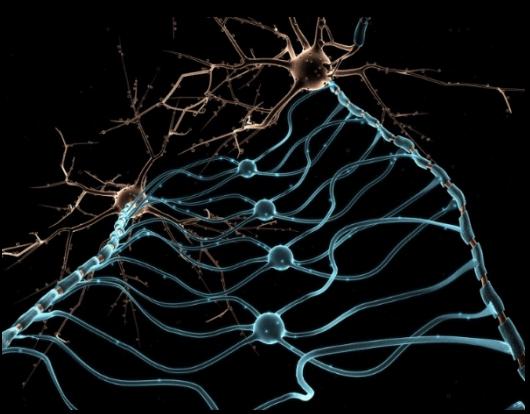
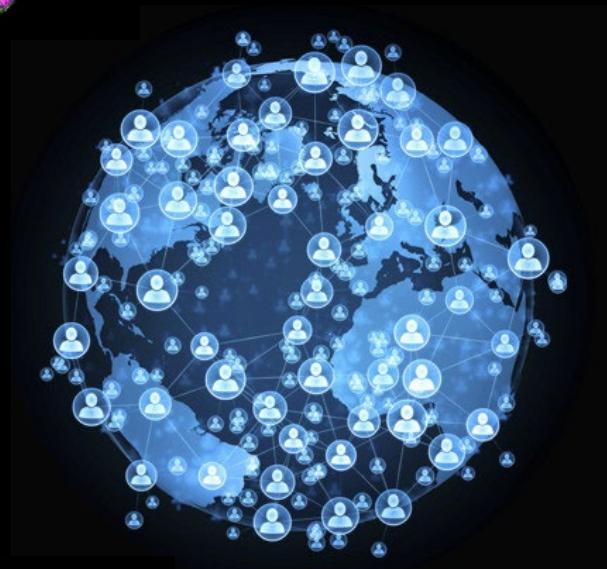
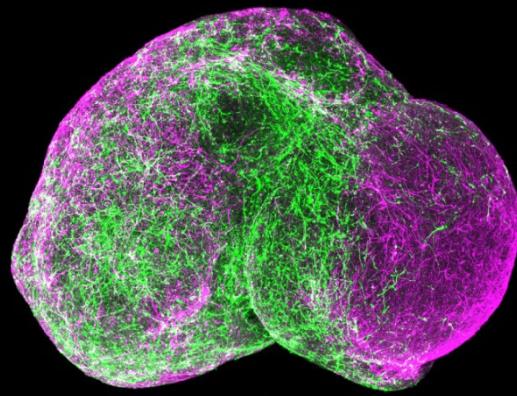
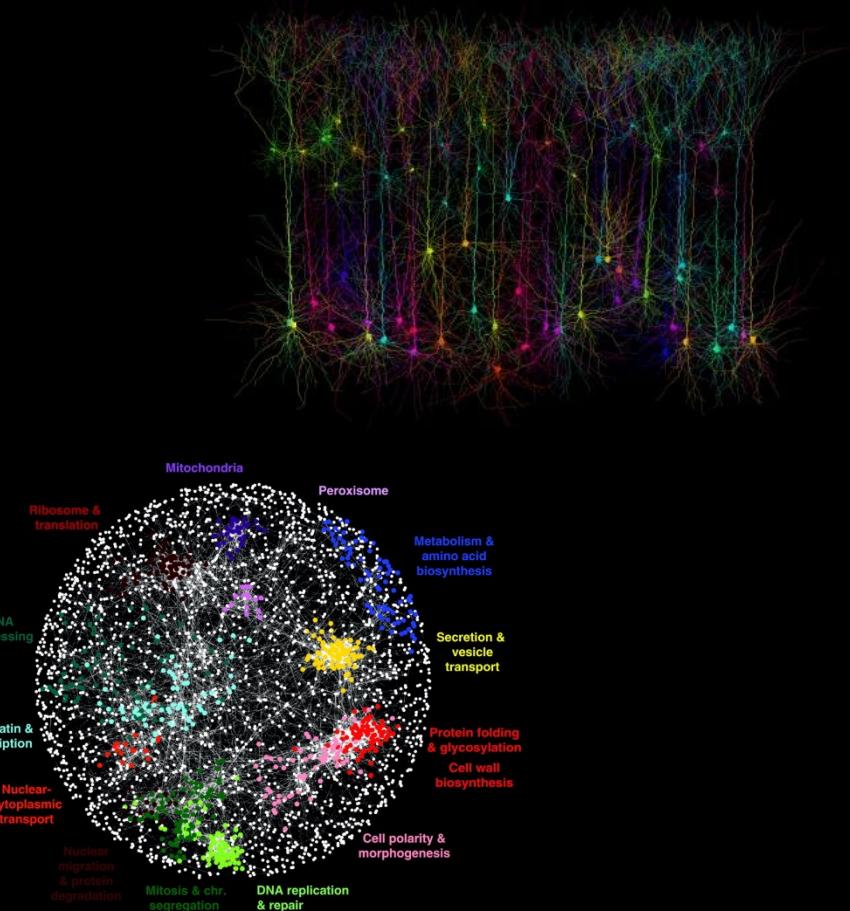


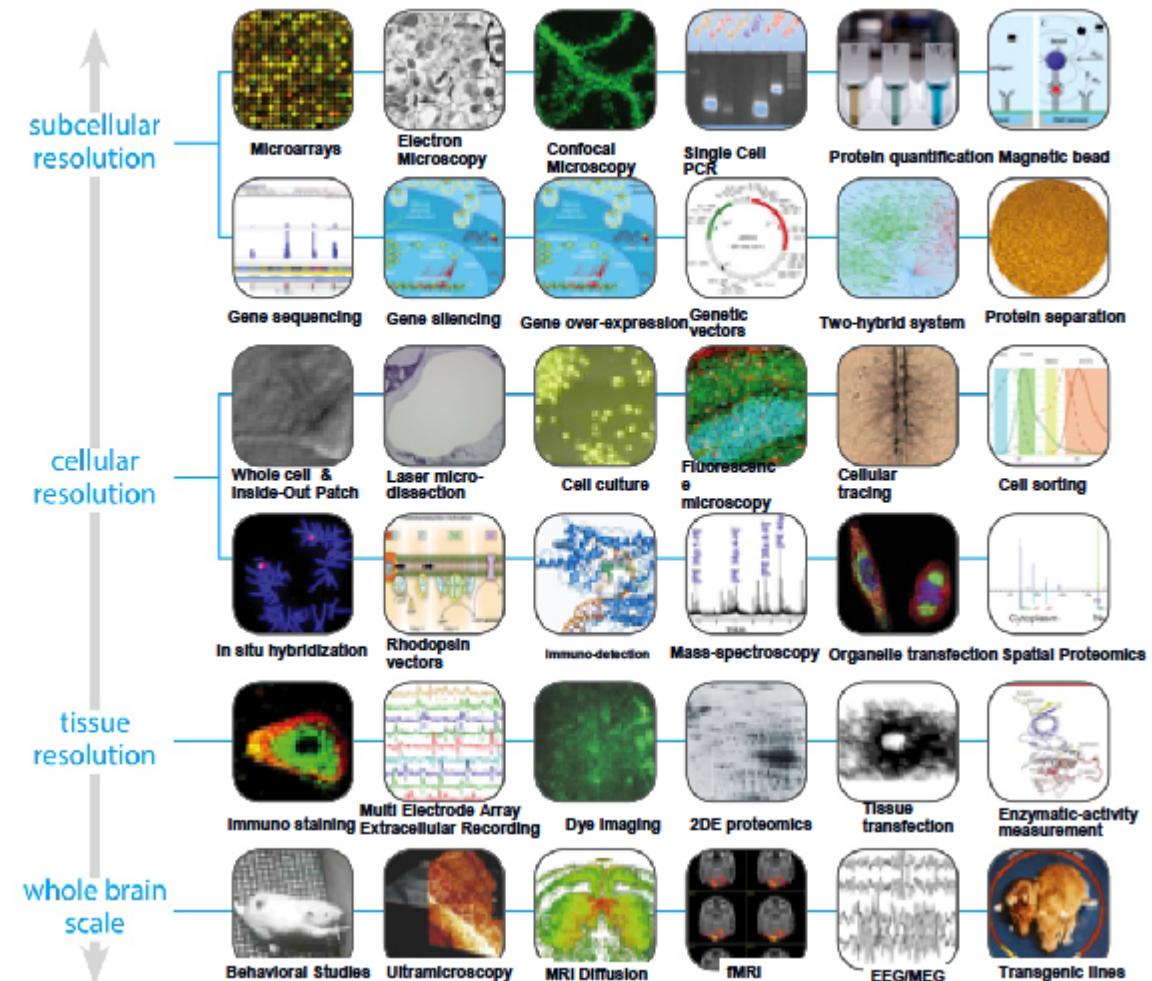
Multivariate Associative Techniques

Bratislav Misic
McConnell Brain Imaging Centre
Montréal Neurological Institute
McGill University

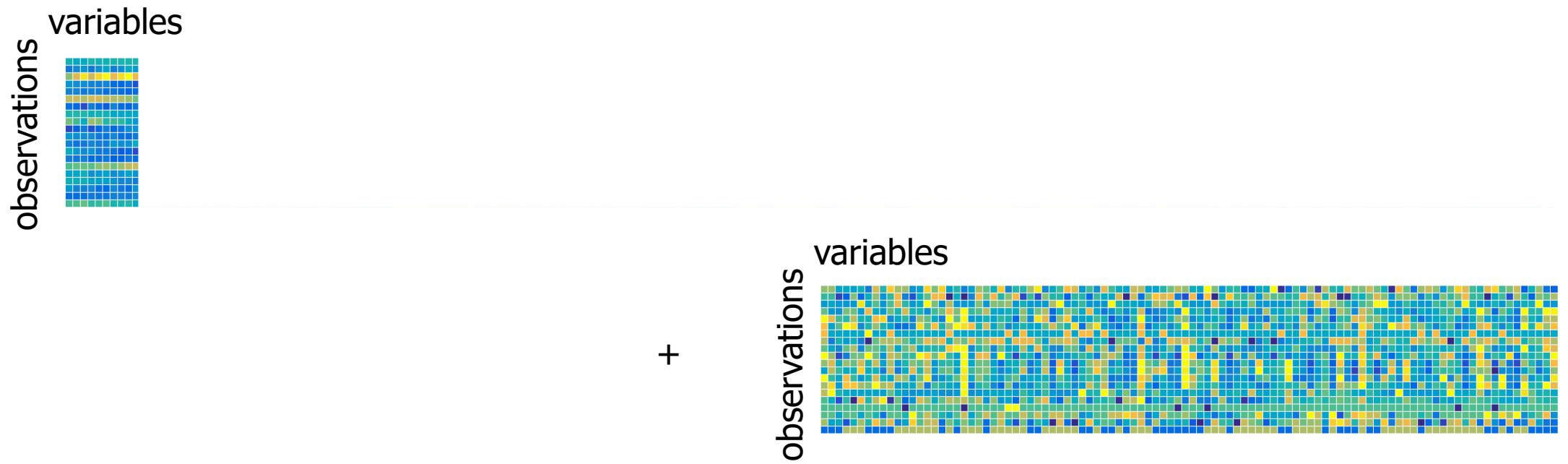


Towards multivariate analysis

- **activation**: mapping individual elements to individual external variables
- **connectivity**: mapping individual structural or functional connections to individual external variables
- **networks**: mapping network attributes to individual external variables
- **multivariate systems**: mapping patterns of elements or connections to patterns of external variables



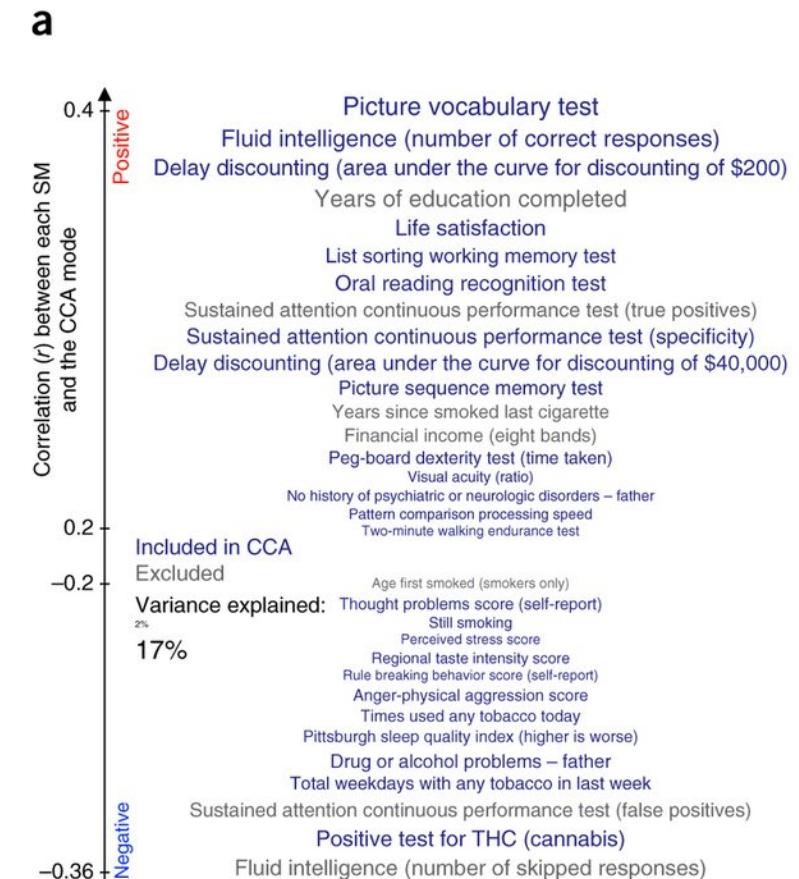
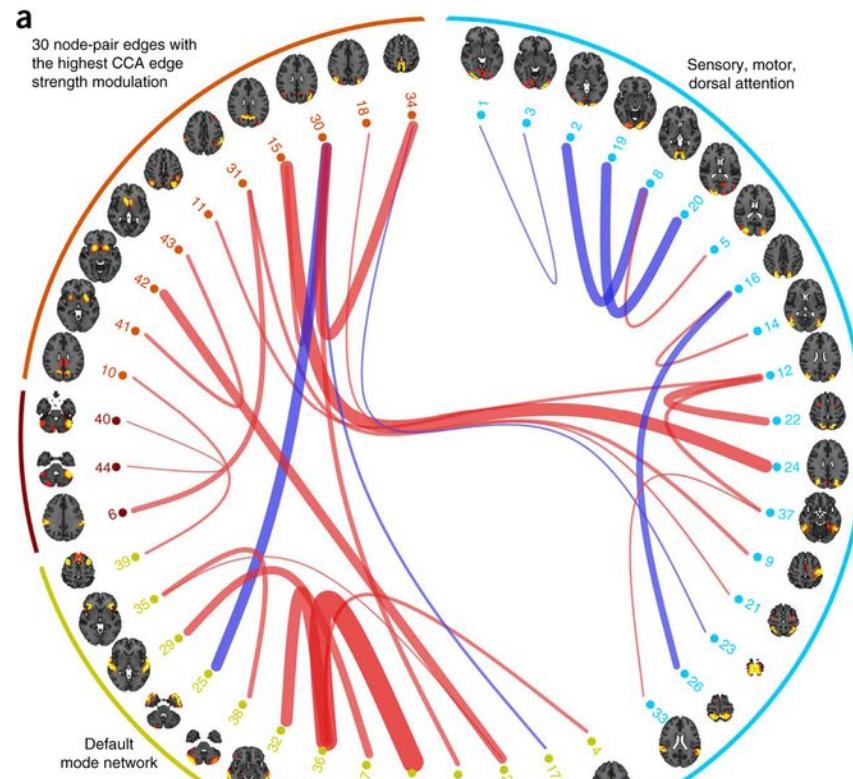
Integrating scales and modalities



- How to deal with **more variables than observations?**
- How to operationalize the **network property?**
- How to relate **multiple data sets** to each other?

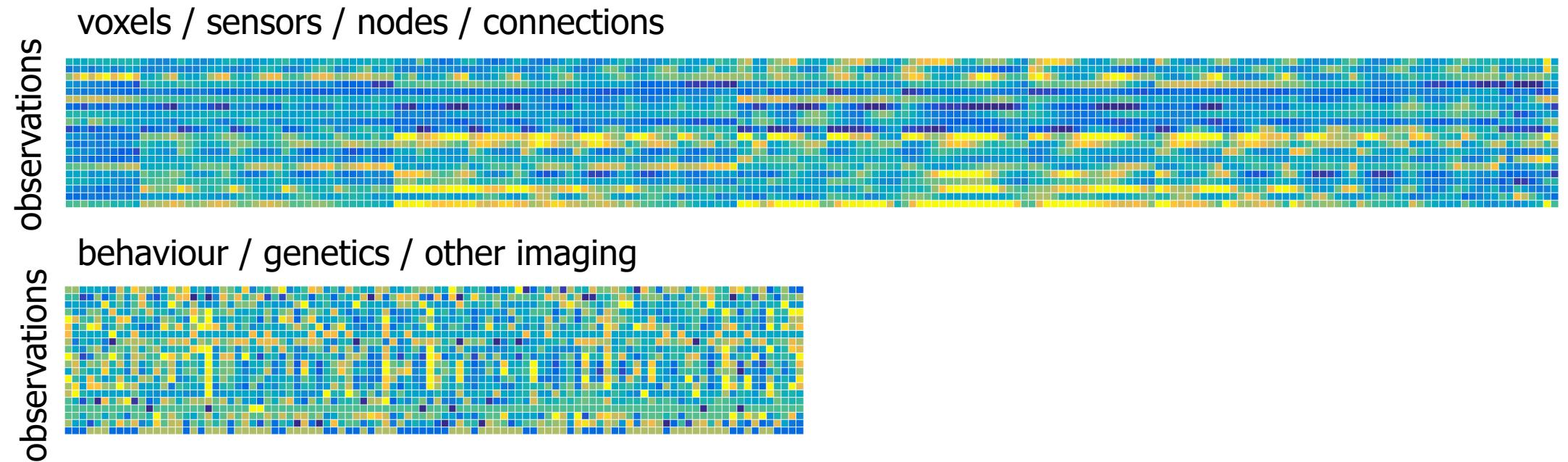
Example: relating connectivity and behaviour

- **dataset:** Human Connectome Project (HCP); n=900 individuals
- **modality #1 :** resting state functional connectivity
- **modality #2:** comprehensive behavioural + demographics battery
- **question:** how do individual differences in functional interactions map onto behaviour?
- **analysis:** Canonical Correlation (CCA)



- **Goal #1:** associative multivariate models are useful and versatile
- **Goal #2:** these techniques are mathematically related
- **Goal #3:** nonparametric statistical inference can be easily applied
- **Goal #4:** examples + demo

A family of techniques



- principal component analysis (**PCA**): $\text{SVD}(\mathbf{X})$
- partial least-squares (**PLS**): $\text{SVD}(\mathbf{X}'\mathbf{Y})$
- canonical correlation analysis (**CCA**): $\text{SVD}((\mathbf{X}'\mathbf{X})^{-1/2}(\mathbf{X}'\mathbf{Y})(\mathbf{Y}'\mathbf{Y})^{-1/2})$
- linear discriminant analysis (**LDA**): $\text{SVD}(\mathbf{W}^{-1}\mathbf{B})$

Singular value decomposition (SVD)

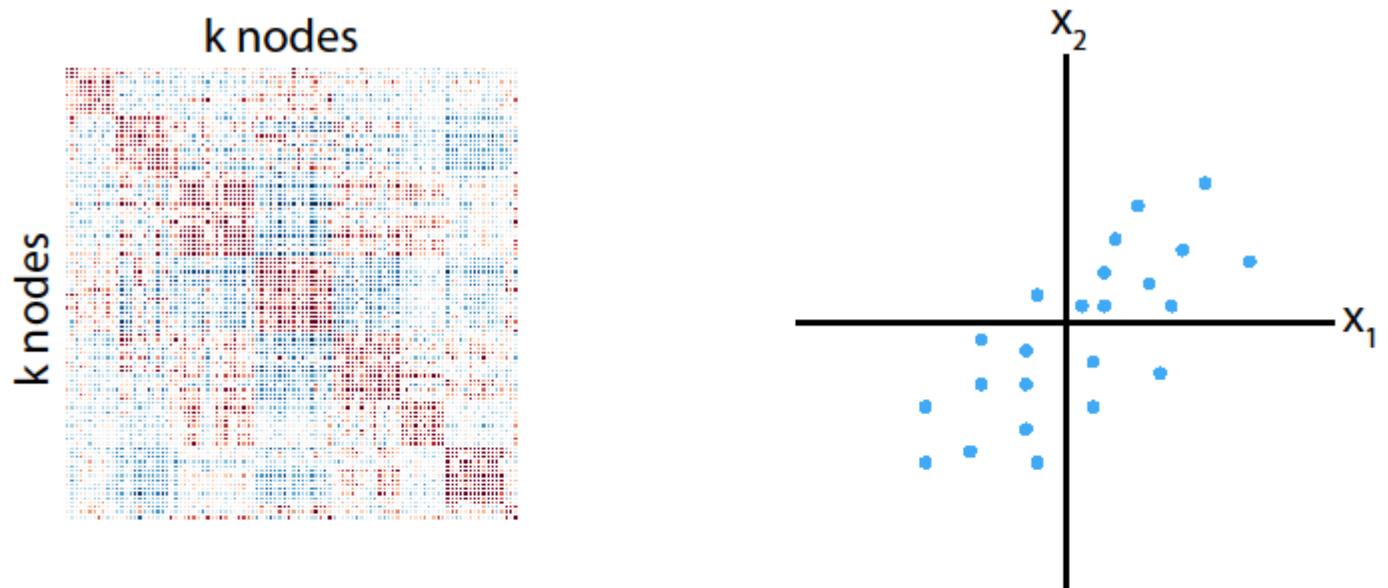
- how can we factorize a data matrix?
 - $SVD(\mathbf{X}) = \mathbf{USV}'$
- this is a generalization of the spectral decomposition:
 - $EIG(\mathbf{X}'\mathbf{X}) = \mathbf{U}\Lambda\mathbf{U}'$
 - $EIG(\mathbf{XX}') = \mathbf{V}\Lambda\mathbf{V}'$
- **U** and **V** are orthonormal **singular vectors**; represent how you should weigh the original variables in **X**
- **S** is a diagonal matrix of **singular values**; represent how strongly paired the **U** and **V** are

$$\begin{matrix} p \\ k \end{matrix} \quad \begin{matrix} p \\ k \end{matrix} \quad \begin{matrix} p \\ p \\ S \end{matrix} \quad \begin{matrix} p \\ p \\ V' \end{matrix}$$

$\mathbf{X} = \mathbf{U} \mathbf{S} \mathbf{V}'$

Principal component analysis (PCA)

- original variables \mathbf{x}_1 and \mathbf{x}_2 are correlated but neither captures the dominant pattern of variance
- want new variables \mathbf{z}_1 and \mathbf{z}_2 that
 - (a) capture as much variance
 - (b) are mutually uncorrelated
- need to find a rotation \mathbf{u} to re-align the original axes (variables)
- \mathbf{u} is chosen to maximize the variance of the new variables \mathbf{z}



Maximizing variance

- want to find a new variable $\mathbf{z} = \mathbf{X}\mathbf{u}$

- choose \mathbf{u} to maximize $var(\mathbf{z})$

under the constraint that $\mathbf{u}'\mathbf{u} = 1$

- $var(\mathbf{z}) = \frac{1}{n-1} \mathbf{u}' \mathbf{X}' \mathbf{X} \mathbf{u} = \mathbf{u}' \mathbf{R} \mathbf{u}$

where $\mathbf{R} = \frac{1}{n-1} \mathbf{X}' \mathbf{X}$

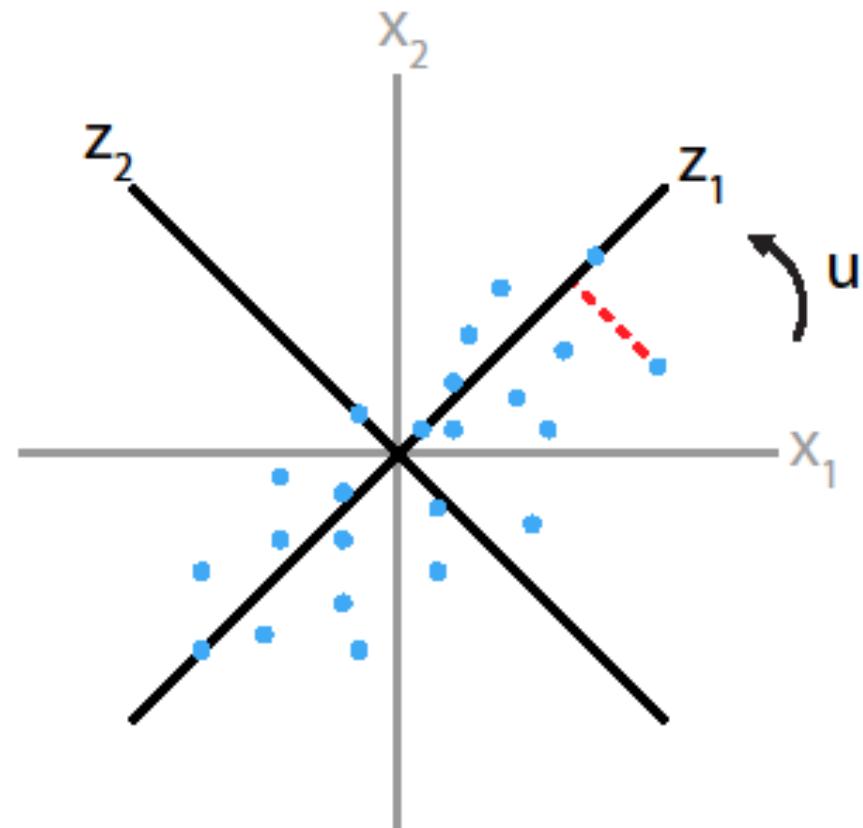
- define Lagrangian: $L = \mathbf{u}' \mathbf{R} \mathbf{u} - \lambda(\mathbf{u}' \mathbf{u} - 1)$

- find maximum: $\frac{\partial L}{\partial \mathbf{u}} = 2\mathbf{R}\mathbf{u} - 2\mathbf{u}\lambda = 0$

$$\mathbf{R}\mathbf{u} = \lambda\mathbf{u} \quad (\mathbf{R} - \lambda\mathbf{I})\mathbf{u} = 0$$

- λ = eigenvalue, \mathbf{u} = eigenvector

- $var(\mathbf{z}) = \mathbf{u}' \mathbf{R} \mathbf{u} = \mathbf{u}' \mathbf{u} \lambda = \lambda$



Singular value decomposition (SVD)

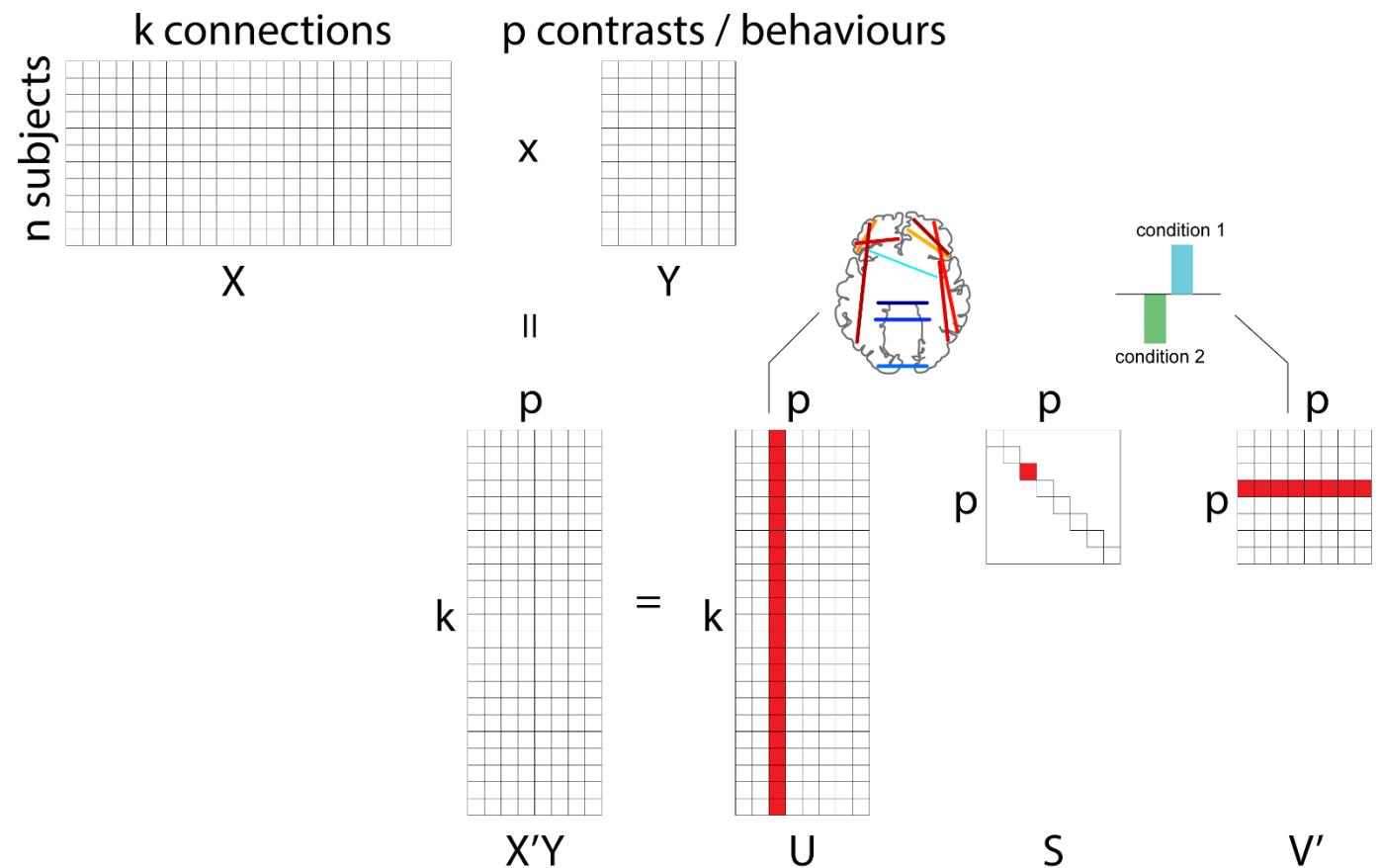
- $SVD(\mathbf{X}) = \mathbf{USV}'$
- $EIG(\mathbf{X}'\mathbf{X}) = \mathbf{U}\Lambda\mathbf{U}'$
- $EIG(\mathbf{XX}') = \mathbf{V}\Lambda\mathbf{V}'$
- $\mathbf{X}'\mathbf{X} = (\mathbf{VS}'\mathbf{U}')(\mathbf{USV}') = \mathbf{VS}'(\mathbf{U}'\mathbf{U})\mathbf{SV}' = \mathbf{V}(\mathbf{S}'\mathbf{S})\mathbf{V}' = \mathbf{V}\Lambda\mathbf{V}$
- $\mathbf{XX}' = (\mathbf{USV})(\mathbf{VS}'\mathbf{U}') = \mathbf{US}(\mathbf{V}'\mathbf{V})\mathbf{S}'\mathbf{U}' = \mathbf{U}(\mathbf{SS}')\mathbf{U}' = \mathbf{U}\Lambda\mathbf{U}'$
- eigenvector of \mathbf{XX}' = left singular vector of \mathbf{X}
- eigenvector of $\mathbf{X}'\mathbf{X}$ = right singular vector of \mathbf{X}
- eigenvalue = squared singular value

$$\begin{matrix} p \\ k \end{matrix} = \begin{matrix} p \\ k \end{matrix} \quad \begin{matrix} X & U \end{matrix}$$

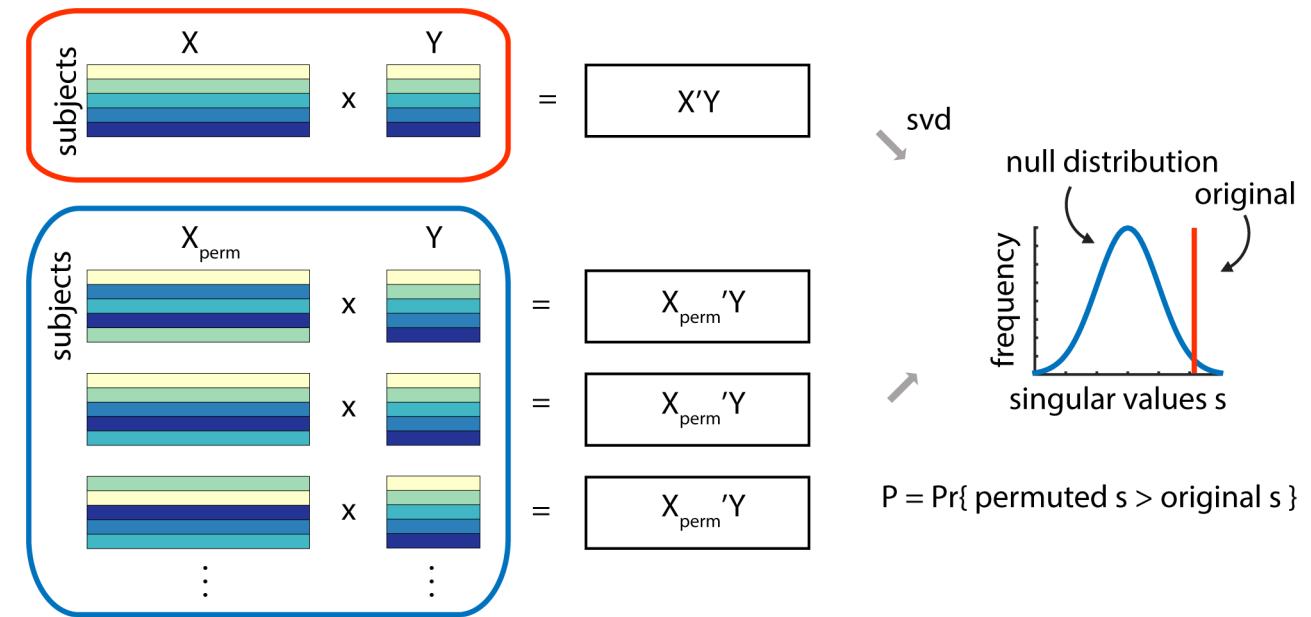
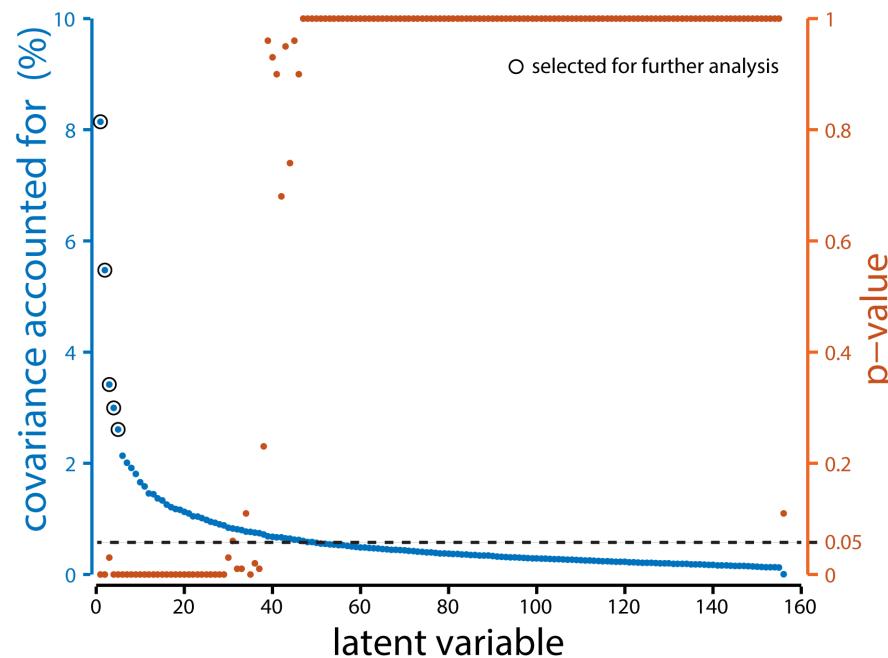
The diagram illustrates the Singular Value Decomposition (SVD) of a matrix X . It shows that X is equal to the product of three matrices: U , S , and V' . Matrix U is a $k \times k$ orthogonal matrix. Matrix S is a $k \times k$ diagonal matrix containing the singular values. Matrix V' is a $p \times p$ orthogonal matrix. The labels p and k are placed above the grids, and X , U , S , and V' are placed below them.

Partial least squares

- in PCA, you decompose the original data matrix \mathbf{X}
- in PLS, you decompose a correlation matrix ($\mathbf{X}'\mathbf{Y}$) between two data matrices \mathbf{X} and \mathbf{Y}
- get multiple **latent variables** (aka components, modes, variates)
 - singular vectors represent the optimal weighing of original variables,
e.g. patterns of connectivity + patterns of behaviour
 - singular value represents the proportion of X-Y covariance accounted for by the pairing



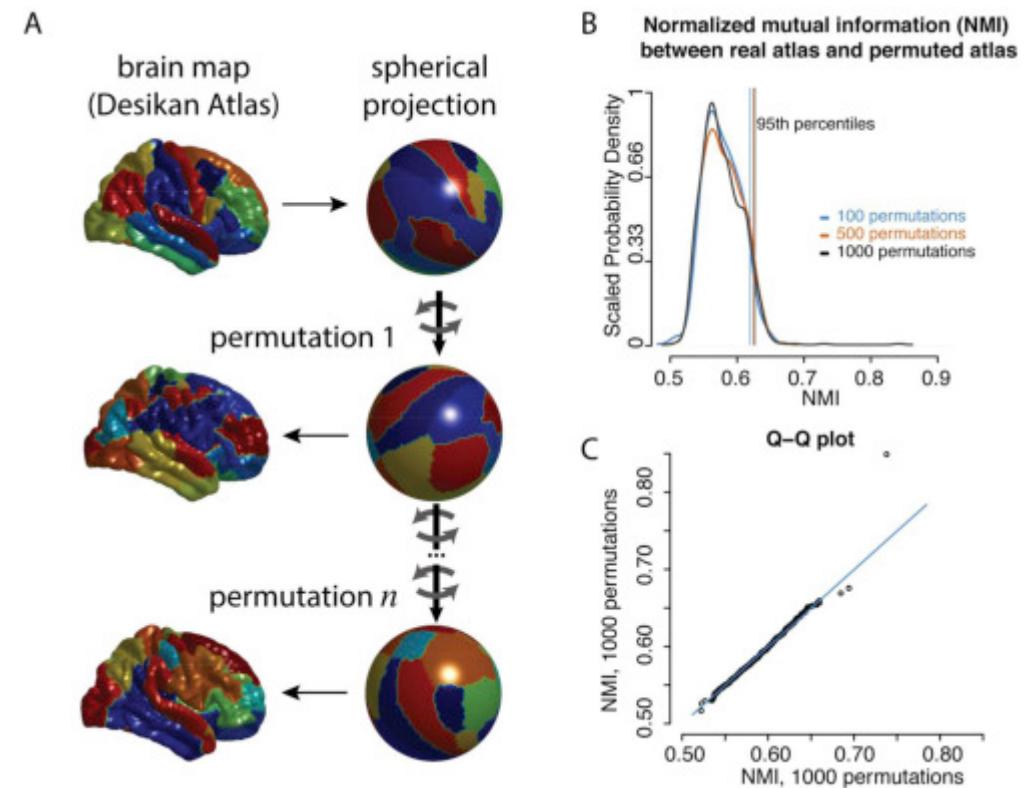
How many components to retain?



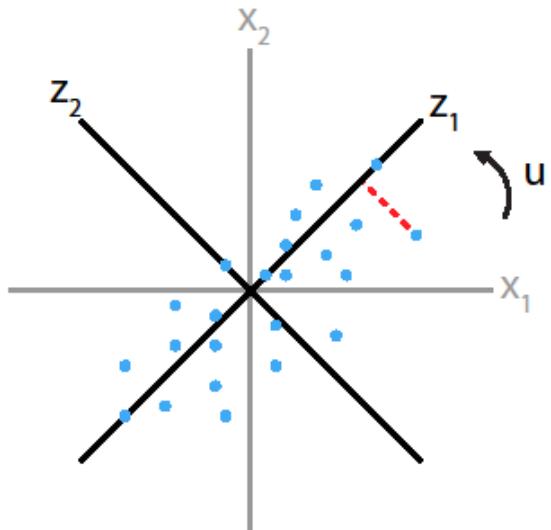
- Kaiser criterion: drop all eigenvalues lower than 1
- “elbow” rule: look for the biggest drop-off in the scree plot
- **permutation test**: which singular values are greater than expected?

Permutation tests when comparing maps

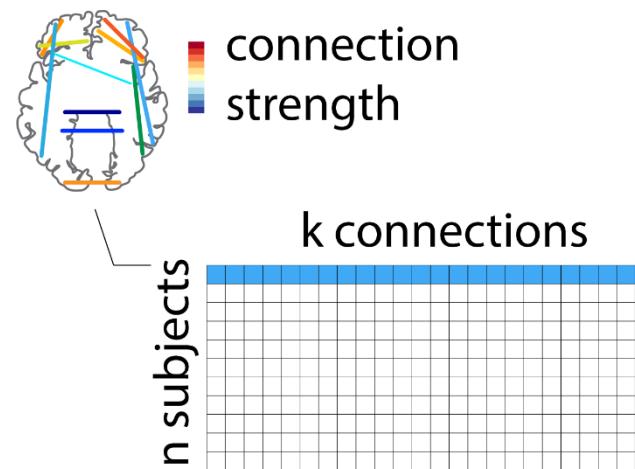
- when comparing topographic maps (i.e. observations = brain regions), simple permutation might not be enough to overcome spatial autocorrelation
- examples: comparing gene expression, morphometry, cytoarchitectonics, etc.
- can use spherical projection nulls instead ("spin tests")



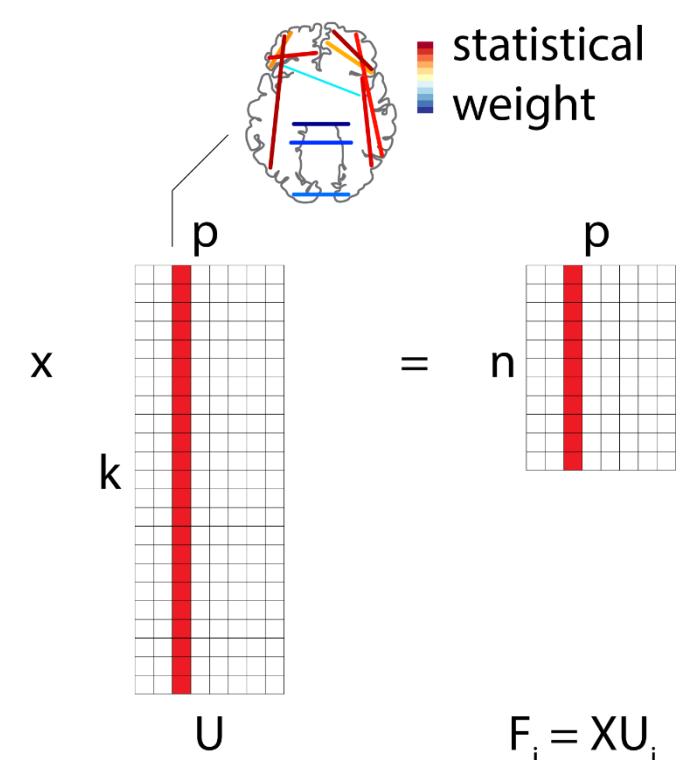
Individual participants



- weights that tell us how much each variable (connection, region, behaviour) contributes to the pattern
- but every participant is different – how do individual participants express the overall pattern?
- participant **scores** – project original data onto the latent variables

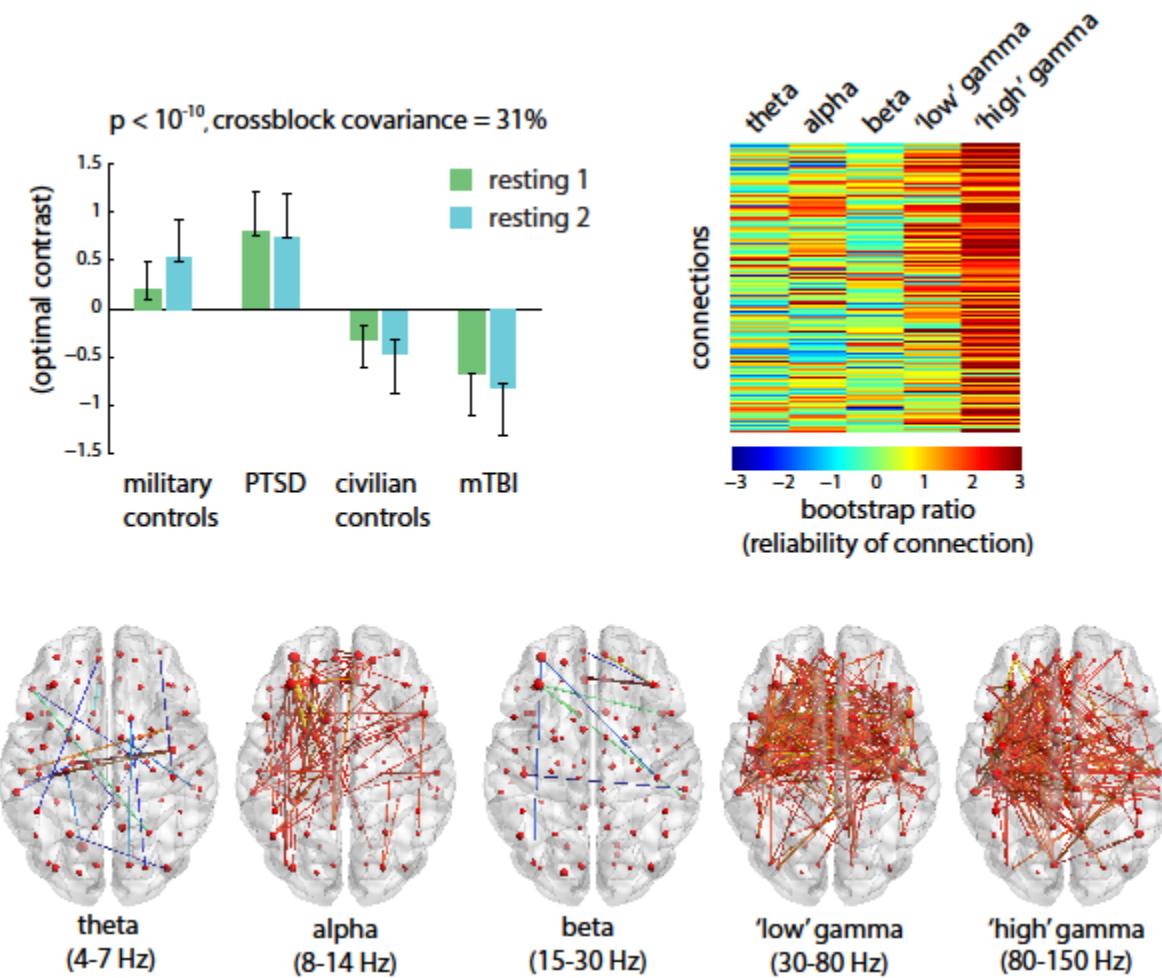


X

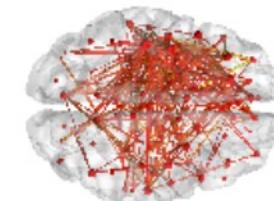
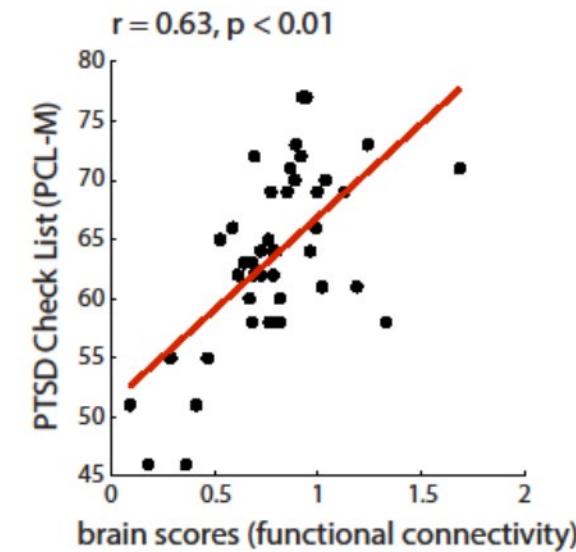


$F_i = XU_i$

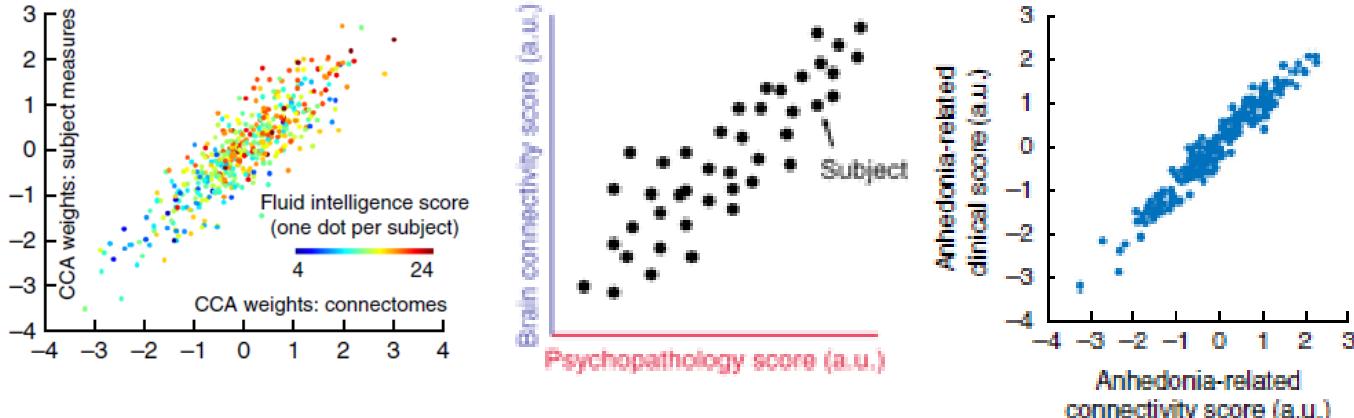
Example: connectivity differences in PTSD and mTBI



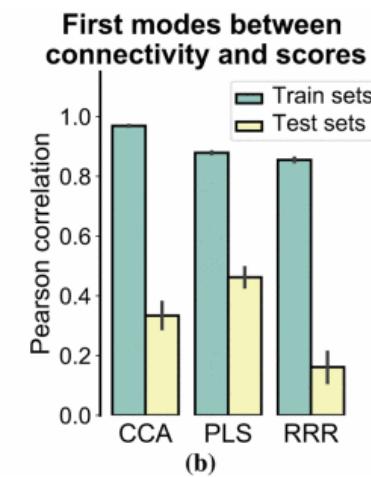
- Do soldiers with PTSD and mTBI differ in patterns of functional connectivity?
- Do individual differences in connectivity relate to symptom severity?



Cross-validation and prediction



- the analysis is explicitly trying to maximize correlations between scores
- need to cross-validate model:
 - do scores correlate out-of-sample?
 - do weights correlate out-of-sample?
 - can you predict individual variables?



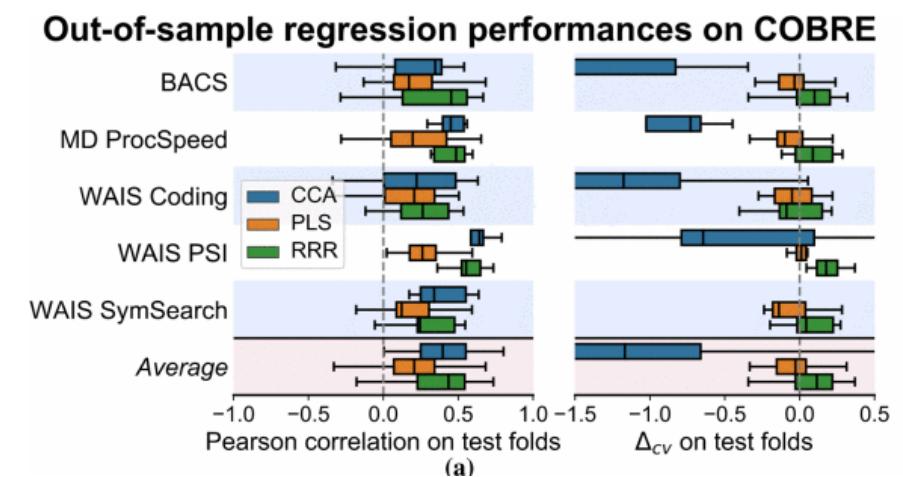
Thomas Yeo
@bttyeo

CCA and PLS are the new voodoo correlations. What do you think? [@GaelVaroquaux](#) [@danilobzdok](#)

7:43 PM · Nov 16, 2017 · Twitter for iPhone

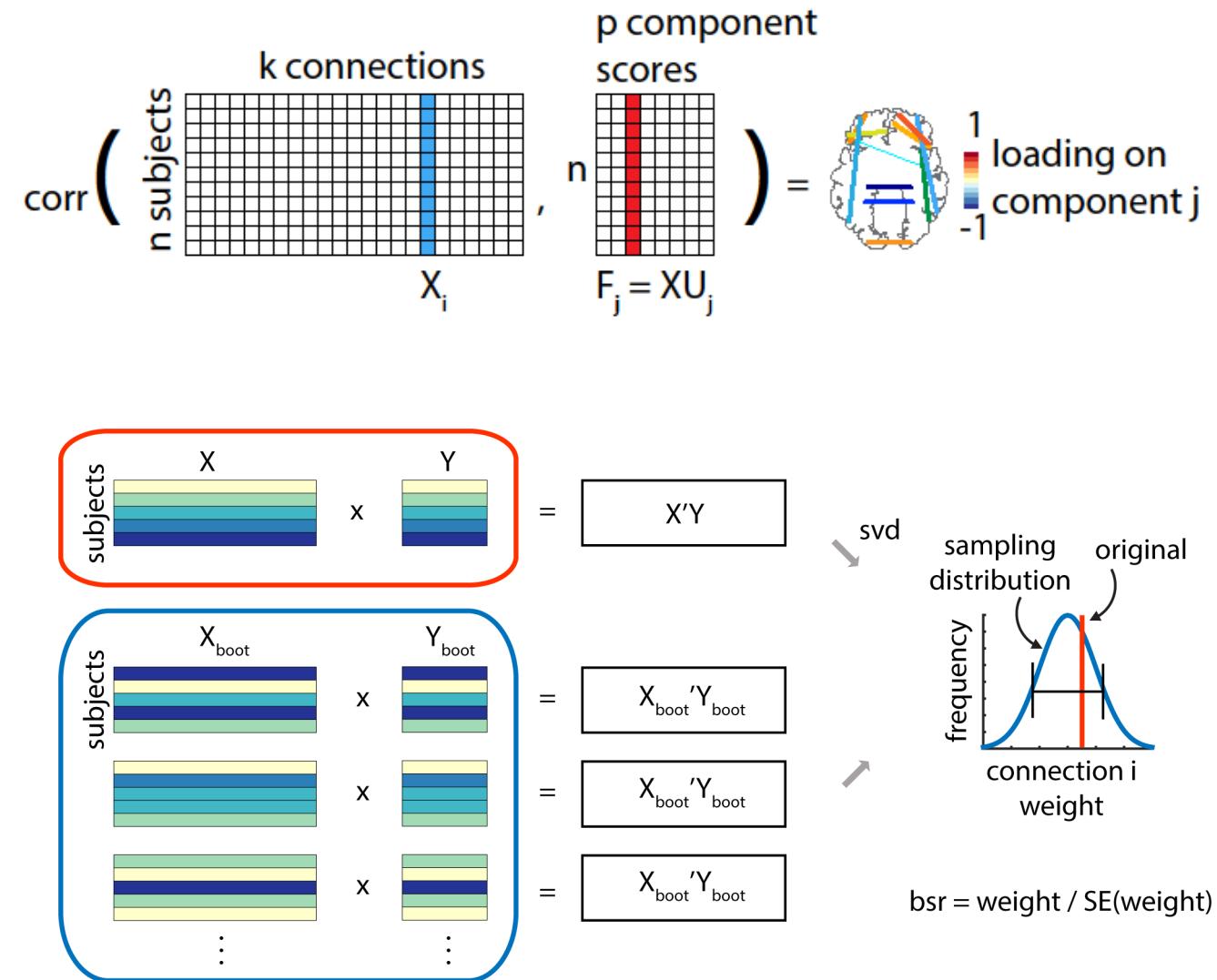
5 Retweets 23 Likes

Reply Retweet Like Share

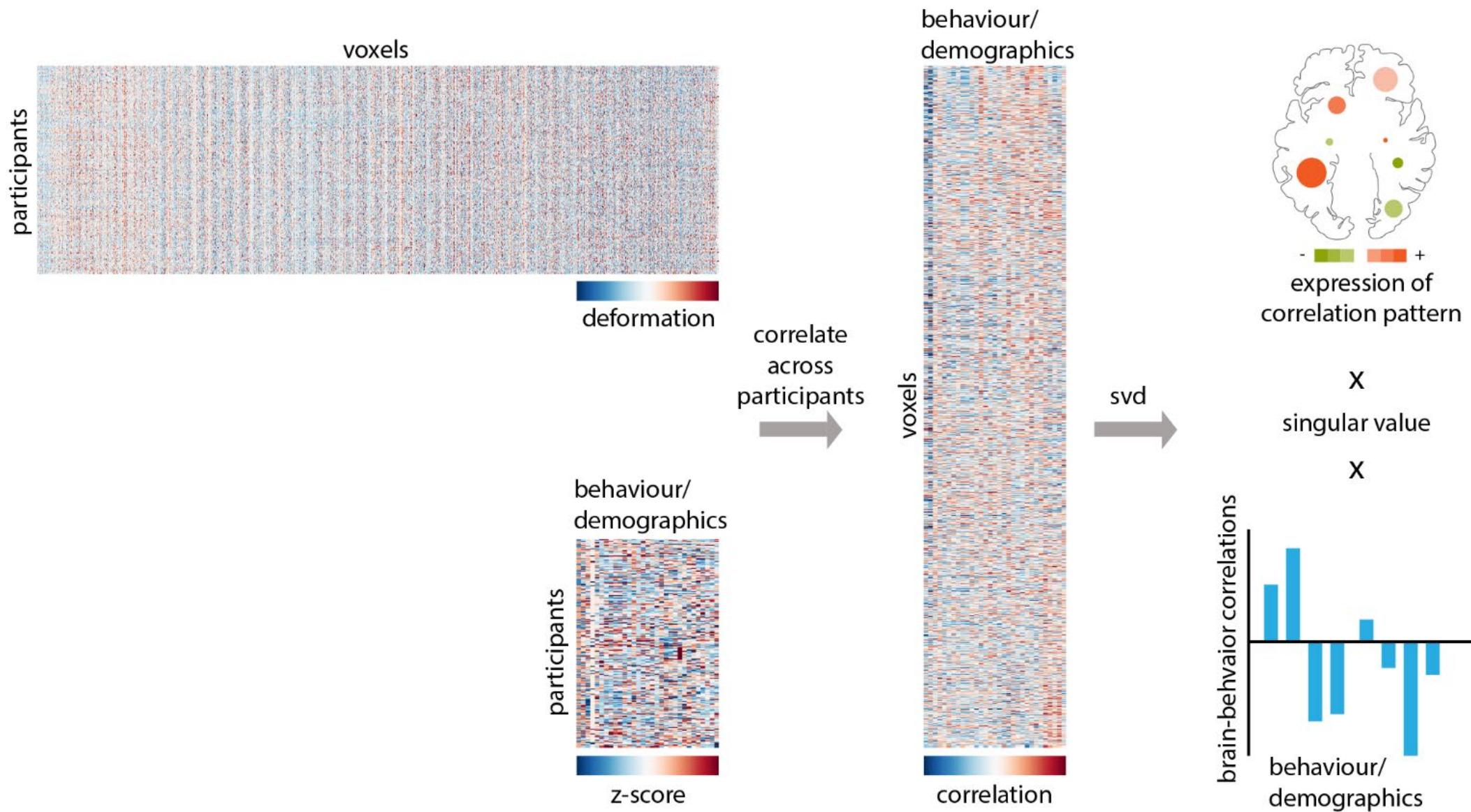


Which variables are important?

- **singular vector weight**: how much should I weigh the variables?
- **loading**: does expression of the original variable correlate with expression of the latent variable?
- **"bootstrap ratio"**: which original variables are weighed highly but also stable across participants?

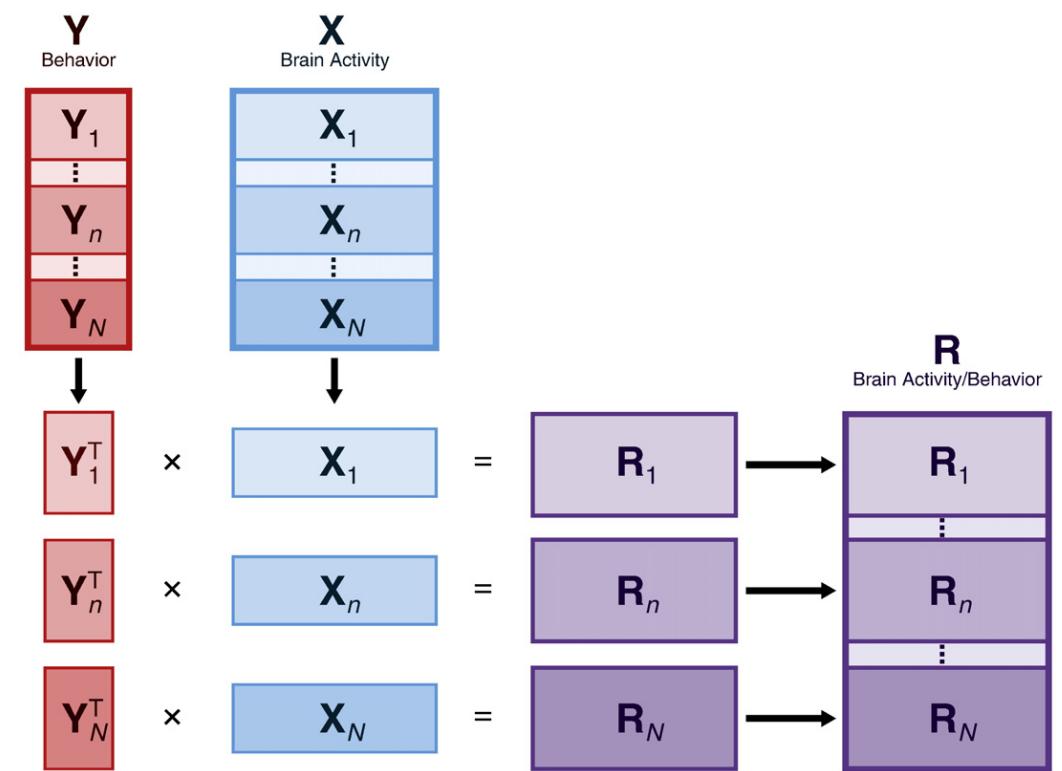
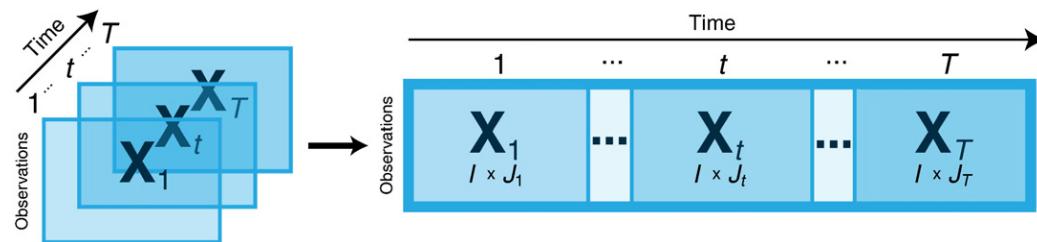


Example: clinical-anatomical signature of Parkinson's disease



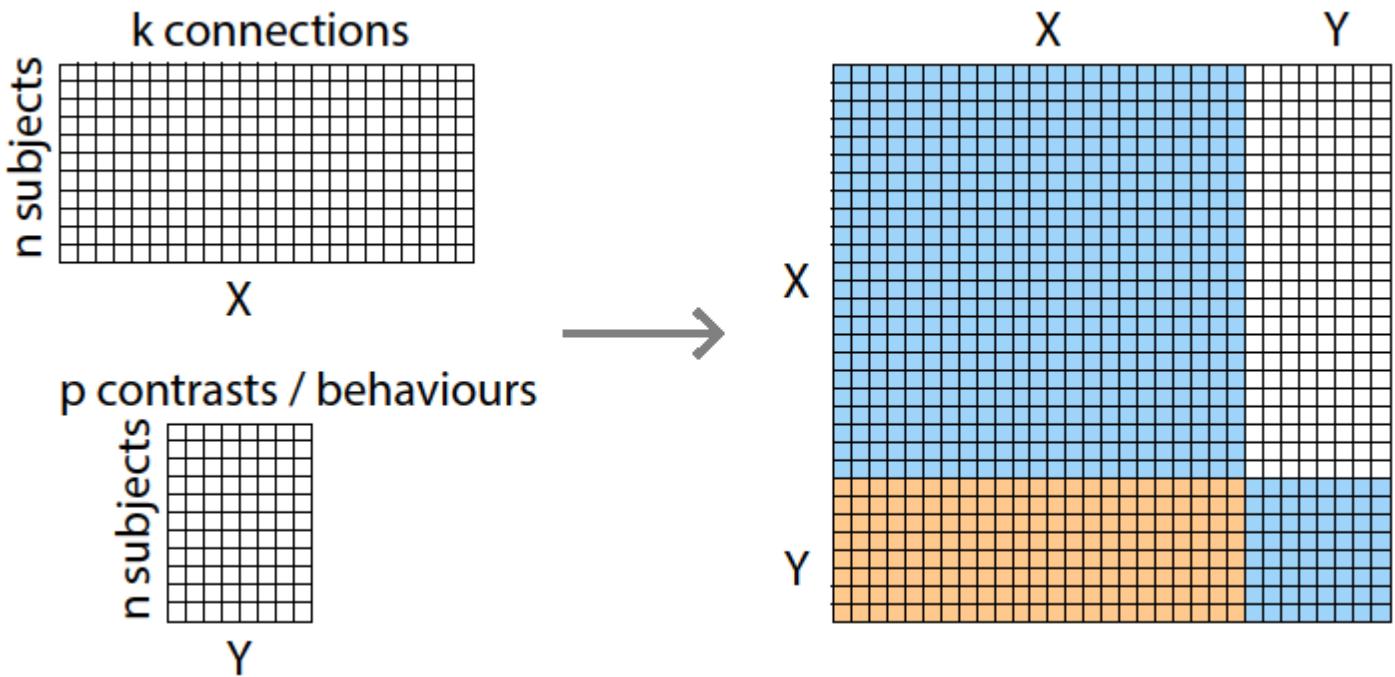
Adding dimensions

- concatenate dimensions (e.g. time + space)
- concatenate within- and between-subject manipulations (e.g. conditions and groups)



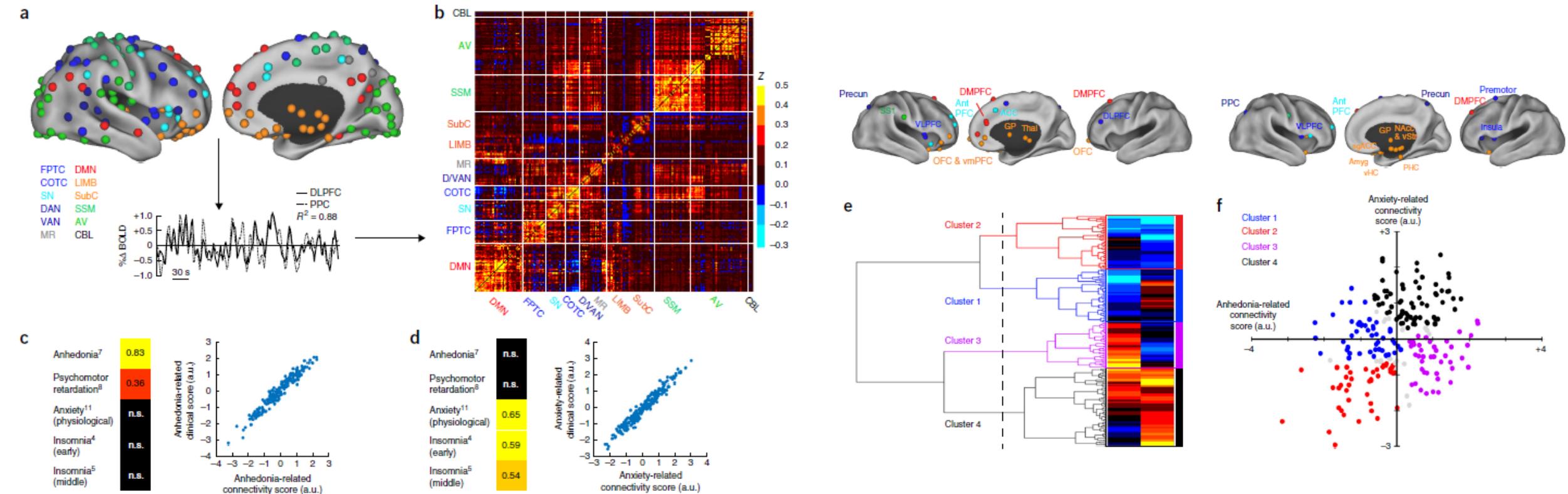
Canonical Correlation Analysis (CCA)

- partial least squares (**PLS**): decompose covariance between **X** and **Y**
 - $SVD(\mathbf{X}'\mathbf{Y})$
- canonical correlation analysis (**CCA**): correct for within-**X** and within-**Y** correlations, then decompose
 - $SVD \left((\mathbf{X}'\mathbf{X})^{-1/2} (\mathbf{X}'\mathbf{Y}) (\mathbf{Y}'\mathbf{Y})^{-1/2} \right)$
- requires #observations > #variables, so typically preceded by some form of dimensionality reduction (PCA, ICA)



Example #1: connectivity- and behaviour-defined biotypes of depression

- **X** = functional connectivity, **Y** = behaviour
- Use CCA to reduce dimensionality, then cluster

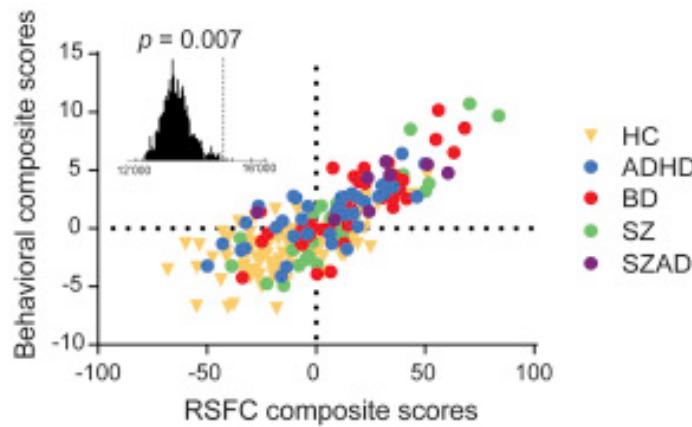


Example #2: transdiagnostic signatures of affective disorders

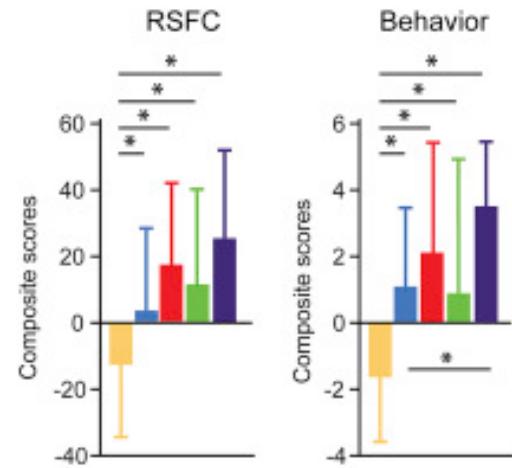
- **X = functional connectivity, Y = behaviour**
- UCLA phenomics consortium:
 - ADHD
 - Bipolar
 - Schizophrenia
 - Schizoaffective

LC1 - General psychopathology

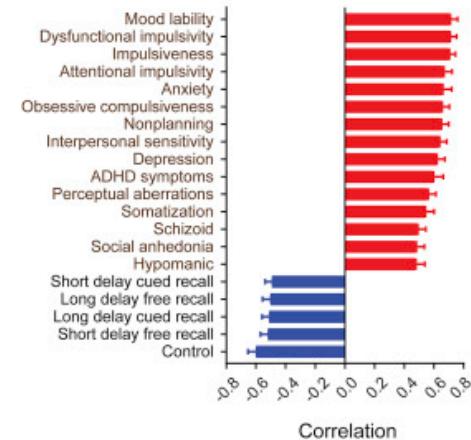
A Correlation between composite scores



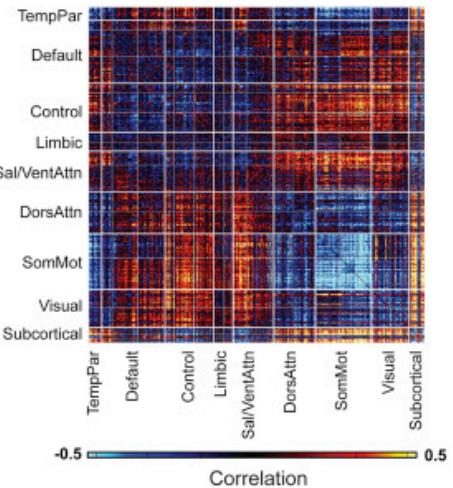
B Group differences in composite scores



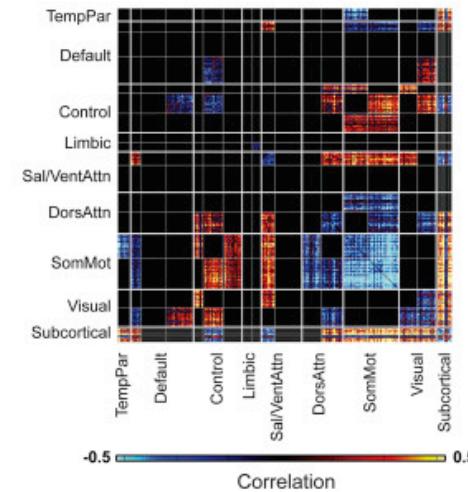
C Behavioral loadings



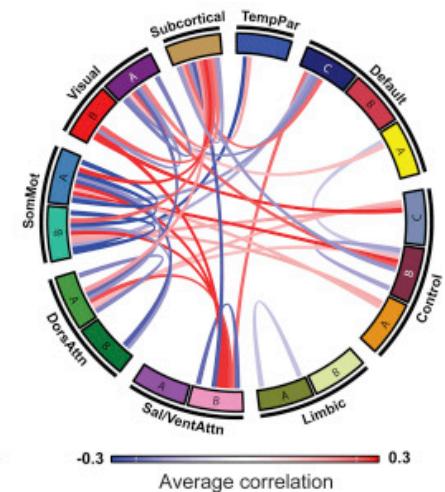
D Unthresholded RSFC loadings



E Thresholded RSFC loadings

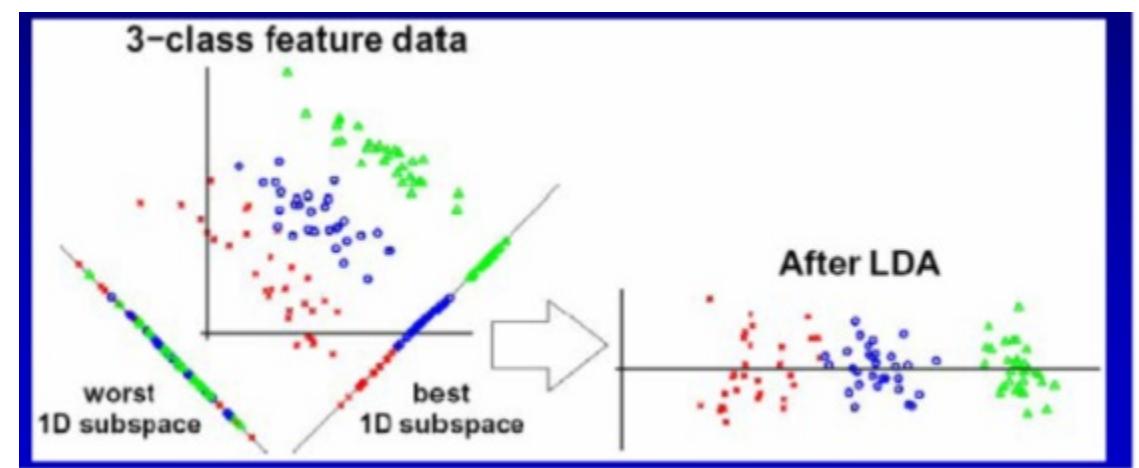


F Significant RSFC network loadings



Linear Discriminant Analysis (LDA)

- \mathbf{X} = continuous, \mathbf{Y} = categorical
- maximize between-group (**B**) vs within-group (**W**) variance
- i.e. find new axes that maximize separation between groups
- $SVD(\mathbf{W}^{-1}\mathbf{B})$



Extensions

- **sparse/regularized** solutions, e.g.

sCCA: Witten & Tibshirani (2009) Stat Appl Genet Mol Biol
 MSNR: Xia et al. (2019) bioRxiv

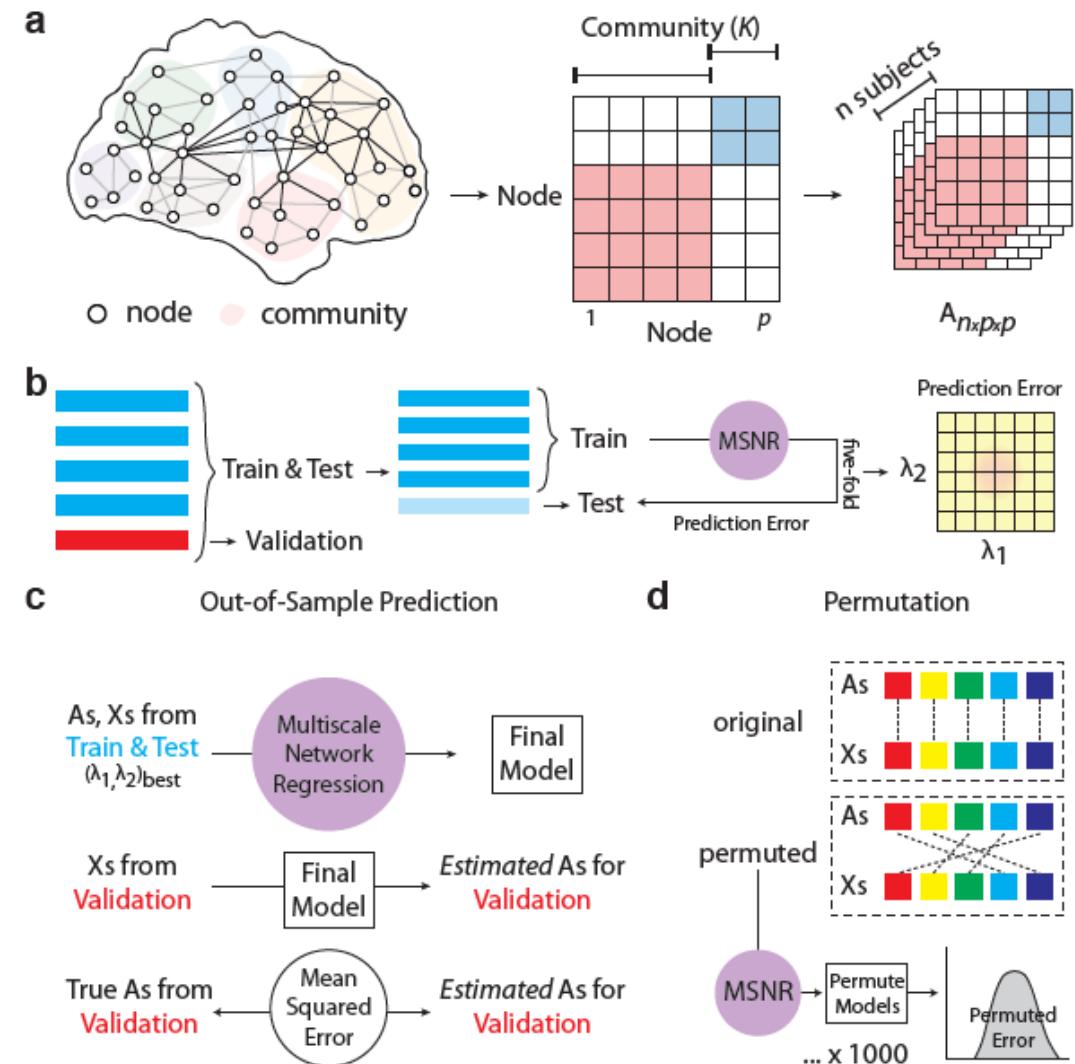
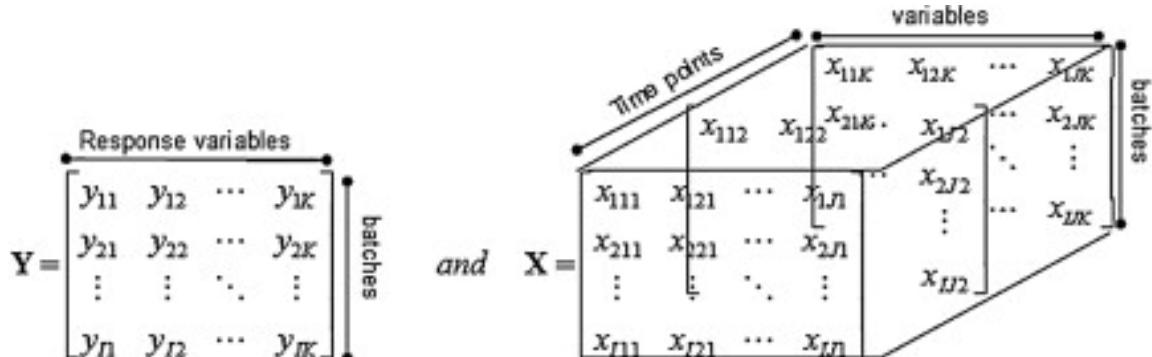
- **3+** datasets, e.g.

PARAFAC: Bro (1997) Chemometr Intell Lab
 Multiway PLS: Wold et al. (1987) J Chemometrics

- **nonlinear** dependencies

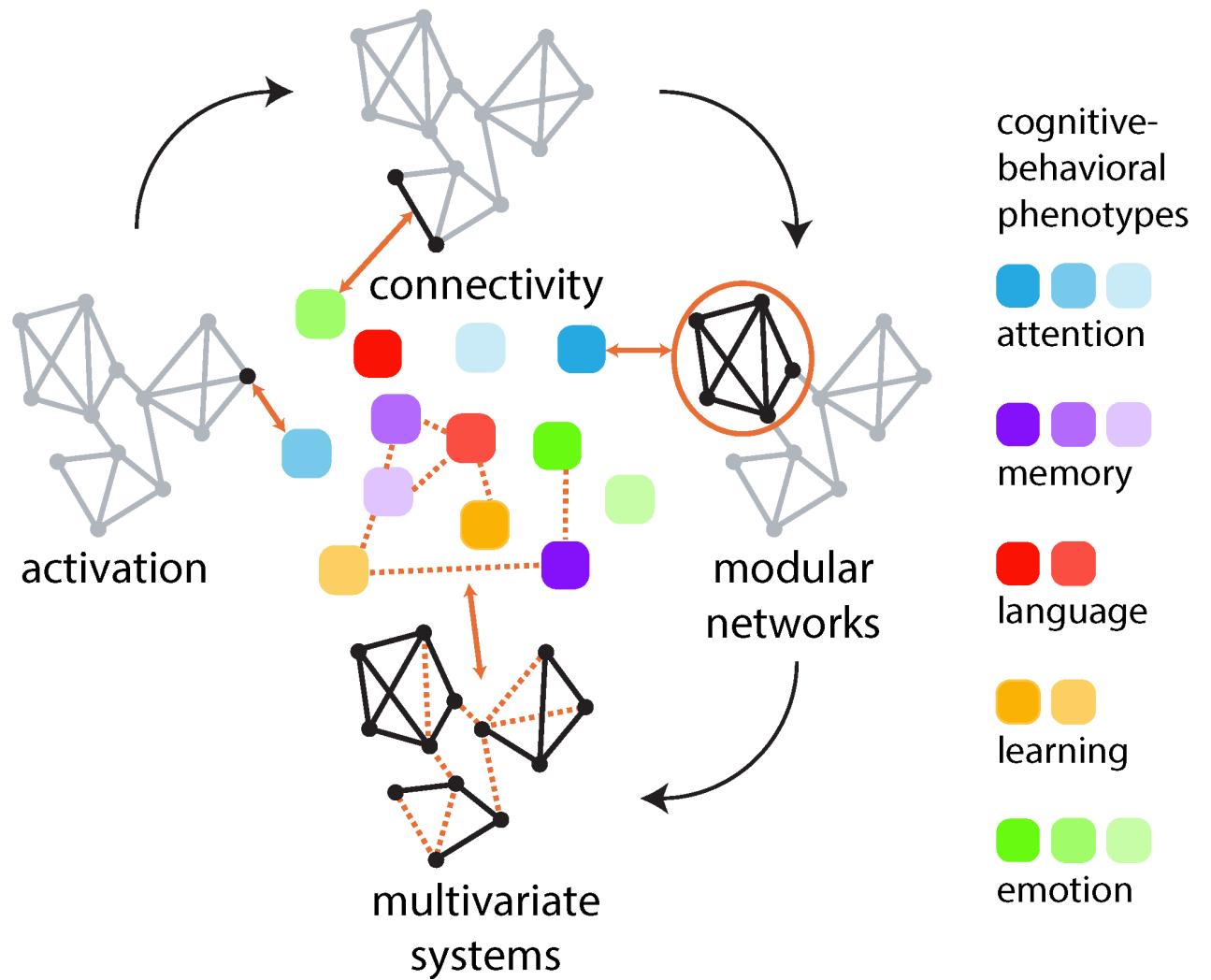
- **bayesian** implementations, e.g.

IBFA: Virtanen et al. (2011) ICML-II



Summary

- Multivariate models embody the **network** property
- Many linear multivariate models are **related**, but entail unique assumptions
- Multivariate models are **versatile** and complementary to other statistical models



- Deep phenotyping in a single individual
- **X** = rs-fMRI functional connectivity
- **Y** = subjective mood (PANAS-X)
- N = 73 sessions
- Question: are connectivity and mood related?

ARTICLE

Received 9 Sep 2015 | Accepted 10 Oct 2015 | Published 9 Dec 2015

DOI: 10.1038/ncomms9885

OPEN

Long-term neural and physiological phenotyping of a single human

Russell A. Poldrack^{1,2,3,4}, Timothy O. Laumann⁵, Oluwasanmi Koyejo⁴, Brenda Gregory³, Ashleigh Hover³, Mei-Yen Chen¹, Krzysztof J. Gorgolewski⁴, Jeffrey Luci^{2,3}, Sung Jun Joo¹, Ryan L. Boyd¹, Scott Hunicke-Smith⁶, Zack Booth Simpson⁷, Thomas Caven⁸, Vanessa Sochat⁹, James M. Shine⁴, Evan Gordon⁵, Abraham Z. Snyder⁵, Babatunde Adeyemo⁵, Steven E. Petersen⁵, David C. Glahn^{10,11}, D. Reese McKay^{10,11}, Joanne E. Curran¹², Harald H.H. Göring¹³, Melanie A. Carless¹³, John Blangero¹², Robert Dougherty¹⁴, Alexander Leemans¹⁵, Daniel A. Handwerker¹⁶, Laurie Frick³, Edward M. Marcotte^{7,17} & Jeanette A. Mumford¹

