


Tutorials

covSTATIS: A multi-table technique for network neuroscience

Giulia Baracchini, Ph.D.¹^a, Ju-Chi Yu, Ph.D.^{2b}, Jenny R. Rieck, Ph.D.³, Derek Beaton, Ph.D.⁴, Vincent Guillemot, Ph.D.⁵, Cheryl L. Grady, Ph.D.^{3,6}, Hervé Abdi, Ph.D.⁷, Robert N. Spreng, Ph.D.^{1,8,9}

¹ Department of Neurology and Neurosurgery, Montreal Neurological Institute and Hospital, McGill University, ² Campbell Family Mental Health Institute, Centre for Addiction and Mental Health, ³ Rotman Research Institute at Baycrest, ⁴ Data Science & Advanced Analytics, Unity Health Toronto, ⁵ Bioinformatics and Biostatistics Hub, Institut Pasteur, Université Paris Cité, ⁶ Departments of Psychiatry and Psychology, University of Toronto, ⁷ School of Behavioral and Brain Sciences, The University of Texas at Dallas, ⁸ McConnell Brain Imaging Centre, Montreal Neurological Institute and Hospital, McGill University, ⁹ Departments of Psychiatry and Psychology, McGill University

Keywords: network neuroscience, similarity analyses, open tools, multivariate methods

<https://doi.org/10.52294/001c.124518>

Aperture Neuro

Vol. 4, 2024

Similarity analyses between multiple correlation or covariance tables constitute the cornerstone of network neuroscience. Here, we introduce covSTATIS, a versatile, linear, unsupervised multi-table method designed to identify structured patterns in multi-table data, and allow for the simultaneous extraction and interpretation of both individual and group-level features. With covSTATIS, multiple similarity tables can now be easily integrated, without requiring *a priori* data simplification, complex black-box implementations, user-dependent specifications, or supervised frameworks. Applications of covSTATIS, a tutorial with Open Data and source code are provided. CovSTATIS offers a promising avenue for advancing the theoretical and analytic landscape of network neuroscience.

Correlation, covariance and distance matrices are among the most commonly used data types in network neuroscience.^{1–4} These matrices are typically built via pairwise comparisons of functional and/or structural neuroimaging data. In these matrices, one entry stores a numerical value quantifying the similarity between two spatial locations (i.e., brain voxels, vertices, regions, channels) and the pattern of these entries reflects an estimate of brain network organization.

In network neuroscience, matrices—also called here data tables—are typically obtained from sets of variables collected on the same individuals (e.g., multiple scans or sessions, multiple imaging modalities),^{5–9} or from the same variables collected on different individuals (e.g., one type of imaging scan on several participants).^{10–13} Data tables are then compared to one another to assess temporal network structure,^{14–17} multi-modal network organization,^{4, 18,19} individual differences,^{20–25} and group or population effects.^{26–30} Data tables are also contrasted with one another to investigate the statistical reliability of patterns derived from network neuroscience methods.^{31,32} Similarity analyses among multiple data tables thus constitute the cornerstone of network neuroscience research.

In network neuroscience, similarity analyses are most often conducted via mass univariate approaches contrasting one edge at a time, or on aggregate information within data tables (e.g., graph theory analyses³³) or across (e.g., categorical groupings), or on single tables with reduced dimensions.³⁴ While these approaches reduce the high dimensionality of network neuroscience data and simplify the analytic landscape, they may limit the statistical robustness of network neuroscience findings and obscure important properties of brain function that could be revealed from the use of multivariate methods applied to full data tables. Multivariate approaches able to align and compare relational information from full data tables across multiple observations can augment the utility, precision, and applicability of network neuroscience data and advance our understanding of brain network organization in health and disease.

Statistical methods exist—called multi-table methods^{35–43}—explicitly designed for multivariate similarity analyses of full data tables. Their goal is to identify structured patterns within preserved high-dimensional multi-table data, and explain and visualize their statistical dependencies. Multi-table methods serve as the basis of network investigations across scientific disciplines,⁴⁴ yet they are

^a Co-first author

Corresponding Author:

Giulia Baracchini (giulia.baracchini@sydney.edu.au)

^b Co-first author

not well known in network neuroscience and therefore remain underused.

Network neuroscience presently counts a few multi-table methods, including machine learning and deep learning tools,^{45,46} graph neural networks,^{47,48} multi-layer and multiplex network approaches,^{18,49} similarity network fusion techniques⁵⁰ and non-linear matrix decomposition algorithms.^{51–53} These approaches have been applied to a variety of research questions about brain network organization both in health and disease.^{54–61} Yet, they often yield complex results challenging to interpret, potentially because these methods rely on complex mathematical implementations^{62,63} and supervised analytical frameworks⁵¹ that do not allow results to be traced back to the original data. There is therefore a pressing need in network neuroscience for multi-table methods that preserve data fidelity and enhance interpretability. CovSTATIS solves this problem by analyzing intact data tables in a linear, unsupervised manner, thus allowing for the simultaneous extraction and interpretation of both individual and group-level features.

The covSTATIS method (and its variant DISTATIS) is a three-way extension of multidimensional scaling and Principal Component Analysis.^{64–66} The name, covSTATIS, combines “covariance” with “STATIS” (a French acronym for “structuring three-way statistical tables”). CovSTATIS takes as input symmetric, positive semi-definite matrices (i.e. symmetric matrices such as cross-product, covariance and correlation matrices, with non-negative eigenvalues) and assesses their similarity.^{67–69} While covSTATIS is specifically designed for correlation/covariance matrices, there exists an equivalent approach for distance matrices called DISTATIS.⁶⁴ CovSTATIS and DISTATIS belong to the STATIS family of multi-table approaches.

In covSTATIS, the pair-wise similarity among I correlation/covariance matrices of dimension $J \times J$ (with J being the number of variables) is quantified by the R_V coefficient – a measure analogous to a squared Product-Moment correlation coefficient with values in the interval $[0,1]$.^{67,70} Given I correlation/covariance matrices, the R_V coefficients are stored in an $I \times I$ similarity matrix \mathbf{C} , where the rows/columns correspond to the I data tables (Figure 1, step 1). Next, covSTATIS performs an eigenvalue decomposition (EVD) on the \mathbf{C} matrix and takes its first eigenvector (of dimension $I \times 1$) to derive weights for each data table. Note that because the R_V is always positive all the entries for the first eigenvector of \mathbf{C} are positive—a consequence of the Perron-Frobenius theorem. The first eigenvector maximally explains the variance in \mathbf{C} and quantifies how similar each table is to the common pattern. Weights for each data table are derived by scaling the first eigenvector to sum to 1. Higher weights identify tables that are more similar to the common pattern, whereas lower weights identify tables less similar to the common pattern. These weights are then used to linearly combine the data tables by multiplying each table by its weight and summing across all the weighted tables. This step generates a $J \times J$ weighted group matrix—called the *compromise matrix* (Figure 1, step 2)—which is next decomposed by EVD. Orthogonal components are extracted from this second EVD and serve as

the main output of covSTATIS (Figure 1, step 3). Components reveal the similarity between J variables with regards to the compromise—the ensemble of the similarity patterns across all I data tables. For each component, *global factor scores* of dimension $J \times L$ (with L being the number of components) capture the relationship between variables with respect to such compromise. For each component and each of the I correlation/covariance matrices, $J \times L$ *partial factor scores* can be obtained to project table-specific relationships between variables onto the same component space (Figure 1, step 4). By quantifying the deviation of each partial factor score from its corresponding global factor scores, we can assess differences between each table and the compromise.

For example, given a covSTATIS analysis of multiple functional connectivity matrices where only positive connectivity values are considered, *global factor scores* represent the brain regions on the component space and illustrate the associations in their connectivity profiles across the whole sample. *Partial factor scores* represent the brain regions of each individual’s connectivity matrix and illustrate how regions are associated with each other in relation to the group pattern. Factor scores of any two components can be used as coordinates to draw scatter plots in the component space, where the distance between two scores represents their similarity. Two *global factor scores* close to each other indicate high similarity in their respective connectivity patterns across the whole sample, while a *partial factor score* close to its corresponding global factor score represents high similarity between an individual’s regional connectivity profile and the regional profile from the whole sample. In sum, covSTATIS provides an unsupervised, linear framework to characterize the similarity among sets of correlation/covariance matrices, and it allows for a one-to-one mapping between input (i.e., whole set of tables and single tables) and output (i.e., *global* and *partial factor scores*). This approach can both identify group-level patterns as well as provide individual-specific expressions of the patterns.

CovSTATIS has both commonalities and differences compared with other dimensionality reduction methods used in neuroscience: (1) Principal Component Analysis (PCA)^{71,72} and Multidimensional Scaling (MDS)⁷³—these two techniques incorporate a single data table; PCA is performed on an observation-by-variable table, and MDS on a dissimilarity matrix between observations. They rely on the same dimensionality reduction technique as covSTATIS, but covSTATIS extends MDS by analyzing multiple similarity matrices; (2) Multiple Factor Analysis (MFA)^{74,75} and STATIS⁶⁴—these are other component-based methods that incorporate multiple data tables, and they are particularly suitable for rectangular data structures. Such data structures are less common in network neuroscience, making MFA and STATIS less desirable methods than covSTATIS. CovSTATIS combines MDS and STATIS to deal with multiple squared, symmetric, correlation/covariance matrices—the standard format of network neuroscience data; (3) Representation Similarity Analysis (RSA)⁷⁶—this method computes a dissimilarity matrix, akin to the matrix of R_V coef-

covSTATIS: a multi-table method for network neuroscience

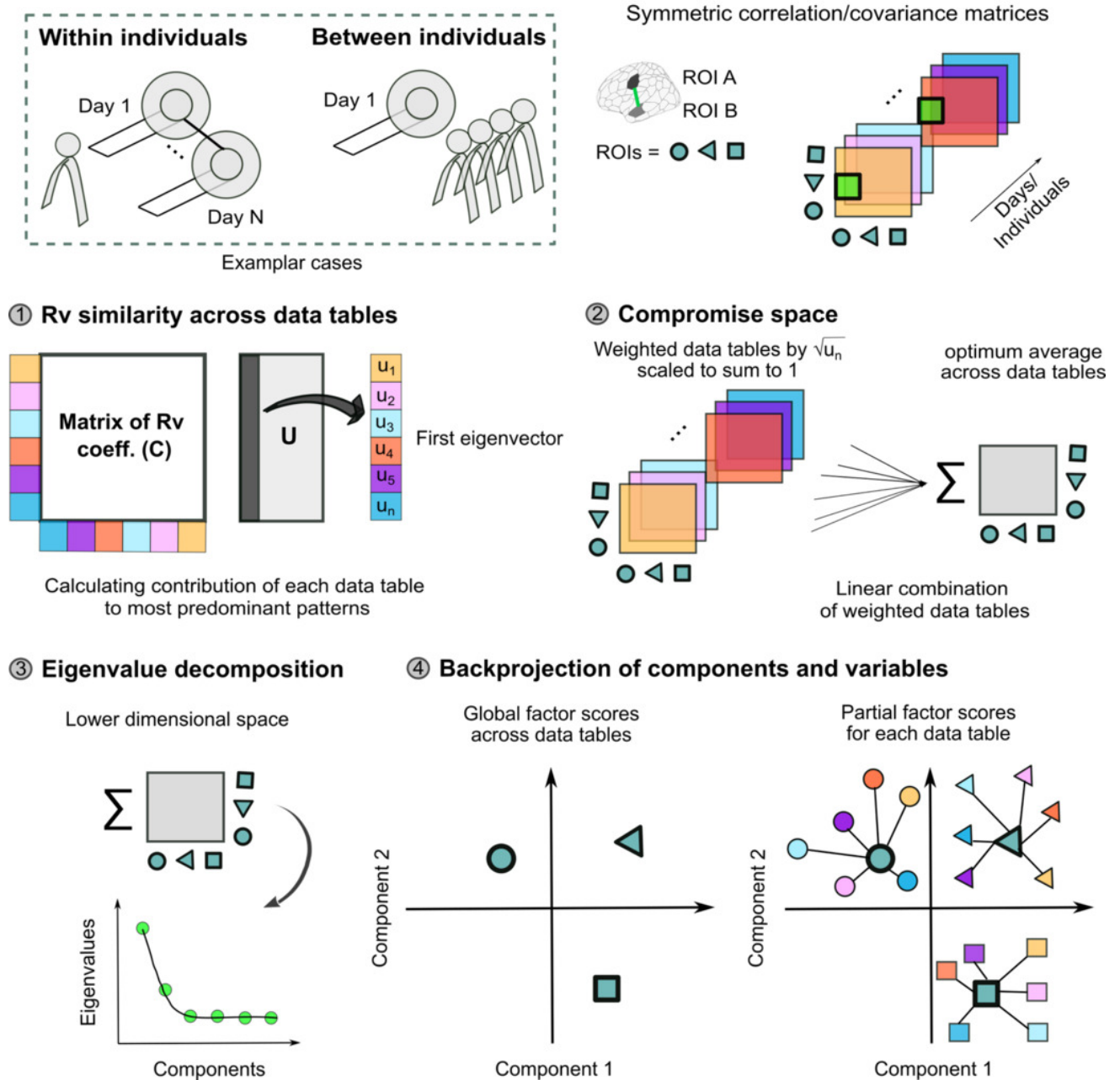


Figure 1. CovSTATIS is used to analyze multiple correlation/covariance matrices obtained either within or between individuals. We provide an example using functional connectivity matrices, collected on several individuals, as input to covSTATIS. (1) First, covSTATIS combines all connectivity matrices by quantifying their overall similarity via their R_v coefficients. These coefficients are then stored in the R_v matrix (C). Next, covSTATIS uses the first eigenvector (u_1) of the R_v matrix to derive weights for each connectivity matrix. (2) With these weights, covSTATIS computes the linear combination of all matrices to generate a common space, the compromise, which best represents the connectivity pattern across the sample. (3) The compromise then undergoes eigenvalue decomposition and orthogonal components are extracted to characterize the variance in the whole-sample connectivity pattern. (4) The variables of the compromise (illustrated by different shapes of green dots; i.e., individual brain regions) are represented as global factor scores in the component space. Global factor scores represent the connectivity pattern of each brain region across the entire sample. The same variables from each individual matrix can also be back projected onto the same space as partial factor scores (indicated by points with the same shape of different colors). Partial factor scores represent the connectivity pattern of each brain region for a specific individual. Importantly, the weighted means of all partial factor scores of a given variable equal to their global factor scores (i.e., barycentric property). In this component space, the distance between factor scores provides meaningful and interpretable information about the similarity in the connectivity profile of any two brain regions. The closer the (global or partial) factor scores of two brain regions, the more similar their connectivity profiles.

ficients in covSTATIS, from a set of correlation/covariance matrices. RSA extracts components from this dissimilarity matrix to characterize differences between the correlation/covariance matrices. These components are the main output of RSA. CovSTATIS, on the other hand, derives weights from these components to compute the compromise—the linear combination of all correlation/covariance matrices—and further extracts components from it. These components are the main outputs of covSTATIS; (4) Graph Theory approaches³³—while used as meaningful statistical descriptors of network neuroscience data, they operate on aggregated pairwise information (e.g., connectivity), not on full data tables; and (5) Gradient Analysis^{51–53}—this method is the most similar to covSTATIS in terms of functionality and implementation. It offers complementary outputs to covSTATIS, ones that are however further removed from the original data. Gradient Analysis relies on non-linear dimensionality reduction methods of affinity matrices of correlation/covariance profiles. Given that Gradient Analysis does not construct a compromise-like space, each correlation/covariance matrix needs to be projected onto the affinity-derived component space to infer similarity/differences in the data. CovSTATIS, instead, obtains the component space from a direct decomposition of the compromise—a linear combination of the original correlation/covariance matrices. As such, covSTATIS provides a multivariate framework that promotes more interpretable links between the output and the original data.

There are numerous other dimensionality reduction techniques, each having its particular strengths and weaknesses. Though almost any dimensionality reduction approach can be applied to neural data, we believe that covSTATIS is particularly well suited for the specific structure and format of network neuroscience data for its flexibility to linearly handle multiple symmetric correlation/covariance tables.

APPLICATIONS OF COVSTATIS IN NETWORK NEUROSCIENCE

In neuroimaging, STATIS-based methods have been applied in a limited capacity.^{77–83} As such, the potential of covSTATIS as a tool for network neuroscience remains largely untapped. Recent work from our group applied covSTATIS to compare spatial patterns of fMRI connectivity across task states,⁸⁴ and to estimate resting-state fMRI connectivity dynamics.⁸⁵ Here, we guide the reader through the former application and provide a step-by-step tutorial with data⁸⁶ and code. The tutorial can be accessed here: https://giuliabaracc.github.io/covSTATIS_netneuro/pages/tutorial.html.

Briefly, in this application of covSTATIS, we used task-based fMRI functional connectivity data from a healthy adult lifespan sample of 144 individuals, to examine how fMRI-derived functional connectivity reconfigures across task conditions with different cognitive load. Note that in the original paper we had 3 different tasks,⁸⁴ but here and in the tutorial, we focus on only one (i.e., n-back with 0-, 1-, 2-back conditions).

Individual-specific functional connectivity matrices were calculated for each task condition. These tables were weighted based on their respective R_V similarity coefficients, and linearly combined to create the *compromise matrix*—the optimum weighted average of all connectivity tables across task conditions and individuals. The compromise was then submitted to an eigenvalue decomposition to assess the similarity of the connectivity patterns across task conditions and subjects. *Global factor scores* represented the average connectivity profile of single brain regions across task conditions and individuals, and *partial factor scores* represented how the regional connectivity profile for each individual and task condition mapped onto the group average. In sum, covSTATIS allowed us to estimate (1) regional differences in functional connectivity patterns across all tasks and individuals, and (2) individual differences in functional connectivity patterns across task conditions. A more detailed breakdown of all steps involved in covSTATIS can be found in our tutorial, including guidelines in the choice of covSTATIS' parameters based on the type of input data.

Other examples of potential applications of covSTATIS include investigations of individual and group differences in spatial and/or temporal network structure in health and disease (**Figure 2, bottom panel**), deep phenotyping of connectivity metrics, and multimodal assessments of network measures within and across individuals (**Figure 2, top panel**). Another promising avenue for covSTATIS is the exploration of brain-behavior relationships within a single framework. Through covSTATIS, participants' correlations—computed from high dimensional brain and behavioral data tables—can be integrated in a unified compromise space from which the shared variance between tables can be extracted.

Importantly, as covSTATIS only requires symmetric positive semi-definite matrices, it provides a general framework to assess shared and distinct patterns across many type of similarity matrices, beyond functional connectivity. With proper matrix preprocessing steps (e.g., taking the Laplacian of a connectivity graph), covSTATIS can analyze other network matrices, such as structural connectivity, similarly to other methods (e.g., network portrait divergence,⁸⁷ net-simile,⁸⁸ and others, see this paper⁸⁹ for a comprehensive list).

While not exhaustive, these applications of covSTATIS highlight the versatility of the method. We hope that the network neuroscience community will benefit from covSTATIS and collectively further refine and expand the approach. Ongoing developments of covSTATIS can be found on our website: https://giuliabaracc.github.io/covSTATIS_netneuro/.

CONCLUSIONS

CovSTATIS serves as a theoretically and computationally accessible tool for similarity analyses, capable of preserving and integrating high dimensional, complex multi-table data typical of network neuroscience. Its linear, unsupervised, user-independent implementation makes covSTATIS

covSTATIS: examples of applications in network neuroscience

Type of research question	Global factor scores	Partial factor scores
Group comparison/ individual differences	Group means of ROIs/networks	ROI/network configuration per group or individual
Deep phenotyping	Session means of ROIs/networks	ROI/network configuration per session
Task/condition differences	Tasks/condition means of ROIs/networks	ROI/network configuration per task/condition
Multimodal neuroimaging	Modality means of ROI/networks	ROI/network configuration per neuroimaging modality
Temporal profiling of network structure (e.g., dynamic functional connectivity)	Means of time chunks (e.g., across sliding windows) of ROI/networks	ROI/network configuration per time chunk (e.g. per window)

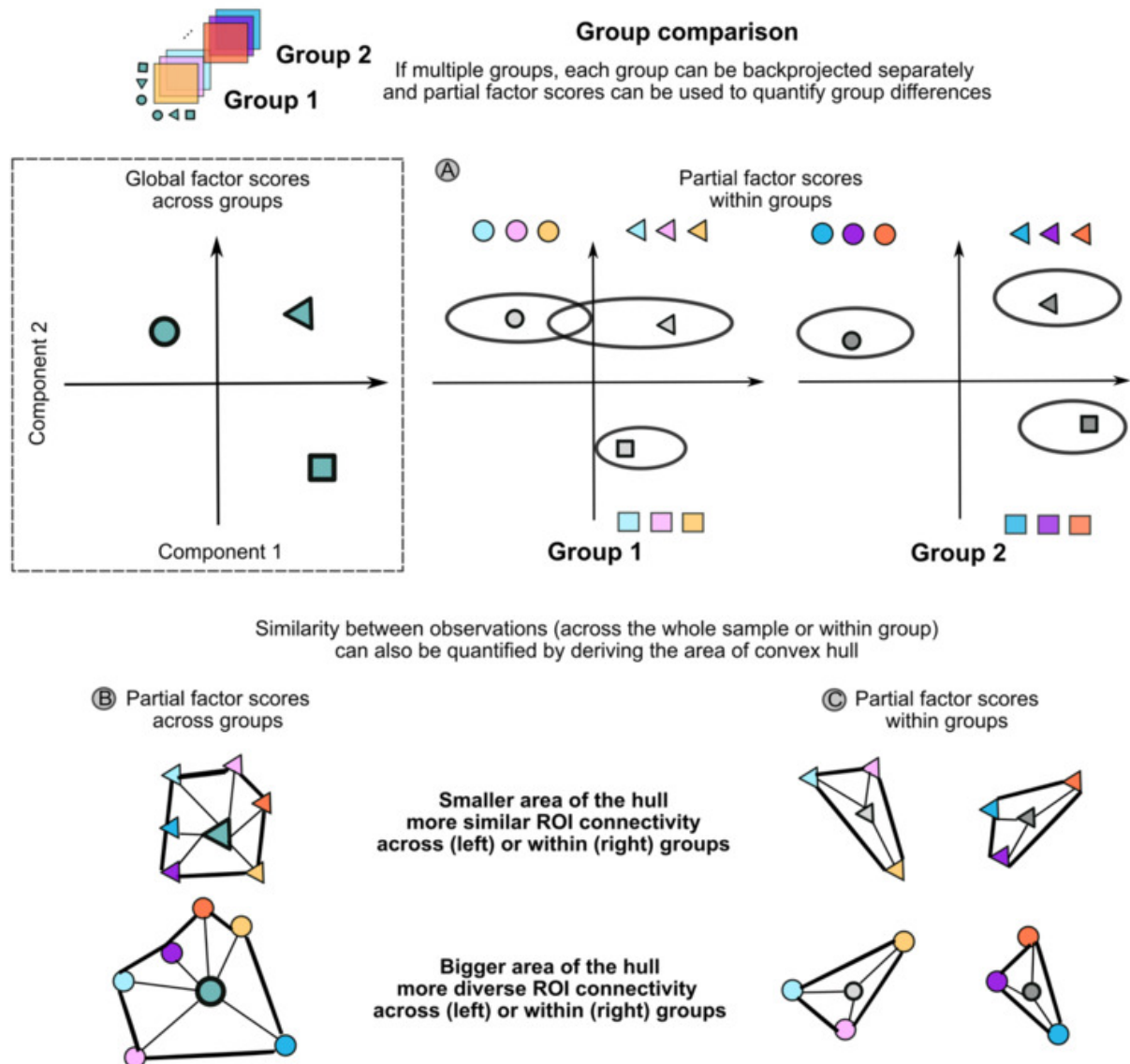
Examples of extractable features

Figure 2. Top panel: examples of applications of covSTATIS in network neuroscience. Bottom panel: examples of extractable features from covSTATIS. For instance, (A) illustrates how we can extract, from global factor scores, group means of partial factor scores, derive their bootstrap confidence intervals, and use them to interpret group differences in network configurations. (B) demonstrates how we can quantify the overall heterogeneity among all partial factor scores via computing the area of the hull. (C) shows how such heterogeneity can also be evaluated for different groups separately.

a more interpretable and versatile tool compared to other commonly used approaches, ultimately paving the path for new discoveries in network neuroscience.

METHODS

NOTATIONS

A matrix is denoted by a bold, uppercase letter (e.g., \mathbf{X}), a vector is denoted by a bold, lowercase letter (e.g., \mathbf{x}), and an element of a matrix is denoted by a lowercase italic letter (e.g., x). The cardinal of a set is denoted by an uppercase italic letter (e.g., I). Given I data tables, we used the subscript i to identify individual data tables (e.g., \mathbf{X}_i). The boldface capital letter \mathbf{I} denotes the identity matrix. The transpose of a matrix is denoted by the superscript \top (e.g., \mathbf{X}^\top).

The j th column of matrix \mathbf{X} is denoted by \mathbf{x}_j , and the value on the k th row and the j th column is denoted by $\mathbf{x}_{i,j}$. For an $I \times J$ matrix, the minimum of I and J is the largest possible rank, denoted by L , of \mathbf{X} . The $\text{trace}(\mathbf{X})$ operator gives the sum of the diagonal elements of the square matrix \mathbf{X} .

COVSTATIS

To generate the compromise space that best represents the common pattern across all data tables (e.g., correlation/covariance matrices), covSTATIS first derives weights from a pairwise similarity matrix, called the \mathbf{R}_V matrix, which quantifies the similarity between data tables via the R_V coefficient. Formally, given two $J \times J$ data tables \mathbf{X}_i and $\mathbf{X}_{i'}$ (e.g., two connectivity matrices with J ROIs from the 2 observations i and i' , e.g., participants or tasks), the R_V coefficient between these two matrices is computed as:

$$R_V = \frac{\text{trace}(\mathbf{X}_i^\top \mathbf{X}_{i'})}{\sqrt{\text{trace}(\mathbf{X}_i^\top \mathbf{X}_i) \text{trace}(\mathbf{X}_{i'}^\top \mathbf{X}_{i'})}} = \frac{\sum_j \sum_{j'} x_{j,j',i} x_{j,j',i'}}{\sqrt{\left(\sum_j \sum_{j'} x_{j,j',i}^2\right) \left(\sum_j \sum_{j'} x_{j,j',i'}^2\right)}} \quad (1)$$

Akin to a squared correlation coefficient, the R_V coefficient takes values in the interval $[0, 1]$. The R_V coefficients between all matrices are then stored in an $I \times I$ \mathbf{R}_V matrix, denoted by \mathbf{C} , where the cell $c_{i,i'}$ stores the value of the R_V coefficient between \mathbf{X}_i and $\mathbf{X}_{i'}$. As \mathbf{C} stores the similarity between data tables, the first component of \mathbf{C} best represents the common pattern across tables, and the first eigenvector of \mathbf{C} (\mathbf{u}_1) quantifies how similar each table is to this common pattern. As a result, to build the compromise space, weights for each data table are derived by \mathbf{u}_1 , rescaled to sum to 1. Formally, \mathbf{C} undergoes an eigenvalue decomposition (EVD):

$$\mathbf{C} = \mathbf{U}\mathbf{\Omega}\mathbf{U}^\top \text{ such that } \mathbf{U}^\top \mathbf{U} = \mathbf{I}, \quad (2)$$

where $\mathbf{\Omega}$ is an $R \times R$ diagonal matrix of eigenvalues of \mathbf{C} with R denoting the rank of \mathbf{C} , and \mathbf{U} is a $I \times R$ matrix of eigenvectors of \mathbf{C} . Next, the weights of \mathbf{X}_i (denoted by α_i) are obtained as:

$$\alpha_i = \sqrt{\frac{u_{i,1}}{\sum_{i=1}^I u_{i,1}}} = \left(u_{i,1}(\mathbf{1}^\top \mathbf{u}_1)^{-1}\right)^{\frac{1}{2}}, \quad (3)$$

where $u_{i,1}$ is the i th value of \mathbf{u}_1 , a value that corresponds to \mathbf{X}_i . The compromise (\mathbf{X}_+) is then computed as the weighted sum of all data matrices, where

$$\mathbf{X}_+ = \sum_{i=1}^I \alpha_i \mathbf{X}_i, \quad (4)$$

and decomposed by an EVD:

$$\mathbf{X}_+ = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^\top \text{ such that } \mathbf{P}^\top \mathbf{P} = \mathbf{I}, \quad (5)$$

where $\mathbf{\Lambda}$ is an $L \times L$ diagonal matrix of eigenvalues of \mathbf{X}_+ with L denoting the rank of \mathbf{X}_+ , and \mathbf{P} is a $J \times L$ matrix of eigenvectors of \mathbf{X}_+ . From EVD, the *global factor scores* \mathbf{F} (i.e., factor scores from the compromise) are computed as:

$$\mathbf{F} = \mathbf{X}_+ \mathbf{P} \mathbf{\Lambda}^{-\frac{1}{2}} \quad (6)$$

and the *partial factor scores* \mathbf{F}_i (i.e., the factor scores derived from the projection of individual tables onto the compromise) are computed as:

$$\mathbf{F}_i = \mathbf{X}_i \mathbf{P} \mathbf{\Lambda}^{-\frac{1}{2}}. \quad (7)$$

It is worth noting that these partial factor scores have a barycentric property, that is their weighted sums equate to the global factor scores:

$$\mathbf{F} = \sum_i \alpha_i \mathbf{F}_i. \quad (8)$$

THE OPTIMIZATION PROBLEM IN COVSTATIS

Optimization in covSTATIS is a two-part problem. The first part is akin to the optimization problem of principal component analysis, the second part is the same as the optimization problem of eigenvalue decomposition.

FIRST PART

First, weights for each data table are obtained to compute the compromise, such that the similarity between the compromise and all input matrices is maximal. Second, components are computed that best explain the compromise's inertia (i.e., variance in more than two dimensions). Formally, the first optimization problem can be written as the following maximization problem:

$$\arg \max_{\boldsymbol{\alpha}} \sum_{i=1}^I \langle \mathbf{X}_i, \underbrace{\sum_{i'=1}^I \alpha_{i'} \mathbf{X}_{i'}}_{\mathbf{X}_+} \rangle^2 \quad \text{with } \boldsymbol{\alpha}^\top \boldsymbol{\alpha} = 1, \quad (9)$$

and the sum of squared scalar products can be developed and simplified:

$$\begin{aligned} \sum_{i=1}^I \langle \mathbf{X}_i, \mathbf{X}_+ \rangle^2 &= \sum_{i=1}^I \sum_{i'=1}^I \alpha_i \alpha_{i'} \sum_{i''=1}^I \text{trace}\{\mathbf{X}_i \mathbf{X}_{i''}\} \\ &\quad \times \text{trace}\{\mathbf{X}_{i''} \mathbf{X}_{i'}\} \\ &= \sum_{i=1}^I \sum_{i'=1}^I \alpha_i \alpha_{i'} \sum_{i''=1}^I c_{i,i''} \times c_{i'',i'} \\ &= \boldsymbol{\alpha}^\top \mathbf{C}^2 \boldsymbol{\alpha} \end{aligned} \quad (10)$$

Therefore, the solution of the optimization problem defined by Equation 9 is the first eigenvector of \mathbf{C}^2 , which is the same as the first eigenvector of \mathbf{C} . According to the Peron-Frobenius theorem, the elements of $\boldsymbol{\alpha}$ will all have the same sign (chosen as positive). These elements are then

scaled to sum to 1 to ensure that partial factor scores will be barycentric for their respective global factor scores (cf. Equation 8). Note that we ignored the denominator of the R_V in the first line as it is a fixed scalar equal to J^2 and has no effect on the maximization problem.

This optimization problem is similar to the optimization problem of Principal Component Analysis (PCA). In PCA, weights are searched *for each variable* to compute factor scores—which are computed as linear combinations of these variables. In covSTATIS, weights are searched *for each data table* to compute the compromise—the linear combination of these data tables.

SECOND PART

The second optimization problem is equivalent to the optimization problem of an eigen-decomposition. In the eigen-decomposition, for each component, weights (i.e., loadings) of each row/column are searched to compute factor scores (F). Factor scores are linear combinations of the loadings that have the largest possible variance (as evaluated by their associated eigenvalues). This optimization problem can be written as the minimization problem of approximating the sum of squared factor scores to the compromise:

$$\begin{aligned} \arg \min_F \|X_+ - FF^T\|_2^2 \quad \text{such that} \quad F^T F = \Lambda \\ \Leftrightarrow \arg \min_P \|X_+ - PAP^T\|_2^2 \quad \text{such that} \quad P^T P = I. \end{aligned} \quad (11)$$

Here, P is the matrix of eigenvectors and Λ is the diagonal matrix of eigenvalues.

The detailed proofs and descriptions of the optimization problems of covSTATIS can be found in the Appendix of a previous publication.⁶⁴

CREDIT AUTHOR STATEMENT

Giulia Baracchini: Conceptualization, Software, Formal Analysis, Data Curation, Writing – Original Draft, Writing – Review and Editing, Visualization. **Ju-Chi Yu:** Conceptualization, Software, Formal Analysis, Writing – Original Draft, Writing – Review and Editing. **Jenny Rieck:** Software, Formal Analysis, Data Curation, Writing – Review and Editing. **Derek Beaton:** Software, Writing – Review and Editing. **Vincent Guillemot:** Software, Writing – Review

and Editing. **Cheryl Grady:** Resources, Writing – Review and Editing. **Hervé Abdi:** Software, Writing – Review and Editing. **R. Nathan Spreng:** Writing – Review and Editing, Supervision.

DATA AVAILABILITY

Data used in the tutorial are available online (<https://osf.io/hnj7s/>) and are described in detail in our previous publication.⁸⁶

CODE AVAILABILITY

The original source code for covSTATIS can be found here: <https://cran.r-project.org/web/packages/DistatisR/> and its helper file here: <https://cran.r-project.org/web/packages/DistatisR/DistatisR.pdf>. All code used to apply covSTATIS in network neuroscience and replicate our tutorial, along with all documentation, can be accessed here: https://giuliabaracc.github.io/covSTATIS_netneuro/pages/tutorial.html. For the tutorial, a downloadable qmd file can be accessed here: https://github.com/giuliabaracc/covSTATIS_netneuro/blob/main/pages/tutorial.qmd. The following GitHub (https://github.com/giuliabaracc/covSTATIS_netneuro/tree/main) and website (https://giuliabaracc.github.io/covSTATIS_netneuro/) links serve as centralized resources for covSTATIS' applications in network neuroscience.

FUNDING SOURCES

This research was supported in part by grants from the Canadian Institutes of Health Research (CIHR) the Natural Sciences and Engineering Research Council of Canada (NSERC) and the National Institutes of Health (NIA R01 AG068563). G.B. and R.N.S. are supported in part by Fonds de recherche du Québec (FRQS). J.-C.Y. receives grant support from the Discovery Fund of the Centre for Addiction and Mental Health (CAMH).

CONFLICTS OF INTEREST

The authors declare no competing interests.

Submitted: June 12, 2024 CST, Accepted: October 02, 2024 CST



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at <http://creativecommons.org/licenses/by/4.0> and legal code at <http://creativecommons.org/licenses/by/4.0/legalcode> for more information.

REFERENCES

1. Zalesky A, Fornito A, Bullmore E. On the use of correlation as a measure of network connectivity. *NeuroImage*. 2012;60:2096-2106. [doi:10.1016/j.neuroimage.2012.02.001](https://doi.org/10.1016/j.neuroimage.2012.02.001)
2. Evans AC. Networks of anatomical covariance. *NeuroImage*. 2013;80:489-504. [doi:10.1016/j.neuroimage.2013.05.054](https://doi.org/10.1016/j.neuroimage.2013.05.054)
3. Pagani M, Bifone A, Gozzi A. Structural covariance networks in the mouse brain. *NeuroImage*. 2016;129:55-63. [doi:10.1016/j.neuroimage.2016.01.025](https://doi.org/10.1016/j.neuroimage.2016.01.025)
4. Bassett DS, Sporns O. Network neuroscience. *Nat Neurosci*. 2017;20:353-364. [doi:10.1038/nn.4502](https://doi.org/10.1038/nn.4502)
5. Poldrack RA et al. Long-term neural and physiological phenotyping of a single human. *Nat Commun*. 2015;6:8885. [doi:10.1038/ncomms9885](https://doi.org/10.1038/ncomms9885)
6. Gordon EM et al. Precision Functional Mapping of Individual Human Brains. *Neuron*. 2017;95:791-807.e7. [doi:10.1016/j.neuron.2017.07.011](https://doi.org/10.1016/j.neuron.2017.07.011)
7. Taylor JR et al. The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) data repository: Structural and functional MRI, MEG, and cognitive data from a cross-sectional adult lifespan sample. *NeuroImage*. 2017;144:262-269. [doi:10.1016/j.neuroimage.2015.09.018](https://doi.org/10.1016/j.neuroimage.2015.09.018)
8. Bycroft C et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562:203-209. [doi:10.1038/s41586-018-0579-z](https://doi.org/10.1038/s41586-018-0579-z)
9. Royer J et al. An Open MRI Dataset For Multiscale Neuroscience. *Sci Data*. 2022;9:569. [doi:10.1038/s41597-022-01682-y](https://doi.org/10.1038/s41597-022-01682-y)
10. Nooner KB et al. The NKI-Rockland Sample: A Model for Accelerating the Pace of Discovery Science in Psychiatry. *Front Neurosci*. 2012;6. [doi:10.3389/fnins.2012.00152](https://doi.org/10.3389/fnins.2012.00152)
11. Pinho AL et al. Individual Brain Charting, a high-resolution fMRI dataset for cognitive mapping. *Sci Data*. 2018;5:180105. [doi:10.1038/sdata.2018.105](https://doi.org/10.1038/sdata.2018.105)
12. Allen EJ et al. A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nat Neurosci*. 2022;25:116-126. [doi:10.1038/s41593-021-00962-x](https://doi.org/10.1038/s41593-021-00962-x)
13. Spreng RN et al. Neurocognitive aging data release with behavioral, structural and multi-echo functional MRI measures. *Sci Data*. 2022;9:119. [doi:10.1038/s41597-022-01231-7](https://doi.org/10.1038/s41597-022-01231-7)
14. Hutchison RM et al. Dynamic functional connectivity: Promise, issues, and interpretations. *NeuroImage*. 2013;80:360-378. [doi:10.1016/j.neuroimage.2013.05.079](https://doi.org/10.1016/j.neuroimage.2013.05.079)
15. Calhoun VD, Miller R, Pearlson G, Adalı T. The Chronnectome: Time-Varying Connectivity Networks as the Next Frontier in fMRI Data Discovery. *Neuron*. 2014;84:262-274. [doi:10.1016/j.neuron.2014.10.015](https://doi.org/10.1016/j.neuron.2014.10.015)
16. Michel CM, Koenig T. EEG microstates as a tool for studying the temporal dynamics of whole-brain neuronal networks: A review. *NeuroImage*. 2018;180:577-593. [doi:10.1016/j.neuroimage.2017.11.062](https://doi.org/10.1016/j.neuroimage.2017.11.062)
17. Lurie DJ et al. Questions and controversies in the study of time-varying functional connectivity in resting fMRI. *Network Neuroscience*. 2020;4:30-69. [doi:10.1162/netn_a_00116](https://doi.org/10.1162/netn_a_00116)
18. Betzel RF, Bassett DS. Multi-scale brain networks. *NeuroImage*. 2017;160:73-83. [doi:10.1016/j.neuroimage.2016.11.006](https://doi.org/10.1016/j.neuroimage.2016.11.006)
19. Suárez LE, Markello RD, Betzel RF, Misic B. Linking Structure and Function in Macroscale Brain Networks. *Trends in Cognitive Sciences*. Published online 2020:S1364661320300267. [doi:10.1016/j.tics.2020.01.008](https://doi.org/10.1016/j.tics.2020.01.008)
20. Finn ES et al. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nat Neurosci*. 2015;18:1664-1671. [doi:10.1038/nn.4135](https://doi.org/10.1038/nn.4135)
21. Zuo XN et al. Human Connectomics across the Life Span. *Trends in Cognitive Sciences*. 2017;21:32-45. [doi:10.1016/j.tics.2016.10.005](https://doi.org/10.1016/j.tics.2016.10.005)
22. Braun U et al. From Maps to Multi-dimensional Network Mechanisms of Mental Disorders. *Neuron*. 2018;97:14-31. [doi:10.1016/j.neuron.2017.11.007](https://doi.org/10.1016/j.neuron.2017.11.007)
23. Tompson SH, Falk EB, Vettel JM, Bassett DS. Network Approaches to Understand Individual Differences in Brain Connectivity: Opportunities for Personality Neuroscience. *Personality Neuroscience*. 2018;1:e5. [doi:10.1017/pen.2018.4](https://doi.org/10.1017/pen.2018.4)

24. Beaty RE, Seli P, Schacter DL. Network neuroscience of creative cognition: mapping cognitive mechanisms and individual differences in the creative brain. *Current Opinion in Behavioral Sciences*. 2019;27:22-30. [doi:10.1016/j.cobeha.2018.08.013](https://doi.org/10.1016/j.cobeha.2018.08.013)
25. Krendl AC, Betzel RF. Social cognitive network neuroscience. *Social Cognitive and Affective Neuroscience*. 2022;17:510-529. [doi:10.1093/scan/nsac020](https://doi.org/10.1093/scan/nsac020)
26. Ferreira LK, Busatto GF. Resting-state functional connectivity in normal brain aging. *Neuroscience & Biobehavioral Reviews*. 2013;37:384-400. [doi:10.1016/j.neubiorev.2013.01.017](https://doi.org/10.1016/j.neubiorev.2013.01.017)
27. Stam CJ. Modern network science of neurological disorders. *Nat Rev Neurosci*. 2014;15:683-695. [doi:10.1038/nrn3801](https://doi.org/10.1038/nrn3801)
28. Medaglia JD, Lynall ME, Bassett DS. Cognitive Network Neuroscience. *Journal of Cognitive Neuroscience*. 2015;27:1471-1491. [doi:10.1162/jocn_a_00810](https://doi.org/10.1162/jocn_a_00810)
29. Fornito A, Zalesky A, Breakspear M. The connectomics of brain disorders. *Nat Rev Neurosci*. 2015;16:159-172. [doi:10.1038/nrn3901](https://doi.org/10.1038/nrn3901)
30. Zhang J et al. What have we really learned from functional connectivity in clinical populations? *NeuroImage*. 2021;242:118466. [doi:10.1016/j.neuroimage.2021.118466](https://doi.org/10.1016/j.neuroimage.2021.118466)
31. Noble S et al. Multisite reliability of MR-based functional connectivity. *NeuroImage*. 2017;146:959-970. [doi:10.1016/j.neuroimage.2016.10.020](https://doi.org/10.1016/j.neuroimage.2016.10.020)
32. Zhang C, Baum SA, Adduru VR, Biswal BB, Michael AM. Test-retest reliability of dynamic functional connectivity in resting state fMRI. *NeuroImage*. 2018;183:907-918. [doi:10.1016/j.neuroimage.2018.08.021](https://doi.org/10.1016/j.neuroimage.2018.08.021)
33. Rubinov M, Sporns O. Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*. 2010;52:1059-1069. [doi:10.1016/j.neuroimage.2009.10.003](https://doi.org/10.1016/j.neuroimage.2009.10.003)
34. Smith SM et al. A positive-negative mode of population covariation links brain connectivity, demographics and behavior. *Nat Neurosci*. 2015;18:1565-1567. [doi:10.1038/nn.4125](https://doi.org/10.1038/nn.4125)
35. Tucker LR. Some mathematical notes on three-mode factor analysis. *Psychometrika*. 1966;31:279-311. [doi:10.1007/BF02289464](https://doi.org/10.1007/BF02289464)
36. Harshman RA et al. Foundations of the PARAFAC procedure: Models and conditions for an “explanatory” multi-modal factor analysis. *UCLA working papers in phonetics*. 1970;16.1:84.
37. Gower JC. Generalized procrustes analysis. *Psychometrika*. 1975;40:33-51.
38. Loan CFV. The ubiquitous Kronecker product. *Journal of Computational and Applied Mathematics*. 2000;123:85-100. [doi:10.1016/S0377-0427\(00\)00393-9](https://doi.org/10.1016/S0377-0427(00)00393-9)
39. Van Deun K, Smilde AK, Van Der Werf MJ, Kiers HA, Van Mechelen I. A structured overview of simultaneous component based data integration. *BMC Bioinformatics*. 2009;10:246. [doi:10.1186/1471-2105-10-246](https://doi.org/10.1186/1471-2105-10-246)
40. Tenenhaus A, Tenenhaus M. Regularized Generalized Canonical Correlation Analysis. *Psychometrika*. 2011;76:257-284. [doi:10.1007/s11336-011-9206-8](https://doi.org/10.1007/s11336-011-9206-8)
41. Wang B et al. Similarity network fusion for aggregating data types on a genomic scale. *Nat Methods*. 2014;11:333-337. [doi:10.1038/nmeth.2810](https://doi.org/10.1038/nmeth.2810)
42. Haxby JV, Guntupalli JS, Nastase SA, Feilong M. Hyperalignment: Modeling shared information encoded in idiosyncratic cortical topographies. *eLife*. 2020;9:e56601. [doi:10.7554/eLife.56601](https://doi.org/10.7554/eLife.56601)
43. Zhou J et al. Graph neural networks: A review of methods and applications. *AI Open*. 2020;1:57-81. [doi:10.1016/j.aiopen.2021.01.001](https://doi.org/10.1016/j.aiopen.2021.01.001)
44. Fried EI et al. Mental disorders as networks of problems: a review of recent insights. *Soc Psychiatry Psychiatr Epidemiol*. 2017;52:1-10. [doi:10.1007/s00127-016-1319-z](https://doi.org/10.1007/s00127-016-1319-z)
45. Zhou T, Thung K, Zhu X, Shen D. Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis. *Human Brain Mapping*. 2019;40:1001-1016. [doi:10.1002/hbm.24428](https://doi.org/10.1002/hbm.24428)
46. Azevedo T et al. A deep graph neural network architecture for modelling spatio-temporal dynamics in resting-state functional MRI data. *Medical Image Analysis*. 2022;79:102471. [doi:10.1016/j.media.2022.102471](https://doi.org/10.1016/j.media.2022.102471)
47. Zhang X et al. Multi-view graph convolutional network and its applications on neuroimage analysis for parkinson's disease. In: *AMIA Annual Symposium Proceedings*. American Medical Informatics Association; 2018:1147.

48. Sebenius I, Campbell A, Morgan SE, Bullmore ET, Lio P. Multimodal Graph Coarsening for Interpretable, MRI-Based Brain Graph Neural Network. In: *2021 IEEE 31st International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE; 2021:1-6. doi:10.1109/MLSP52302.2021.9690626
49. Casas-Roma J et al. Applying multilayer analysis to morphological, structural, and functional brain networks to identify relevant dysfunction patterns. *Network Neuroscience*. 2022;6:916-933. doi:10.1162/netn_a_00258
50. Markello RD et al. Multimodal phenotypic axes of Parkinson's disease. *npj Parkinsons Dis*. 2021;7:6.
51. Margulies DS et al. Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proc Natl Acad Sci USA*. 2016;113:12574-12579. doi:10.1073/pnas.1608282113
52. Tian Y, Margulies DS, Breakspear M, Zalesky A. Topographic organization of the human subcortex unveiled with functional connectivity gradients. *Nat Neurosci*. 2020;23:1421-1432. doi:10.1038/s41593-020-00711-6
53. Vos de Wael R et al. BrainSpace: a toolbox for the analysis of macroscale gradients in neuroimaging and connectomics datasets. *Commun Biol*. 2020;3:103. doi:10.1038/s42003-020-0794-7
54. De Domenico M. Multilayer modeling and analysis of human brain networks. *GigaScience*. 2017;6. doi:10.1093/gigascience/gix004
55. Muldoon SF, Bassett DS. Network and Multilayer Network Approaches to Understanding Human Brain Dynamics. *Philos of Sci*. 2016;83:710-720.
56. Marquand AF, Haak KV, Beckmann CF. Functional corticostriatal connection topographies predict goal-directed behaviour in humans. *Nat Hum Behav*. 2017;1:0146. doi:10.1038/s41562-017-0146
57. Huntenburg JM et al. A Systematic Relationship Between Functional Connectivity and Intracortical Myelin in the Human Cerebral Cortex. *Cerebral Cortex*. 2017;27:981-997. doi:10.1093/cercor/bhx030
58. Larivière S et al. Multiscale Structure-Function Gradients in the Neonatal Connectome. *Cerebral Cortex*. 2020;30:47-58. doi:10.1093/cercor/bhz069
59. Betzel RF et al. The community structure of functional brain networks exhibits scale-specific patterns of inter- and intra-subject variability. *NeuroImage*. 2019;202:115990. doi:10.1016/j.neuroimage.2019.07.003
60. Paquola C et al. Microstructural and functional gradients are increasingly dissociated in transmodal cortices. *PLoS Biol*. 2019;17:e3000284. doi:10.1371/journal.pbio.3000284
61. Royer J et al. Gradients of brain organization: Smooth sailing from methods development to user community. *Neuroinformatics*. Published online 2024:1-12.
62. Coifman RR et al. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *Proceedings of the National Academy of Sciences*. 2005;102:7426-7431. doi:10.1073/pnas.0500334102
63. Lafon S, Lee AB. Diffusion maps and coarse-graining: a unified framework for dimensionality reduction, graph partitioning, and data set parameterization. *IEEE Trans Pattern Anal Mach Intell*. 2006;28:1393-1403.
64. Abdi H, Williams LJ, Valentin D, Bennani-Dosse M. STATIS and DISTATIS: optimum multitable principal component analysis and three way metric multidimensional scaling. *WIREs Computational Stats*. 2012;4:124-167. doi:10.1002/wics.198
65. Abdi H, Valentin D, Chollet S, Chrea C. Analyzing assessors and products in sorting tasks: DISTATIS, theory and applications. *Food Quality and Preference*. 2007;18:627-640. doi:10.1016/j.foodqual.2006.09.003
66. Abdi H. Congruence: Congruence coefficient, RV coefficient, and Mantel Coefficient. In: Salkind NJ, Dougherty DM, Frey B, eds. *Encyclopedia of Research Design*. Sage; 2010:222-229.
67. Kherif F. Group analysis in functional neuroimaging: selecting subjects using similarity measures. *NeuroImage*. 2003;20:2197-2208. doi:10.1016/j.neuroimage.2003.08.018
68. Shinkareva SV, Ombao HC, Sutton BP, Mohanty A, Miller GA. Classification of functional brain images with a spatio-temporal dissimilarity map. *NeuroImage*. 2006;33:63-71. doi:10.1016/j.neuroimage.2006.06.032
69. Abdi H, Dunlop JP, Williams LJ. How to compute reliability estimates and display confidence and tolerance intervals for pattern classifiers using the Bootstrap and 3-way multidimensional scaling (DISTATIS). *NeuroImage*. 2009;45:89-95. doi:10.1016/j.neuroimage.2008.11.008
70. Escoufier Y. Le Traitement des Variables Vectorielles. *Biometrics*. 1973;29:751. doi:10.2307/2529140

71. Abdi H, Williams LJ. Principal component analysis. *WIREs Computational Stats*. 2010;2:433-459. [doi:10.1002/wics.101](https://doi.org/10.1002/wics.101)
72. Hotelling H. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*. 1933;24:417-441. [doi:10.1037/h0071325](https://doi.org/10.1037/h0071325)
73. Abdi H. Metric Multidimensional Scaling. In: Frey B, ed. *The SAGE Encyclopedia of Research Design*. Sage; 2022:950-958.
74. Escofier B, Pagès J. Multiple factor analysis (AFMULT package). *Computational Statistics & Data Analysis*. 1994;18:121-140. [doi:10.1016/0167-9473\(94\)90135-X](https://doi.org/10.1016/0167-9473(94)90135-X)
75. Abdi H, Williams LJ, Valentin D. Multiple factor analysis: principal component analysis for multitable and multiblock data sets. *WIREs Computational Stats*. 2013;5:149-179. [doi:10.1002/wics.1246](https://doi.org/10.1002/wics.1246)
76. Kriegeskorte N. Representational similarity analysis – connecting the branches of systems neuroscience. *Front Sys Neurosci*. Published online 2008. [doi:10.3389/neuro.06.004.2008](https://doi.org/10.3389/neuro.06.004.2008)
77. Churchill NW et al. Optimizing preprocessing and analysis pipelines for single-subject fMRI. I. Standard temporal motion and physiological noise correction methods. *Human Brain Mapping*. 2012;33:609-627. [doi:10.1002/hbm.21238](https://doi.org/10.1002/hbm.21238)
78. Yourganov G et al. Pattern classification of fMRI data: Applications for analysis of spatially distributed cortical networks. *NeuroImage*. 2014;96:117-132. [doi:10.1016/j.neuroimage.2014.03.074](https://doi.org/10.1016/j.neuroimage.2014.03.074)
79. Sha L et al. The Animacy Continuum in the Human Ventral Vision Pathway. *Journal of Cognitive Neuroscience*. 2015;27:665-678. [doi:10.1162/jocn_a_00733](https://doi.org/10.1162/jocn_a_00733)
80. St-Laurent M, Abdi H, Buchsbaum BR. Distributed Patterns of Reactivation Predict Vividness of Recollection. *Journal of Cognitive Neuroscience*. 2015;27:2000-2018. [doi:10.1162/jocn_a_00839](https://doi.org/10.1162/jocn_a_00839)
81. Connolly J et al. Identification of Resting State Networks Involved in Executive Function. *Brain Connect*. 2016;6:365-374. [doi:10.1089/brain.2015.0399](https://doi.org/10.1089/brain.2015.0399)
82. Mitchell DJ, Cusack R. Semantic and emotional content of imagined representations in human occipitotemporal cortex. *Sci Rep*. 2016;6:20232. [doi:10.1038/srep20232](https://doi.org/10.1038/srep20232)
83. Rundle MM, Coch D, Connolly AC, Granger RH. Dissociating frequency and animacy effects in visual word processing: An fMRI study. *Brain and Language*. 2018;183:54-63. [doi:10.1016/j.bandl.2018.05.005](https://doi.org/10.1016/j.bandl.2018.05.005)
84. Rieck JR, Baracchini G, Nichol D, Abdi H, Grady CL. Reconfiguration and dedifferentiation of functional networks during cognitive control across the adult lifespan. *Neurobiology of Aging*. 2021;106:80-94. [doi:10.1016/j.neurobiolaging.2021.03.019](https://doi.org/10.1016/j.neurobiolaging.2021.03.019)
85. Baracchini G et al. The Biological Role of Local and Global fMRI BOLD Signal Variability in Human Brain Organization. Published online 2023. [doi:10.1101/2023.10.22.563476](https://doi.org/10.1101/2023.10.22.563476)
86. Rieck JR, Baracchini G, Nichol D, Abdi H, Grady CL. Dataset of functional connectivity during cognitive control for an adult lifespan sample. *Data in Brief*. 2021;39:107573. [doi:10.1016/j.dib.2021.107573](https://doi.org/10.1016/j.dib.2021.107573)
87. Bagrow JP, Bollt EM. An information-theoretic, all-scales approach to comparing networks. *Appl Netw Sci*. 2019;4:45. [doi:10.1007/s41109-019-0156-x](https://doi.org/10.1007/s41109-019-0156-x)
88. Berlingerio M et al. Netsimile: A scalable approach to size-independent network similarity. *arXiv preprint arXiv:12092684*. Published online 2012.
89. Faskowitz J et al. Connectome topology of mammalian brains and its relationship to taxonomy and phylogeny. *Front Neurosci*. 2023;16:1044372.