolrich, S. M. Smith. FSL. *NeuroImage*, vol. 62, no. 2, pp. 782–790, 2012.

[40] A. M. Dale, B. Fischl, M. I. Sereno. Cortical surface-based analysis: I. Segmentation and surface reconstruction. *NeuroImage*, vol. 9, no. 2, pp. 179–194, 1999.

[41] D. S. Marcus, M. P. Harms, A. Z. Snyder, M. Jenkinson, J. A. Wilson, M. F. Glasser, D. M. Barch, K. A. Archie, G. C. Burgess. Human connectome project informatics: Quality control, database services, and data visualization. *NeuroImage*, vol. 80, no. 8, pp. 202–219, 2013.

[42] L. Griffanti, G. Salimi-Khorshidi, C. F. Beckmann, E. J. Auerbach, G. Douaud, C. E. Sexton, E. Zsoldos, K. P. Ebmeier, N. Filippin, C. E. Mackay, S. Moeller, J. Q. Xu, E. Yacoub, G. Baselli, K. Ugurbil, K. L. Miller, S. M. Smith. ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. *NeuroImage*, vol. 95, no. 4, pp. 232–247, 2014.

[43] G. Salimi-Khorshidi, G. Douaud, C. F. Beckmann, M. F. Glasser, L. Griffanti, S. M. Smith. Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers. *NeuroImage*, vol. 90, pp. 449–468, 2014.

[44] M. F. Glasser, S. N. Sotiropoulos, J. A. Wilson, T. S. Coalson, B. Fischl, J. L. Andersson, J. Q. Xu, S. Jbabd. The minimal preprocessing pipelines for the human connectome project. *NeuroImage*, vol. 80, no. 3, pp. 105–124, 2013.

[45] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, M. Joliot. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, vol. 15, no. 1, pp. 273–289, 2002.

[46] M. L. Stanley, M. N. Moussa, B. M. Paolini, R. G. Lyday, J. H. Burdette, P. J. Laurienti. Defining nodes in complex brain networks. *Frontiers in Computational Neuroscience*, vol. 7, Article number 169, 2013.

[47] S. B. Eickhoff, B. Thirion, G. Varoquaux, D. Bzdok. Connectivity-based parcellation: Critique and implications. *Human Brain Mapping*, vol. 36, no. 12, pp. 4771–4792, 2015.

[48] S. M. Smith. The future of fMRI connectivity. *NeuroImage*, vol. 62, no. 2, pp. 1257–1266, 2012.

[49] C. Y. Wee, P. T. Yap, D. Q. Zhang, L. H. Wang, D. G. Shen. Group-constrained sparse fMRI connectivity modeling for mild cognitive impairment identification. *Brain Structure and Function*, vol. 219, no. 2, pp. 641–656, 2014.

[50] M. R. Xia, J. H. Wang, Y. He. BrainNet Viewer: A network visualization tool for human brain connectomics. *PLoS One*, vol. 8, no. 7, Article number e68910, 2013.

[51] Z. N. Fu. A Study of Dynamic Functional Brain Connectivity Using Functional Magnetic Resonance Imaging (fMRI): Method and Applications, Ph. D. dissertation, The University of Hong Kong, China, 2016.

[52] E. S. Finn, X. L. Shen, D. Scheinost, M. D. Rosenberg, J. Huang, M. M. Chun, X. Papademetris, R. T. Constable. Functional connectome fingerprinting: Identifying individuals using patterns of brain connectivity. *Nature Neuroscience*, vol. 18, no. 11, pp. 1664–1671, 2015.

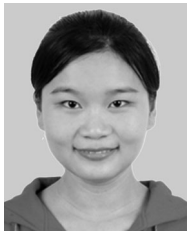
[53] L. Y. Chen, J. Yang, G. G. Xu, Y. Q. Liu, J. T. Li, C. S. Xu. Biomarker identification of rat liver regeneration via adaptive logistic regression. *International Journal of Automation and Computing*, vol. 13, no. 2, pp. 191–198, 2016.

Lei Nie received the B. Sc. degree in information and computing science from Sichuan University, China in 2009, and received the Ph. D. degree in computer science and technology from University of the Chinese Academy of Sciences, China in 2015. He was a visiting Ph. D. student in the Department of Computing, Imperial College London, UK from 2012 to 2014. He is currently a research associate in the Department of Computing, Imperial College London, UK.

His research interests include neuroimaging and machine learning.

E-mail: l.nie13@imperial.ac.uk

ORCID iD: 0000-0003-1115-5027



Xian Yang received the B. Eng. degree in electronic information engineering from Huazhong University of Science and Technology, China in 2008, received the M. Sc. degree in digital communication from University of Bath, UK in 2009, and received the Ph. D. degree in computing from Imperial College London, UK in 2016. She is currently a research associate in the Department of Computing, Imperial College London, UK.

Her research interests include bioinformatics, system biology, neuroimaging, data mining and health informatics.

E-mail: xian.yang08@imperial.ac.uk

Paul M. Matthews received the B. A. degree in chemistry from University of Oxford, UK in 1978, the D. Phil degree in biochemistry from University of Oxford, UK in 1982, and the M. D. degree from Stanford University School of Medicine, USA in 1987. He is the Edmond and Lily Safra Chair and head of the Division of Brain Sciences at Imperial College London, UK. Amongst other external activities, he is chair of the Imaging Enhancement Working Group and a member of the steering group for UK Biobank (https://www.ukbiobank.ac.uk/), which has initiated a programme to image the brain, heart, carotids, bones and body of 100 000 people to understand disease risk in later life. He was the founding director of the Centre for Functional Magnetic Resonance Imaging of the Brain (FMRIB) (http://www.fmrib.ox.ac.uk/) and of the GSK Clinical Imaging Centre at the Hammersmith Hospital, for which he was a lead in spinning out Imanova Ltd., which is run as a public-private partnership between Imperial College, UCL, Kings College and the Medical Research Council (http://www.imanova.co.uk/).

His research interests include innovative translational applications of clinical imaging for the neurosciences. His broad area of research interest has been in molecular and functional neuroimaging and in neurological therapeutics development. A particular focus of his work has involved close collaboration with colleagues in computing and engineering to encourage effective translation of advanced imaging data to information.

E-mail: p.matthews@imperial.ac.uk

Zhi-Wei Xu received the B. Sc. degree from University of Electronic Science and Technology of China, in 1982, received the M. Sc. degree from Purdue University, USA in 1984, and received the Ph. D. degree from the University of Southern California, USA in 1987. He is a professor and CTO of the Institute of Computing Technology (ICT) of the Chinese Academy of Sciences (CAS). His prior industrial experience included chief engineer of Dawning Corporation (now Sugon as listed in Shanghai Stock Exchange), a leading high-performance computer vendor in China. He currently leads "Cloud-Sea Computing Systems", a strategic priority research project of the Chinese Academy of Sciences that aims at developing billion-thread computers with elastic processors.

His research interests include high-performance computer architecture and network computing science.

E-mail: zxu@ict.ac.cn

Yi-Ke Guo received the B. Sc. degree, the M. Sc. degree in computer science and technology from Tsinghua University, China in 1985, and the Ph. D. degree in computational logic from Imperial College London, UK in 1993. He is the founding director of the Data Science Institute, Imperial College London, UK. He is also a professor in the Department of Computing, Imperial College London, UK. During last 15 years, he has been leading a data science group to carry out many research projects, including discovery net on grid based data analysis for scientific discovery, MESSAGE on wireless mobile sensor network for environment monitoring, BAIR on system biology for diabetes study, iHealth on modern informatics infrastructure for healthcare decision making, UBIOPRED on large informatics platform for translational medicine research, digital city exchange on sensor information-based urban dynamics modelling, IC Cloud system for large scale collaborative scientific research. He is now the principal investigator of the eTRIKS project, a 23M Euro project in building a cloud-based translational informatics platform for global medical research.

His research interests include data mining, machine learning and bioinformatics

E-mail: y.guo@imperial.ac.uk (Corresponding author)

ORCID iD: 0000-0002-3075-2161