ROBUST IDENTIFICATION OF PARTIAL-CORRELATION BASED NETWORKS WITH APPLICATIONS TO CORTICAL THICKNESS DATA

D. Wheland[1], *A. Joshi*[1,2], *K. McMahon*[4], *N. Hansell*[5], *N. Martin*[5], *M. Wright*[5], *P. Thompson*[3], *D. Shattuck*[3], *R. Leahy*[1]

[1]Signal & Image Processing Inst., University of Southern California, Los Angeles, CA
[2]Brain Creativity Institute, University of Southern California, Los Angeles CA, USA
[3]Laboratory of Neuro Imaging, Dept. of Neurology, UCLA School of Medicine, Los Angeles, CA, USA
[4]Center for Magnetic Resonance, School of Psychology, University of Queensland, Brisbane, Australia
[5]Queensland Institute of Medical Research, Brisbane, Australia

ABSTRACT

Insight into brain development and organization can be gained by computing correlations between structural and functional measures in parcellated cortex. Partial correlations can often reduce ambiguity in correlation data by identifying those pairs of regions whose similarity cannot be explained by the influence of other regions with which they may both interact. Consequently a graph with edges indicating non-zero partial correlations may reveal important subnetworks obscured in the correlation data. Here we describe and investigate PC*, a graph pruning algorithm for identification of the partial correlation network in comparison to direct calculation of partial correlations from the inverse of the sample correlation matrix. We show that PC* is far more robust and illustrate its use in the study of covariation in cortical thickness in ROIs defined on a parcellated cortex.

Index Terms— brain networks, human connectome, graphical Gaussian model, partial correlation, PC algorithm

1. INTRODUCTION

Brain imaging technology enables the exploration of brain anatomy and structural and functional connectivity at increasingly high resolutions. These techniques can be used to investigate direct connectivity (through diffusion imaging of white matter tracts) and functional connectivity (through event related and resting state fMRI). Additional insight into the organization of the brain can be gained by studying how the morphology of anatomical structures covary across populations [1] and the influence of genetic factors on these covariations.

Common to all three modes of investigating large scale interactions (structural connectivity, functional connectivity, morphological covariation) is the need for computational tools to robustly analyze these direct or indirect measures of connectivity between brain regions [2].

In this paper we explore methods for computing connectivity using partial correlations. The partial correlation between two regions of interest (ROIs) is the component of the correlation between the regions that cannot be explained through correlation with other ROIs to which both regions are connected. Equivalently, the partial correlation is the correlation coefficient between signals recorded at the two ROIs after subtracting from each the regression on the signals from all other ROIs. The partial correlation therefore naturally admits a graphical network representation in which the nodes are the ROIs and edges exist only between pairs of ROIs for which the partial correlation is nonzero. This network contains connections between ROIs for which any correlation between the structural or functional similarity between them cannot be explained through indirect correlations with other ROIs.

Partial correlation is typically estimated from the concentration matrix (the inverse of the sample covariance matrix [3]), but this method is suboptimal in statistical power for sparse networks and suffers from a large sample size requirement that increases linearly with the size of the number of ROIs. This sample size requirement can be a problem when working with pathological brains for which the patient population is small or when the number of ROIs (nodes) in the network is large. We propose a new algorithm based on the PC algorithm [4] for obtaining statistical brain networks that similarly improves connectivity testing power and reduces sample size requirements by leveraging sparsity of the concentration matrix.

We choose cortical gray matter thickness as a morphological feature with which to study (indirect) brain connectivity. The biological basis for cortical thickness correlations is unknown, but it has been argued that the covariation of the morphological features, such as gray matter thickness, may result from mutually tropic influences, the contribution of genes or common experience-related plasticity and extensive learning [2]. Gray matter thickness has been implicated in many important brain phenomena, such as aging and Alzheimer's disease, and studies have indicated that intercorrelated regions of morphological features may be a part of a functional, neuroanatomically interconnected system [5]. Furthermore, significant correspondence between anatomical connectivity and functional connectivity of functional data has been shown [6].

In this study, we use a morphometry-based connection concept with cortical gray matter thickness derived from 645 automatically parcellated cortical volumes to study the underlying architecture of the brain network. The connectivity matrices from the concentration method of calculating partial correlation and a modified version of the PC algorithm are discretized to adjacency matrices at different sparsity levels and compared in terms of stability on the thickness data. We believe this is the first time that a PC-like algorithm has been applied to structural

This work is supported by NIH grants R01 NS074980 and P41 RR01364

brain data and show that it outperforms the standard concentration method.

2. GRAPHICAL MODELS

A graphical model combines a multivariate distribution with a graphical representation that encodes the conditional independence structure between the random variables characterized by the distribution [3]. The random variables are depicted as nodes or vertices in a graph, and their conditional dependence structure is encoded through their edges where a lack of an edge between two nodes denotes conditional independence. In this way, graphs are a useful tool for visualizing and investigating conditional independence structure between a collection of random variables, or the processes they represent. Graphs with directed edges, such as the directed acyclic graphs used in Bayesian networks [4], convey a notion of causality, but here we are not concerned with causality and restrict ourselves to undirected graphs.

An undirected graph, $G = (V, E)$, is represented by a set of vertices, $V = \{V_1, V_2, \ldots, V_P\}$ and a set of edges E, each represented by by unordered pairs of vertices, where V_i corresponds to the i-th element of P-length random vector $X = (X_1, X_2, \ldots, X_P)$, which we assume here to follow a multivariate distribution P_X. $\{V_i, V_j\} \in E$ means that V_i and V_j are adjacent, denoted $V_i \sim_G V_j$. The edges of the graph convey information by encoding a subset of the conditional independence relationships implied by P_X according to the following properties:

Global Markov Property - Let V_A, V_B, and V_S be disjoint subsets of V. If V_S is a *separator* of V_A and V_B in G (all paths between any vertex in V_A and any vertex in V_B pass through at least one vertex in V_S), then the random variables corresponding to set A are conditionally independent from those corresponding to set B given S [3]. That is, $V_A \perp_G V_B \mid V_S \Longrightarrow X_A \perp\!\!\!\perp_{P_X} X_B \mid X_S$ where $\perp_G$ and $\perp\!\!\!\perp_{P_X}$ denote separation in G and conditional independence with respect to P_X respectively.

Faithfulness - If the converse of the global Markov property is *also* true such that all conditional independence relationships are represented by the graph ($V_A \perp_G V_B \mid V_S \Longleftrightarrow X_A \perp\!\!\!\perp_{P_X} X_B \mid X_S$), then P_X and G are said to be faithful to one another and G is termed a perfect map of P_X [4].

Faithfulness greatly increases the benefit one gets from the graph by ensuring that *all* independence information is encoded in the graph. In this paper we assume that our thickness data has a MVN distribution: $P_X = N_P(\mu, \Sigma)$, for which faithfulness is guaranteed with probability 1 [7].

3. CONCENTRATION MATRIX THRESHOLDING

To find the graph G associated with data drawn from the MVN distribution P_X, we need to find the set of edges that define the graph. Because zero partial correlation between two variables is equivalent to their conditional independence under the MVN distribution [3], the edges are given by the non-zero partial correlations between each pair of random variables (or nodes in the graph). The simplest approach is to compute partial correlations directly from the concentration matrix: $\rho_{ij|\text{rest}} = \frac{-k_{ij}}{\sqrt{k_{ii}k_{jj}}}$ where k_{ij} denotes element (i, j) of the concentration matrix, K, and conditioning on "rest" means conditioning on all of the variables in X except X_i and X_j [3]. Thus, $k_{ij} = 0 \Longleftrightarrow X_i \perp\!\!\!\perp_{P_X} X_j \mid \text{rest}$, and the output adjacency matrix is zero when k_{ij} is zero and one otherwise. As noted above, for the MVN distribution, the corresponding graph is faithful to P_X with probability 1.

In practice, the true concentration matrix is unknown, but can be estimated by taking the inverse of the sample covariance matrix. However, invertibility of the covariance matrix requires that the number of observations, N, be greater than the number of variables plus 1: $\text{N} \geq P + 1$. Satisfying this condition can be difficult when the number of nodes (ROIs) is large and the population size is small.

4. PC* ALGORITHM

The PC* algorithm presented in this paper is a variation of part of the PC algorithm (named after its authors' first initials; [4]) that optimizes the algorithm's speed for undirected graphs. To facilitate the following description of the algorithm, we assume that all partial correlations are known a priori. The PC* algorithm assumes P_X is faithful to some undirected, target graph $G = (V, E)$ and starts with a completely connected undirected graph $\hat{G} = (V, \hat{E})$ and prunes its edges to reach G. Edges are pruned by testing for independence between adjacent vertices in $\hat{G}$ conditioned on sets that seek to separate the vertices in G. For a graph satisfying the faithfulness property, the separation of two vertices by a set is equivalent to zero correlation conditioned on that set. Therefore, when zero partial correlation occurs for any conditioning set, that set is a separator of G, implying that the corresponding edge cannot be in E, and so the PC* algorithm removes it from $\hat{E}$. The conditioning set size starts at 0 and is increased until it is guaranteed that every false edge ($\{e \mid e \in \hat{E}, e \notin E\}$) has been conditioned on a separator and in turn removed. The conditioning sets of size m for testing the edge between arbitrary vertices V_A and V_B is every subset of $\text{adj}(V_A, \hat{G}) - V_B$ of size m where $\text{adj}(V_k, G) = \{V_l \mid V_l \sim_G V_k\}$ (the set of immediate neighbors of V_A in G). Thus, as m increases, if $\{V_A, V_B\} \notin E$, one of the conditioning sets is guaranteed to separate V_A and V_B in G. Of course, in practice the partial correlations are not known a priori, but conditional independence can be determined using their estimates to perform hypothesis tests. In particular we transform the correlation coefficients to pivotal Student's t-scores [8].

The advantage of PC*'s conditioning scheme relates to the estimates of the partial correlations and is two-fold. 1.) It is more accurate to condition on smaller separating subsets when dealing with realizations of random variables because extra conditioning variables increase the variance of the estimated partial correlations. 2.) For each variable removed from a conditioning subset, one less sample is required. Whereas the concentration method requires at least P samples for mean-zero data, the PC algorithm requires only the cardinality of the separator plus 2, which can be substantially less for sparse

graphs.

Following Li and Wang [9], we incorporate the Benjamini-Hochberg procedure [10] for controlling the fase discovery rate (FDR) into PC*. Also, whereas the original PC algorithm infers directed acyclic graphs under the assumption that the underlying distribution is faithful to such a graph and when testing an edge must condition on the neighbors of both corresponding vertices, PC* assumes faithfulness to an undirected graph and need only condition on the neighbors of one of the vertices to infer an undirected graph in approximately half the time. Algorithm 1 gives pseudocode for PC*.

Algorithm 1 PC* Algorithm

Require: the target false discovery rate, q, and an N×P data matrix, X, comprised of N observations $\sim N_P(\mu, \Sigma)$ faithful to a CIG $G = (V, E)$.
Ensure: the estimate of the CIG, $\hat{G} = (V, \hat{E})$

$\hat{G} = (V, \hat{E}) \leftarrow$ complete graph of P vertices
p_max (matrix storing the maximum estimated p-value for each possible edge) $\leftarrow -1$
$m \leftarrow 0$
repeat
  for each edge $e = \{V_i, V_j\} \in \hat{E}$ **do**
    if $|\text{adj}(V_i, \hat{G})| \leq |\text{adj}(V_j, \hat{G})|$ **then**
      neighbors $\leftarrow \text{adj}(V_i, \hat{G})$
    else
      neighbors $\leftarrow \text{adj}(V_j, \hat{G})$
    end if
    for each conditioning set $S \subseteq \text{neighbors}\backslash\{V_i, V_j\}$ such that $|S| = m$ **while** $e \in \hat{E}$ **do**
      Calculate $\rho_{ij|S}$ from X and transform to a t-score to get p-value $p_{ij|S}$.
      if $p_{ij|S} >$ p_max for edge $\{V_i, V_j\}$ **then**
        p_max for edge $\{V_i, V_j\} \leftarrow p_{ij|S}$
        if All edges have been tested at least once (p_max ≥ 0 for all edges) **then**
          p_edges_max $\leftarrow$ the p-values from p_max of only the edges that have not been removed from $\hat{G}$
          rem_edges $\leftarrow$ FDR_BH(p_edges_max,q) (perform Benjamini-Hochberg procedure to determine edges to remove.)
          $\hat{E} \leftarrow \hat{E} -$ rem_edges
        end if
      end if
    end for
  end for
  m $\leftarrow$ m+1
until $|\text{adj}(V_i, \hat{G})\backslash V_i| < m$ for all $V_i \in V$

5. DATA COLLECTION AND PREPROCESSING

Our empirical data consisted of 3D structural brain MRI scans of 668 normal, right handed subjects (age range: 22-25 years). The scans were collected using a 4 Tesla Bruker Medspec whole body scanner (Bruker Medical, Ettlingen, Germany) at the Center for Magnetic Resonance (University of Queensland, Australia). Three-dimensional T1-weighted images were acquired with a magnetization prepared rapid gradient echo (MP-RAGE) sequence to resolve anatomy at high resolution. Acquisition parameters were: inversion time (TI) /repetition time (TR) /echo time (TE) = 1500 / 2500 / 3.83 msec; flip angle = 15 degree; slice thickness = 0.9mm with a $256 \times 256 \times 256$ acquisition matrix. BrainSuite's automated processing pipeline [11, 12] was used for automatic skull stripping, tissue classification, surface extraction, gray matter thickness estimation and cortical and subcortical parcellation of the data. Figure 1 shows the lateral and medial views of the template's left cortical surface after being parcellated and color coded into the 25 ROIs used in this paper.

For each subject, the multiple gray matter thickness measurements in each of the parcellated cortical regions were averaged to form a summary statistic for each region. The log transform of cortical thickness data was taken such that the resulting data approximately follows the MVN distribution [13].

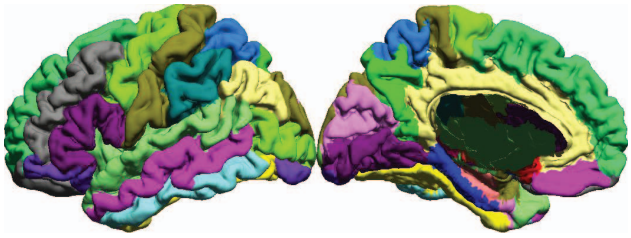


Fig. 1: Lateral and medial views of the 25 parcellated cortical regions per hemisphere displayed on the template's left hemisphere

	Sample Size = 645		Sample Size = 161	
Method	SE	AEV(Std)	SE	AEV(Std)
Concentration	83	0.059(0.069)	24	0.068(0.048)
PC*	111	0.042(0.068)	54	0.049(0.063)

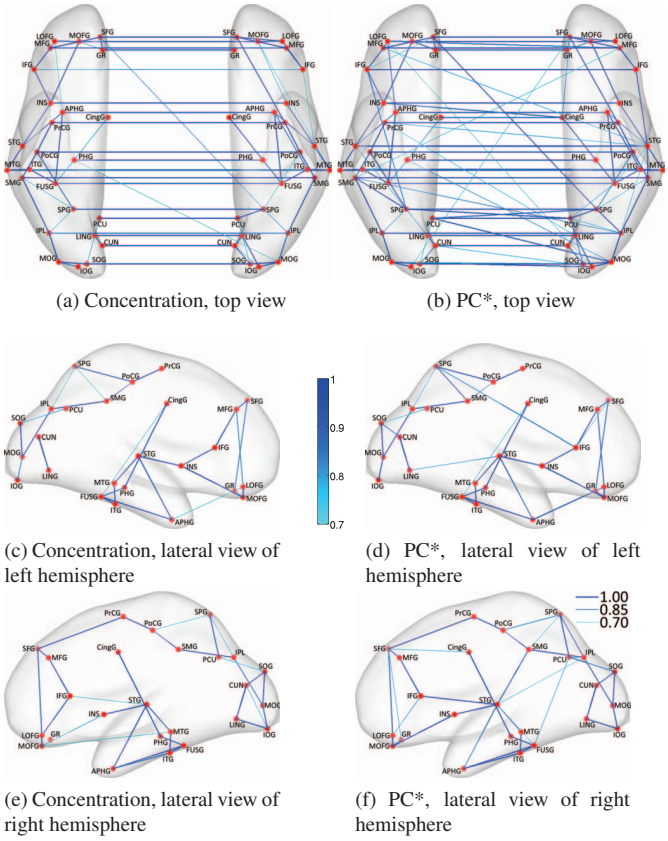
Table 1: The number of stable edges (SE), average edge variance (AEV) taken over all possible edges, and edge variance standard deviation (Std) also taken over all possible edges for each method and sample size.

Any subjects with a log-transformed thickness outside 3 standard deviations with respect to a parcellated region were rejected, resulting in a final 645 subject by 50 region data matrix.

6. SAMPLING STABILITY

Bootstrapping was performed for both the concentration-matrix and the PC* method to estimate the edge variances and occurrence probabilities based on the thickness dataset's empirical distribution [14]. Bootstrapping was run twice, 500 trials in each run: once randomly sampling 645 subjects at a time to estimate the stability for the entire dataset and a second time randomly sampling 161 subjects at a time to illustrate stability when fewer samples are available. For each trial, PC*'s FDR was set to 15% and the concentration method's threshold was chosen to match the number of edges given by PC*. An edge was deemed stable if it occurred for at least 70% of the bootstrap trials, and Table 1 summarizes the results. For sample size = 645, Figure 2 shows the stable network for both methods formed solely from stable edges.

In another comparison of method stability under the empirical distribution, we started by running both methods on the full-sample dataset with parameters such that each method found 159 edges (the number of edges found by PC* with FDR = 15% on the full dataset) and took the resulting graphs as ground truths for their respective methods. For 15 trials, the graphs were regenerated on resampled datasets of various sizes, and for each method, the Dice coefficients were calculated between these graphs and their corresponding ground truth graph. (The Dice coefficient between a true edge set E and a method's estimated edge set $\hat{E}$ is $\text{dice}(E, \hat{E}) = \frac{2|E \cap \hat{E}|}{|E| + |\hat{E}|}$ and ranges between 0 when the two sets have no edges in common and 1 when $E = \hat{E}$.) Tuning PC*'s FDR to achieve the target 159 edges proved overly difficult for the 3 smallest sample sizes (20, 15, and 10) which failed to meet the target by 41.1, 55.5, and 100 edges on average respectively. The mean Dice coefficient for each method is plotted below as a function of sample size in Figure 3.



(a) Concentration, top view

(b) PC*, top view

(c) Concentration, lateral view of left hemisphere

(d) PC*, lateral view of left hemisphere

(e) Concentration, lateral view of right hemisphere

(f) PC*, lateral view of right hemisphere

Fig. 2: 3 views of the stable, bootstrapped graphs for both methods with sample size = 645. Edge darkness and thickness increase with edge stability. In the top view, left and right hemispheres in the figures correspond to anatomical left and right hemispheres respectively. In lateral views only intrahemispheric connections are shown.

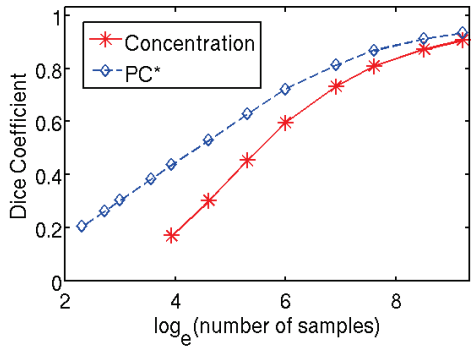


Fig. 3: A comparison of each method's consistency with their full-sample result under the thickness dataset's empirical distribution for various sample sizes. Similarity is measured by Dice coefficient over 15 trials for each sample size. Note: the concentration method's curve is undefined for sample sizes less than 51.

7. DISCUSSION

The bootstrap results show that PC* is significantly more stable than the standard concentration method, particularly as the sample size decreases. Table 1 reports that PC* finds significantly more stable edges and has less edge variance than the concentration method. For example, for sample size = 161, PC* finds more than twice as many stable edges and has approximately 72% of the average edge variance as the concentration method. Furthermore, Figure 3 shows that the graphs returned by PC* are much more consistent than those returned by the concentration method. At a sample size of 51, PC* has a Dice coefficient approximately equal to the Dice coefficient for the concentration method at a sample size of 200.

The stable PC* and concentration networks shown in Figure 2 are consistent with previous findings and largely similar with some notable differences. The long range interhemispheric connections between homologous regions present in both networks are expected and lend credibility to the results. The intrahemispheric connections are largely identical between the two methods with a few exceptions and the caveat that in general they are stronger in the PC* network. Curiously, the extra stable edges found by the PC* method largely manifest as long range, interhemispheric connections to non-homologous regions, of which the concentration network has comparatively very few.

8. REFERENCES

[1] J.P. Lerch et al., "Mapping anatomical correlations across cerebral cortex (MACACC) using cortical thickness from MRI," *Neuroimage*, vol. 31, no. 3, pp. 993–1003, 2006.

[2] Sporns et al., "The human connectome: a structural description of the human brain," *PLoS Comput Biol*, vol. 1, no. 4, pp. e42, 2005.

[3] J. Whittaker, *Graphical models in applied multivariate statistics*, vol. 16, Wiley New York, 1990.

[4] P. Spirtes et al., *Causation, prediction, and search*, vol. 81, The MIT Press, 2000.

[5] Yong He et al., "Small-world anatomical networks in the human brain revealed by cortical thickness from MRI," *Cerebral Cortex*, vol. 17, no. 10, pp. 2407–2419, 2007.

[6] Honey et al., "Network structure of cerebral cortex shapes functional connectivity on multiple time scales," *Proceedings of the National Academy of Sciences*, vol. 104, no. 24, pp. 10240–10245, 2007.

[7] C. Meek, "Strong-completeness and faithfulness in belief networks," Tech. Rep., Department of Philosophy, Carnegie Mellon University, Paper 205.

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

[9] Li et al., "Controlling the false discovery rate of the association/causality structure learned with the PC algorithm," *The Journal of Machine Learning Research*, vol. 10, pp. 475–514, 2009.

[10] Benjamini et al., "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal Statatistical Society. Series B (Methodological)*, pp. 289–300, 1995.

[11] A. Joshi et al., "A framework for brain registration via simultaneous surface and volume flow," in *Information Processing in Medical Imaging*. Springer, 2009, pp. 576–588.

[12] D.W. Shattuck et al., "BrainSuite: an automated cortical surface identification tool," *Medical Image Analysis*, vol. 6, no. 2, pp. 129–142, 2002.

[13] A.A. Joshi et al., "Bayesian approach for network modeling of brain structural features," in *Proc. of SPIE*, 2010, vol. 7626, p. 762607.

[14] B. Efron et al., *An introduction to the bootstrap*, vol. 57, Chapman & Hall/CRC, 1993.