Research Article

A Computationally Efficient, Exploratory Approach to Brain Connectivity Incorporating False Discovery Rate Control, *A Priori* Knowledge, and Group Inference

Aiping Liu,[1] Junning Li,[2] Z. Jane Wang,[1] and Martin J. McKeown[1, 3]

[1] *Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, Canada V6T 1Z4*
[2] *Laboratory of Neuro Imaging, Department of Neurology, UCLA School of Medicine, Los Angeles, CA 90095, USA*
[3] *Division of Neurology, Department of Medicine and Pacific Parkinson's Research Centre, University of British Columbia, Vancouver, BC, Canada V5Z 1M9*

Correspondence should be addressed to Z. Jane Wang, zjanew@ece.ubc.ca

Graphical models appear well suited for inferring brain connectivity from fMRI data, as they can distinguish between direct and indirect brain connectivity. Nevertheless, biological interpretation requires not only that the multivariate time series are adequately modeled, but also that there is accurate error-control of the inferred edges. The PC_{fdr} algorithm, which was developed by Li and Wang, was to provide a computationally efficient means to control the false discovery rate (FDR) of computed edges asymptotically. The original PC_{fdr} algorithm was unable to accommodate *a priori* information about connectivity and was designed to infer connectivity from a single subject rather than a group of subjects. Here we extend the original PC_{fdr} algorithm and propose a multisubject, error-rate-controlled brain connectivity modeling approach that allows incorporation of prior knowledge of connectivity. In simulations, we show that the two proposed extensions can still control the FDR around or below a specified threshold. When the proposed approach is applied to fMRI data in a Parkinson's disease study, we find robust group evidence of the disease-related changes, the compensatory changes, and the normalizing effect of L-dopa medication. The proposed method provides a robust, accurate, and practical method for the assessment of brain connectivity patterns from functional neuroimaging data.

1. Introduction

The interaction between macroscopic brain regions has been increasingly recognized as being vital for understanding the normal brain function and the pathophysiology of many neuropsychiatric diseases. Brain connectivity patterns derived from neuroimaging methods are therefore of great interest, and several recently published reviews have described different modeling methods for inferring brain connectivity from fMRI data [1, 2]. Specifically, graphical models which represent statistical dependence relationships between time series derived from brain regions, such as structural equation models [3], dynamic causal models [4], and Bayesian networks [5], appear to be well suited for assessing connectivity between brain regions.

Graphical models, when applied to functional neuroimaging data, represent brain regions of interest (ROIs) as nodes and the stochastic interactions between ROIs as edges. However, in most nonbrain imaging graphical model applications, the primary goal is to create a model that fits the overall multivariate data well, does not necessarily accurately reflect the particular connections between nodes. Yet in the applications of graphical models to brain connectivity, the neuroscientific interpretation is largely based on the pattern of connections inferred by the model. This places a premium on accurately determining the "inner workings" of the model such as accounting for the error rate of the edges in the model.

The false discovery rate (FDR) [6, 7], defined as the expected ratio of spurious connections to all learned

connections, has been suggested as a suitable error-rate control criterion when inferring brain connectivity. Compared with traditional type I and type II statistical error rates, the FDR is more informative in bioinformatics and neuroimaging, since it is directly related with the uncertainty of the reported positive results. When selecting candidate genes for genetic research, for example, researchers may want 70% of selected genes to be truly associated with the disease, that is, an FDR of 30%.

Naively controlling traditional type I and type II error rates at specified levels may not necessarily result in reasonable FDR rates, especially in the case of large, sparse networks. For example, consider an undirected network with 40 nodes, with each node interacting, on average, with 3 other nodes; that is, there are 60 edges in the network. An algorithm with the *realized* type I error rate of 5% and the *realized* power of 90% (i.e., the *realized* type II error rate = 10%) will recover a network with $60 \times 90\% = 54$ correct connections and $[40 \times (40 - 1)/2 - 60] \times 5\% = 36$ false connections, which means that $36/(36 + 54) = 40\%$ of the claimed connections actually would not exist in the true network! This example, while relatively trivial, demonstrates that the FDR may not be kept suitably low by simply controlling traditional type I and type II error rates.

Recent work in the machine learning field has started to investigate controlling the FDR in network structures using a generic Bayesian approach and classical FDR assessment [8]. This work was subsequently extended to look specifically at graphical models where the FDR was assessed locally at each node [9].

Li and Wang proposed a network-learning method that allows asymptotically control of the FDR globally. They based their approach on the PC algorithm (named after Peter Spirtes and Clark Glymour), a computationally efficient and asymptotically reliable Bayesian network-learning algorithm. The PC algorithm assesses the (non)existence of an edge in a graph by determining the conditional dependence/independence relationships between nodes [10]. However, different from the original PC algorithm, which controls the type I error rate individually for each edge during conditional independence testing, the Li and Wang algorithm, referred as the PC_{fdr} algorithm, is capable of asymptotically controlling the FDR under prespecified levels [11]. The PC_{fdr} algorithm does this by interpreting the learning of a network as testing the existence of edges, and thus the FDR control of edges becomes a multiple-testing problem, which has a strong theoretical basis and has been extensively studied by statisticians [11].

Beside giving an introduction of these recent advancements, this paper will present two extensions to the original PC_{fdr} algorithm, the combination of which leads to a multi-subject brain connectivity modeling approach incorporating FDR control, *a priori* knowledge and group inference. One extension is an adaptation of *a priori* knowledge, allowing users to specify which edges must appear in the network, which cannot and which are to be learned from data. The resulting algorithm is referred to as PC_{fdr}^{+} algorithm in this paper. Many applications require imposing prior knowledge into network learning. For example, analyzing

causal relationship in time series may forbid backward connections from time $t + 1$ to t, such as that in dynamic Bayesian networks. In some situations, researchers may want to exclude some impossible connections based on anatomical knowledge. Incorporating *a priori* knowledge into PC_{fdr} algorithm allows for more flexibility in using the method and potentially leads to greater sensitivity in accurately discovering the true brain connectivity.

The second extension to PC_{fdr} algorithm is a combination of the PC_{fdr} algorithm and a mixed-effect model to robustly deal with intersubject variability. As neuroimaging research typically involves a group of subjects rather than focusing on an individual subject, group analysis plays an important role in final biological interpretations. However, compared with the extensive group-level methods available for analysis of amplitude changes in blood-oxygen-level-dependent (BOLD) signals (e.g., Worsley et al. [12], Friston et al. [13]), the problem of group-level brain connectivity analysis is less well studied. This is likely due to the fact that it requires not only accommodating the variances and the correlations across subjects, but also accounting for the potentially different structures of subject-specific brain connectivity networks. The proposed group-level exploratory approach for brain connectivity inference combines the PC_{fdr} algorithm (or the extended PC_{fdr}^{+} algorithm if *a priori* knowledge is available) and a mixed-effect model, a widely used method for handling intersubject variability.

Several methods have been proposed to infer group connectivity in neuroimaging. Bayesian model selection [14] handles intersubject variability and error control; however, its current proposed implementation does not scale well, making it more suitable for confirmatory, rather than an exploratory research. Varoquaux et al. [15] propose a data-driven method to estimate large-scale brain connectivity using Gaussian modeling and deals with the variability between subjects by using optimal regularization schemes. Ramsey et al. [16] describe and evaluate a combination of a multisubject search algorithm and the orientation algorithm.

The major distinguishing feature of the proposed approach compared to these aforementioned approaches is that the current data-driven approach aims at controlling the FDR directly at the group-level network. We demonstrate that in simulations that, with a sufficiently large subject size, the proposed group-level algorithm is able to reliably recover network structures and still control the FDR around prespecified levels. When the proposed approach, referred as the gPC_{fdr}^{+} algorithm, is applied to real fMRI data with Parkinson's disease, we demonstrate evidence of direct and indirect (i.e., compensatory) disease-related connectivity changes, as well as evidence that L-dopa provides a "normalizing" effect on connectivity in Parkinson's disease, consistent with its dramatic clinical effect.

2. Materials and Methods

2.1. Preliminaries. Graphical models, such as Bayesian networks, encode conditional independence/dependence relationships among variables graphically with nodes and edges

according to the Markov properties [17]. The concept of conditional (in)dependence is very important for the inference of brain connectivity, as it assists in distinguishing between direct and indirect connectivity. For example, the activities in two brain regions are initially correlated, but become independent after all possible influences from other brain regions are removed, then this is an example of indirect connectivity, as the initial activity was actually induced by common input from another region(s). On the other hand, if the activities of two brain regions are correlated even after all possible influences from other regions are removed, then very likely there is a direct functional connection between them and hence is an example of direct connectivity. Conditional dependence is the real interest in learning brain connectivity because it implies that two brain regions are directly connected.

Since a graphical model is a graphical representation of conditional independence/dependence relationships, the nonadjacency between two nodes is tested by inspecting their conditional independence given all other nodes. As multiple edges are tested simultaneously, FDR-control procedures should be applied to correct the effect of multiple testing.

Given two among N random variables, there are 2^{N-2} possible subsets of the other $N - 2$ variables upon which the two variables could be conditionally independent. To avoid exhaustively testing such an exponential number of conditional independence relationships, the following proposition [10] can be employed [9, 11].

Proposition 1. *Given a multivariate probability distribution whose conditional independence relationships can be perfectly encoded as a Bayesian network according to the Markov property, two nodes a and b are nonadjacent if and only if there is a subset C of nodes either all in the neighbors of a or all in the neighbors of b such that a and b are conditionally independent on given C.*

Based on Proposition 1, nodes a and b can be disconnected once they are found conditionally independent upon a conditional node set C. As the tests of adjacency progress for every node pair, the neighborhood of nodes keeps shrinking, so an exhaustive search of the conditional node set C is avoided. This greatly reduces computation, especially for a sparse network.

2.2. Brain Connectivity Inference Incorporating False Discovery Rate Control and A Priori Knowledge

2.2.1. PC_{fdr}^+ Algorithm.
The initial version of Li and Wang's [11] method, called the PC_{fdr} algorithm, was proved to be capable of asymptotically controlling the FDR. Here we present an extension of the PC_{fdr} algorithm which can incorporate *a priori* knowledge, which was not specified in the original PC_{fdr} algorithm. We name the extension as the PC_{fdr}^+ algorithm where the superscript "+" indicates that it is an extension. The pseudocode of the PC_{fdr}^+ algorithm is given in Algorithm 1, and its Matlab implementation is downloadable at http://www.junningli.org/software. It handles prior knowledge with two inputs: E_{must}, the set of edges assumed to appear in the true graph, and E_{test}, the set of edges to be tested from the data. The original PC_{fdr} algorithm can thus be regarded as a special case of the extended algorithm, by setting $E_{must} = \varnothing$ and $E_{test} = \{$all possible edges$\}$.

2.2.2. Asymptotic Performance.
Before we present theorems about the asymptotic performance of the PC_{fdr}^+ algorithm and its heuristic modification, let us first introduce the assumptions related to the theorems.

(A1) The multivariate probability distribution P is faithful to a directed acyclic graph (DAG) whose skeleton is G_{true}.

(A2) The number of vertices is fixed.

(A3) Given a fixed significance level of testing conditional independence, the power of detecting conditional dependence approaches 1 at the limit of large sample sizes.

(A4) The union of E_{must}, the edges assumed to be true, and E_{test}, the edges to be tested, covers E_{true}, all the true edges; that is, $E_{test} \cup E_{must} \supseteq E_{true}$.

Assumption (A1) is generally assumed when graphical models are applied, and it restricts the probability distribution P to a certain class. Assumption (A2) is usually implicitly stated, but here we emphasize it because it simplifies the proof. Assumption (A3) may seem overly restrictive, but actually can be easily satisfied by standard statistical tests, such as the likelihood ratio test introduced by Neyman and Pearson [18] and the partial-correlation test by Fisher [19], if the data are identically and independently sampled. Assumption (A4) relates to prior knowledge, which interestingly does not require that the assumed "true" edges E_{must} be a subset of the true edges E_{true}, but just that all true edges are included in the union of the assumed "true" edges and the edges to be tested.

The detection power of the PC_{fdr}^+ algorithm and its heuristic modification at the limit of large sample sizes is elucidated in Theorem 2.

Theorem 2. *Assuming (A1), (A2), and (A3), both the PC_{fdr}^+ algorithm and its heuristic modification, the PC_{fdr*}^+ algorithm, are able to recover all the true connections in E_{test} with probability one as the sample size approaches infinity:*

$$\lim_{m \to \infty} P\left(E'_{true} \subseteq E'_{stop}\right) = 1, \tag{1}$$

where E'_{true} denotes the set of true edges in E_{test}; that is, $E'_{true} = E_{true} \cap E_{test}$, E'_{stop} denotes the set of edges inferred by the algorithm about E_{test}; that is, $E'_{stop} = E_{stop} \cap E_{test}$, and m denotes the sample size.

It should be noted that Theorem 2 does not need Assumption (A4), which implies that the true edges in E_{test} are still able to be recovered by the algorithms with probability one at the limit of large sample sizes, even if

Input: the data D, the undirected edges E_{must} that are assumed to exist in the true undirected graph G_{true} according to prior knowledge, the undirected edges E_{test} ($E_{\text{must}} \cap E_{\text{test}} = \varnothing$) to be tested from the data D, and the FDR level q for making inference about E_{test}.
Output: an undirected graph G_{stop}, that is, the value of G when the algorithm stops, or equivalently, E_{stop}, the edges in G_{stop}.
Notations: D denotes the multivariate input data. a, b denote the vertices. E, C denote the vertex sets. $a \sim b$ denotes an undirected edge. $\text{adj}(a, G)$ denotes vertices adjacent to a in graph G. $a \perp b|C$ denotes the conditional independence between a and b given C.

(1) Form an undirected graph G from $E_{\text{test}} \cup E_{\text{must}}$.
(2) Initialize the maximum p values associated with the edges in E_{test} as $P^{\max} = \{p_{a \sim b}^{\max} = -1 \mid a \sim b \in E_{\text{test}}\}$.
(3) Let depth $d = 0$.
(4) **repeat**
(5)   **for** each ordered pair of vertices a and b that $a \sim b \in E \cap E_{\text{test}}$ and $|\text{adj}(a, G) \setminus \{b\}| \geq d$ **do**
(6)   **for** each subset $C \subseteq adj(a, G) \setminus \{b\}$ and $|C| = d$ **do**
(7)   Test hypothesis $a \perp b \mid C$ and calculate the p value $p_{a \perp b|C}$.
(8)   **if** $p_{a \perp b|C} > p_{a \sim b}^{\max}$, **then**
(9)   Let $p_{a \sim b}^{\max} = p_{a \perp b|C}$.
(10)   **if** every element of $P^{\max}$ has been assigned a valid p value by step 9, **then**
(11)   Run the FDR procedure, Algorithm 2, with $P^{\max}$ and q as the input.
(12)   **if** the non-existence of certain edges are accepted, **then**
(13)   Remove these edges from G.
(14)   Update G and E.
(15)   **if** $a \sim b$ is removed, **then**
(16)   **break** the **for loop** at line 6.
(17)   **end if**
(18)   **end if**
(19)   **end if**
(20)   **end if**
(21)   **end for**
(22)   **end for**
(23)   Let $d = d + 1$.
(24) **until** $|\text{adj}(a, G) \setminus \{b\}| < d$ for every ordered pair of vertices a and b that $a \sim b$ is in $E \cap E_{\text{test}}$.
A heuristic modification, named the $\text{PC}_{\text{fdr}*}^+$ algorithm, at step 14 removes $p_{a \sim b}^{\max}$ from $P^{\max}$ as well once $a \sim b$ is removed from G.

ALGORITHM 1: The PC_{fdr}^+ algorithm.

the edges assumed to be present by users are not completely correctly specified.

The FDR of the PC_{fdr}^+ algorithm at the limit of large sample sizes is elucidated in Theorem 3.

Theorem 3. *Assuming (A1), (A2), (A3), and (A4), the FDR of the set of edges inferred by the PC_{fdr}^+ algorithm about E_{test} approaches a value not larger than the user-specified level q as the sample size m approaches infinity:*

$$\limsup_{m \to \infty} FDR\left(E'_{stop}, E'_{true}\right) \leq q, \qquad (2)$$

where $FDR(E'_{stop}, E'_{true})$ is defined as

$$FDR\left(E'_{stop}, E'_{true}\right) = E\left[\frac{\left|E'_{stop} \setminus E'_{true}\right|}{\left|E'_{stop}\right|}\right],$$

$$\text{Define } \frac{\left|E'_{stop} \setminus E'_{true}\right|}{\left|E'_{stop}\right|} = 0, \quad \text{if } E'_{stop} = \varnothing. \qquad (3)$$

Theorem 3 concerns the PC_{fdr}^+ algorithm, and it requires Assumption (A4). We are still not sure whether similar

FDR performance can be proved for the $\text{PC}_{\text{fdr}*}^+$ algorithm. Assumption (A4) does not require that the assumed "true" edges E_{must} is a subset of the true edges E_{true} but only that all true edges are included in the union of the assumed "true" edges and the edges to be tested. This is particularly useful in practice, since it does not require users' prior knowledge to be absolutely correct, but allows some spurious edges to be involved in E_{must}, once all true edges have been included in either E_{must} or E_{test}. Assumption (A4) can be satisfied by making $E_{\text{test}} \cup E_{\text{must}}$ large enough to cover all the true edges, but as shown in (4) this will increase the computational cost of the algorithm.

Theorems 2 and 3 address the performance of the PC_{fdr}^+ algorithm and its heuristic modification at the limit of large sample sizes. Because the PC_{fdr}^+ algorithm is derived from the PC_{fdr} algorithm, its performance should be very similar. The numerical examples of the PC_{fdr} algorithm in Li and Wang's [11] work may provide helpful and intuitive understanding on the performance of the PC_{fdr}^+ algorithm with moderate sample sizes.

The detailed proofs of Theorems 2 and 3 are provided in Appendix A.

Input: a set of p values $\{p_i \mid i = 1, \ldots, H\}$, and the threshold of the FDR q
Output: the set of rejected null hypotheses
(1) Sort the p-values of H hypothesis tests in the ascendant order as $p_{(1)} \leq \cdots \leq p_{(H)}$.
(2) Let $i = H$, and $H^* = H$ (or $H^* = H(1 + 1/2, \ldots, +1/H)$, depending on the assumption of the dependency among the test statistics).
(3) **while**

$$\frac{H^*}{i} p_{(i)} > q \text{ and } i > 0, \quad (\star)$$

do
(4)   Let $i = i - 1$.
(5) **end while**
(6) Reject the null hypotheses associated with $p_{(1)}, \ldots, p_{(i)}$, and accept the null hypotheses associated with $p_{(i+1)}, \ldots, p_{(H)}$.

ALGORITHM 2: FDR setup [6].

2.2.3. Computational Complexity. The majority of the computational effort in the PC_{fdr}^+ is utilized in performing statistical tests of conditional independence at step 7 and the FDR at step 11. If the algorithm stops at the depth $d = d_{max}$, then the number of conditional independence tests required is bounded by

$$T = 2|E_{test}| \sum_{d=0}^{d_{max}} C_{\Delta-1}^d \leq |E_{test}| 2^{\Delta}, \quad (4)$$

where $|E_{test}|$ is the number of edges to be tested, Δ is the maximum degree of graph G_{init} (the graph formed at step 1) whose edges are $E_{must} \cap E_{test}$, and $C_{\Delta-1}^d$ is the number of combinations of choosing d unordered and distinct elements from $\Delta - 1$ elements. The bound usually is very loose, because it assumes that no edge has been removed until $d = d_{max}$.

The computational complexity of the FDR procedure, Algorithm 2, invoked at step 11 of the PC_{fdr}^+ algorithm is $O(H \log(H))$ when it is invoked for the first time, where $H = |E_{test}|$ is the number of input p values and is $O(H)$ later, with the optimization suggested in Appendix B. In the worst case that $p_{a \perp b|C}$ is always larger than $p_{a \sim b}^{max}$, the complexity of the computation spent on the FDR control in total is bounded by $O(|E_{test}| \log(|E_{test}|) + T|E_{test}|)$ where T is the number of performed conditional independence tests (see (4)). This is a very loose bound because it is rare that $p_{a \perp b|C}$ is always larger than $p_{a \sim b}^{max}$.

In practice, the PC_{fdr}^+ algorithm runs very quickly, especially for sparse networks. In our experiments (see Section 3.1), it took about 10 seconds to infer the structure of a first-order dynamic network with 20 nodes from data of 1000 time points.

2.2.4. Miscellaneous Discussions. It should be noted that controlling the FDR locally is not equivalent to controlling it globally. For example, if it is known that there is only one connection to test for each node, then controlling the FDR locally in this case will degenerate to controlling the point-wise error rate, which cannot control the FDR globally.

Listgarten and Heckerman [8] proposed a permutation method to estimate the number of spurious connections in a graph learned from data. The basic idea is to repetitively apply a structure learning algorithm to data simulated from the null hypotheses with permutation. This method is generally applicable to any structure learning method, but permutation may make the already time-consuming structure learning problem even more computationally cumbersome, limiting its use in practical situations.

2.3. FDR-Controlled Group Brain Connectivity Inference with or without A Priori Knowledge. In this section, we propose another extension to the PC_{fdr} algorithm: from the single subject level to the group level. Assessing group-level activity is done by considering a mixed-effect model (Step 7 of Algorithm 3), and we name it the gPC_{fdr} algorithm where "g" indicates that it is an extension at the group level. When also incorporating *a priori* knowledge, the resulting algorithm is named the gPC_{fdr}^+ algorithm.

Suppose we have m subjects within a group. Then for subject i, the conditional independence between the activities of two brain regions a and b given other regions C can be measured by the partial correlation coefficient between $X_a(i)$ and $X_b(i)$ given $X_C(i)$, denoted as $r_{ab|C}(i)$. Here $X_\bullet$ denotes variables associated with a vertex or a vertex set, and index i indicates that these variables are for subject i. By definition, the partial correlation coefficient $r_{ab|C}(i)$ is the correlation coefficient between the residuals of projecting $X_a(i)$ and $X_b(i)$ onto $X_C(i)$ and can be estimated by the sample correlation coefficient as

$$\hat{r}_{ab|C}(i) = \frac{\text{Cov}[Y_{a|C}(i), Y_{b|C}(i)]}{\sqrt{\text{Var}[Y_{a|C}(i)] \text{Var}[Y_{b|C}(i)]}}, \quad (5)$$

where

$$\beta_{a|C}(i) = \arg\min_{\beta} |X_a(i) - X_C(i)\beta|^2,$$

$$\beta_{b|C}(i) = \arg\min_{\beta} |X_b(i) - X_C(i)\beta|^2,$$

$$Y_{a|C}(i) = X_a(i) - X_C(i)\beta_{a|C}(i),$$

$$Y_{b|C}(i) = X_b(i) - X_C(i)\beta_{b|C}(i). \quad (6)$$

For clarity, in the following discussion we omit the subscript "$ab \mid C$" and simply use index "i" to emphasize that a variable is associated with subject i.

Input: the multisubject data D, undirected complete graph G, complete vertex set E and the FDR controlled
level q for making inference about E.
Output: the recovered undirected graph G_{stop}.
Notations: a, b denote the vertices. E, C denote the vertex set. $a \sim b$ denotes an undirected edge. $\text{adj}(a, G)$
denotes vertices adjacent to a in graph. $a \perp b \mid C$ denotes the conditional independence between a and b given C.

(1) Form an undirected graph G on the vertex set E.
(2) Initialize the maximum p values associated with the edges in E as $P^{\max} = \{p_{a \sim b}^{\max} = -1 \mid a \sim b \in E\}$.
(3) Let depth $d = 0$.
(4) **repeat**
(5)   **for** each ordered pair of vertices a and b that $a \sim b \in E \cap E$ and $|\text{adj}(a, G) \setminus \{b\}| \geq d$ **do**
(6)   **for** each subset $C \subseteq \text{adj}(a, G) \setminus \{b\}$ and $|C| = d$ **do**
(7)   Test hypothesis $a \perp b \mid C$ for each subject and calculate the p value $p_{a \perp b \mid C}$ at the **group level**.
(8)   **if** $p_{a \perp b \mid C} > p_{a \sim b}^{\max}$, **then**
(9)   Let $p_{a \sim b}^{\max} = p_{a \perp b \mid C}$.
(10)   **if** every element of $P^{\max}$ has been assigned a valid p value by step 9, **then**
(11)   Run the FDR procedure, Algorithm 2, with $P^{\max}$ and q as the input.
(12)   **if** the non-existence of certain edges are accepted, **then**
(13)   Remove these edges from G.
(14)   Update G and E.
(15)   **if** $a \sim b$ is removed, **then**
(16)   **break** the **for loop** at line 6.
(17)   **end if**
(18)   **end if**
(19)   **end if**
(20)   **end if**
(21)   **end for**
(22)   **end for**
(23)   Let $d = d + 1$.
(24) **until** $|\text{adj}(a, G) \setminus \{b\}| < d$ for every ordered pair of vertices a and b that $a \sim b$ is in E.
Note: When *a priori* knowledge is available, we can also incorporate the prior knowledge into the gPC_{fdr}
algorithm to obtain the $\text{gPC}_{\text{fdr}}^+$ algorithm, where the inputs are updated as follows: the multisubject data D,
the undirected edges E_{must} that are assumed to appear in the true undirected graph G_{true} according to prior
knowledge, the undirected edges E_{test} whose existences are to be tested from the data, and the FDR controlled
level q for making inference about E_{test}.

ALGORITHM 3: The gPC_{fdr} algorithm.

To study the group-level conditional independence relationships, a group-level model should be introduced for r_i. Since partial correlation coefficients are bounded and their sample distributions are not Gaussian, we apply Fisher's z-transformation to convert (estimated) partial correlation coefficients r to a Gaussian-like distributed z-statistic z, which is defined as

$$z = Z(r) = \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right), \qquad (7)$$

where r is a (estimated) partial correlation coefficient and z is its z-statistic.

The group model we employ is

$$z_i = z_g + e_i, \qquad (8)$$

where e_i follows a Gaussian distribution $N(0, \sigma_g^2)$ with zero mean and σ_g^2 variance. Consequently, the group-level testing of conditional independence is to be used to test the null hypothesis $z_g = 0$.

Because z_i is unknown and can only be estimated, the inference of z_g should be conducted with $\hat{z}_i = Z(\hat{r}_i)$. If

$X_a(i)$, $X_b(i)$, and $X_C(i)$ jointly follow a multivariate Gaussian distribution, then $\hat{z}_i$ asymptotically follows a Gaussian distribution $N(z_i, \sigma_i^2)$ with $\sigma_i^2 = 1/(N_i - p - 3)$, where N_i is the sample size of subject i's data and p represents the number of variables in $X_C(i)$. Therefore, based on (8), we have

$$\hat{z}_i = z_g + e_i + \epsilon_i, \qquad (9)$$

where ϵ_i follows $N(0, \sigma_i^2)$ and e_i follows $N(0, \sigma_g^2)$. This is a mixed-effect model where ϵ_i denotes the intrasubject randomness and e_i denotes the intersubject variability. At the group level, $\hat{z}_i$ follows a Gaussian distribution $N(z_g, \sigma_i^2 + \sigma_g^2)$. Note that unlike regular mixed-effect models, the intrasubject variance σ_i^2 in this model is known, because N_i and p are known given the data $X(i)$ and C. In general, $\sigma_i^2 = 1/(N_i - p - 3)$ is not necessarily equal to σ_j^2 for $i \neq j$, and the inference of z_g should be conducted in the manner of mixed models, such as estimating σ_g^2 with the restricted maximum likelihood (ReML) approach. However, if the sample size of each subject's data is the same, then σ_i^2 equals σ_j^2. For this balanced case, which is typically true in fMRI applications
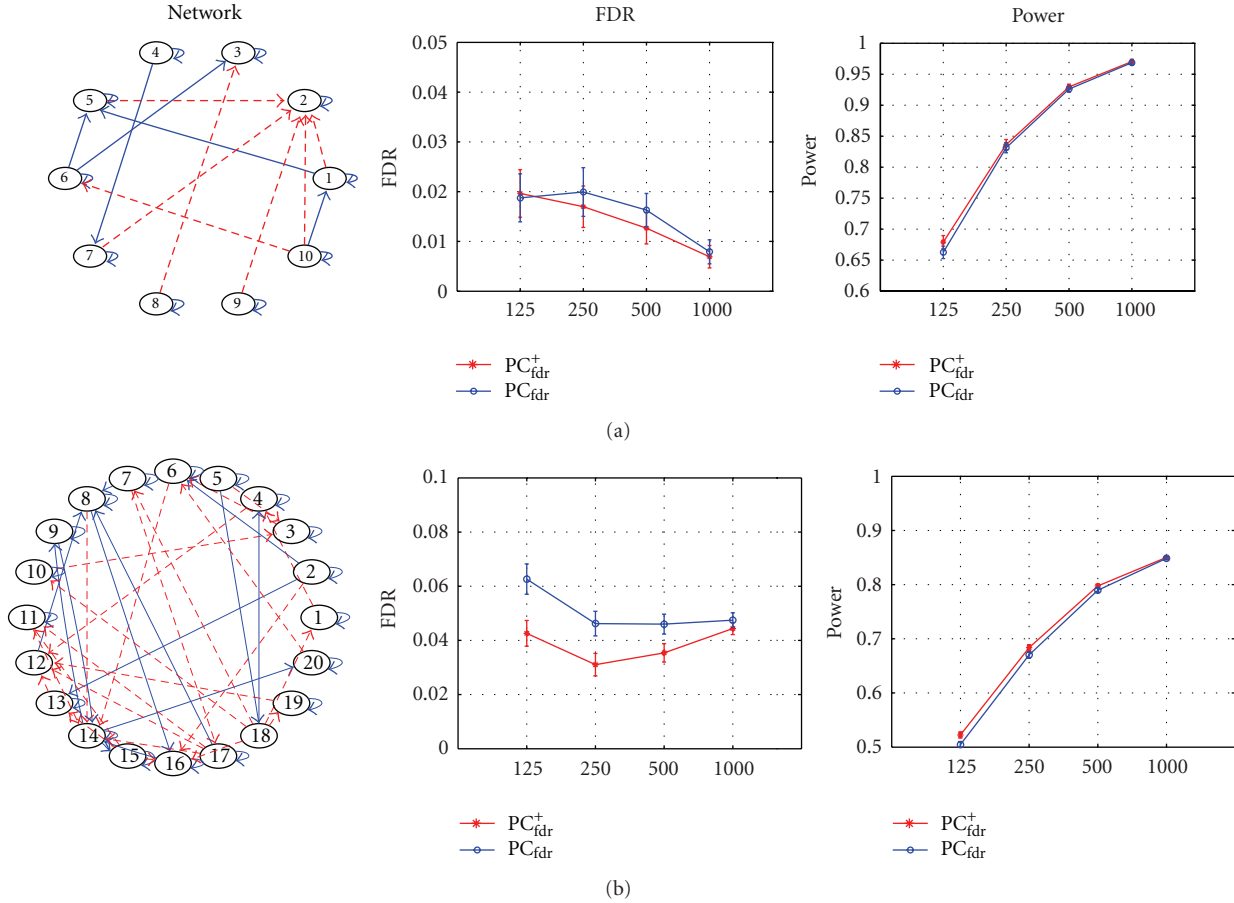
Figure 1: Simulation results for the PC$_{fdr}^+$ algorithm. (a) Simulation results for the network with 10 nodes and 23 edges. (b) Simulation results for the network with 20 nodes and 56 edges. In the networks, solid arrows represent edges from time t to $t + 1$, and dashed arrows represent edges with no time lag (i.e., from time t to t). For the FDR and detection power curves, the blue solid lines represent the PC$_{fdr}$ algorithm, the red solid lines represent the PC$_{fdr}^+$ algorithm, the x-axis means the sample sizes, and the y-axis means the FDR or detection power.

and as well the case in this paper, we can simply apply a t-test to $\hat{z}_i$'s to test the null hypothesis $z_g = 0$.

Replacing Step 7 of the single-subject PC$_{fdr}$ algorithm (i.e., the intrasubject hypothesis test) with the test of $z_g = 0$, we can extend the single-subject version of the algorithm to its group-level version. We will employ this t-test in our simulations and in the real fMRI data analysis presented later in this paper. Such a testing approach significantly simplifies the estimation process, and our simulation results presented later demonstrate that this method can still control the FDR at a user specified error rate level.

3. Experiments

3.1. Simulations for the PC$_{fdr}^+$ Algorithm. Here we compare the performances of the proposed PC$_{fdr}^+$ algorithm and the original PC$_{fdr}$ algorithm, using time series generated from two dynamic Bayesian networks in Figure 1. One network has 20 nodes (10 channels) and 23 edges, and the other has 40 nodes (20 channels) and 56 edges. The dynamic Bayesian networks are assumed Gaussian, with connection coefficients

uniformly distributed in $[0.2, 0.6]$ with Gaussian noise whose amplitudes are uniformly distributed in $[0.5, 1.1]$. We use partial correlation coefficients to test conditional independence relationships. The target FDR for both methods is set as 5%. For the PC$_{fdr}^+$ algorithm, one-third of the nonexisting connections are excluded as prior knowledge.

Figure 1 shows the estimated FDR and detection power results, at sample sizes of 125, 250, 500, and 1000 time points and with 50 repetitive trials for each sample size. As shown in graphs (a) and (b), the PC$_{fdr}^+$ and PC$_{fdr}$ algorithms can both control the FDR under or around 5%. For both methods, the detection power increases as the sample size increases. However, we can see that the PC$_{fdr}^+$ algorithm yields higher detection power and lower FDR than the original PC$_{fdr}$ algorithm does. As mentioned earlier in the Introduction Section, the PC$_{fdr}^+$ algorithm has the advantage of providing researchers more flexibility in using the method and higher accuracy in discovering brain connectivity.

3.2. Simulations for the gPC$_{fdr}$ Algorithm. The simulations here serve two purposes: first, to verify whether the proposed gPC$_{fdr}$ algorithm for modeling brain connectivity can control

the FDR at the group level, and second, to compare the gPC$_{fdr}$ algorithm with the single-subject PC$_{fdr}$ algorithm proposed in [11] and the state-of-art IMaGES algorithm investigated in Ramsey et al. [16] for inferring the structure of the group connectivity network.

The simulations were conducted as follows. First, a connectivity network is generated as the group-level model. Individual subject-level networks are then derived from the group-level model by randomly adding or deleting connections with a small probability, and subject-specific data are generated according to individual subject networks. Next, the network-learning methods, that is, the proposed gPC$_{fdr}$ algorithm, the single-subject PC$_{fdr}$ method with pooling together the data from all subjects, and the IMaGES algorithm, are applied to the simulated data. Finally, the outputs of the algorithms are compared with the true group-level network to evaluate their accuracy.

The data generation process is as follows.

(1) Randomly generate a directed acyclic graph (DAG) as the group-level network and associate each connection with a coefficient. The DAG is generated by randomly connecting nodes with edges and then orienting the edges according to a random order of the nodes. The connection coefficients are assigned as random samples from the uniform distribution $U(\beta_1, \beta_2)$, where β_1 and β_2 characterize the coefficient strength.

(2) For each subject, a subject-level network is derived from the group-level network by randomly adding and deleting connections. More specifically, for each of the existing connections, the connection is deleted with probability 0.05, and for each of the absent connections, a connection is added with probability 0.01. The corresponding connection coefficients are randomly sampled from the uniform distribution $U(\beta_1, \beta_2)$.

(3) Given a subject-level network, the subject-specific data are generated from a Gaussian Bayesian network, with the additional Gaussian noise following the standard Gaussian distribution $N(0, 1)$.

In the first simulation, we compare the performances of the proposed gPC$_{fdr}$ algorithm, the original PC$_{fdr}$ algorithm, and the IMaGES algorithm [16], when using different connection coefficient strengths. In this example, the group-level network is the DAG in Figure 2(a). From this model, twenty subject-level models are derived, and for each subject, data with three hundred samples are simulated. To test the performances of the algorithms with a range of connection strengths, we vary the connection coefficient generating distribution $U(\beta_1, \beta_2)$ gradually from $U(0.2, 0.3)$ to $U(0.7, 0.8)$. At the network-learning stage, we set the target FDR to be 5% for the gPC$_{fdr}$ algorithm. For reliable assessment, this procedure is repeated thirty times.

Figures 2(b), 2(c), and 2(d) show the FDR and the type I error rate, and the detection power results as a function of connection strength. We note that all methods are relatively invariant to connection strength. The proposed

gPC$_{fdr}$ algorithm steadily controls the FDR below or around the desired level and accurately makes the inference at the group level. The detection power of IMaGES algorithm is higher than that of gPC$_{fdr}$ algorithm, but it fails to control the FDR under the specified 5% level. Its higher detection power is achieved by sacrificing FDR. This is reasonable, since IMaGES is not specifically designed to control the FDR error rate.

In the second simulation, we test the performances of the algorithms as a function of the number of subjects within the group. The group-level network is the DAG in Figure 3(a), and the number of subjects increases from eight to twenty-five. At the network-learning stage, we set the target FDR to be 5%. This procedure is repeated thirty times.

Figure 3(b) demonstrates the FDR results as a function of the number of subjects within the group. It is noted that the proposed gPC$_{fdr}$ algorithm is able to keep the FDR below or around the specified level. The detection power gradually increases as the number of subjects increases. When there are more than 15 subjects, the gPC$_{fdr}$ algorithm seems that it can achieve higher (better) detection power and lower (better) FDR and type I error rate than the IMaGES algorithm does. It suggests that when the number of subjects is large enough, the proposed gPC$_{fdr}$ algorithm can jointly address efficiency, accuracy, and intersubject variability. The original PC$_{fdr}$ algorithm of simply pooling the data together fails to control the FDR, and the resulting FDR does not decrease as the number of subject increases, probably due to the increasing heterogeneity within the group. In order to investigate the effects of the number of ROIs, we also investigate two networks with 15 and 25 nodes, respectively, and repeat the simulations (not shown here). The results are qualitatively similar to what we show here.

3.3. fMRI Application. In order to assess the real-world application performance of the proposed method, we apply the gPC$_{fdr}^+$ algorithm for inferring group brain connectivity network to fMRI data collected from twenty subjects. All experiments were approved by the University of British Columbia Ethics Committee. Ten normal people and ten Parkinson's disease (PD) patients participated in the study. During the fMRI experiment, each subject was instructed to squeeze a bulb in their right hand to control an "inflatable" ring so that it smoothly passed through a vertically scrolling a tunnel. The normal controls performed only one trial, while Parkinson's subjects performed twice, once before L-dopa medication and the other approximately an hour later, after taking medication.

Three groups were categorized: group N for the normal controls, group P$_{pre}$ for the PD patients before medication, and group P$_{post}$ for the PD patients after taking L-dopa medication. For each subject, 100 observations were used in the network modeling. For details of the data acquisition and preprocessing, please refer to Palmer et al. [20]. 12 anatomically defined regions of interest (ROIs) were chosen based on prior knowledge of the brain regions associated with motor performance (Table 1).

We utilized the two extensions of the PC$_{fdr}$ algorithm and learned the structures of first-order group dynamic Bayesian
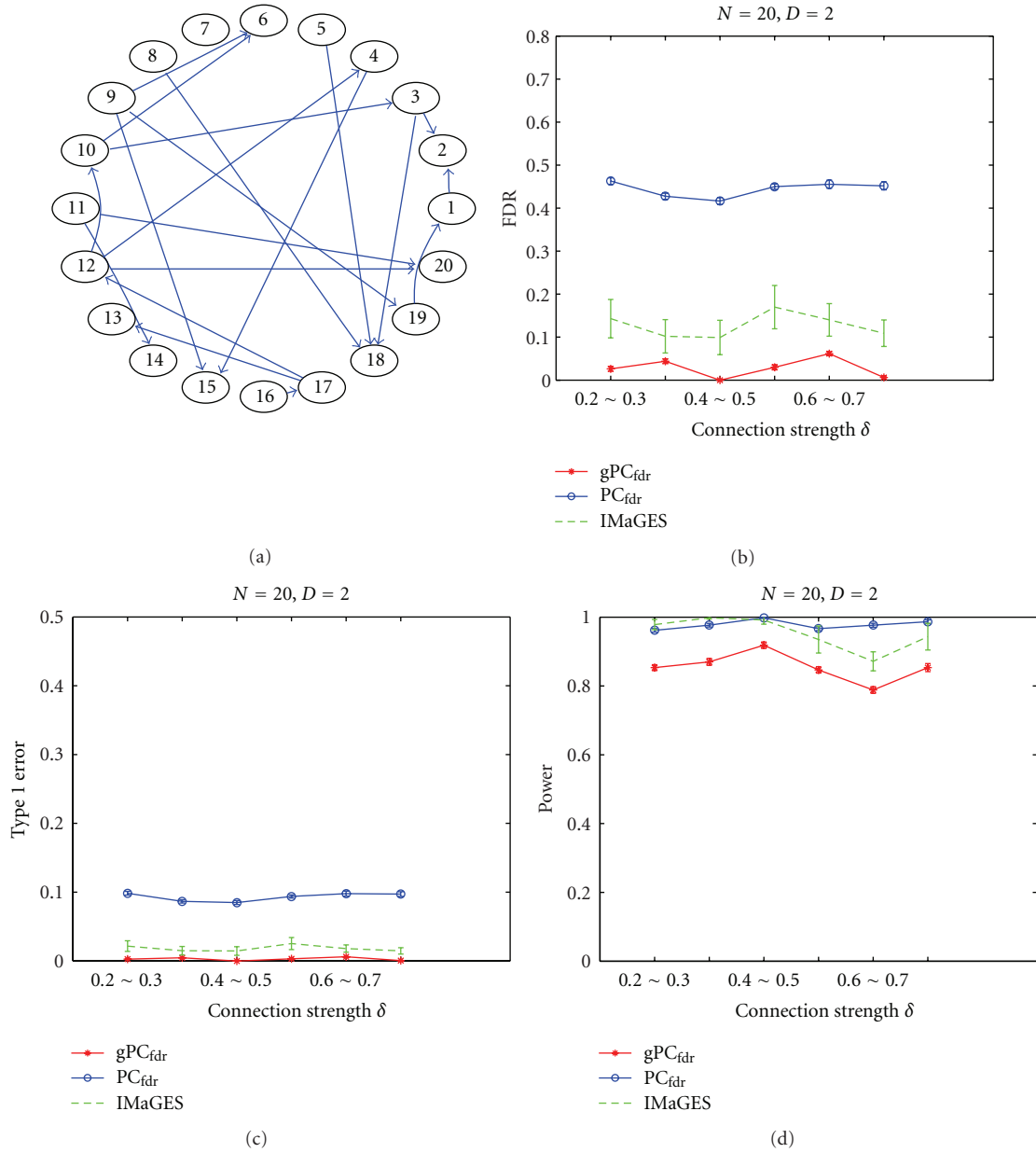
(a)

(b)

(c)

(d)

FIGURE 2: Simulation 1: assessing the effects of connection strength on the learned group networks. (a) The group-level network, with 20 nodes and an average of two connections per node. (b) The FDR curves (with standard deviation marked) of the gPC_{fdr} algorithm, the original PC_{fdr} algorithm by pooling all subject data together, and the IMaGES algorithm. (c) The type I error rate curves. (d) The detection power curves. The x-axis represents the generating distribution $U(\beta_1, \beta_2)$ for sampling the connection coefficients.

networks from fMRI data. Because the fMRI BOLD signal can be considered as the convolution of underlying neural activity with a hemodynamic response function, we assumed that there must be a connection from each region at time t to its mirror at time $t + 1$. We also assumed that there must be a connection between each region and its homologous region in the contralateral hemisphere. The TR interval (i.e., sampling period) was a relatively long, 1.985 seconds; we restricted ourselves to learn only connections between ROIs without time lags. In total, there are $12 + 6 = 18$ pre-defined connections and $12 \times (12 - 1) \div 2 - 6 = 60$ candidate connections to be tested. The brain connectivity

networks (with the target FDR of 5%) learned for the normal (group N) and PD groups before (group P_{pre}) and after (group P_{post}) medication are compared in Figure 4. Note the connection between the cerebellar hemisphere and contralateral thalamus in the normal subjects and between the supplementary motor area (SMA) and the contralateral putamen, consistent with prior knowledge. Interestingly, in P_{pre} subjects, the left cerebellum now connects with the right SMA, and the right SMA $\leftrightarrow$ left putamen connection is lost. Also, there are now bilateral primary motor cortex (M1) $\leftrightarrow$ putamen connections seen in the P_{pre} group, presumably as a compensatory mechanism. After medication
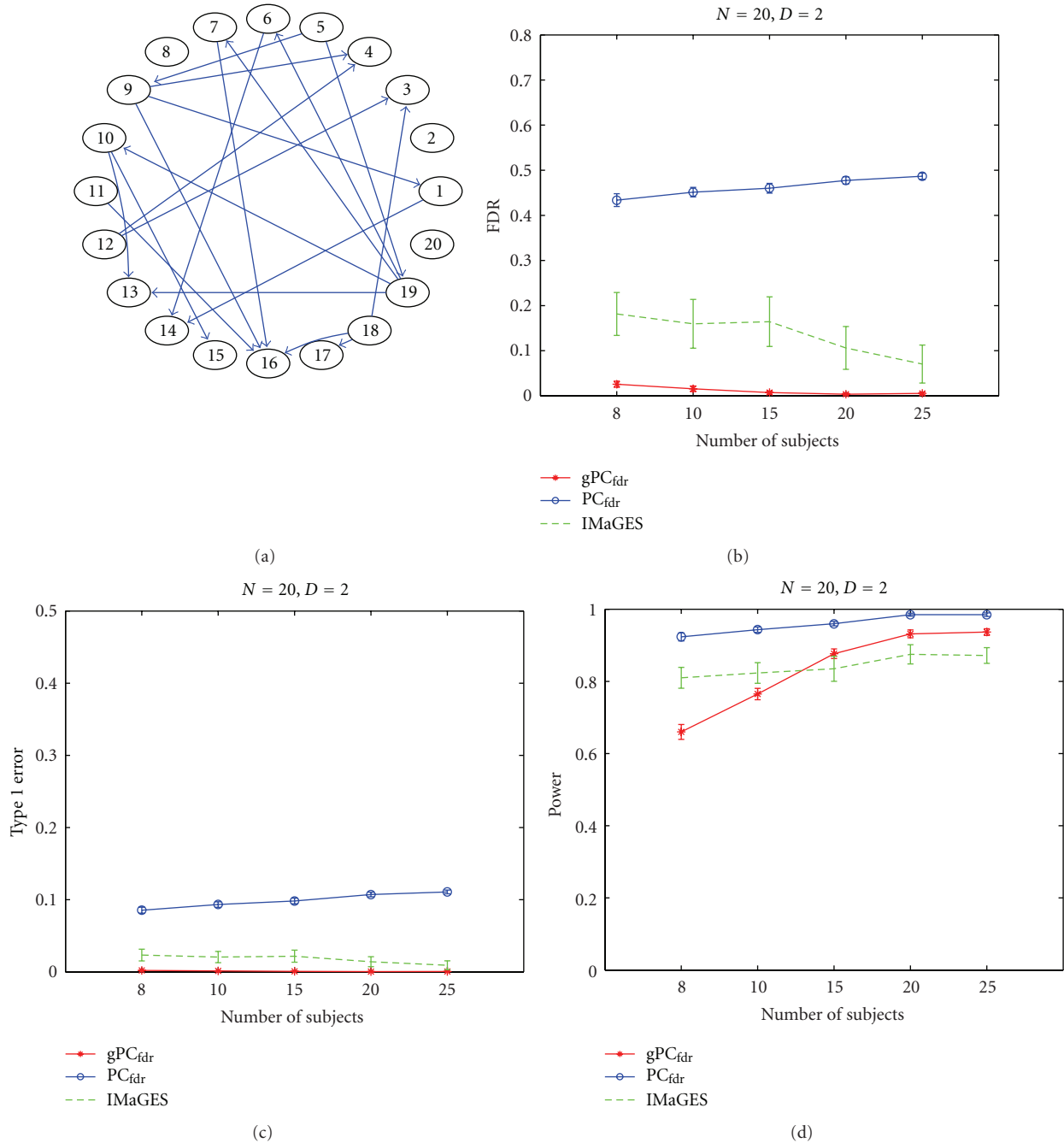
(a)

(b)

(c)

(d)

FIGURE 3: Simulation 2: assessing the effects of increasing the number of subjects on the learned group networks. (a) The group-level network, with 20 nodes and an average of two connections per node. (b) The FDR curves (with standard deviation marked) of the proposed gPC$_{fdr}$ algorithm, the original PC$_{fdr}$ algorithm by pooling all subject data together, and the IMaGES algorithm. (c) The type I error rate curves. (d) The detection power curves. The x-axis represents the number of subjects within the group.

(P$_{post}$), the left SMA ↔ left thalamus connection is restored back to be normal.

4. Discussion

Up to now, graphical models to infer brain connectivity from fMRI data have implicitly relied on the unrealistic assumption that if a model accurately represented the overall activity in several ROIs, the internal connections of such a model would accurately reflect underlying brain connectivity. The PC$_{fdr}$ algorithm was designed to loosen this overly restrictive assumption and asymptotically control the FDR of network connections inferred from data.

In this paper, we first presented the PC$_{fdr}^+$ algorithm, an extension of the PC$_{fdr}$ algorithm, which allows for incorporation of prior knowledge of network structure into
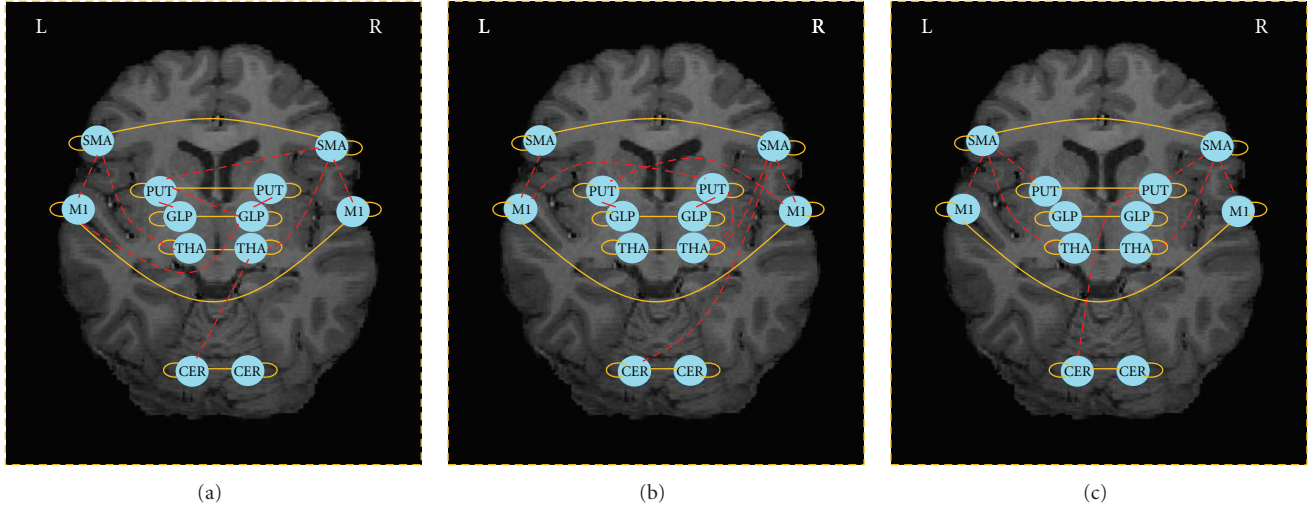
FIGURE 4: (a) Learned brain connectivity for the normal group (group N). (b) Learned brain connectivity for the PD group before medication (group P_{pre}). (c) Learned brain connectivity for the PD group after medication (group P_{post}). Here "L" and "R" refer to the left and right sides, respectively. The solid lines are predefined connectivity, and the dashed lines are learned connectivity.

TABLE 1: Brain regions of interest (ROIs).

Full name of brain region	Abbreviation
Left/right lateral cerebellar hemispheres	lCER, rCER
Left/right globus pallidus	lGLP, rGLP
Left/right putamen	lPUT, rPUT
Left/right supplementary motor cortex	lSMA, rSMA
Left/right thalamus	lTHA, rTHA
Left/right primary motor cortex	lM1, rM1

"l" or "r" in the abbreviations stands for "Left" or "Right," respectively.

the learning process, greatly enhancing its flexibility in practice. The PC_{fdr}^+ algorithm handles prior knowledge with two inputs: E_{must}, which is the set of edges that are assumed to appear in the true graph, and E_{test}, the set of edges that are to be tested from the observed data. We proved that, with mild assumptions and at the limit of large samples, the PC_{fdr}^+ algorithm is able to recover all the true edges in E_{test} and also curb the FDR of the edges inferred about E_{test}.

It is interesting that the PC_{fdr}^+ algorithm does not require the assumed "true" edges E_{must} to be a subset of the true edges E_{true}, but only that all true edges are included in the union of the assumed "true" edges and the edges to test. This is very useful in research practice, since it allows some spurious edges to be involved in E_{must}, as long as all the true edges have been included in either E_{must} or E_{test}. Users can satisfy this requirement by making $E_{test} \cup E_{must}$ large enough to cover all the true edges.

When we compared the PC_{fdr}^+ algorithm with the original PC_{fdr} algorithm, both of them successfully controlled the FDR under the target threshold in simulations, providing a practical tradeoff between computational complexity and accuracy. However, the PC_{fdr}^+ algorithm achieved better detection power and better FDR than the original PC_{fdr} algorithm. Incorporating prior knowledge into PC_{fdr} algorithm

therefore enhances inference accuracy and improves the flexibility in using the method.

Another extension to PC_{fdr} algorithm we described here was the ability to infer brain connectivity patterns at the group level, with intersubject variance explicitly taken into consideration. As a combination of the PC_{fdr} algorithm and a mixed-effect model, the gPC_{fdr} algorithm takes advantage of the error control ability of the PC_{fdr} algorithm and the capability of handling intersubject variance. The simulation results suggest that the proposed method was able to accurately discover the underlying group network and steadily control the false discovery rate. Moreover, the gPC_{fdr} algorithm was shown to be much more reliable than simply pooling together the data from all subjects. This may be especially important in disease states and older subjects. Compared with the IMaGES algorithm, gPC_{fdr} demonstrated better control of the FDR.

As with all group models, a limitation of the proposed gPC_{fdr} algorithm is the requirement of a sufficient number of subjects. While it is appreciated that in many biomedical applications data collection is resource intensive, and if the number of subjects is insufficient, the gPC_{fdr} algorithm may give unreliable results. Nevertheless, the group extension to the PC_{fdr} algorithm is one attempt to make brain connectivity inference using error-rate-controlled exploratory modeling.

When applying the proposed gPC_{fdr}^+ to fMRI data collected from PD subjects performing a motor tracking task, we found group evidence of disease changes (e.g., loss of left cerebellar ↔ SMA connectivity), compensatory changes in PD (e.g., bilateral M1 ↔ contralateral putamen connectivity), and evidence of restoration of connectivity after medication (left SMA ↔ left thalamus). The tremendous variability in clinical progression of PD is likely due to variability not only in disease rate progression, but also in variability in the magnitude of compensatory changes.

This highlights the importance of the proposed method, as it allows robust estimation of disease effects, compensatory effects, and effects of medication, all with a reasonable sample size, despite the enhanced intersubject variability seen in PD.

Appendices

A. Proof of Theorems

To assist the reading, we list below notations frequently used in the proof.

V: all the nodes in a graph,

G_{true}: the skeleton of the true underlying directed acyclic graph (DAG),

$\mathcal{A}_{a\sim b}$: the event that edge $a \sim b$ is in the graph recovered by the PC_{fdr}^+ algorithm,

$\mathcal{A}_{E'_{\text{true}}}$: $\mathcal{A}_{E'_{\text{true}}} = \bigcap_{a\sim b\in E'_{\text{true}}} \mathcal{A}_{a\sim b}$, the joint event that all the edges in E'_{true}, the true edges in E_{test}, are recovered by the PC_{fdr}^+ algorithm,

$p_{a\sim b}$: the value of $p_{a\sim b}^{\max}$ when the PC_{fdr}^+ algorithm stops,

$C_{a\sim b}^*$: a certain vertex set that d-separates a and b in the true DAG and that is also a subset of either $\text{adj}(a, G_{\text{true}}) \setminus \{b\}$ or $\text{adj}(b, G_{\text{true}}) \setminus \{a\}$, according to Proposition 1. $C_{a\sim b}^*$ is defined only for vertex pairs that are not adjacent in G_{true},

$p_{a\sim b}^*$: the p value of testing $X_a \perp X_b \mid X_{C_{a\sim b}^*}$. The conditional independence relationship may not be really tested during the process of the PC_{fdr}^+ algorithm, but $p_{a\sim b}^*$ can still denote the value as if the conditional independence relationship was tested,

H^*: the value in ($\star$) in Algorithm 2 that is either H or $H(1 + 1/2, \ldots, +1/H)$, depending on the assumption of the dependency of the p values.

Lemma A.1. *If $\mathcal{A}_i(m), \ldots, \mathcal{A}_K(m)$ are a finite number of events whose probabilities each approach 1 as m approaches infinity*

$$\lim_{m\to\infty} P(\mathcal{A}_i(m)) = 1, \tag{A.1}$$

then the probability of the joint of all these events approaches 1 as m approaches infinity:

$$\lim_{m\to\infty} P\left(\bigcap_{i=1}^{K} \mathcal{A}_i(m)\right) = 1. \tag{A.2}$$

For the proof of this lemma, please refer to Li and Wang's [11] work.

Lemma A.2. *If there are F ($F \geq 1$) false hypotheses among H tested hypotheses and the p values of the all the false hypotheses are smaller than or equal to $(F/H^*)q$, where H^* is either H or $H(1 + 1/2, \ldots, +1/H)$ depending on the assumption of the dependency of the p values, then all the F false hypotheses will be rejected by the FDR procedure, Algorithm 2.*

For the proof of this lemma, please refer to Li and Wang's [11] work.

Proof of Theorem 2. If there is not any true edge in E'_{true}, that is, $E'_{\text{true}} = \varnothing$, then the proof is trivially $E'_{\text{true}} = \varnothing \subseteq E'$.

In the following part of the proof, we assume $E'_{\text{true}} \neq \varnothing$. For the PC_{fdr}^+ algorithm and its heuristic modification, whenever the FDR procedure, Algorithm 2, is invoked, $p_{a\sim b}^{\max}$ is always less than $\max_{C\in V\setminus\{a,b\}}\{p_{a\perp b\mid C}\}$, and the number of p values input to the FDR algorithm is always not more than $|E_{\text{test}}|$. Thus, according to Lemma 5, if

$$\max_{a\sim b\in E'_{\text{true}}}\left\{\max_{C\in V\setminus\{a,b\}}\{p_{a\perp b\mid C}\}\right\} \leq \frac{|E'_{\text{true}}|}{|E_{\text{test}}|\sum_{i=1}^{|E_{\text{test}}|}(1/i)}q, \tag{A.3}$$

then all the true connections will be recovered by the PC_{fdr}^+ algorithm and its heuristic modification.

Let $\mathcal{A}'_{a\perp b\mid C}$ denote the event

$$p_{a\perp b\mid C} \leq \frac{|E'_{\text{true}}|}{|E_{\text{test}}|\sum_{i=1}^{|E_{\text{test}}|}(1/i)}q, \tag{A.4}$$

$\mathcal{A}'_{E'_{\text{true}}}$ denote the event of (A.3), and $\mathcal{A}_{E'_{\text{true}}}$ denote the event that all the true connections in E_{test} are recovered by the PC_{fdr}^+ algorithm and its heuristic modification.

$\because$ $\mathcal{A}'_{E'_{\text{true}}}$ is a sufficient condition for $\mathcal{A}_{E'_{\text{true}}}$, according to Lemma 5.

$\therefore$ $\mathcal{A}_{E'_{\text{true}}} \supseteq \mathcal{A}'_{E'_{\text{true}}}$.

$\therefore$ $P(\mathcal{A}_{E'_{\text{true}}}) \geq P(\mathcal{A}'_{E'_{\text{true}}})$.

$\because$ $\mathcal{A}'_{E'_{\text{true}}}$ is the joint of a limited number of events as

$$\mathcal{A}'_{E'_{\text{true}}} = \bigcap_{a\sim b\in E'_{\text{true}}}\bigcap_{C\subseteq V\setminus\{a,b\}}\mathcal{A}'_{a\perp b\mid C}, \tag{A.5}$$

and $\lim_{m\to\infty}P(\mathcal{A}'_{a\perp b\mid C}) = 1$ according to Assumption (A3).

$\therefore$ According to Lemma 4, $\lim_{m\to\infty}P(\mathcal{A}'_{E'_{\text{true}}}) = 1$.

$\therefore$ $1 \geq \lim_{m\to\infty}P(\mathcal{A}_{E'_{\text{true}}}) \geq \lim_{m\to\infty}P(\mathcal{A}'_{E'_{\text{true}}}) = 1$.

$\therefore$ $\lim_{m\to\infty}P(\mathcal{A}_{E'_{\text{true}}}) = 1$. $\square$

Lemma A.3. *Given any FDR level $q > 0$, if the p value vector $P = [p_1, \ldots, p_H]$ input to Algorithm 2 is replaced with $P' = [p'_1, \ldots, p'_H]$, such that (1) for the those hypotheses that are rejected when P is the input, p'_i is equal to or less than p_i, and (2) for all the other hypotheses, p'_i can be any value between 0 and 1, then the set of rejected hypotheses when P' is the input is a superset of those rejected when P is the input.*

For the proof of this lemma, please refer to Li and Wang's [11] work.

Corollary A.4. *Given any FDR level $q > 0$, if the p value vector $P = [p_1, \ldots, p_H]$ input to Algorithm 2 is replaced with $P' = [p'_1, \ldots, p'_H]$ such that $p'_i \leq p_i$ for all $i = 1, \ldots, H$, then the set*

of rejected hypotheses when P' is the input is a superset of those rejected when P is the input.

The corollary can be easily derived from Lemma 6.

Proof of Theorem 3. Let $P_{\text{stop}} = \{p_{a \sim b}\}$ denote the value of P^{max} when the PC$_{\text{fdr}}$-skeleton algorithm stops.

$\because$ The FDR procedure is invoked whenever P^{max} is updated, and P^{max} keeps increasing as the algorithm progresses.

$\therefore$ According to Corollary 7, E'_{stop} is the same as the edges recovered by directly applying the FDR procedure to P_{stop}.

The theorem is proved through comparing the result of the PC$_{\text{fdr}}^+$ algorithm with that of applying the FDR procedure to a virtual p value set constructed from P_{stop}. The virtual p value set P^* is defined as follows.

For a vertex pair $a \sim b$ that is not adjacent in G_{true}, let $C_{a \sim b}^*$ denote a certain vertex set that d-separates a and b in the true graph and that is also a subset of either $\text{adj}(a, G_{\text{true}}) \setminus \{b\}$ or $\text{adj}(b, G_{\text{true}}) \setminus \{a\}$. Let us define $P^* = \{p_{a \sim b}^* \mid a \sim b \in E_{\text{test}}\}$ as

$$p_{a \sim b}^* = \begin{cases} p_{a \perp b \mid C_{a \sim b}^*} : & a \sim b \notin E'_{\text{true}}, \\ p_{a \sim b} : & a \sim b \in E'_{\text{true}}. \end{cases} \quad (A.6)$$

Though $p_{a \perp b \mid C_{a \sim b}^*}$ may not be actually calculated during the process of the algorithm, $p_{a \perp b \mid C_{a \sim b}^*}$ still can denote the value as if it was calculated.

Let us design a virtual algorithm, called *Algorithm**, that infers true edges in E_{test} by just applying the FDR procedure to P^*, and let E^* denote the edges in E_{test} claimed to be true by this virtual algorithm. This algorithm is virtual and impracticable because the calculation of P^* depends on the unknown E'_{true}, but this algorithm exists because E'_{true} exists.

For any vertex pair a and b that is not adjacent in G_{true}, we have the following.

$\because$ X_a and X_b are conditional independent given $X_{C_{a \sim b}^*}$.

$\therefore$ $p_{a \perp b \mid C_{a \sim b}^*}$ follows the uniform distribution on $[0, 1]$.

$\therefore$ The FDR of *Algorithm** is under q.

When all the true edges in the test set are recovered by the PC$_{\text{fdr}}^+$ algorithm, that is, $E'_{\text{true}} \subseteq E'_{\text{stop}}$, all the edges in G_{true} are included in E_{stop} due to Assumption (A4). In this case, the conditional independence between X_a and X_b given $X_{C_{a \sim b}^*}$ is tested for all the falsely recovered edges $a \sim b \in E'_{\text{stop}} \setminus E'_{\text{true}}$, because for these edges, subsets of $\text{adj}(a, G_{\text{true}}) \setminus \{b\}$ and subsets of $\text{adj}(a, G_{\text{true}}) \setminus \{b\}$ have been exhaustively searched and $C_{a \sim b}^*$ is one of them. Therefore, $p_{a \sim b} \geq p_{a \sim b}^*$ for all $a \sim b \in E'_{\text{stop}}$ when event $\mathcal{A}_{E'_{\text{true}}}$ happens. Consequently, according to Lemma 6,

$$\text{if event } \mathcal{A}_{E'_{\text{true}}} \text{ happens}, E'_{\text{stop}} \subseteq E^*. \quad (A.7)$$

Let $q(E')$ denote the *realized* FDR of reporting E' as the set of true edges in E_{test}:

$$q(E') = \begin{cases} \dfrac{|E' \setminus E'_{\text{true}}|}{|E'|} : & E' \neq \varnothing, \\ 0 : & E' = \varnothing. \end{cases} \quad (A.8)$$

The FDRs of the PC$_{\text{fdr}}^+$ algorithm and *Algorithm** are $E[q(E'_{\text{stop}})]$ and $E[q(E^*)]$, respectively. Here $E[x]$ means the expected value of x.

$\because$ $E[q(E'_{\text{stop}})] = E[q(E'_{\text{stop}}) \mid \mathcal{A}_{E'_{\text{true}}}] P(\mathcal{A}_{E'_{\text{true}}}) + E[q(E'_{\text{stop}}) \mid \overline{\mathcal{A}}_{E'_{\text{true}}}] P(\overline{\mathcal{A}}_{E'_{\text{true}}}) \leq Q + P(\overline{\mathcal{A}}_{E'_{\text{true}}})$, where $Q = E[q(E'_{\text{stop}}) \mid \mathcal{A}_{E'_{\text{true}}}] P(\mathcal{A}_{E'_{\text{true}}})$.

$\therefore$ $\limsup_{m \to \infty} E[q(E'_{\text{stop}})] \leq \limsup_{m \to \infty} Q + \limsup_{m \to \infty} P(\overline{\mathcal{A}}_{E'_{\text{true}}})$.

$\because$ $\lim_{m \to \infty} P(\mathcal{A}_{E'_{\text{true}}}) = 1$, according to Theorem 2.

$\therefore$ $\limsup_{m \to \infty} P(\overline{\mathcal{A}}_{E'_{\text{true}}}) = \lim_{m \to \infty} P(\overline{\mathcal{A}}_{E'_{\text{true}}}) = 0$.

$\therefore$ $\limsup_{m \to \infty} E[q(E'_{\text{stop}})] \leq \limsup_{m \to \infty} Q$.

$\because$ $Q \leq E[q(E'_{\text{stop}})]$.

$\therefore$ $\limsup_{m \to \infty} Q \leq \limsup_{m \to \infty} E[q(E'_{\text{stop}})]$.

$\therefore$ $\limsup_{m \to \infty} E[q(E'_{\text{stop}})] = \limsup_{m \to \infty} Q = \limsup_{m \to \infty} E[q(E'_{\text{stop}}) \mid \mathcal{A}_{E'_{\text{true}}}] P(\mathcal{A}_{E'_{\text{true}}})$.

Similarly, $\limsup_{m \to \infty} E[q(E^*)] = \limsup_{m \to \infty} E[q(E^*) \mid \mathcal{A}_{E'_{\text{true}}}] P(\mathcal{A}_{E'_{\text{true}}})$.

$\because$ Event $\mathcal{A}_{E'_{\text{true}}}$ implies $E'_{\text{true}} \subseteq E'_{\text{stop}} \subseteq E^*$.

$\therefore$ Given event $\mathcal{A}_{E'_{\text{true}}}$,

$$q(E'_{\text{stop}}) = \frac{|E'_{\text{stop}}| - |E'_{\text{true}}|}{|E'_{\text{stop}}|} = 1 - \frac{|E'_{\text{true}}|}{|E'_{\text{stop}}|}$$

$$\leq 1 - \frac{|E'_{\text{true}}|}{|E^*|} = \frac{|E^*| - |E'_{\text{true}}|}{|E^*|} = q(E^*). \quad (A.9)$$

$\therefore$ $\limsup_{m \to \infty} E[q(E'_{\text{stop}}) \mid \mathcal{A}_{E'_{\text{true}}}] P(\mathcal{A}_{E'_{\text{true}}}) \leq \limsup_{m \to \infty} E[q(E^*) \mid \mathcal{A}_{E'_{\text{true}}}] P(\mathcal{A}_{E'_{\text{true}}})$.

$\therefore$ $\limsup_{m \to \infty} E[q(E'_{\text{stop}})] \leq \limsup_{m \to \infty} E[q(E^*)]$.

$\because$ *Algorithm** controls the FDR under q.

$\therefore$ $E[q(E^*)] \leq q$.

$\therefore$ $\limsup_{m \to \infty} E[q(E^*)] \leq q$.

$\therefore$ $\limsup_{m \to \infty} E[q(E'_{\text{stop}})] \leq q$.

$\square$

B. Computational Complexity

The PC$_{\text{fdr}}^+$ algorithm spends most of its computation on performing statistical tests of conditional independence at step 7 and controlling the FDR at step 11. If the algorithm stops at the depth $d = d_{\text{max}}$, then the number of conditional independence tests required is bounded by

$$T = 2|E_{\text{test}}| \sum_{d=0}^{d_{\text{max}}} C_{\Delta-1}^d, \quad (B.1)$$

where $|E_{\text{test}}|$ is the number of edges to test, Δ is the maximum degree of graph G_{init} (the graph formed at step 1 of the PC$_{\text{fdr}}^+$ algorithm) whose edges are $E_{\text{must}} \cap E_{\text{test}}$, and $C_{\Delta-1}^d$ is the number of combinations of choosing d unordered and

distinct elements from $\Delta - 1$ elements. In the worst case that $d_{\max} = \Delta - 1$, the complexity is bounded by $2|E_{\text{test}}|2^{\Delta-1} = |E_{\text{test}}|2^{\Delta}$. The bound usually is very loose, because it assumes that no edge has been removed until $d = d_{\max}$. In real-world applications, the algorithm is very fast for sparse networks.

The computational complexity of the FDR procedure, Algorithm 2, invoked at step 11 of the PC_{fdr}^{+} algorithm, in general is $O(H \log(H) + H) = O(H \log(H))$ where $H = |E_{\text{test}}|$ is the number of input p values. The main complexity $H \log(H)$ is at the sorting (step 1). However, if it is recorded the sorted $P^{\max}$ of the previous invocation of the FDR procedure, then the complexity of keeping the updated $P^{\max}$ sorted is only $O(H)$. With this optimization, the complexity of the FDR-control procedure is $O(H \log(H))$ at its first operation and is $O(H)$ later. The FDR procedure is invoked only when $p_{a\perp b|C} > p_{a\sim b}^{\max}$. In the worst case that $p_{a\perp b|C}$ is always larger than $p_{a\sim b}^{\max}$, the complexity of the computation spent on the FDR control in total is bounded by $O(|E_{\text{test}}| \log(|E_{\text{test}}|) + T|E_{\text{test}}|)$ where T is the number of performed conditional independence tests. This is a very loose bound because it is rare that $p_{a\perp b|C}$ is always larger than $p_{a\sim b}^{\max}$.

Acknowledgment

References

[1] S. M. Smith, K. L. Miller, G. Salimi-Khorshidi et al., "Network modelling methods for FMRI," *NeuroImage*, vol. 54, no. 2, pp. 875–891, 2011.

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

[3] A. R. McIntosh and F. Gonzalez-Lima, "Structural equation modeling and its application to network analysis in functional brain imaging," *Human Brain Mapping*, vol. 2, no. 1-2, pp. 2–22, 1994.

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

[5] J. Li, Z. J. Wang, J. J. Eng, and M. J. McKeown, "Bayesian network modeling for discovering "dependent synergies" among muscles in reaching movements," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 1, pp. 298–310, 2008.

[6] Y. Benjamini and D. Yekutieli, "The control of the false discovery rate in multiple testing under dependency," *Annals of Statistics*, vol. 29, no. 4, pp. 1165–1188, 2001.

[7] J. D. Storey, "A direct approach to false discovery rates," *Journal of the Royal Statistical Society B*, vol. 64, no. 3, pp. 479–498, 2002.

[8] J. Listgarten and D. Heckerman, "Determining the number of non-spuriousarcs in a learned DAG model: investigation of a Bayesian and a frequentist approach," in *Proceedings of the 23rd Conference on Uncertainty in Artificial Intelligence*, 2007.

[9] I. Tsamardinos and L. E. Brown, "Bounding the false discovery rate in local bayesian network learning," in *Proceedings of the 23rd AAAI Conference on Artificial Intelligence (AAAI '08)*, pp. 1100–1105, July 2008.

[10] P. Spirtes, C. Glymour, and R. Scheines, *Causation, Prediction, and Search*, The MIT Press, 2001.

[11] J. Li and Z. J. Wang, "Controlling the false discovery rate of theassociation/causality structure learned with the pc algorithm," *Journal of Machine Learning Research*, vol. 10, pp. 475–514, 2009.

[12] K. J. Worsley, C. H. Liao, J. Aston et al., "A general statistical analysis for fMRI data," *NeuroImage*, vol. 15, no. 1, pp. 1–15, 2002.

[13] K. J. Friston, K. E. Stephan, T. E. Lund, A. Morcom, and S. Kiebel, "Mixed-effects and fMRI studies," *NeuroImage*, vol. 24, no. 1, pp. 244–252, 2005.

[14] K. E. Stephan, W. D. Penny, J. Daunizeau, R. J. Moran, and K. J. Friston, "Bayesian model selection for group studies," *NeuroImage*, vol. 46, no. 4, pp. 1004–1017, 2009.

[15] G. Varoquaux, A. Gramfort, J. B. Poline, and B. Thirion, "Brain covariance selection: better individual functional connectivity models using population prior," *Advances in Neural Information Processing Systems*, vol. 23, pp. 2334–2342, 2010.

[16] J. D. Ramsey, S. J. Hanson, and C. Glymour, "Multi-subject search correctly identifies causal connections and most causal directions in the DCM models of the Smith et al. simulation study," *NeuroImage*, vol. 58, no. 3, pp. 838–848, 2011.

[17] S. L. Lauritzen, *Graphical Models*, Oxford University Press, 1996.

[18] J. Neyman and E. S. Pearson, "On the use and interpretation of certaintest criteria for purposes of statistical inference: part I," *Biometrika*, vol. 20A, pp. 175–240, 1928.

[19] R. A. Fisher, "Frequency distribution of the values of the correlation 40 coefficients in samples from an indefinitely large population," *Biometrika*, vol. 10, no. 4, pp. 507–521, 1915.

[20] S. J. Palmer, B. Ng, R. Abugharbieh, L. Eigenraam, and M. J. McKeown, "Motor reserve and novel area recruitment: amplitude and spatial characteristics of compensation in Parkinson's disease," *European Journal of Neuroscience*, vol. 29, no. 11, pp. 2187–2196, 2009.