Ming Hsieh Department of Electrical Engineering

USC Viterbi School of Engineering

Robust Identification of Partial-correlation Based Networks with Applications to Cortical Thickness Data

David Wheland[1], Anand Joshi[1], Richard Leahy[1]

Contribution

- Explore brain connectivity through morphological covariation of cortical thickness data
- Propose modified PC[2] algorithm, PC*, towards this end
- Compare performance of PC* against established concentration matrix method for calculating partial correlations

Graphical Models

- Combines a multivariate distribution (in this case corresponding to regions of interest (ROIs) on the cortical surface) with a graphical representation
  - Direct dependence between random variables is denoted by an edge between the corresponding nodes in the graph as follows
- **Faithfulness property** implies for any nodes A and B in the graph, all paths between A and B contain at least one node in control set Z if and only if X_A is conditionally uncorrelated with X_B given X_Z, assuming a multivariate normal distribution for the X's where X_A is the random variable corresponding to node A.
  - True for randomly selected positive distribution with probability 1.

Concentration Method

- Partial correlation between two ROIs A and B controlling for ROI Z, denoted $\rho_{AB|Z}$, is the component of the correlation between A and B that cannot be explained through correlation with Z.
- Let Σ denote the ROI covariance matrix and $K = \Sigma^{-1}$, then $\rho_{AB|\text{rest}} = \frac{k_{ij}}{\sqrt{k_{ii}k_{jj}}}$ where rest denotes all other ROIs.

PC* Algorithm

- Uses assumed faithfulness property to identify control sets Z which are equivalent to controlling for "rest" but contain fewer elements.
- For example, if X_A is independent of X_B given X_Z then X_A is independent of X_B given any superset of X_Z
- Controlling for fewer elements increases statistical power of partial correlation based independence tests and reduces sample size requirements

References

[1] Signal and Image Processing Inst., University of Southern California
[2] P. Spirtes et al., *Causation, prediction, and search*, The MIT Press, 2000.
[3] D.W. Shattuck et al., "BrainSuite: an automated cortical surface identification tool," *Medical Image Analysis*, vol. 6, no. 2, 2002.
[4] A. Joshi et al., "Bayesian approach for network modeling of brain structural features," in *Proc. of SPIE*, 2010.

Contact: wheland@usc.edu
This work is supported by NIH grants R01 NS074980 and P41 RR01364

Dataset and Preprocessing

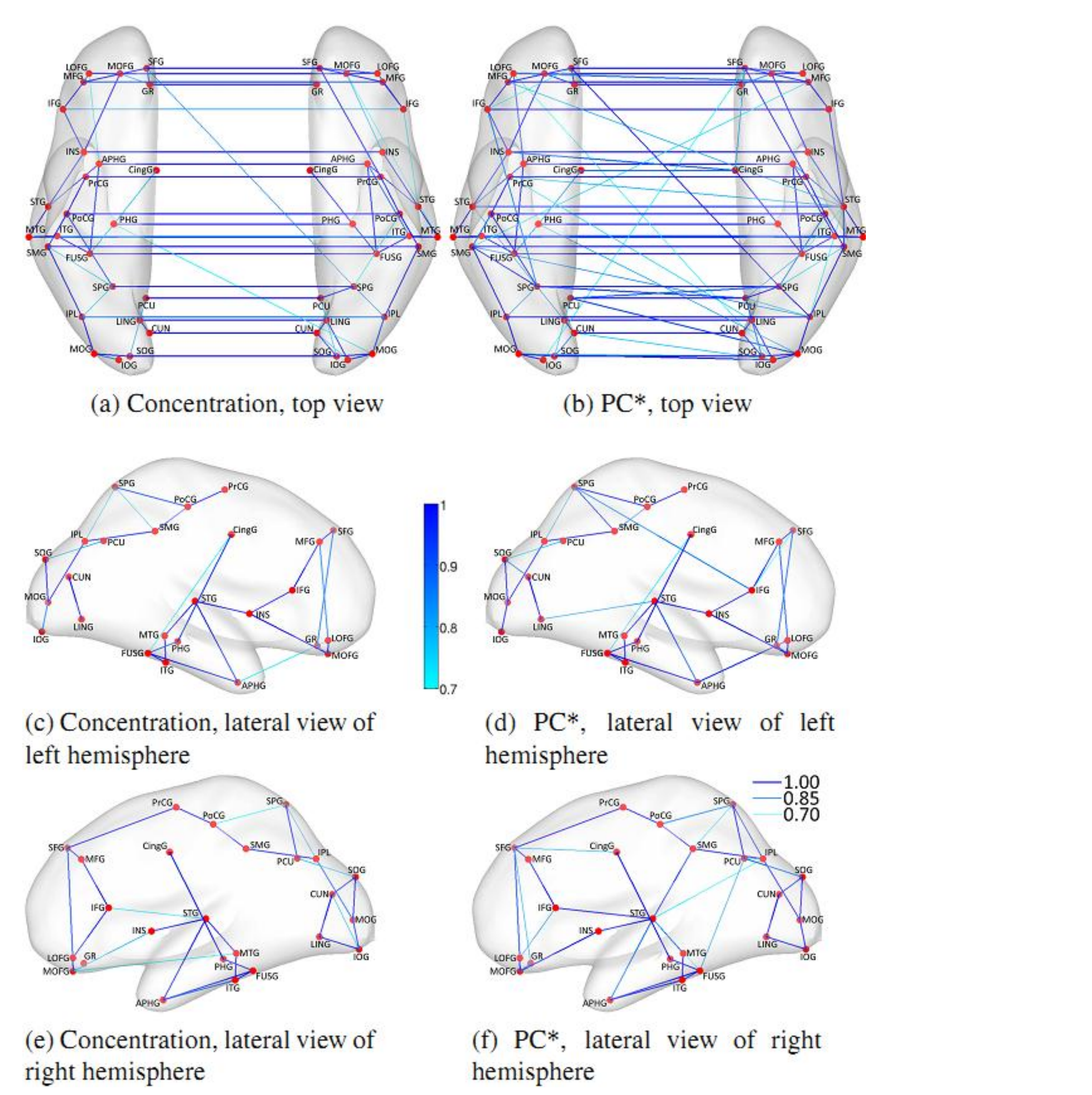
- 668 normal, right handed subjects.
- Three-dimensional T1-weighted MRI volumes acquired using a 4 Tesla scanner running MP-RAGE
- Automated post processing by BrainSuite[3]: from MRI volume to gray matter thickness on extracted and ROI-parcellated surface.
- For each subject averaged gray matter thickness over each ROI.
- Log transform taken such that resulting data approximately follows a normal distribution[4]

Results

- Bootstrapped data to estimate edge stability. 2 Cases:
  - 500 trials, randomly sampling 645 subjects
    500 trials, randomly sampling 161 subjects

	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)

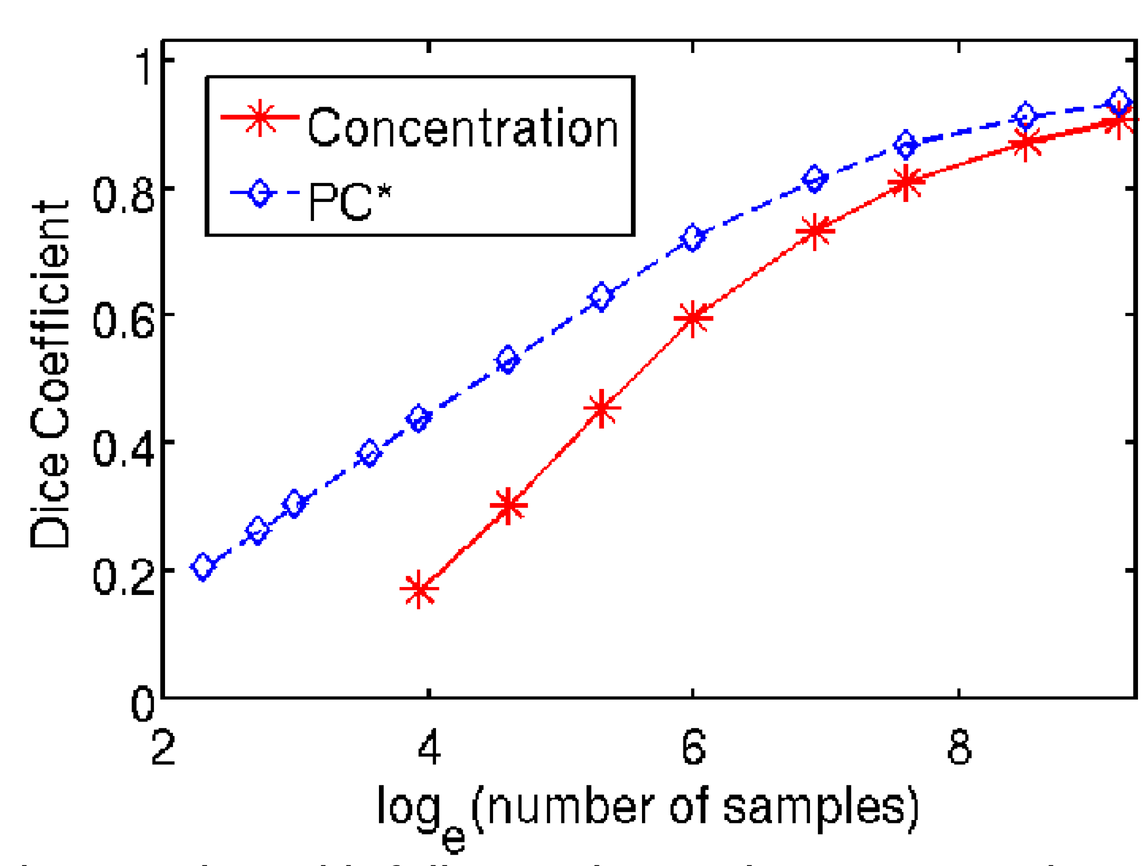
The number of stable edges (SE), average edge variance (AEV) and edge variance standard deviation (Std). An edge was deemed stable if it occurred in at least 70% of the bootstrap trials.



(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

3 views of each method's stable networks formed solely from stable edges.

- Ran each method on the full sample dataset with parameters tuned to find 159 edges and set result as the ground truth for the method.
  - Reran methods on reduced samples of dataset and evaluated performance based on percent edge overlap with full-sample result.



Percent edge overlap with full-sample result versus number of samples