Structural Equation Modeling of Neuroimaging Data: Exhaustive Search and Markov Chain Monte Carlo

S. Hanson†, T. Matsuka†, C. Hanson†, D. Rebbechi†, Y. Halchenko†, A. Zaimi & B. Pearlmutter‡

† Psychology Department, Rutgers-Newark, NJ 07102, USA ‡ Hamilton Institute, NUI Maynooth, Ireland

Graphical Models and Brain Interactivity

Graphical linear statistical modeling, such as path analysis and structural equation modeling (Jennings, McIntosh, Kapur,1998; McIntosh and Gonzalez-Lima, 1994), have been used with some success to investigate the interactivity of brain areas that are recruited during a specific cognitive or perceptual task. As the neuroimaging field matures, theories of cognitive neuroscience naturally involve hypothesis about interactions between brain areas, hence our ability to model underlying networks depends critically on detecting larger data structures (e.g. networks) than local regions of interest.

Nonetheless, past work in network detection has had the following problems:

1. Confirmatory not Exploratory: given high uncertainty and noise in Neuroimaging data, Exploratory methods may be more desirable
2. Lack of sensitivity in alternative hypotheses (as graph space increases; see table)
3. Interpreting Graph interactivity despite slowness of hemodynamic response
4. Greedy Search methods that are slow and can be biased (Bullmore et al, 1999).

BAD NEWS: GRAPH space increases faster than exponential. Graphs of interest in Cognitive Neuroscience: Oriented (direction) Labeled (unique) (ROI) pairwise acyclic (covariance): Number of graphs = $3^{(n*n-1)/2}$ where edges can be {+1,0,-1} hence:

Number of Nodes	Number of Graphs
2	3
3	27
4	729
5	59049
6	143448907
**7	10460353203
8	2287679000000

*Six nodes produces a number of graphs takes 1 day on a 14 node OPTERON Cluster where 1M graphs are searched per node. Seven ** Nodes would take approximately 2 years, while 8 Nodes would take roughly 4300 years (43 centuries) assuming linear computational time with graph number. This is particularly disturbing for Cognitive Neuroscience, in which theories of working memory (Baddely, 1987), meta-memory (Johnson, 2004) category learning (Poldrack & Gluck 2002), language processing, and executive control and many other cognitive or perceptual process may involve a small number of areas of the brain that could easily exceed SEVEN ROIs.*

New methods for Graphical Modeling of Brain Interactivity

This new approach is based on a simple observation that most of these graphs up to SEVEN are relatively easy to EXHAUSTIVELY search. In those cases we do the following:

1 Nodes of the graphs are first located using GLM or some other noise abatement methods

2 Region location-clustering in individuals to maximize brain interactivity interpretation and Clustering for region analysis (see Poline et al, 2004, for K-Means region mapping).

3 Determine time series per ROI. Three possibilities compared here:

- m by m by m smoothed voxels over time (reduce noise, increase signal consistency)
- **EIGEN-Time Series:** do PCA on V by V voxel space over time producing an eigenvector over time for each ROI
- **MDS-Time Series:** do MDS on V by V voxel space computing distances over time points of each time series from the all others (more similarity based-less linear)

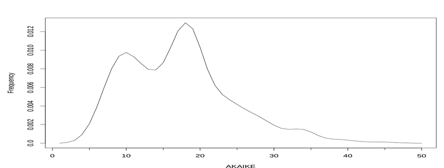
4 Project Voxels in reduced Projection space (5x5x5 smoothing)-drop 10-20% of variance (reducing noise and increasing correlation structure in time series). Pick dimensions with highest variance from MDS.

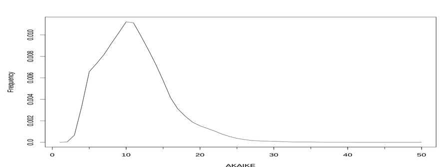
5 Compute COVARIANCE/CORRELATION between EIGEN-Time Series or MDS Time Series (note in PCA case composite score may not be invariant under rotation)

6 EXHAUSTIVE SEARCH. Fit ALL N-node ORIENTED LABELED graphs to covariance-LISREL. However we do not fit any saturated graphs, i.e. those that are guaranteed to be overparameterized

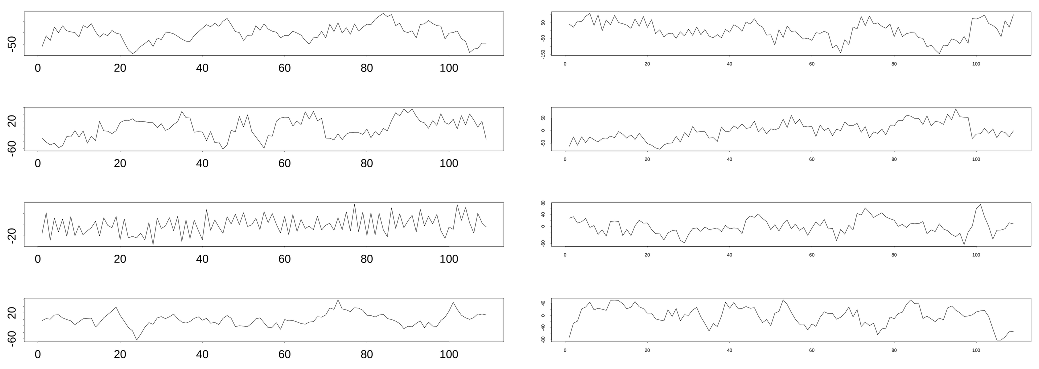
7 Sort and find distribution of goodness of fit over all graphs; Should be SINGLE peaked and + skewed (see below)

8 Select best graph (AIC), or if there is a cluster of equivalent graphs pick composite of best graph (voting).


House


Study

For graphs larger (some) N (based on today's computational resources, see table before) use focused sampling or Monte Carlo Markov Chain.

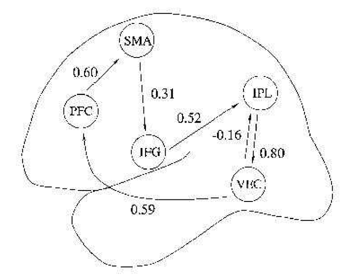


ROI Timeseries for PCA, MDS

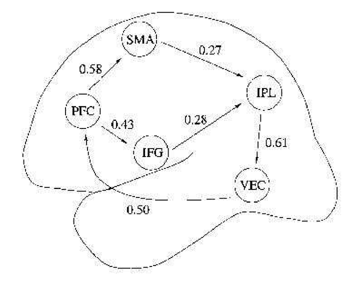
Graphs we can Exhaustively Search

Working Memory 5 (59049 Graphs) Nodes

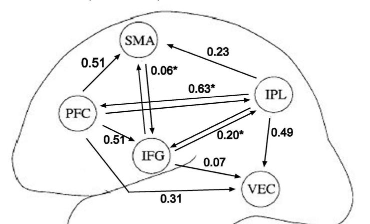
Bullmore et al (2000) and Working Memory Graphs: Directed Search in Graph Space: Greedy method, poor tradeoff in breadth vs depth. Comparison of Bullmore et al with Exhaustive Search (59049) Best graph in terms of AIC or P is neither close to graph (II) found by Bullmore nor close to theoretical graph (I). VEC: ventral extrastriate cortex; PFC: prefrontal cortex, SMA: Supplementary motor area; IFG: Inferior frontal gyrus; Inferior parietal lobule (ILP)



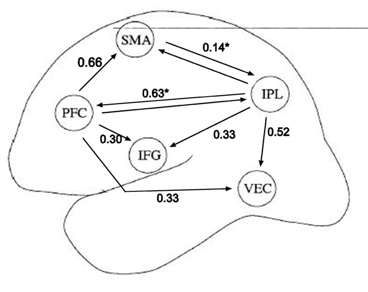
I: theoretical model of working memory Bullmore et al (2000)



II: Best graph found by Bullmore et al. (2000)

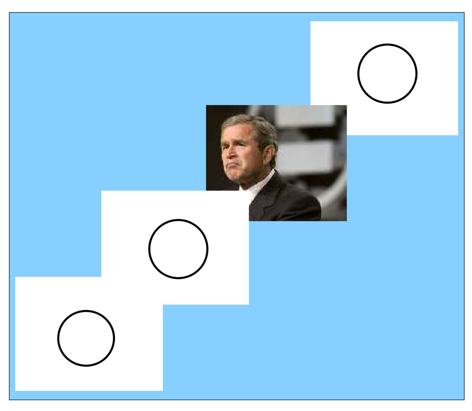


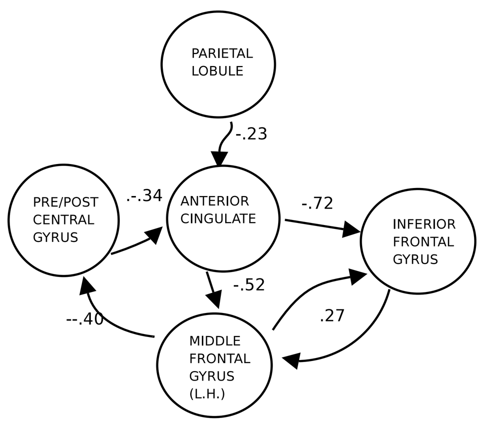
IV: Best possible search based on exhaustive search and P value * reciprocal direction fitted as a single degree of freedom, after examining best graphs that are contained edges in both directions



III: Best possible search based on exhaustive search and AIC

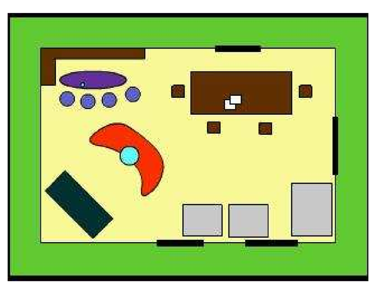
Oddball Tasks 5 (59049) Nodes





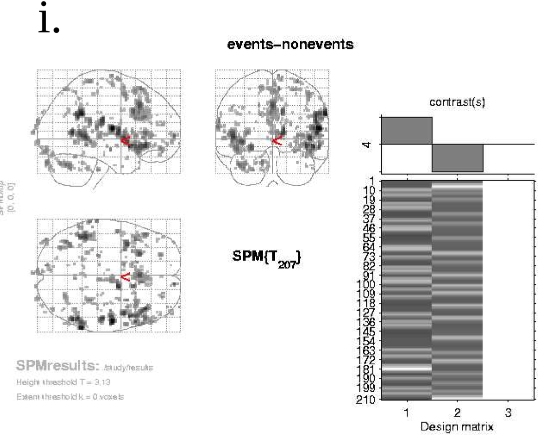
GLM was used to identify typical areas in a visual oddball task; subjects were fixated and asked to push a button on each trial when they saw an oddball (see left figure) and covariances were fit with all possible graphs-the best AIC on the right.

Event Perception 4 (728) and 6 (143448907) Nodes

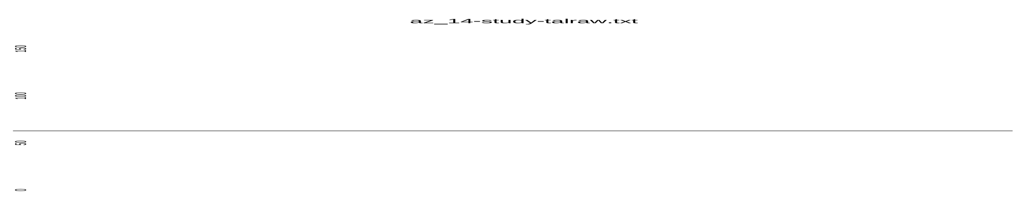


i.

ii.

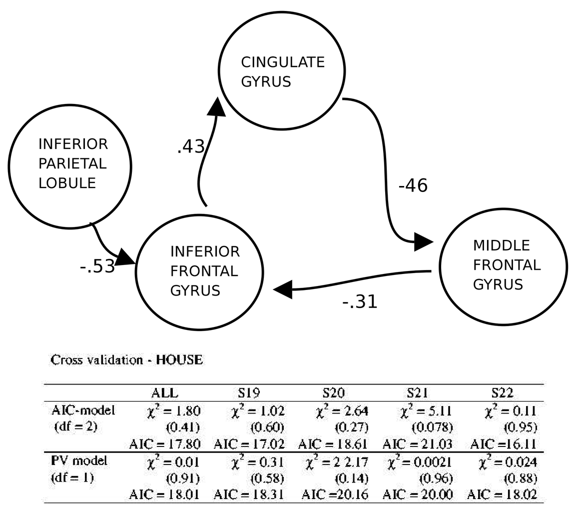


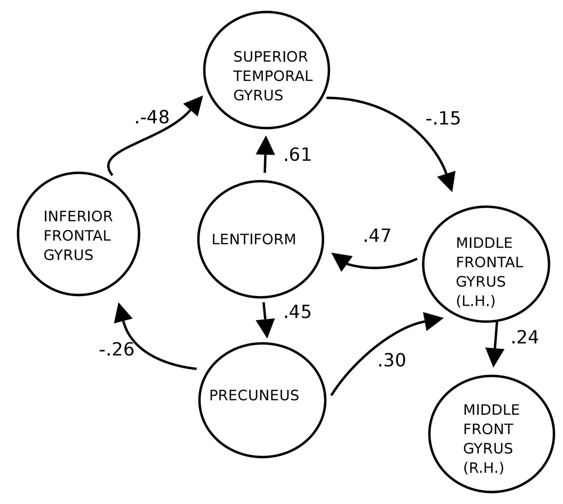
iii.



iv.

Examples of movies used in task (i) schema-poor and (ii) schema-rich. Thresholded brain activity of events change points vs background (iii) and cluster analysis of brain activity (iv) for each subject; four common areas are found over all (8) subject for all cluster analyses.





Left Panel: House stimulus Graph resulting from searching 728 possible graphs with highest probability and lowest AIC. Note ROIs associated with "alerting attentional" functions working memory and spatial, sequential ordering. Note similarity to Oddball graph.

Right Panel: The study stimulus graph. Note increased complexity of graph/nodes and overlap of areas. This is based on 14.3M graphs; ≈ 2days on the OPTERON cluster.

What do you do when it takes 43 Centuries?

Methods of Search when N≥7: MCMC

Search in large graph spaces have been explored in the NIPS (Neural Information Processing) Community for more than 10 years; MCMC was one of those search methods. The purpose of MCMC is to find a mean graph model defined as $\bar{\mathbf{G}} = \sum_{\mathbf{G}} p(\mathbf{G}|\mathbf{D})\mathbf{G}$. Such model averaging is known (Lee, 1999) to produce more reliable and stable results comparing to the single chosen best model. This graph $\bar{\mathbf{G}}$ is defined by a set of edges whenever $\mathbf{G}_{ij} = 1$, thus $\bar{\mathbf{G}}$ represents the probability of edge ij being present in the graph. For the graph averaging we use posterior probability $p(\mathbf{G}|\mathbf{D})$ as defined in (Burnham & Anderson, 2002). Such definition uses Akaike Information Criterion (AIC) in order to define pdf

$$p(\mathbf{G}|\mathbf{D}) \approx \frac{e^{-1/2\Delta AIC(\mathbf{G},\mathbf{D})}}{\sum_{\mathbf{G}} e^{-1/2\Delta AIC(\mathbf{G},\mathbf{D})}},$$

where $\Delta AIC(\mathbf{G},\mathbf{D})$ is the difference between $AIC(\mathbf{G},\mathbf{D})$ and minimal AIC for any graph $\mathbf{G}$ given our data $\mathbf{D}$.

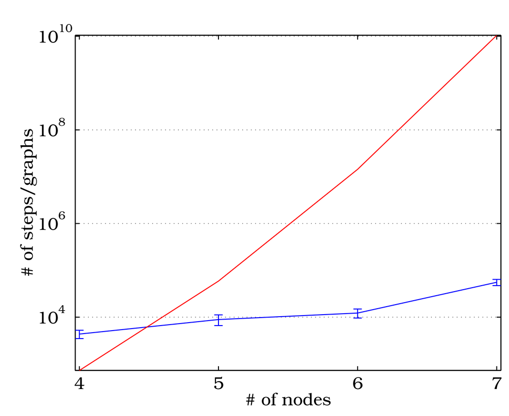
Such pdf is based on an exponential prior on the complexity of graphs:

$$\log p(\mathbf{G}|\mathbf{D}) \approx -1/2\Delta AIC(\mathbf{G},\mathbf{D}) - const = -\chi^2 - |\mathbf{G}| + const$$

MCMC sampling stops once a stable solution is found. In order to decide whether the solution is stable we run multiple MCMCs in parallel and stop whenever their solutions are similar, i.e. maximal standard deviation from estimated mean for each parameter across all MCMC runs doesn't exceed 0.04 which gives 95% confidence that none of the parameters estimation errors exceeds 0.08.

To test the algorithm we used covariance data from the previous cases (4,5,6 nodes) and one simulated 7 node (≈10 Billion graphs) in which the covariance and graph were constructed. The algorithm estimated mean/average graph which describes a probability of any edge being present in the original structure. Further thresholding of such probability matrix lead to the most probable graph structure in all cases. As shown in the figure on the right accurate graph estimates were made with only super-linear graph samples.

```
N                    # number of graph nodes
K=20                 # number of independent MCMC processes
DEVTHRESH=0.04       # maximal deviation threshold to stop
MCMCs

initialize K independent MCMC processes with random
graphs
initialize K avgmodel's  Ḡ  with 0s
initialize minAIC=2*N*(N+1)/2

N=0
while ( max(deviation( Ḡ )) > DEVTHRESH ) do
  foreach k-th MCMC process do
    G*= Gk with random edge i->j changed randomly to 1 of
          3 states: non-present, i->j, j->i

    run LISREL on G*

    if AIC(G*) < minAIC
    then
        minAIC = 0.99 * AIC(G*)
        re-initialize K avgmodel's  Ḡ
        N = 0
        restart 'while loop' with new minAIC

    u = uniform random from [0,1] interval

    if  u≤{1, p(G)/p(Gk)}
    then
        Gk = G*

    Ḡk = Ḡk + Gk
  N=N+1
Ḡk = Ḡk / N
```



Pseudocode of MCMC Algorithm on the left made possible to get accurate graph estimates with only super linear graph sample

References

http://www.rumba.rutgers.edu