High-level cognition is supported by at least second order dynamic correlations in neural activity patterns

Lucy L. W. Owen[1], Thomas H. Chang[1,2], and Jeremy R. Manning[1,†]

[1]Department of Psychological and Brain Sciences,
Dartmouth College, Hanover, NH
[3]Amazon.com, Seattle, WA
[†]Address correspondence to jeremy.r.manning@dartmouth.edu

August 28, 2019

Abstract

Our thoughts arise from coordinated patterns of interactions between brain structures that change with our ongoing experiences. High-order dynamic correlations in brain activity patterns reflect different subgraphs of the brain's connectome that display homologous lower-level dynamic correlations. We tested the hypothesis that high-level cognition is supported by high-order dynamic correlations in brain activity patterns. We developed an approach to estimating high-order dynamic correlations in timeseries data, and we applied the approach to neuroimaging data collected as human participants either listened to a ten-minute story or a temporally scrambled version of the story, or underwent a resting state scan. We trained across-participants pattern classifiers to decode (in held-out data) when in the session each activity snapshot was collected. We found that classifiers trained to decode from high-order dynamic correlations yielded better performance on data collected as participants listened to the (unscrambled) story. By contrast, classifiers trained to decode data from scrambled versions of the story or during the resting state scan yielded the best performance when they were trained using first-order dynamic correlations or raw activity patterns. We suggest that as our thoughts become more complex, they are supported by higher-order patterns of dynamic network interactions throughout the brain.

Introduction

A central goal in cognitive neuroscience is to elucidate the *neural code*: the mapping between (a) mental states or cognitive representations and (b) neural activity patterns. One means of testing models of the neural code is to ask how accurately that model is able to "translate" neural activity patterns into known (or hypothesized) mental states or cognitive representations (e.g., Haxby et al., 2001; Huth et al., 2016, 2012; Kamitani & Tong, 2005; Mitchell et al., 2008; Nishimoto et al., 2011; Norman et al., 2006; Pereira et al., 2018; Tong & Pratte, 2012). Training decoding models on different types of neural features can also help to elucidate which specific aspects of neural activity patterns are informative about cognition– and, by extension, which types of neural activity patterns might comprise the neural code. For example, prior work has used region of interest analyses to estimate the anatomical locations of specific neural representations (e.g., Etzel et al., 2009), or to compare the relative contributions to the neural code of multivariate activity patterns versus patterns of dynamic correlations between neural activity patterns (e.g., Fong et al., 2019; Manning et al., 2018). An emerging theme in this literature is that cognition is mediated by complex dynamic interactions between brain structures (Bassett et al., 2006; Demertzi et al., 2019; Sporns & Honey, 2006; Turk-Browne, 2013).

Studies of the neural code to date have primarily focused on univariate or multivariate neural patterns (for review see , NormEtal06) or (more recently) on patterns of dynamic first-order correlations (i.e., interactions between pairs of brain structures; Fong et al., 2019; Manning et al., 2018). We wondered what the future of this line of work might hold. For example, is the neural code mediated by higher-order interactions between brain structures? Second-order correlations reflect *homologous* patterns of correlation. In other words, if the changing patterns of correlations between two regions, A and B, are similar to those between two other regions, C and D, this would be reflected in the second-order correlations between (A–B) and (C–D). In this way, second-order correlations identify similarities and differences between subgraphs of the brain's connectome. Analogously, third-order correlations reflect homologies between second-order correlations– i.e., homologous patterns of homologous interactions between brain regions. More generally, higher-order correlations reflect homologies between patterns of lower-order correlations. We can then ask: which "orders" of interaction are most reflective of high-level cognitive processes?

Another central question pertains to the extent to which the neural code is carried by activity patterns that directly reflect ongoing cognition (e.g., following Haxby et al., 2001; Norman et al., 2006), versus the dynamic properties of the network structure itself, independent of specific activity patterns in any given set of regions (e.g., following Bassett et al., 2006). For example, graph theoretic measures such as centrality and degree (Bullmore & Sporns, 2009) may be used to estimate how a given brain structure is "communicating" with other structures, independently of the specific neural representations carried by those structures. If one considers a brain region's graph theoretic position in the network (e.g., its eigenvector centrality) as a dynamic property, one can compare how the positions of different regions are correlated, and/or how those patterns of correlations change over time. We can also compute higher-order patterns in these correlations to characterize homologous subgraphs in the connectome that display similar changes in their constituent brain structures' interactions with the rest of the brain.

To gain insights into the above aspects of the neural code, we developed a computational framework for estimating dynamic high-order correlations in timeseries data. This framework provides an important advance, in that it enables us to examine patterns in higher-order correlations that are computationally intractable to estimate via conventional methods. Given a multivariate timeseries, our framework provides timepoint-by-timepoint estimates of the first-order correlations, second-order correlations, and so on (up to tenth-order correlations in this manuscript). Our approach combines a kernel-based method for computing dynamic correlations in timeseries data with a dimensionality reduction step that projects the resulting dynamic correlations into a low-dimensional space. We explored two dimensionality reduction approaches: principle components analysis (PCA; Pearson, 1901), which preserves an approximately invertable transformation back to the original data; and a second non-invertible algorithm that explored patterns in eigenvector centrality (Landau, 1895). This latter approach characterizes correlations between each feature dimension's relative *position* in the network in favor of the specific activity histories of different features.

We validated our approach using synthetic data where the underlying correlations were known. We then applied our framework to a neuroimaging dataset collected as 125 participants listened to either an audio recording of a ten-minute story or a temporally scrambled version of the story, or underwent a resting state scan (Simony et al., 2016). We used a subset of the data to train across-participant classifiers to decode listening times using a blend of neural features (comprising neural activity patterns, as well as different orders of correlations between those patterns that were inferred using our computational framework). We found that both the PCA-based and eigenvector centrality-based approaches yielded neural patterns that could be used to decode accurately. Both approaches also yielded the best decoding accuracy for data collected during (intact) story listening when high-order (PCA: second-order; eigenvector centrality: fourth-order) dynamic correlation patterns were included as features. When we trained classifiers on the scambled stories or resting state data, only lower-order dynamic patterns were informative to the decoders. Taken together, our results indicate that high-level cognition is supported by high-order dynamic patterns of communication between brain structures.

Methods

Our general approach to comprises four general steps (Fig. 1). First, we derive a kernel-based approach to computing dynamic pairwise correlations in a T (timepoints) by K (features) multivariate timeseries, $\mathbf{X}_0$. This yields a T by $O(K^2)$ matrix of dynamic correlations, $\mathbf{Y}_1$, where each row comprises the upper triangle of the correlation matrix at a single timepoint, reshaped into a row vector (this reshaped vector is $(\frac{K^2-K}{2})$-dimensional). Second, we apply a dimensionality reduction step to project the matrix of dynamic correlations back onto a K-dimensional space. This yields a T by K matrix, $\mathbf{X}_1$, that reflects an approximation of the dynamic correlations reflected in the original data. Third, we use repeated applications of the kernel-based dynamic correlation step to $\mathbf{X}_n$ and the dimensionality reduction step to the resulting $\mathbf{Y}_{n+1}$ to estimate high-order dynamnic correlations. Each application of these steps to a T by K time series $\mathbf{X}_n$ yields a T by K matrix, $\mathbf{X}_{n+1}$, that reflects the dynamic correlations between the columns of $\mathbf{X}_n$. In this way, we refer to n as the *order* of the timeseries, where $\mathbf{X}_0$ (order 0) denotes the original data and $\mathbf{X}_n$ denotes n^{th}-order dynamic

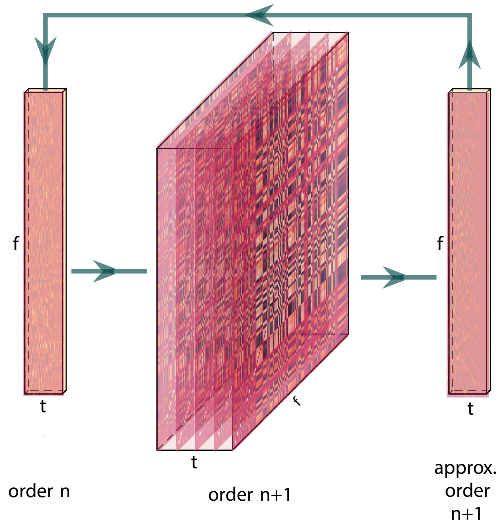

Figure 1: **Estimating dynamnic high-order correlations.** Given a T by K matrix of multivariate timeseries data, $\mathbf{Y}_n$ (where $n \in \mathbb{N}, n \geq 0$), we use Equation 5 to compute a timeseries of K by K correlation matrices, $\mathbf{Y}_{n+1}$. We then approximate $\mathbf{Y}_{n+1}$ with the T by K matrix $\mathbf{X}_{n+1}$. This process may be repeated to scalably estimate iteratively higher-order correlations in the data.

correlations between the columns of $\mathbf{X}_0$. Finally, we use a cross-validation–based decoding approach to evaluate how well information contained in a given order (or weighted mixture of orders) may be used to decode relevant cognitive states. If including a given $\mathbf{X}_n$ in the feature set yields higher classification accuracy on held-out data, we interpret this as evidence that the given cognitive states are reflected in patterns of n^{th}-order correlations.

Kernel-based approach for computing dynamic correlations

Given a matrix of observations, we can compute the (static) Pearson's correlation between any pair of columns, $\mathbf{X}(\cdot, i)$ and $\mathbf{X}(\cdot, j)$ using (Pearson, 1901):

$$\mathrm{corr}(\mathbf{X}(\cdot, i), \mathbf{X}(\cdot, j)) = \frac{\sum_{t=1}^{T} \left(\mathbf{X}(t, i) - \mathbf{X}(\bar{\cdot}, i)\right)\left(\mathbf{X}(t, j) - \mathbf{X}(\bar{\cdot}, j)\right)}{\sqrt{\sum_{t=1}^{T} \sigma^2_{\mathbf{X}(\cdot, i)} \sigma^2_{\mathbf{X}(\cdot, j)}}}, \text{ where} \tag{1}$$

$$\mathbf{X}(\bar{\cdot}, k) = \sum_{t=1}^{T} \mathbf{X}(t, k), \text{ and} \tag{2}$$

$$\sigma^2_{\mathbf{X}(\cdot, k)} = \sum_{t=1}^{T} \left(\mathbf{X}(t, k) - \mathbf{X}(\bar{\cdot}, k)\right)^2 \tag{3}$$

We can generalize this formula to compute time-varying correlations by incorporating a *kernel function* that takes a time t as input, and returns how much the observed data at each timepoint $\tau \in [-\infty, \infty]$ contributes

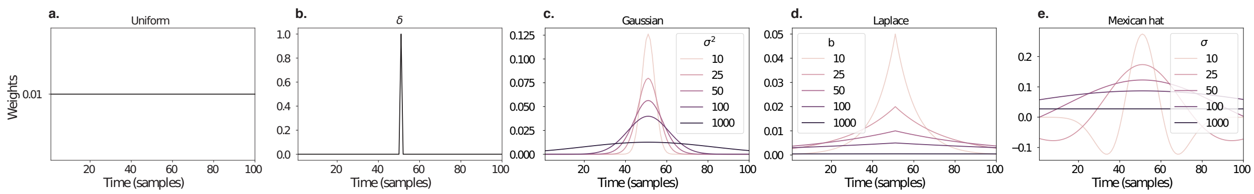

Figure 2: **Examples of kernel functions.** Each panel displays per-timepoint weights at $t = 50$, evaluated for 100 timepoints ($\tau \in [1, ..., 100]$). **a. Uniform kernel.** The weights are timepoint-invariant; observations at all timepoints are weighted equally, and do not change as a function of τ. This is a special case of weight function that reduces dynamic correlations to static correlations. **b. Dirac δ kernel.** Only the observation at timepoint t is given a non-zero weight (of 1). **c. Gaussian kernels.** Each kernel's weights fall off in time according to a Gaussian probability density function centered on time t. Weights derived using several different example variance parameters (σ^2 are displayed. **d. Laplace kernels.** Each kernel's weights fall off in time according to a Laplace probability density function centered on time t. Weights derived using several different example scale parameters (b) are displayed. **e. Mexican hat (Ricker wavelet) kernels.** Each kernel's weights fall off in time according to a Ricker wavelet centered on time t. This function highlights the *contrasts* between local versus surrounding activity patterns in estimating dynamic correlations. Weights derived using several different example width parameters (σ) are displayed.

to the estimated instantaneous correlation at time t (Fig. 2).

Given a kernel function $\kappa_t(\cdot)$ for timepoint t, evaluated at timepoints $\tau \in [1, ..., T]$, we can update the static correlation formula in Equation 2 to estimate the *instantaneous correlation* at timepoint t:

$$\text{timecorr}_{\kappa_t}\left(\mathbf{X}(\cdot, i), \mathbf{X}(\cdot, j)\right) = \frac{\sum_{\tau=1}^{T}\left(\mathbf{X}(\tau, i) - \widetilde{\mathbf{X}}_{\kappa_t}(i)\right)\left(\mathbf{X}(\tau, j) - \widetilde{\mathbf{X}}_{\kappa_t}(j)\right)}{\sqrt{\sum_{\tau=1}^{T}\widetilde{\sigma}^2_{\kappa_t}(\mathbf{X}(\cdot, i))\widetilde{\sigma}^2_{\kappa_t}(\mathbf{X}(\cdot, j))}}, \text{ where} \tag{4}$$

$$\widetilde{\mathbf{X}}_{\kappa_t}(t, k) = \sum_{\tau=1}^{T}\kappa_t(\tau, k)\mathbf{X}(\tau, k), \tag{5}$$

$$\widetilde{\sigma}^2_{\kappa_t}(\mathbf{X}(\cdot, k)) = \sum_{\tau=1}^{T}\left(\mathbf{X}(\tau, k) - \widetilde{\mathbf{X}}_{\kappa_t}(t, k)\right)^2. \tag{6}$$

Here $\text{timecorr}(\mathbf{X}(\cdot, i), \mathbf{X}(\cdot, j), \kappa_t)$ reflects the correlation at time t between columns i and j of $\mathbf{X}$, estimated using the kernel κ_t.

Dynamic inter-subject functional connectivity (DISFC)

Equation 5 provides a means of taking a single observation matrix, $\mathbf{X}_n$ and estimating the dynamic correlations from moment to moment, $\mathbf{Y}_{n+1}$. Suppose that one has access to a set of multiple observation matrices that reflect the same phenomenon. For example, one might collect neuroimaging data from several experimental participants, as each participant performs the same task (or sequence of tasks). Let $\mathbf{X}^1_n$, $\mathbf{X}^2_n$, ..., $\mathbf{X}^P_n$ reflect the T by K observation matrices ($n = 0$) or reduced correlation matrices ($n > 0$) for each of P participants in an experiment. We can use *inter-subject functional connectivity* (ISFC; Simony et al., 2016) to compute the degree of stimulus-driven correlations reflected in the multi-participant dataset at a given timepoint t using:

$$\bar{\mathbf{C}}(t) = M\left(R\left(\frac{1}{2P}\sum_{p=1}^{P} Z\left(Y_n^p(t)\right)^{\mathrm{T}} + Z\left(Y_n^p(t)\right)\right)\right), \tag{7}$$

where M extracts and vectorizes the diagonal and upper triangle of a symmetric matrix, Z is the Fisher z-transformation (Zar, 2010):

$$Z(r) = \frac{\log(1+r) - \log(1-r)}{2} \tag{8}$$

R is the inverse of Z:

$$R(z) = \frac{\exp(2z-1)}{\exp(2z+1)}, \tag{9}$$

and $\mathbf{Y}_n^p(t)$ denotes the correlation matrix (Eqn. 2) between each column of $\mathbf{X}_n^p$ and each column of the average $\mathbf{X}_n$ from all *other* participants, $\bar{\mathbf{X}}_n^{\setminus p}$:

$$\bar{\mathbf{X}}_n^{\setminus p} = R\left(\frac{1}{P-1}\sum_{q\in\setminus p} Z\left(\mathbf{X}_n^q\right)\right), \tag{10}$$

where $\setminus p$ denotes the set of all participants other than participant p. In this way, the T by $\frac{K^2-K}{2}$ DISFC matrix $\bar{\mathbf{C}}$ provides a time-varying extension of the ISFC approach developed by Simony et al. (2016).

Low-dimensional representations of dynamic correlations

Given a T by $\frac{K^2-K}{2}$ matrix of dynamic correlations, $\mathbf{Y}_n$, we propose two general approaches to computing a T by K low-dimensional representation of these correlations, $\mathbf{X}_n$. The first approach uses dimensionality reduction algorithms to project $\mathbf{Y}_n$ onto a K-dimensional space. The second approach uses graph-theoretic measures to characterize the relative positions of each feature ($k \in [1, ..., K]$) in the network defined by the correlation matrix at each timepoint.

Dimensionality reduction-based approaches to computing $\mathbf{X}_n$

The modern library of dimensionality reduction algorithms include Principal Components Analysis (PCA; Pearson, 1901), Probabilistic PCA (PPCA; Tipping & Bishop, 1999), Exploratory Factor Analysis (EFA; Spearman, 1904), Independent Components Analysis (ICA; Comon et al., 1991; Jutten & Herault, 1991), *t*-Stochastic Neighbor Embedding (*t*-SNE; van der Maaten & Hinton, 2008), Uniform Manifold Approximation and Projection (UMAP; McInnes & Healy, 2018), non-negative matrix factorization (NMF; Lee & Seung, 1999), Topographic Factor Analysis (TFA) Manning et al. (2014), Hierarchical Topographic Factor analysis (HTFA) Manning et al. (2018), Topographic Latent Source Analysis (TLSA) Gershman et al. (2011), Dictionary learning (J. Mairal et al., 2009; J. B. Mairal et al., 2009), deep autoencoders (Hinton & Salakhutdinov, 2006), among others. While complete characterizations of each of these algorithms is beyond the scope of the present manuscript, the general intuition driving these approaches is to compute the T by I matrix, $\mathbf{X}$, that is closest to the original T by J matrix, $\mathbf{Y}$, where (typically) $I \ll J$. The different approaches place different constraints on what properties $\mathbf{X}$ must satisfy and which aspects of the data are compared (and how) to characterize the match between $\mathbf{X}$ and $\mathbf{Y}$.

Applying dimensionality reduction algorithms to $\mathbf{Y}$ yields a $\mathbf{X}$ whose columns reflect weighted combinations (or nonlinear transformations) of the original columns of $\mathbf{Y}$. This has two main consequences. First, with each repeated dimensionality reduction, the resulting $\mathbf{X}_n$ has lower and lower fidelity (with respect to what the "true" $\mathbf{Y}_n$ might have looked like without using dimensionality reduction to maintain scalability). In other words, computing $\mathbf{X}_n$ is a lossy operation. Second, whereas each columns of $\mathbf{Y}_n$ may always be mapped directly onto specific pairs of columns of $\mathbf{Y}_{n-1}$, the columns of $\mathbf{X}_n$ reflect weighted combinations and/or nonlinear transformations of the columns of $\mathbf{Y}_n$. Many dimensionality reduction algorithms are invertable (or approximately invertable). However, attempting to map a given $\mathbf{X}_n$ back onto the original feature space of $\mathbf{Y}_0$ will usually require $O(TK^{2n})$ space and therefore quickly becomes intractable as n or K grow large.

Graph theoretic approaches to computing $\mathbf{X}_n$

The above dimensionality reduction approaches to approximating a given $\mathbf{Y}_n$ with a lower-dimensional $\mathbf{X}_n$ preserve a (potentially recombined and transformed) mapping back to the original data in $\mathbf{Y}_0$. We also explore graph theoretic approaches that forgo a preserved mapping back to the original data in favor of preserving each feature's relative *position* in the broader network of interactions and connections. To illustrate the distinction between the two general approaches we explore, suppose a network comprises nodes A, B, and C. If A and B exhibit uncorrelated activity patterns, the functional connection between them will be (by definition) close to 0. However, if A and B each interact with C in similar ways, we might attempt to capture those similarities using a measure that reflects the how A and B interact in the network.

In general, graph theoretic measures take as input a matrix of interactions (e.g., using the above notation, an K by K correlation matrix or binarized correlation matrix reconstituded from a single timepoint's row of

$\mathbf{Y}$) and return as output a set of K measures describing how each node (feature) sits within that correlation matrix with respect to the rest of the population. Widely used measures include betweeness centrality (the proportion of shortest paths between each pair of nodes in the population that involves the given node in question; e.g., Barthélemy, 2004; Freeman, 1977; Geisberger et al., 2008; Newman, 2005; Opsahl et al., 2010); diversity and dissimilarity (characterizations of how differently connected a given node is from others in the population; e.g., Lin, 2009; Rao, 1982; Ricotta & Szeidl, 2006); Eigenvector centrality and pagerank centrality (measures of how influential a given node is within the broader network; e.g., Bonacich, 2007; Halu et al., 2013; Lohmann et al., 2010; Newman, 2008); transfer entropy and flow coefficients (a measure of how much information is flowing from a given node to other nodes in the network; e.g., Honey et al., 2007; Schreiber, 2000); k-coreness centrality (a measure of the connectivity of a node within its local sub-graph; e.g., Alvarez-Hamelin et al., 2005; Christakis & Fowler, 2010); within-module degree (a measure of how many connections a node has to its close neighbors in the network; e.g., Rubinov & Sporns, 2010); participation coefficient (a measure of the diversity of a node's connections to different sub-graphs in the network; e.g., Rubinov & Sporns, 2010); and sub-graph centrality (a measure of a node's participation in all of the network's sub-graphs; e.g., Estrada & Rodríguez-Velázquez, 2005); among others.

For a given graph theoretic measure, $\eta : \mathcal{R}^{K \times K} \to \mathbb{R}^K$, we can use η to tranform each row of $\mathbf{Y}_n$ in a way that characterizes the corresponding graph-theoretic properties of each column. This results in a new T by K matrix, $\mathbf{X}_n$, that reflects how the features reflected in the columns of $\mathbf{Y}_n$ participate in the network during each timepoint (row).

Higher-order correlations

Because $\mathbf{X}_n$ has the same shape as the original data $\mathbf{X}_0$, approximating $\mathbf{Y}_n$ with a lower-dimensional $\mathbf{X}_n$ enables us to estimate high-order dynamic correlations in a scalable way. Given a T by K input matrix, the output of Equation 5 requires $O(TK^2)$ space to store. Repeated applications of Equation 5 (i.e., computing dynamic correlations between the columns of the outputted dynamic correlation matrix) each require exponentially more space; in general the n^{th}-order dynamic correlations of a T by K timeseries occupies $O(TK^{2n})$ space. However, when we approximate or summarize the output of Equation 5 with a T by K matrix (as described above), it becomes feasible to compute even very high-order correlations in high-dimensional data. Specifically, approximating the n^{th}-order dynamic correlations of a T by K timeseries requires only $O(TK^2)$ additional space– the same as would be required to compute first-order dynamic correlations. In other words, the space required to store $n + 1$ multivariate timeseries reflecting up to n^{th} order correlations in the original data scales linearly with n using our approach (Fig. 1). **JRM STOPPED HERE**

Evaluation metrics

We evaluate our approach to extracting dynamic correlations and higher-order correlations using several metrics detailed next. First, we generated synthetic data using known time-varying correlations, and then we evaluated the fidelity with which Equation 5 could recover those correlations (for synthetic datasets with different properties, and using different kernels to define the weights; Fig. 2). We then turned to a series of analyses on a (real) neuroimaging dataset where the ground truth correlations were *not* known. We evaluated whether the recovered correlations could be used to accurately label held-out neuroimaging data with the time at which it was collected. We used this latter evaluations (using timepoint decoding) as a proxy for gauging how much explanatory power the recovered correlations held with respect to the observed data.

Generating synthetic data

To explore recovery of a constant covariance (Fig. 3, a.), we generated syntheic data sampled from a constant covariance matrix. To do this, we created one random covariance matrix, K, with 50 features, and for each of the 300 timepoints we sampled from a Gaussian distribution centered on K. Similarly, we generated synthetic data sampled from a random covariance matrix (Fig. 3, b.) by creating a new random covariance matrix $K(t)$, for each of the 300 timepoints and sampled from a Gaussian distribution centered on $K(t)$.

To generate synthetic data from a dynamically changing covariance matrix (Fig. 3, c.), we generated two random covariance matrices, K_1 and K_2. We then computed a weighted average covariance matrix for each of the 300 timepoint, $K(t)$, by taking the linearly spaced weights (w) of the two random matrices,

$$K(t) = w(t) * K_1 + (1 - w(t)) * K_2, \tag{11}$$

$$\tag{12}$$

and for each of the 300 timepoints sampled from a Gaussian distribution centered on $K(t)$.

Lastly, for the synthetic data containing block structure (Fig. 3, d.), we followed the same process of creating synthetic data sampled from a constant covariance matrix (see above) but sampled from a new random covariance matrix after 60 consecutive timepoints. We then pieced the blocks together to create a sythetic dataset with 300 total timepoints but drawn from 5 separate covariance matrices.

Recovery of ground truth parameters from synthetic data

We applied timecorr, using delta and gaussian (width = 10) kernels Fig. 2) to each of these synthetic datasets, then correlated each recovered correlation matrix with the ground truth. We repeated this process 10 times and explored how recovery varies with the kernel and the specific structure of the data. For the ramping synthetic dataset (Fig. 3, c.) and for the block synthetic dataset (Fig. 3, d.) we made further comparisons of the timecorr recovered correlation matrices. We compared the ramping recovered correlation matrices to only the first random covariance matrix K_1 (First, Fig. 3, c.) and to only the last random covariance matrix K_2 (Last, Fig. 3, c.) from Equation 12. We also compared the block recovered correlation matrices in to the block specific covariance matrix (Block 1-5, Fig. 3, d.).

Timepoint decoding

To explore how higher-order structure varies with stimulus structure and complexity, we used a previous neuroimaging dataset Simony et al. (2016) in which participants listened to an audio recording of a story; 36 particpants listen to an intact version of the story, 17 participants listen to time-scrambled recordings of the same story where paragraphs were scrambled, 36 particpants listen to word-scrambled version and 36 participants lay in rest condition.

Prior work has shown participants share similar neural responses to richly structured stimuli when compared to stimuli with less structure. To assess whether the moment-by-moment higher order correlations were reliably preserved across participants, we used inter-subject functional connectivity (ISFC) to isolate the time-varying correlational structure (functional connectivity patterns that were specifically driven by the story participants listened to. Following the analyses conducted by (HTFA) Manning et al. (2018), we first applied *hierarchical topographic factor analysis* (HTFA) to the fMRI datasets to obtain a time series of 700 node activities for every participant. We then computed the dynamic weighted ISFC using a range of kernels and widths. Specifically, we used Gaussian, Laplace, and mexican hat kernels, as well as widths of 5, 10, 20, and 50. We then approximated these dynamic correlation using two reduction measures, PCA and eigenvector centrality, and computed the dynamic weighted ISFC on the approximations. We repeated this process up to 10th order approximated correlations.

To assess decoding accuracy, we randomly divided participants for each stimulus into training and testing groups. For the zeroth order, we computed the mean factor activity for each group. For all subsequent orders up to the tenth order, we computed the mean approximated dynamic ISFC of factor activity for each group. To assess how additional higher-order correlations contribute to decoding accuracy, for each order we included a weighted-mixutre (described below) of the activity patterns of all previous orders. For each

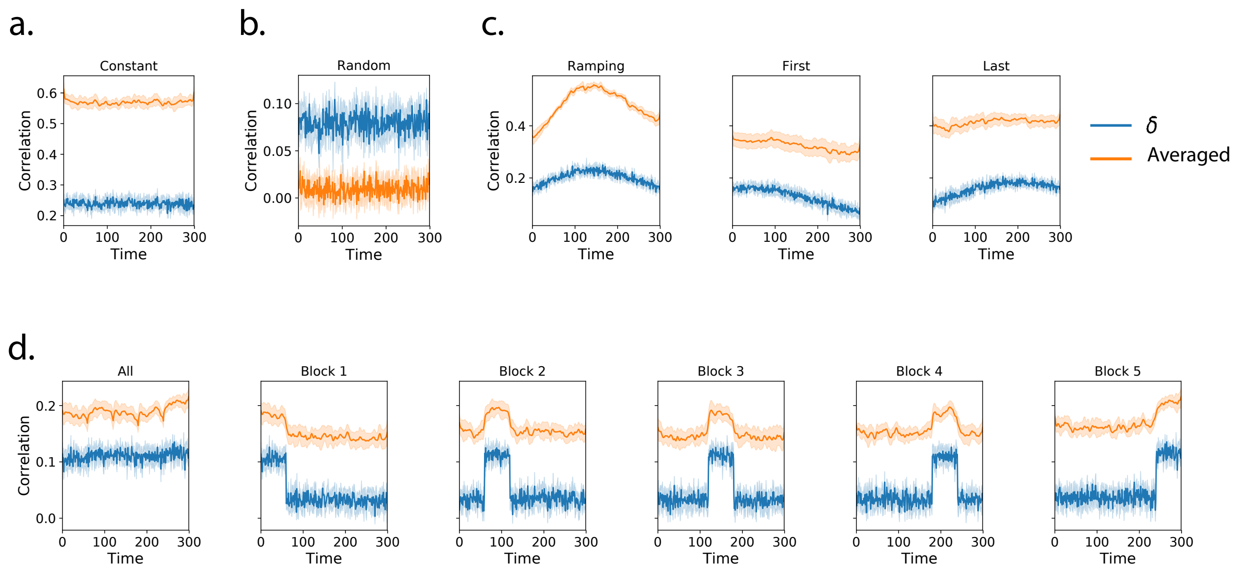

Figure 3: **Dynamic correlation recovery with synthetic data.** Using synthetic data containing different underlying correlational structure, we test how well we can recover dynamic correlation matrices using different kernels when compared to groud truth. We compare the results using a delta kernel with averaged results from several kernels (Gaussian, Laplace, and mexican hat) and several widths (5, 10, 20, and 50). We plot recovery using of datasets containing the following underlying structure: **a. Constant. b. Random. c. Ramping. d. Block.**

group of participants in turn, we compared these activity patterns (using Pearson correlations) to estimate the story times each pattern corresponded to. Specifically, we asked, for each timepoint: what are the correlations between the first group's and second group's activity patterns at each order. We note that the decoding test we used is a conservative in which we count a timepoint label as incorrect if it is not an exact match.

For each order we obtained the weighted-mixture of the correlation matrices for the current order and all previous orders using mixing parameter, ϕ, where $0 < \phi < 1$ reflects a weighted mixture of order based decoding Fig. 4 Panel c.). We calculated ϕ, by subdividing the training group and using the quasi-Newton method of Broyden, Fletcher, Goldfarb, and Shanno (BFGS (Nocedal & Wright, 2006)) for optimization. We repeated this cross-validation process 10 times for each parameter set.

Results

Synthetic data

To assess the performance of dynamic correlation recovery using timecorr, we varied width the kernel and the specific structure of the data. We applied timecorr, using delta and gaussian kernels Fig. 2) to each of the following synthetic datasets: constant, random, ramping, and block. We then correlated each recovered correlation matrix with the ground truth.

For the constant synthetic dataset, a gaussian kernel (width=10) outperformed the delta kernel (Fig. 3, a.). This is in contrast with the random synthetic dataset, for which the delta kernel best captures the rapidly changing structure (Fig. 3, b.). For the ramping synthetic dataset, the slow changing strucutre within the data is best captured by the gaussian kernel and the best recovery occurs in the middle (Ramping, Fig. 3, c.). In addition to comparing the timecorr recovered correlation matrices to the ground truth, we further compared the ramping recovered correlation matrices to only the first random covariance matrix K_1 (First, Fig. 3, c.) and to only the last random covariance matrix K_2 (Last, Fig. 3, c.), both of which perform best at the beginning and end respectively.

Similary for the block sythetic dataset, we compared the timecorr recovered correlation matrices to the ground truth as well as to each block-specific covariance matrix (Block 1-5, Fig. 3, d.). Although the structure is changing by block, the gaussian kernel once again outperforms the delta kernel. Performance does however drop near even boundaries for when using the gaussian kernel.

Neuroimaging dataset (Simony et al., 2016)

For our decoding analysis, we used HTFA-derived node activities Manning et al. (2018) from fMRI data collected as participants listened to an audio recording of a story (intact condition; 36 participants), listened to time scrambled recordings of the same story (17 participants in the paragraph-scrambled condition listened to the paragraphs in a randomized order and 36 in the word-scrambled condition listened to the words in a randomized order), or lay resting with their eyes open in the scanner (rest condition; 36 participants). We sought to demonstrate how higher-order correlations may be used to examine dynamic interactions of brain patterns in (real) multi-subject fMRI datasets. This story listening dataset was collected as part of a separate study, where the full imaging parameters, image preprocessing methods, and experimental details may be found (Simony et al., 2016). The dataset is available at http://arks. princeton.edu/ark:/88435/dsp015d86p269k.

We next evaluated if our model of high-order correlations in brain activity can capture cognitively relevant brain patterns. We performed a decoding analysis, using cross validation to estimate (using other participants' data) which parts of the story each weighted-mixture of higher-order brain activity pattern corresponded to (see *Materials and methods*). We note that our primary goal was not to achieve perfect decoding accuracy, but rather to use decoding accuracy as a benchmark for assessing whether different neural features specifically capture cognitively relevant brain patterns.

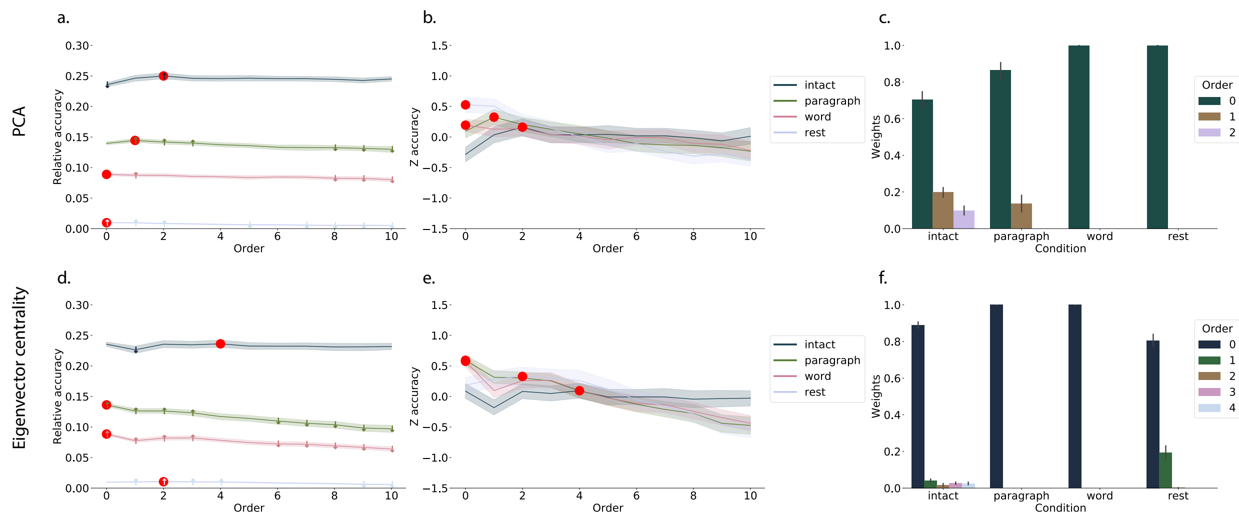

Figure 4: **Decoding by order. a.&d. Relative decoding accuracy by order.** Ribbons of each color display cross-validated decoding performance averaged over all parameters for each condition (intact, paragraph, word, and rest) using PCA (a.) or eigenvector centrality (d.) to approximate correlations. Decoders were trained using increasingly more higher-order information and this ribbons are displayed relative to chance at 0. The red dots indicates maximum decoding accuracy for each condition. **b.&e. Z-transformed decoding accuracy by order.** We Z-transformed the decoding accuracy by order to better visualize the order with the maximum decoding accuracy for each condition using PCA (b.) or eigenvector centrality (e.) to approximate correlations. **c.&f. Optimized weights.** Bar heights indicate the optimized mixing paramete ϕ of each contributing order up to and including the order with the maximum decoding accuracy for each contributing order using PCA (c.) or eigenvector centrality (f.) to approximate correlations. For the order with maximum decoding accuracy by condition, we show barplots of the optimized weights ϕ for each contributing order.

Separately for each experimental condition, we divided participants into two groups. For the zeroth order, we computed the mean factor activity for each group. For all subsequent orders up to the tenth order, we computed the mean approximated dynamic ISFC of factor activity for each group (see *Materials and methods*), and combined in a weighted mixutre with all previous orders (i.e. cross-validation for the second order contained a weighted-mixture of zeroth, first, and second order (Fig. 4, c.&f.). For each order, we correlated the group 1 activity patterns with group 2 activity patterns. We then subdivided the group 1 to obtain an optimal weighting parameter for each order's correlation matrix using the same cross validation method. We used the optimal weighting parameters to obtain a weighted-mixture (see *Materials and methods*) of each order's correlation matrix. Using these correlations, we labeled the group 1 timepoints using the group 2 timepoints with which they were most highly correlated; we then computed the proportion of correctly labeled group 1 timepoints. (We also performed the symmetric analysis whereby we labeled the group 2 timepoints using the group 1 timepoints as a template.) We repeated this procedure 100 times (randomly re-assigning participants to the two groups each time) to obtain a distribution of decoding accuracies for each experimental condition. There were 272 timepoints for paragraph condition, 300 timepoints for intact and word conditions, and 400 timepoints for rest condition, so chance performance on this decoding test is was $\frac{1}{272}$, $\frac{1}{300}$, and $\frac{1}{400}$ respectively.

We repeated this process for each set of parameters, varying kernel type and width, and averaged over the reduction technique used to approximate the higher-order correlations (PCA Fig. 4, a.-c. and eigenvector centrality Fig. 4, d.-f.). Since there is no ground truth in these analyses, and we did not know which parameters best capture the data, we instead report a robustness search by averaging over the parameters and reporting which results consistently showed up across all parameters.

The two methods used to approximate the higher-order correlations (PCA Fig. 4, a.-c. and eigenvector centrality Fig. 4, d.-f.) capture different facets of the activity patterns. Using PCA, the higher-order correlations are all linked to the original activity patterns, whereas eigenvectory centrality breaks the immediate link with specific brain areas and instead characterizes the position of the nodes in the network that are similar over time.

We found for both PCA and eigenvector centrality, during the intact condition in the experiment, classifiers that incorporated higher-order correlations yielded consistently higher accuracy than classifiers trained only on lower-order patterns (Fig. 4, a.&d.). We plot the average correlations for up to the fourth order for the intact condition (Fig. 5) representing the degree of agreement by location pair over time. By contrast, we found that incorporating higher-order (greater than first order) correlations did not further improve decoding accuracy for the other listening conditions or rest condition. This suggests that the cognitive processing that supported the most cognitively rich condition involved higher-order network

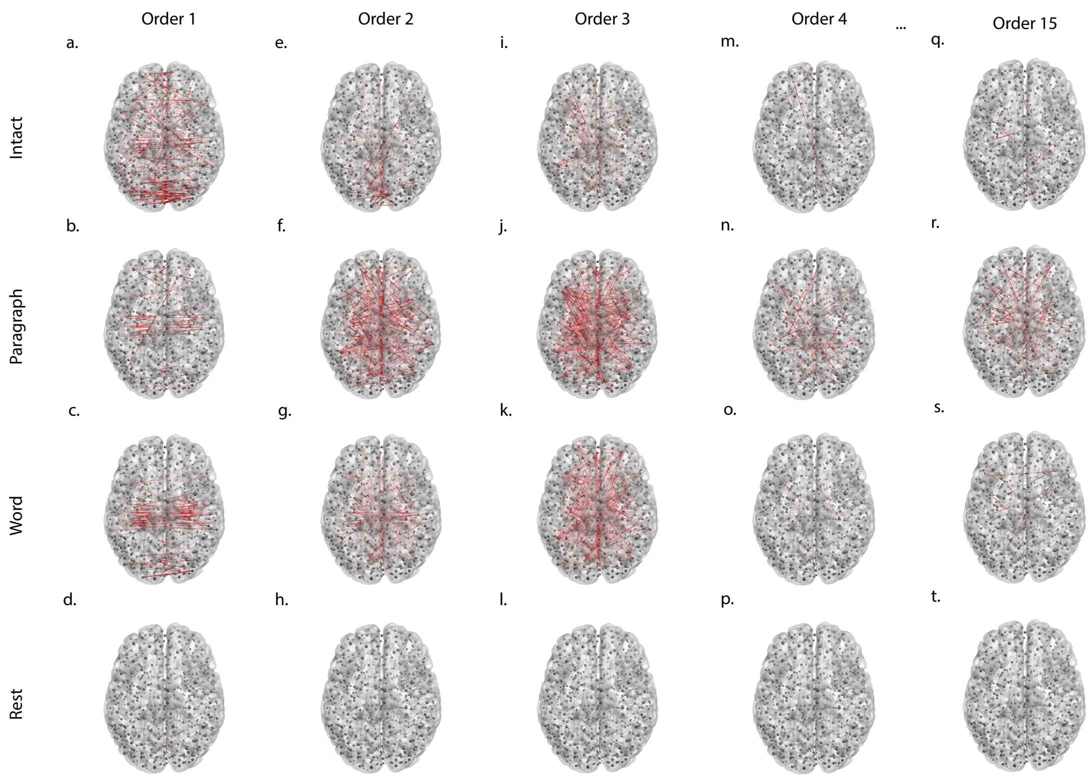

Figure 5: **Average correlations by order for the intact listening condition.** Using eigenvector centrality to approximate higher-order correlations for the intact, paragraph scrambled, word scrambled, and rest condition. We plot the strongest 50% absolute value mean correlation for **a.-d. first order, e.-h. second order, i.-l. third order, and m.-p. fourth order** , representing the degree of agreement by location pair over time. To demonstrate how this method is computationally scalable, we also approximated **a.-d. fifteenth order** dynamic correlation, which would be possible to compute using conventional methods since it would require more bits to represent the solution than there are molecules in the universe.

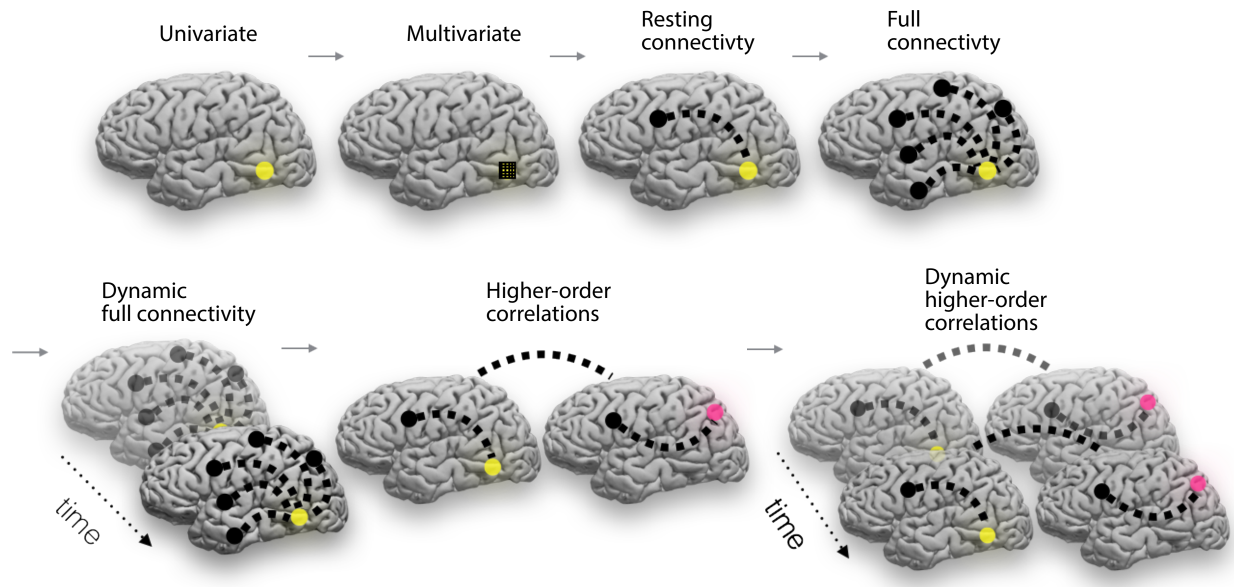

Figure 6: **Direction of the field (adapted from (Turk-Browne, 2013)).** The evolution of fMRI analyses started with univariate activation, which refers to the average amplitude of BOLD activity evoked by events of an experimental condition. Next, multivariate classifiers are trained on patterns of activation across voxels to decode distributed representations for specific events. The next, resting connectivity, is the temporal correlation of one or more seed regions with the remainder of the brain during rest. Additionally, task-based connectivity examines how these correlations differ by cognitive state. Following this increasing trajectory of increasing complexity, full connectivity considers all pairwise correlations in the brain, most commonly at rest. Next, dynamic full connectivity considers how full connectivity changes over time. Continuing this line of reasoning, we expect higher-order network dynamics might provide even richer insights into the neural basis of cognition.

dynamics.

Discussion

Based on prior work (Demertzi et al., 2019) and following the direction of the field (Turk-Browne, 2013) we think our thoughts might be encoded in dynamic network patterns, and possibly higher order network patterns (Fig. 6). We sought to test this hypothesis by developing an approach to inferring high-order network dynamics from timeseries data.

One challenge in studying dynamic interactions is the computational resources required to calculate higher-order correlations. We developed a computationally tractable model of network dynamics (Fig. 1) that takes in a feature timeseries and outputs approximated first-order dynamics (i.e., dynamic functional correlations), second-order dynamics (reflecting homologous networks that dynamically form and disperse), and higher-order network dynamics (up to tenth-order dynamic correlations).

We first validated our model using synthetic data, and explored how recovery varied with different underlying data structures and kernels. We then applied the approach to an fMRI dataset (Simony et al., 2016) in which participants listened to an audio recording of a story, as well as scrambled versions of the same story (where the scrambling was applied at different temporal scales). We trained classifiers to take the output of the model and decode the timepoint in the story (or scrambled story) that the participants were listening to. We found that, during the intact listening condition in the experiment, classifiers that incorporated higher-order correlations yielded consistently higher accuracy than classifiers trained only on lower-order patterns (Fig. 4, a.&d.). By contrast, these higher-order correlations were not necessary to support decoding the other listening conditions and (minimally above chance) during a control rest condition. This suggests that the cognitive processing that supported the most cogntively rich listening conditions involved second-order (or higher) network dynamics.

Although we found decoding accuracy was best when incorporating higher-order network dynamics for all but rest condition, it is unclear if this is a product of the brain or the data collection technique. It could be that the brain is second-order or it could be that fMRI can only reliably give second-order interactions. Exploring this method with other data collection technique will be important to disentangle this question.

Concluding remarks

How can we better understand how brain patterns change over time? How can we quantify the potential network dynamics that might be driving these changes? One way to judge the techniques of the future is to look at the trajectory of the fMRI field so far has taken so far (Fig. 1). The field started with univariate activation, measuring the average activity for each voxel. Analyses of multivariate activation followed, looking at spatial patterns of activity over voxels. Next, correlations of activity were explored, first with measures like resting connectivity that take temporal correlation between a seed voxel and all other voxels then with full connectivty that measure all pairwise correlations. Additionally, this path of increasing complexity also moved from static to dynamic measurements. One logical next step in this trajectory would be dynamic higher-order correlations. We have created a method to support these calculations by scalably approximating dynamic higher-order correlations.

Acknowledgements

We acknowledge discussions with Luke Chang, Hany Farid, Paxton Fitzpatrick, Andrew Heusser, Eshin Jolly, Qiang Liu, Matthijs van der Meer, Judith Mildner, Gina Notaro, Stephen Satterthwaite, Emily Whitaker, Weizhen Xie, and Kirsten Ziman. Our work was supported in part by NSF EPSCoR Award Number 1632738 to J.R.M. and by a sub-award of DARPA RAM Cooperative Agreement N66001-14-2-4-032 to J.R.M. The content is solely the responsibility of the authors and does not necessarily represent the official views of our supporting organizations.

Author contributions

Concept: J.R.M. Implementation: T.H.C., L.L.W.O., and J.R.M. Analyses: L.L.W.O and J.R.M.

References

Alvarez-Hamelin, I., Dall'Asta, L., Barrat, A., & Vespignani, A. (2005). k-corr decomposition: a tool for the visualiztion of large scale networks. *arXiv*, cs/0504107v2.

Barthélemy, M. (2004). Betweenness centrality in large complex networks. *European Physical Journal B*, *38*, 163–168.

Bassett, D., Meyer-Lindenberg, A., Achard, S., Duke, T., & Bullmore, E. (2006, December). Adaptive reconfiguration of fractal small-world human brain functional networks. *Proceedings of the National Academy of Sciences, USA*, *103*(51), 19518-23.

Bonacich, P. (2007). Some unique properties of eigenvector centrality. *Social Networks*, *29*(4), 555–564.

Bullmore, E., & Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, *10*(3), 186–198.

Christakis, N. A., & Fowler, J. H. (2010). Social network sensors for early detection of contagious outbreaks. *PLoS One*, *5*(9), e12948.

Comon, P., Jutten, C., & Herault, J. (1991). Blind separation of sources, part II: Problems statement. *Signal Processing*, *24*(1), 11 - 20.

Demertzi, A., Tagliazucchi, E., Dehaene, S., Deco, G., Barttfeld, P., Raimondo, F., ... Sitt, J. D. (2019). Human consciousness is supported by dynamic complex patterns of brain signal coordination. *Science Advances*, *5*(2), eaat7603.

Estrada, E., & Rodríguez-Velázquez, J. A. (2005). Subraph centrality in complex networks. *Physical Review E*, *71*(5), 056103.

Etzel, J. A., Gazzola, V., & Keysers, C. (2009). An introduction to anatomical ROI-based fMRI classification. *Brain Research*, *1282*, 114–125.

Fong, A. H. C., Yoo, K., Rosenberg, M. D., Zhang, S., Li, C.-S. R., Scheinost, D., ... Chun, M. M. (2019). Dynamic functional connectivity during task performance and rest predictsindividual differences in attention across studies. *NeuroImage*, *188*, 14–25.

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, *40*(1), 35–41.

Geisberger, R., Sanders, P., & Schultes, D. (2008). Better approximation of betweenness centrality. *Proceedings of the meeting on Algorithm Engineering and Experiments*, 90–100.

Gershman, S., Blei, D., Pereira, F., & Norman, K. (2011). A topographic latent source model for fMRI data. *NeuroImage*, *57*, 89–100.

Halu, A., Mondragón, R. J., Panzarasa, P., & Bianconi, G. (2013). Multiplex PageRank. *PLoS One*, *8*(10), e78293.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*, 2425–2430.

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, *313*(5786), 504–507.

Honey, C. J., Kötter, R., Breakspear, M., & Sporns, O. (2007). Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proceedings of the National Academy of Science USA*, *104*(24), 10240–10245.

Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, *532*, 453–458.

Huth, A. G., Nisimoto, S., Vu, A. T., & Gallant, J. L. (2012). A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron*, *76*(6), 1210-1224.

Jutten, C., & Herault, J. (1991). Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, *24*(1), 1–10.

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, *8*, 679–685.

Landau, E. (1895). Zur relativen Wertbemessung der Turnierresultate. *Deutsches Wochenschach*, *11*, 366–369.

Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, *401*, 788–791.

Lin, J. (2009). Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory*, *37*(1), 145–151.

Lohmann, G., Margulies, D. S., Horstmann, A., Pleger, B., Lepsien, J., Goldhahn, D., . . . Turner, R. (2010). Eigenvector centrality mapping for analyzing connectivity patterns in fMRI data of the human brain. *PLoS One*, *5*(4), e10232.

Mairal, J., Ponce, J., Sapiro, G., Zisserman, A., & Bach, F. R. (2009). Supervised dictionary learning. *Advances in Neural Information Processing Systems*, 1033–1040.

Mairal, J. B., Bach, F., Ponce, J., & Sapiro, G. (2009). Online dictionary learning for sparse coding. *Proceedings of the 26th annual international conference on machine learning*, 689–696.

Manning, J. R., Ranganath, R., Norman, K. A., & Blei, D. M. (2014). Topographic factor analysis: a Bayesian model for inferring brain networks from neural data. *PLoS One*, *9*(5), e94914.

Manning, J. R., Zhu, X., Willke, T. L., Ranganath, R., Stachenfeld, K., Hasson, U., . . . Norman, K. A. (2018). A probabilistic approach to discovering dynamic full-brain functional connectivity patterns. *NeuroImage*, *180*, 243–252.

McInnes, L., & Healy, J. (2018). UMAP: Uniform manifold approximation and projection for dimension reduction. *arXiv*, *1802*(03426).

Mitchell, T., Shinkareva, S., Carlson, A., Chang, K., Malave, V., Mason, R., & Just, M. (2008). Predicting human brain activity associated with the meanings of nouns. *Science*, *320*(5880), 1191.

Newman, M. E. J. (2005). A measure of betweenness centrality based on random walks. *Social Networks*, *27*, 39–54.

Newman, M. E. J. (2008). The mathematics of networks. *The New Palgrave Encyclopedia of Economics*, *2*, 1–12.

Nishimoto, S., Vu, A., Naselaris, T., Benjamini, Y., Yu, B., & Gallant, J. (2011). Reconstructing visual experience from brain activity evoked by natural movies. *Current Biology*, *21*, 1 - 6.

Nocedal, J., & Wright, S. J. (2006). *Numerical optimization*. New York: Springer.

Norman, K. A., Newman, E., Detre, G., & Polyn, S. M. (2006). How inhibitory oscillations can train neural networks and punish competitors. *Neural Computation*, *18*, 1577–1610.

Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: generalizing degree and shortest paths. *Social Networks*, *32*, 245–251.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *2*, 559-572.

Pereira, F., Lou, B., Pritchett, B., Ritter, S., Gershman, S. J., Kanwisher, N., . . . Fedorenko, E. (2018). Toward a universal decoder of linguistic meaning from brain activation. *Nature Communications*, *9*(963), 1–13.

Rao, C. R. (1982). Diversity and dissimilarity coeficients: a unified approach. *Theoretical Population Biology*, *21*(1), 24–43.

Ricotta, C., & Szeidl, L. (2006). Towards a unifying approach to diversity measures: Bridging the gap between the Shannon entropy and Rao's quadratic index. *Theoretical Population Biology*, *70*(3), 237–243.

Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *NeuroImage*, *52*, 1059–1069.

Schreiber, T. (2000). Measuring information transfer. *Physical Review Letters*, *85*(2), 461–464.

Simony, E., Honey, C. J., Chen, J., & Hasson, U. (2016). Uncovering stimulus-locked network dynamics during narrative comprehension. *Nature Communications*, *7*(12141), 1–13.

Spearman, C. (1904). General intelligence, objectively determined and measured. *Americal Journal of Psychology*, *15*, 201–292.

Sporns, O., & Honey, C. J. (2006). Small worlds inside big brains. *Proceedings of the National Academy of Science USA*, *103*(51), 19219–19220.

Tipping, M. E., & Bishop, C. M. (1999). Probabilistic principal component analysis. *Journal of Royal Statistical Society, Series B*, *61*(3), 611–622.

Tong, F., & Pratte, M. S. (2012). Decoding patterns of human brain activity. *Annual Review of Psychology*, *63*, 483–509.

Turk-Browne, N. B. (2013). Functional interactions as big data in the human brain. *Science*, *342*, 580–584.

van der Maaten, L. J. P., & Hinton, G. E. (2008). Visualizing high-dimensional data using t-SNE. *Journal of Machine Learning Research*, *9*, 2579-2605.

Zar, J. H. (2010). *Biostatistical analysis*. Prentice-Hall/Pearson.