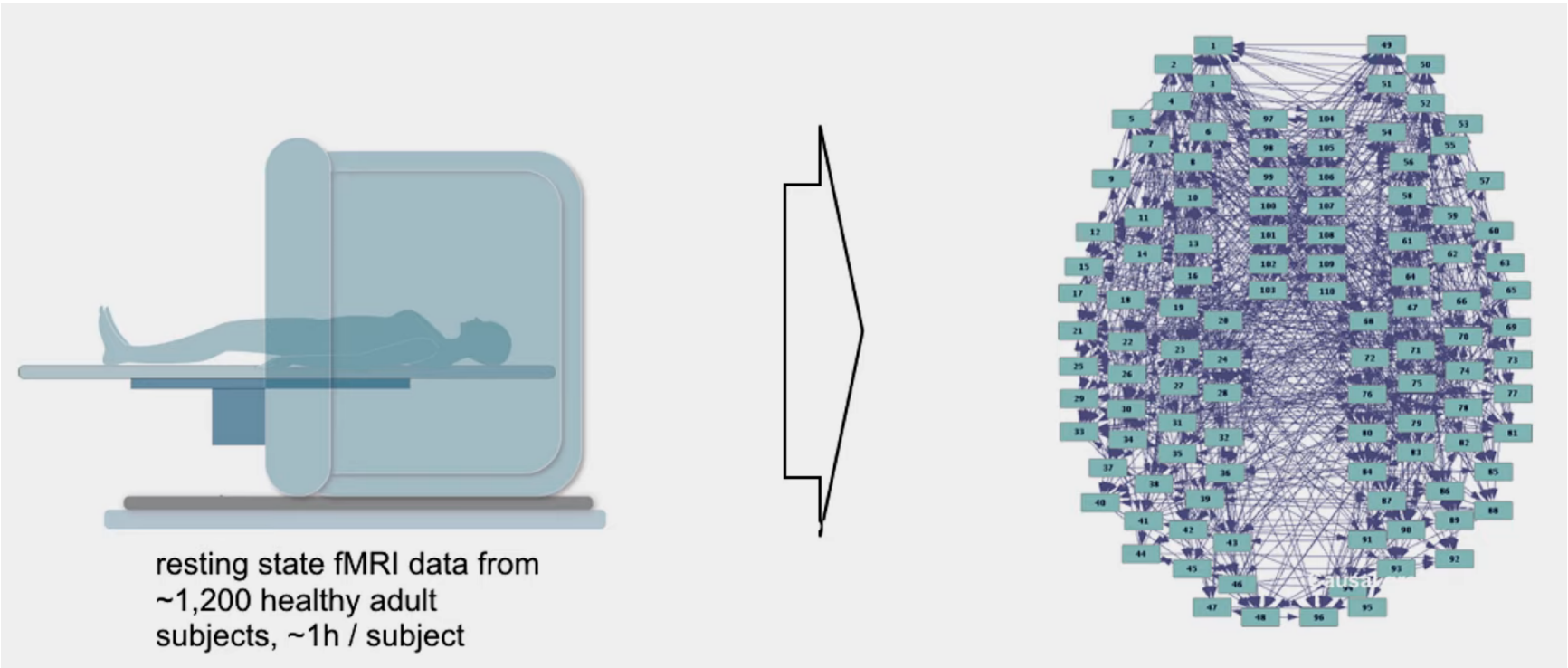
Graphical Models For Complex Health Data (P8124)

Daniel Malinsky

Columbia University
dsm2128@cumc.columbia.edu

Some Applications in Neuroscience

From fMRI to connectivity networks: Dubois et al. (2020)



(from slides by F. Eberhardt)

What is the main scientific question?

Dubois et al. are interested in "causal mechanisms behind emotions" in the brain, specifically how the amygdala may functionally connect to other brain regions/structures. Which brain regions are "involved" in emotion processing, and how are they causally related to one another?

What are the variables?

- Brain consists of billions of neurons, fMRI typically measures BOLD signals in $\sim$ 1million voxels $\Rightarrow$ how do you get to nodes in a graph?
- **So much work** goes into defining ROIs, mapping them across scans/subjects, removing motion artifacts, processing $\sim$ 1million voxels into small # of stable variables...
- In this paper, use volumetric parcellation (based on "standard" brain) into 110 regions
- Using mean BOLD signal in each ROI
- Lots of work goes into this, lots of open questions about how to do this best for "functional" analysis, etc.

Aggregation issues

ROIs are aggregates of time-varying signals in underlying regions/voxels/neurons. Exactly how you perform an aggregation like this has *very important* consequences for estimating associations. Why?

Aggregation issues

ROIs are aggregates of time-varying signals in underlying regions/voxels/neurons. Exactly how you perform an aggregation like this has *very important* consequences for estimating associations. Why?

Think about X_1, X_2 and their association with Y. Now consider $Z \equiv \frac{1}{2}(X_1 + X_2)$ and the association between Z and Y. There may be association between each components and Y but no association between the aggregate and Y. $Cov(X_1 + X_2, Y) = Cov(X_1, Y) + Cov(X_2, Y)$.

Aggregation/sub-sampling happens both in space and time with fMRI. In some cases even spurious assocations can be induced.

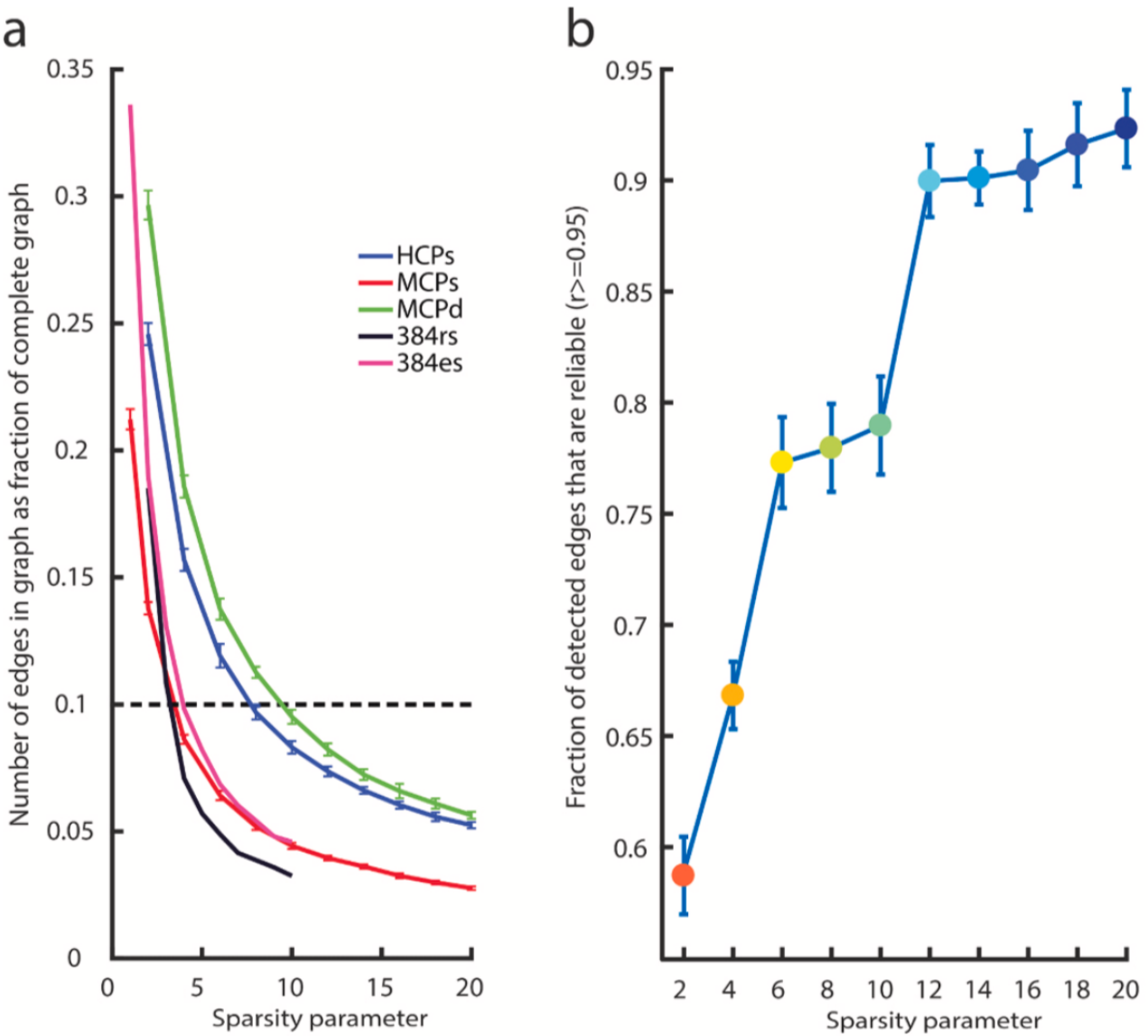
Multiple datasets

- Resting-state fMRI from 880 unique subjects (HCP_s)
- Two versions of resting-state fMRI from 1 subject over 80 sessions (MCP)
- Data from neurosurgical patients (epileptic) who underwent electrical stimulation

Structure learning algorithm: FGES

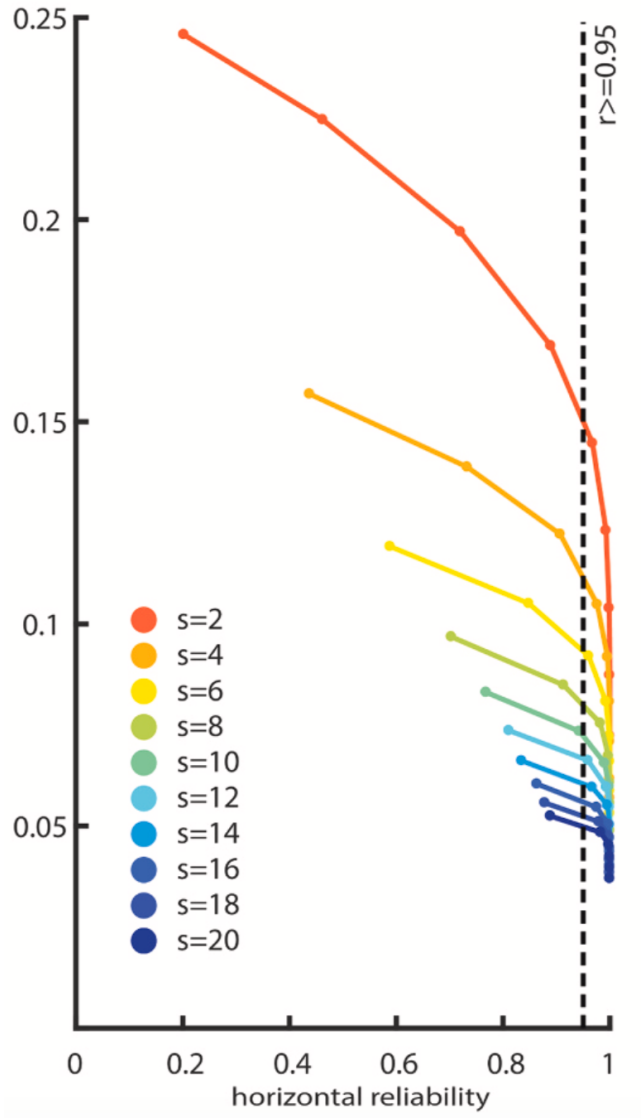
- FGES is a faster, parallelized version of Greedy Equivalence Search
- Objective: (greedy) optimization of the BIC score $\ell(D, \widehat{\theta}_{\mathcal{G}}) - s\frac{d}{2}\log(n)$ (assuming Gaussian likelihood)
- This score is consistent for any $s > 0$ (additional sparsity penalty), treated as a tuning parameter in this paper
- How to choose s?

Varying the sparsity parameter



Horizontal reliability

- For each s, estimate $\widehat{\mathcal{G}}_s^1, ..., \widehat{\mathcal{G}}_s^{11}$ over 11 datasets
- Count # of times each adjacency appears across 11 graphs (co-occurrence)
- Based on simulation (1000 sets of 11 random graphs w/ fixed sparsity) count how often co-occurence score would occur by chance
- HR defined as proportion of adjacencies that have a lower (or equal) co-occurrence count if 11 graphs (of fixed density) were generated by chance than the co-occurrence count observed
- Note: "null" distribution is complicated/approximate here, "random" graphs can be generated in many different ways
- Note: does not capture reliability of non-adjacencies


0.25
0.2
0.15
0.1
0.05
r>=0.95
s=2
s=4
s=6
s=8
s=10
s=12
s=14
s=16
s=18
s=20
0
0.2
0.4
0.6
0.8
1
horizontal reliability

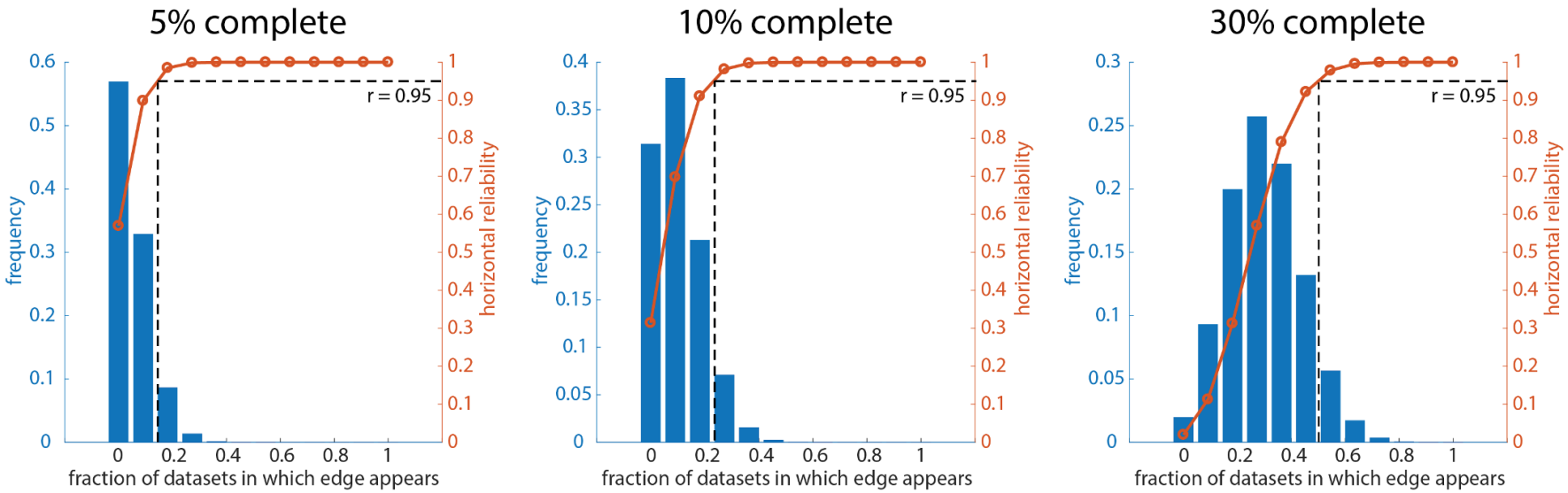
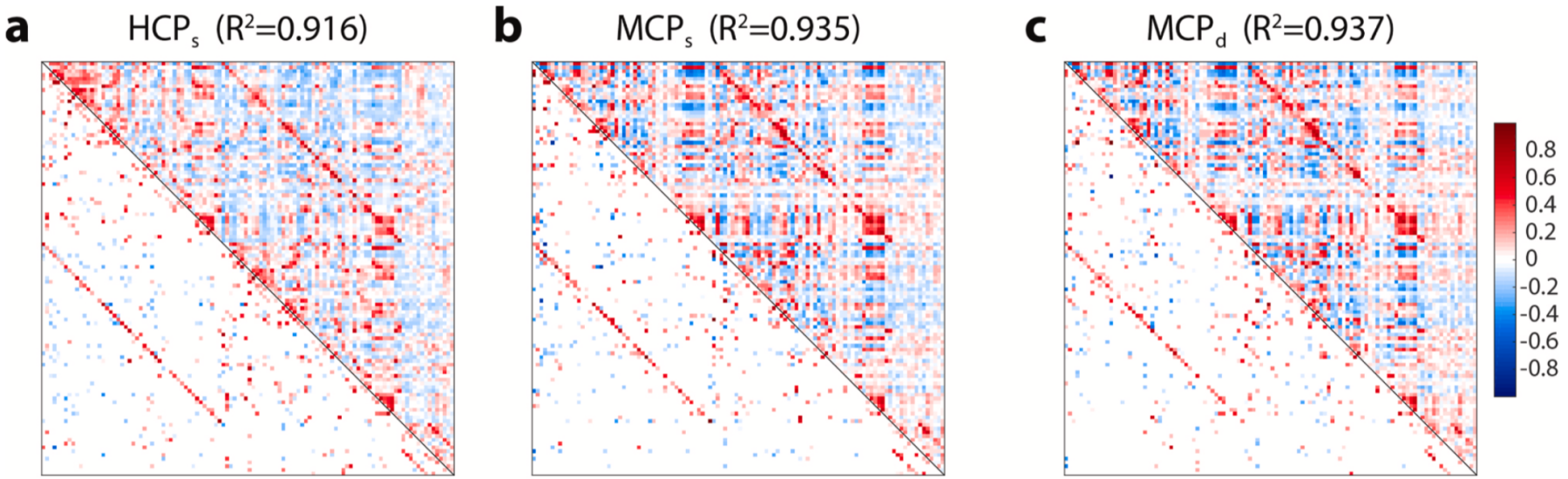


Figure S3. A horizontal reliability metric that can be compared across graphs of different densities. We simulated 1000 sets of 11 random graphs with three adjacency densities: 5%, 10% and 30%, from left to right. The blue histogram shows the distribution of co-occurrence scores (how many times an adjacency is repeated in the 11 subsets, from 0 to 11) that would occur by chance. The histograms differ for graphs of different densities (e.g., compare leftmost and rightmost plots). To account for this, we defined "horizontal reliability" of an adjacency A as the proportion of adjacencies that have a lower (or equal) co-occurrence count if 11 graphs (of fixed density) were generated by chance than the co-occurrence count observed for A (orange curves). Reliable edges are defined as edges with a horizontal reliability of .95 or higher (horizontal dashed lines), which corresponds to a different actual count of repeats for each graph density (vertical dashed lines).

Correlation matrices vs (weighted) adjacency matrices



Comparing sample correlation matrix w/ correlation matrix implied by graph

Linear structural equation model:

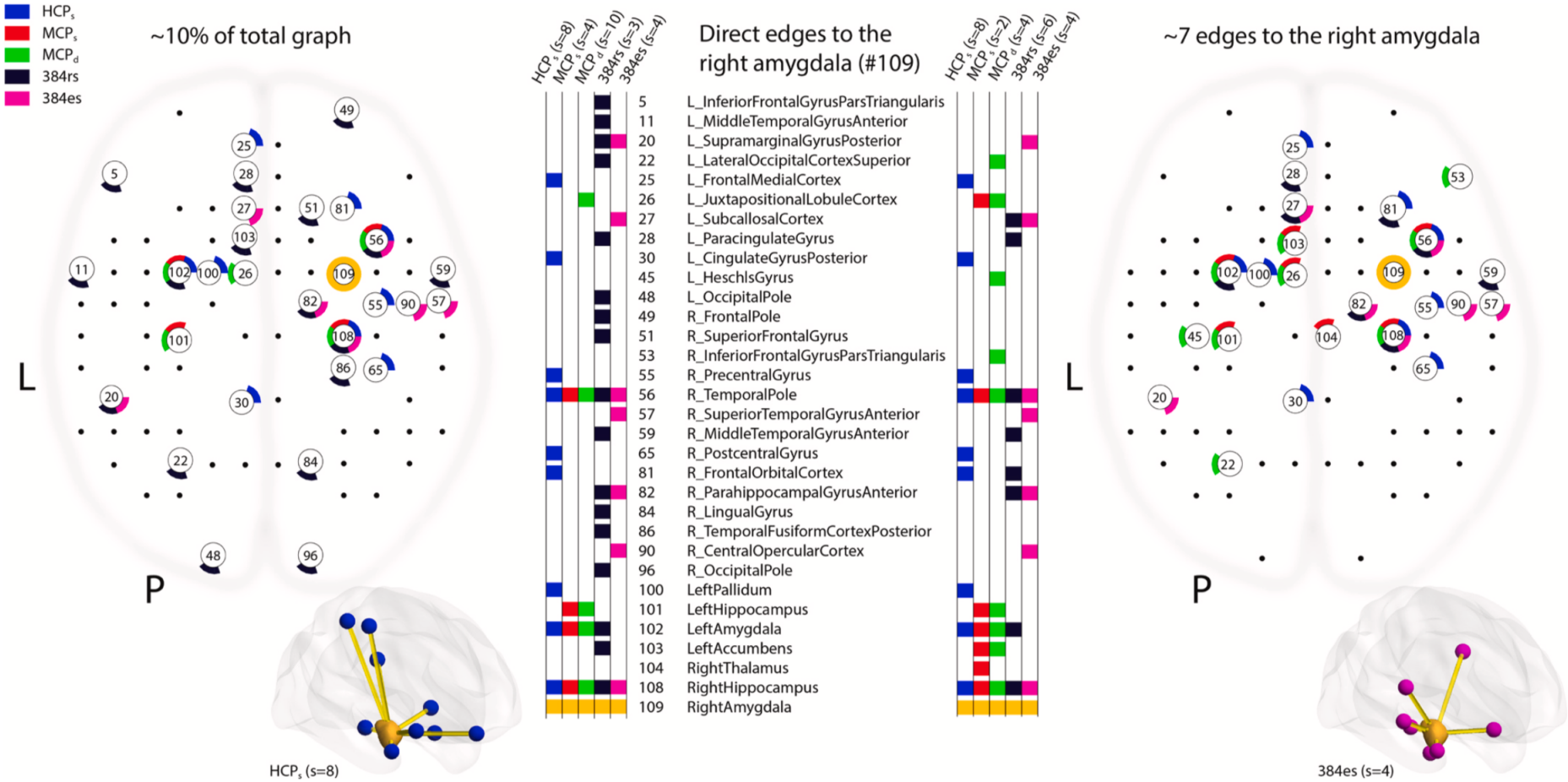
$$X = BX + \epsilon$$

$\Rightarrow$ w/ $B_{ij} = 0$ if no edge $X_i \not\to X_j$

$$Cov(X, X) = (I - B)^{-1}\,\mathbb{E}[\epsilon, \epsilon](I - B)^{-1} = (I - B)^{-1}\Sigma_\epsilon(I - B)^{-1}$$

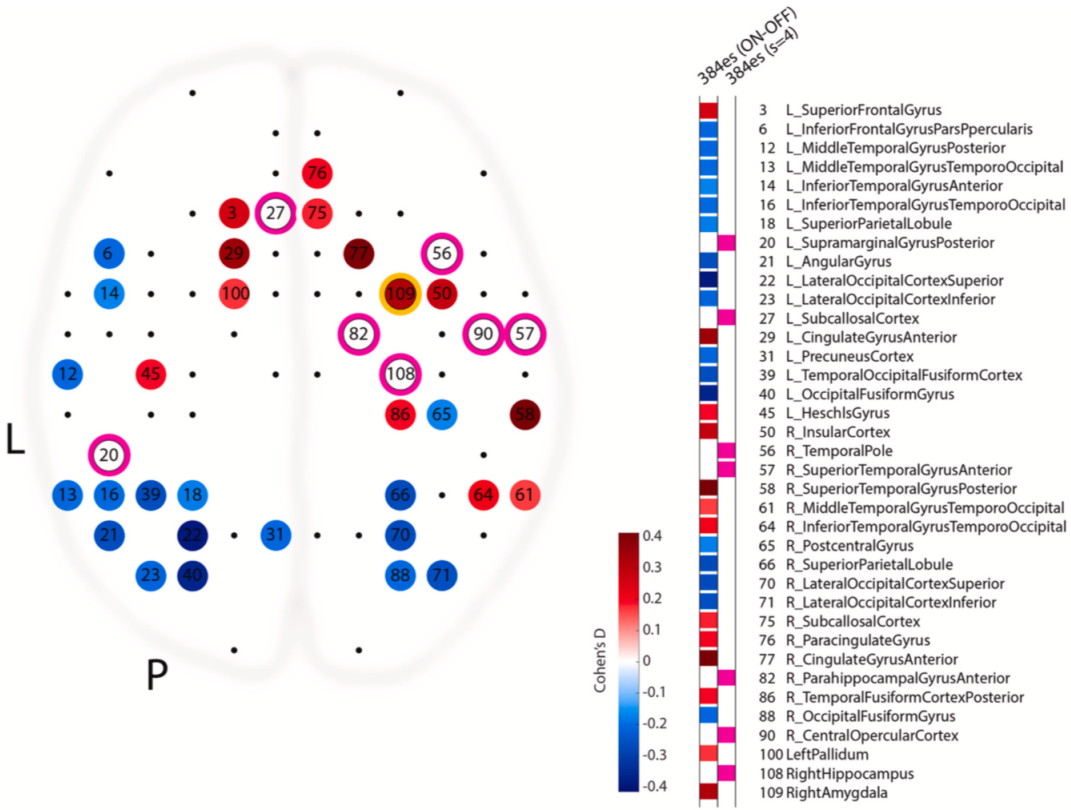
$\Rightarrow$ standardize this for "graph-based" correlation matrix

Authors compare this to the "raw" correlation matrix using R^2 to see how much the sparse graphs "captures" raw correlation pattern



Across data sets, "right amygdala was reliably found to be directly connected to the ipsilateral temporal pole and hippocampus, and the contralateral amygdala"

Electrostimulated patient: connections to amygdala vs ON/OFF analysis



No overlap btw ON/OFF analysis and direct adjacencies from estimated graph ($s = 4$). Why?

Conclusions/limitations

- Fascinating combination of resting state data w/ electrical stimulation data, multiple data sets
- Comprehensive attempt to assess reliability/stability across data sets and tuning param values
- Some "known" /believable structure is reproduced by FGES, some structure is unclear or highly variable
- Inconsistency btw learned graph and ON/OFF analysis
- Assumptions known to be violated: no feedback, no unmeasured confounding, linear Gaussian, i.i.d. samples

Heterogeneity in connectomes: Dajani et al. (2019)

What is the main scientific question?

"The goal of this study was to parse heterogeneity in a core executive function, cognitive flexibility, in children with a range of abilities [...] using directed functional connectivity profiles"

"... explore whether alternate categories (rather than DSM diagnosis or behavioral profile) might better parse the data, and to identify subgroups of children with similar brain network connectivity profiles or 'connectomes'"

What is the data?

fMRI scans during "cognitive flexibility" task from $n = 132$ (later $n = 99$) children 8-13 yrs, who have been classified as ASD, ADHD, or TC according to various behavioral scales and diagnostics.

Various preprocessing to remove motion artifacts, etc.

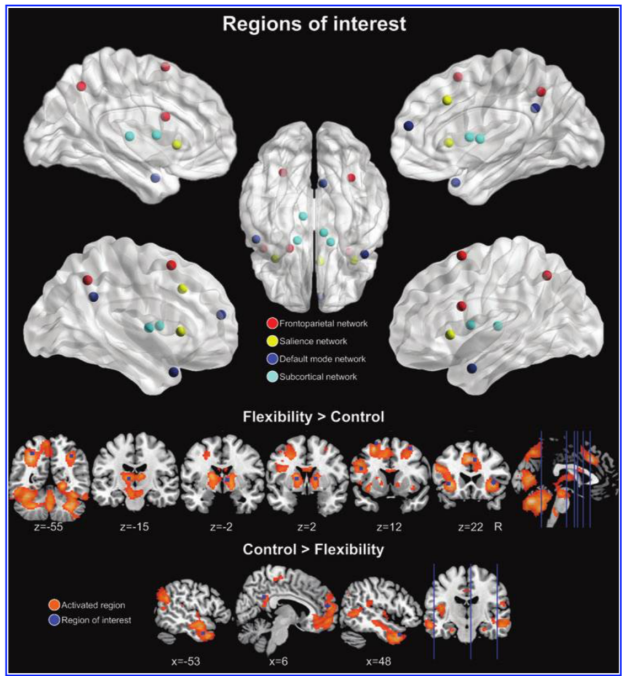
In this case, using the individual time series of BOLD signals rather than averages over time, though still averaging over voxels within an ROI (contrast to Dubois et al.).

What are the variables?

Select 16 ROIs based on brain regions that are associated with the Flexible Item Selection Task (FIST), mostly based on prior experiments w/ adults.

Use an atlas to map location coordinates for ROIs across different brain scans.

Excluded various regions to reduce the total number of ROIs



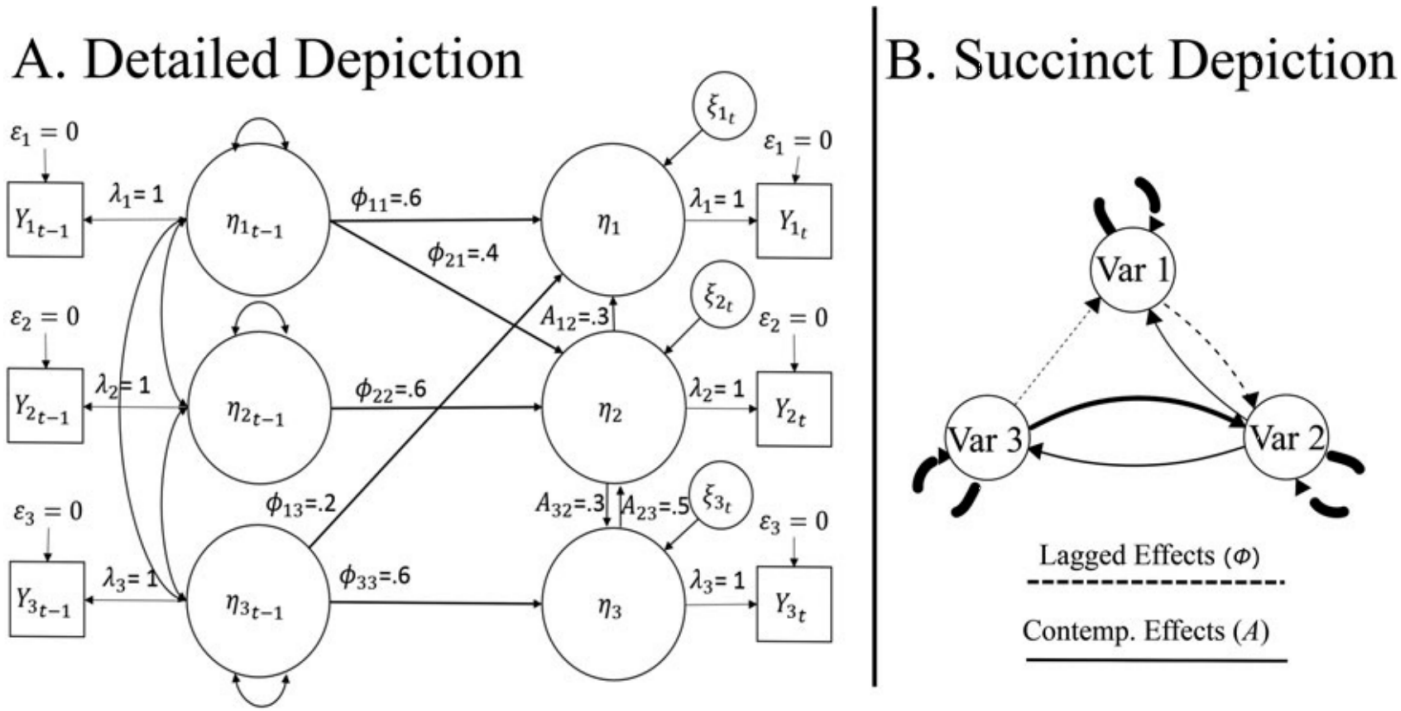
Structure learning algorithm: GIMME

GIMME (Group Iterative Multiple Model Estimation) is a greedy score-based algorithm that is very similar to GES, but combines info from multiple subjects (group-level vs individual-level connections).

- Starting from the null graph, greedily add edges which improve "fit" for the largest number of subjects in the sample. Continue until no new edges improve fit for majority.
- The measure of model "fit" is somewhat different from GES, but similarly likelihood-based.
- Next, add additional edges if they improve fit for individuals (or subgroups as defined by the Walktrap community detection alg). Stops after meeting criteria for model fit for two of four indices: CFI $\geq$0.95, NNFI $\geq$0.95, RMSEA $\leq$0.05, and SRMR $\leq$0.05
- Somewhat heuristic (no proof of consistency)

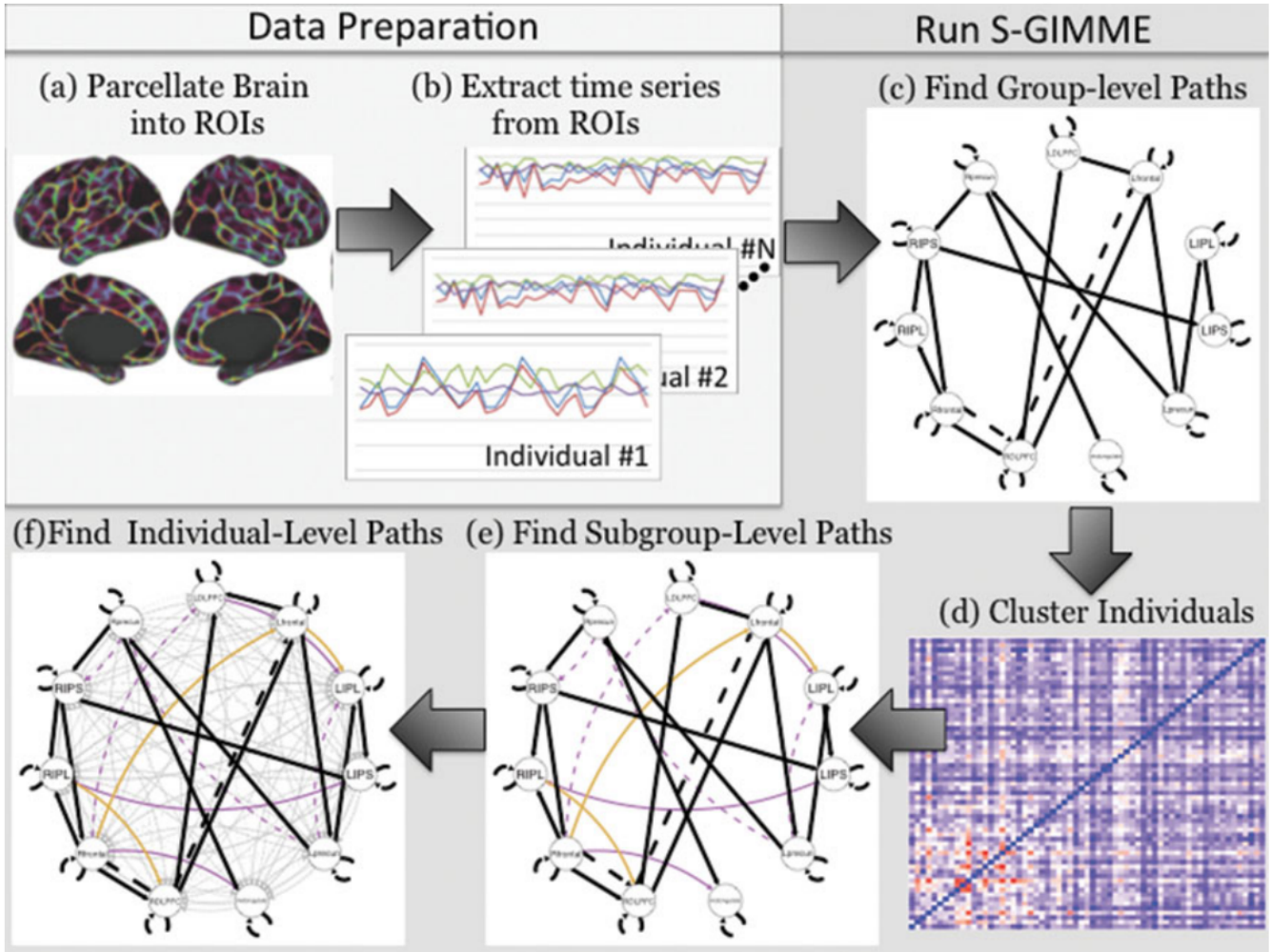
Model

$$\eta_t = A\eta_t + F\eta_{t-1} + \xi_t$$



From Gates et al. 2017

Overview



From Gates et al. 2017

Example result

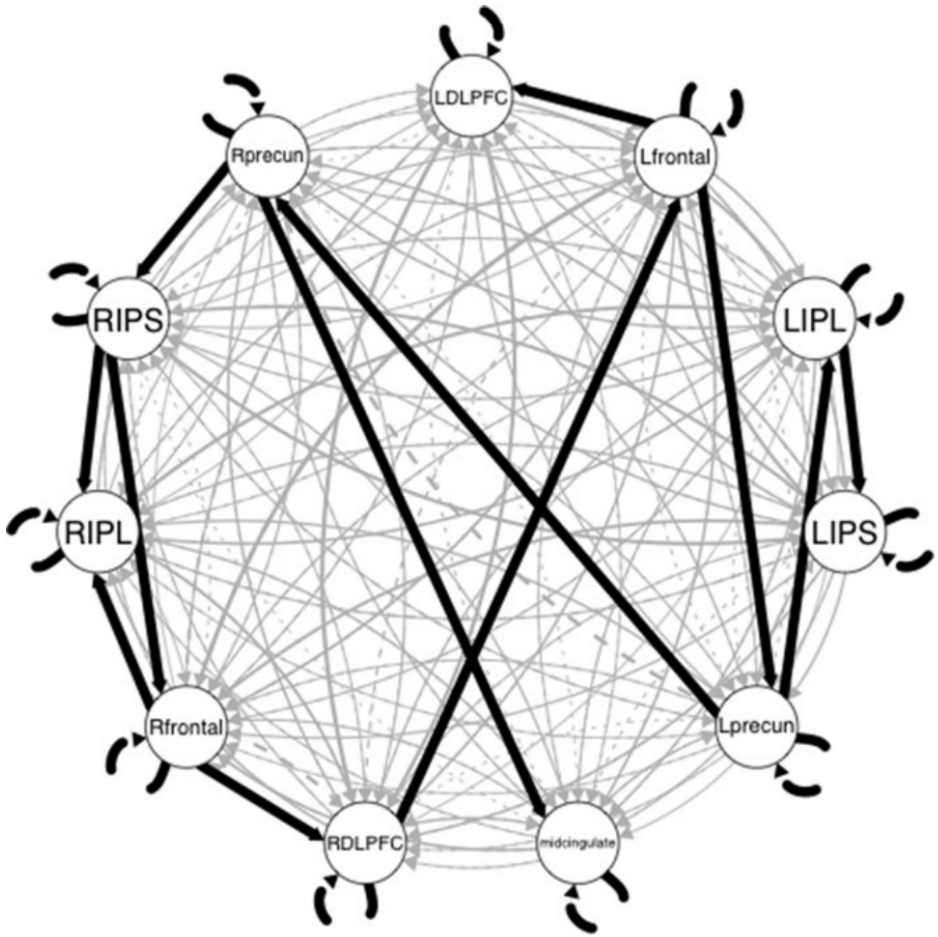


Figure 2. Original GIMME results obtained from the empirical example. Black lines indicate group-level effects; gray lines indicate individual-level effects. Line width corresponds to proportion of individuals having the effect. Dashed lines are lagged; solid lines are contemporaneous relations.

From Gates et al. 2017

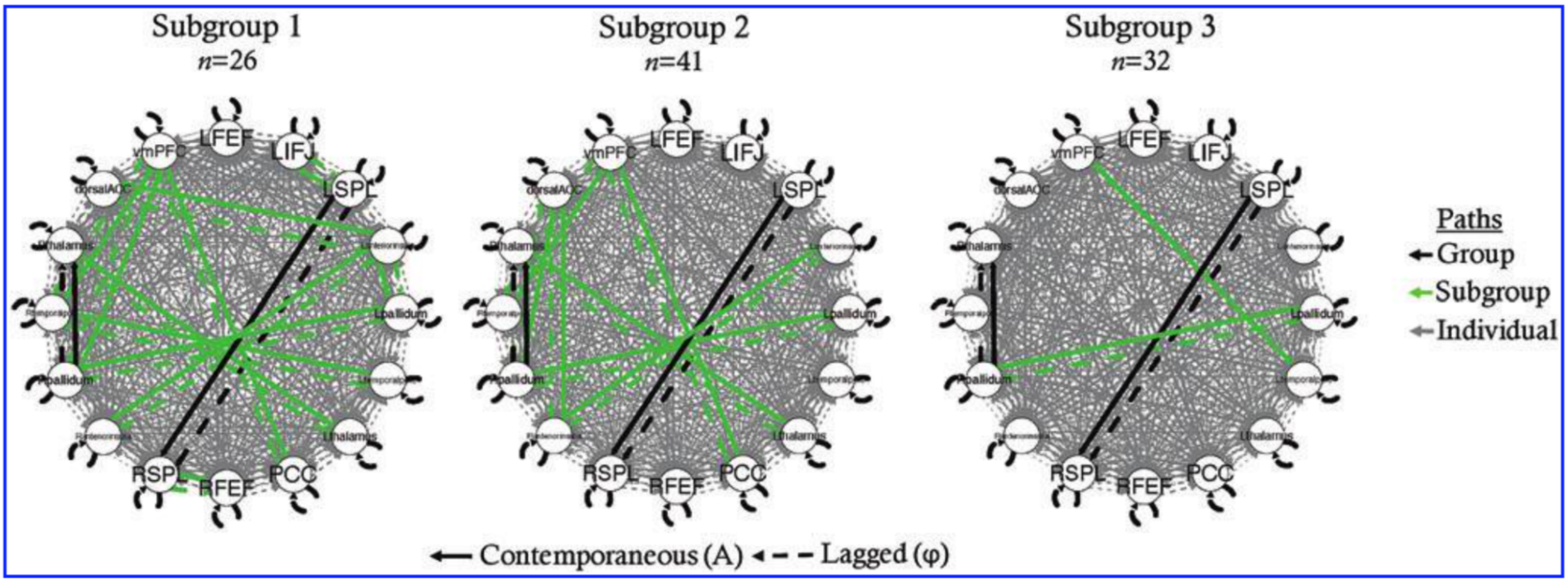
Stability of subgroups

Are clusters detected by GIMME "stable"? How to evaluate this?

- "Clustering algorithms are prone to producing false positives, meaning clusters are produced even in cases where none truly exists"
- "Quantitatively, a cluster solution is said to be stable if the graph had 20% or more of its edges perturbed before the cluster solution for the rewired graph is as different as when 20% of the nodes are randomly placed into different clusters."
- "Cluster solution considered valid if modularity for orig sol is $\geq$ 95th percentile of modularity obtained from random graphs"

$\Rightarrow$ compare to other "stability" notions we've discussed (this doesn't involve subsampling or testing on "new" data)

Subgroups


Subgroup 1
n=26
Subgroup 2
n=41
Subgroup 3
n=32
vmPFC
LFEF
LIFJ
LSPL
dorsalACC
Lthalamus
Lpallidum
RSPL
RFEF
PCC
Rthalamus
Paths
Group
Subgroup
Individual
Contemporaneous (A)
Lagged (φ)

Stability

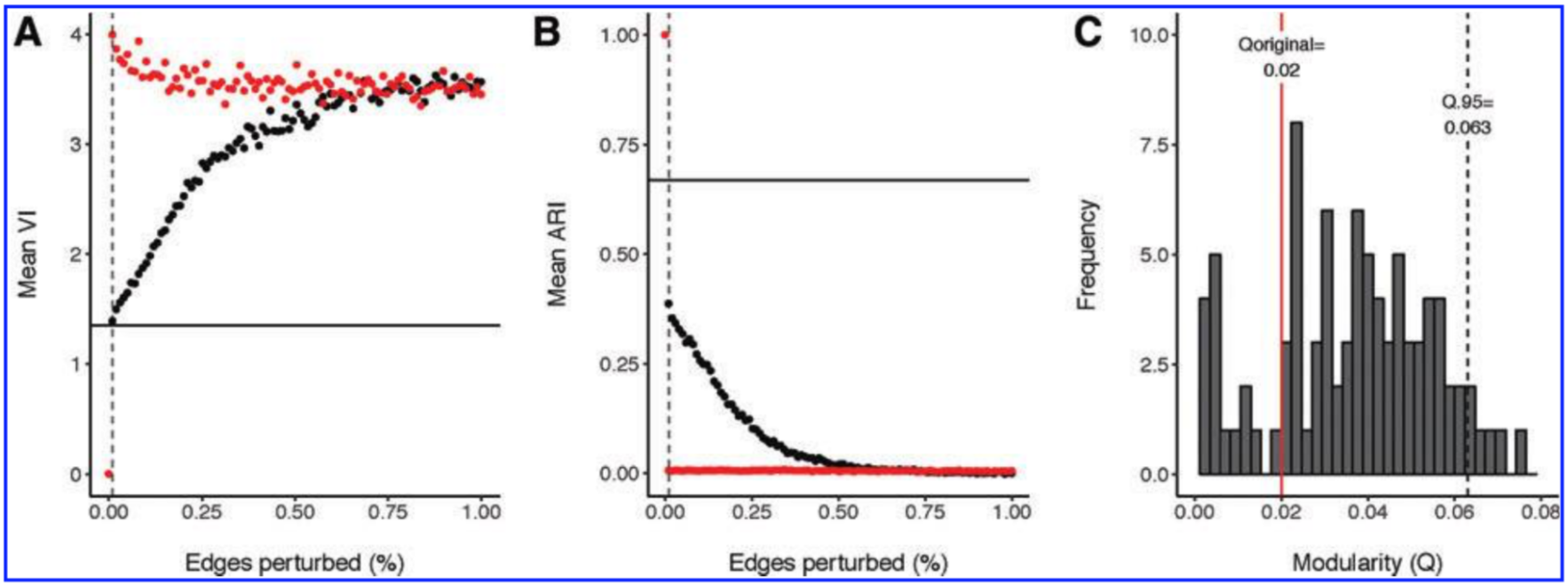


FIG. 3. Cluster validation. Results from both VI and ARI demonstrate that the clustering solution is not stable. **(A, B)** The black horizontal line represents the point at which 20% of participants were placed into different clusters than the original solution (20% of nodes perturbed). The dashed vertical line identifies the point at which the perturbed graph reached 20% of nodes perturbed. Black dots represent the perturbed graph based on the original clustering solution, while the red dots represent a perturbed random graph. **(C)** Demonstrates that modularity for the original clustering solution (0.02) was not better than expected by chance (>0.06), suggesting this clustering solution is not valid. ARI, Adjusted Rand Index; VI, Variation of Information. Color images are available online.

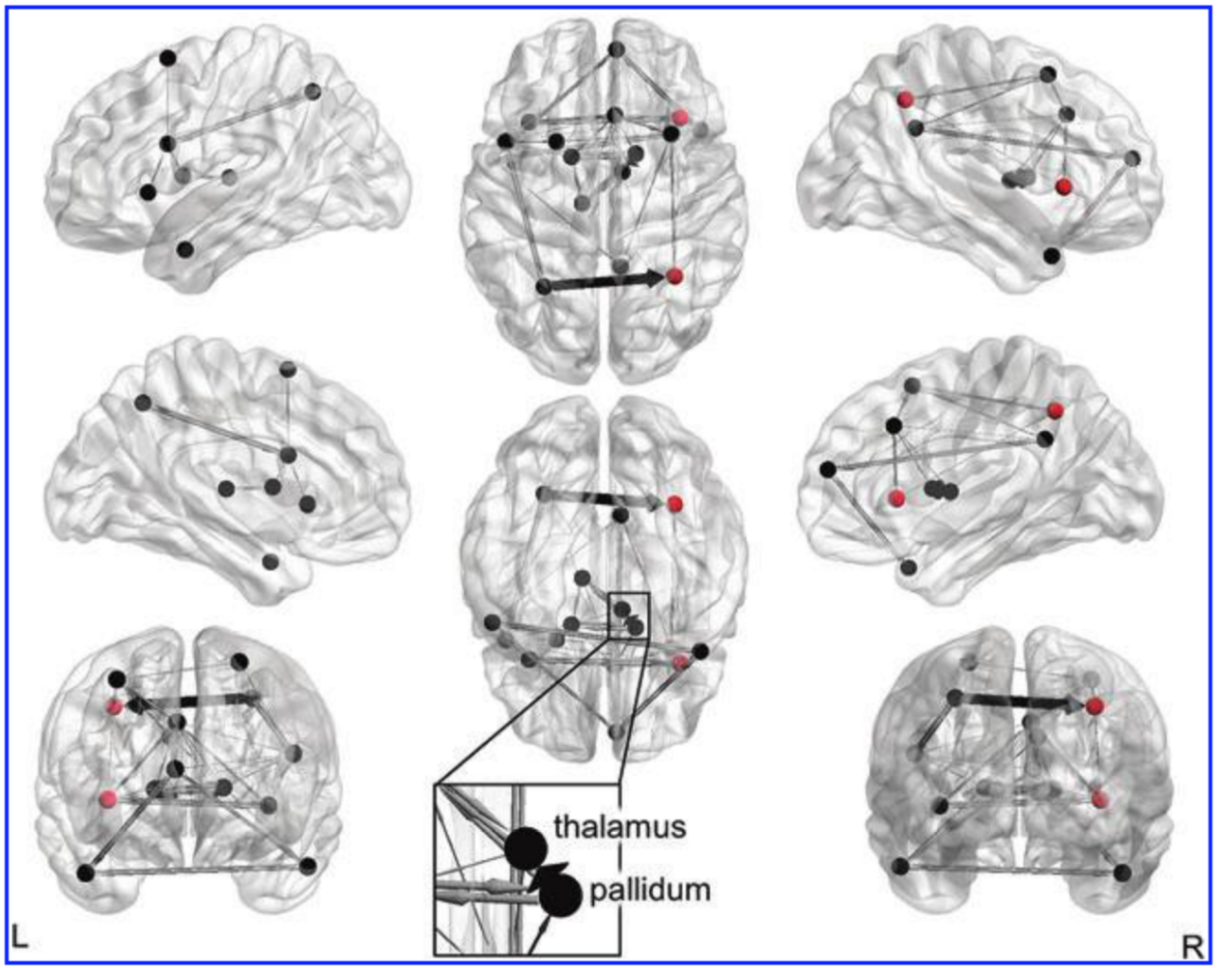
No stable subgroups

"[Lack of stability] suggests that the neural substrates of cognitive flexibility in children may not differ categorically"

Find significant heterogeneity $\Rightarrow$ unexpected, may be reasons to revise diagnostic categories

"The network-level heterogeneity apparent here is in contrast with many group-based studies of the development of the *undirected* functional connectome, which conclude that network topology is stable by about 8 years of age"

Homogenous results



Associations w/ symptoms

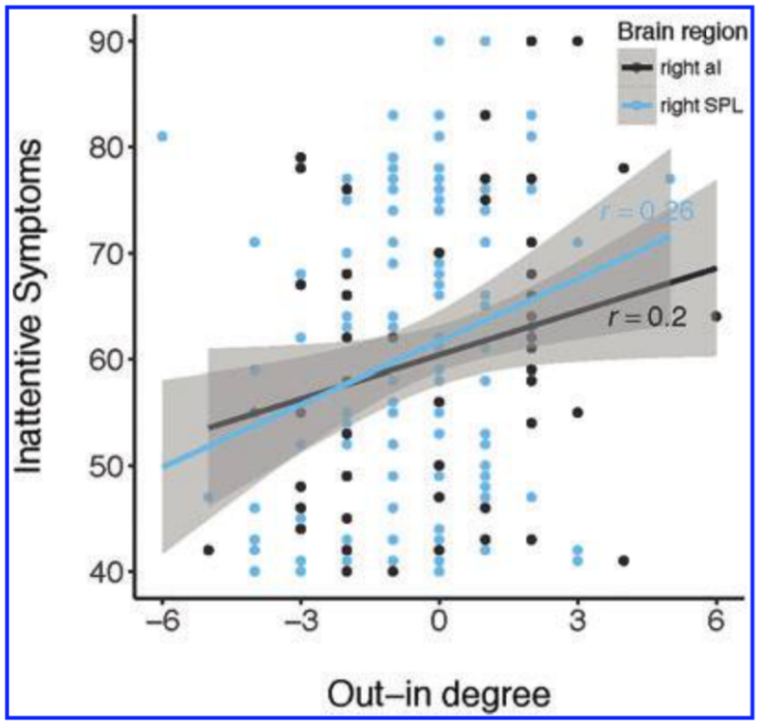


FIG. 5. Out/in degree relates dimensionally to inattentive symptoms. Net inflow to the rAI and rSPL is related to fewer inattentive symptoms across children in clinical and nonclinical samples ($N = 132$). Color images are available online.

Associations w/ symptoms

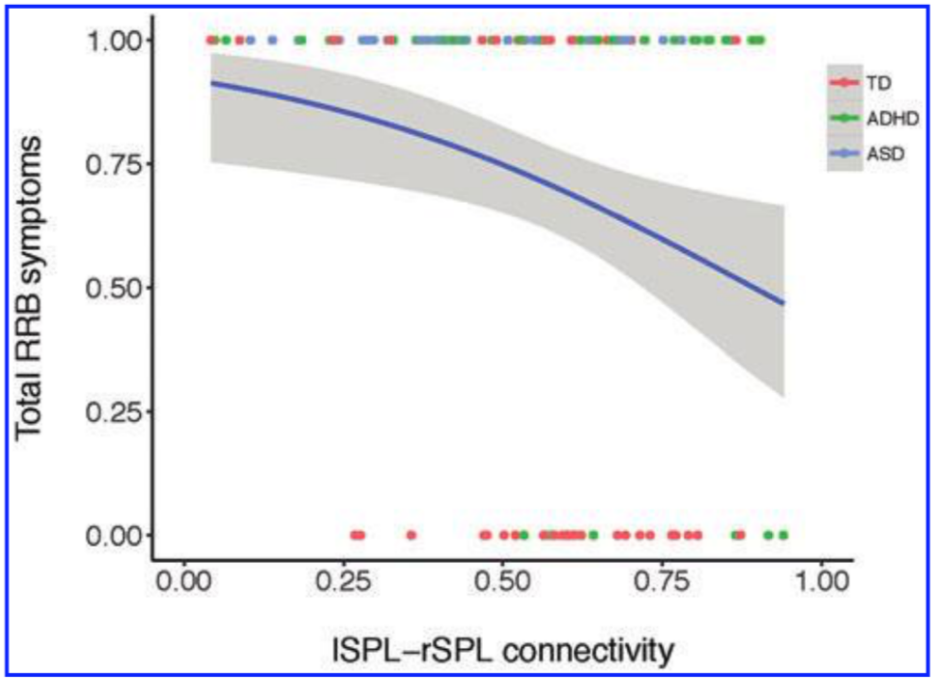


FIG. 6. Stronger lSPL → rSPL connectivity is associated with greater odds of having zero parent-reported total restricted and RRBs. Measured with the total score of the Repetitive Behavior Scale-Revised across children in clinical and nonclinical samples ($N = 132$). Total RRB symptoms: 0 represents no symptoms reported, 1 represents a score of 1 or more. Estimated logit function presented in blue. ADHD, attention-deficit/hyperactivity disorder; ASD, autism spectrum disorder; RRB, repetitive behavior; TD, typically developing. Color images are available online.

Conclusions/limitations

- No evidence for stable subgroups according to diagnostic categories – surprising
- Some interesting/suggestive associations btw graph metrics and ADHD symptoms
- Limited by small number of ROIs, lots of omitted variables. Only $n = 99$ subjects after excluding patients with high head motion (which is associated with symptoms).
- No independent data sample on which to validate results
- Algorithmic procedure depends on various parameters/thresholds which are basically heuristic, though have been observed to do well in previous simulation studies