ORIGINAL RESEARCH ARTICLE

Constructing brain connectivity group graphs from EEG time series

A. T. Walden and L. Zhuang

Department of Mathematics, Imperial College London, London SW7 2BZ, UK.

ARTICLE HISTORY
Compiled January 7, 2019

ABSTRACT
Graphical analysis of complex brain networks is a fundamental area of modern neuroscience. Functional connectivity is important since many neurological and psychiatric disorders, including schizophrenia, are described as 'dys-connectivity' syndromes. Using EEG time series collected on each of a group of 15 individuals with a common medical diagnosis of positive syndrome schizophrenia we seek to build a single, representative, brain functional connectivity group graph. Disparity/distance measures between spectral matrices are identified and used to define the normalized graph Laplacian enabling clustering of the spectral matrices for detecting 'outlying' individuals. Two such individuals are identified. For each remaining individual, we derive a test for each edge in the connectivity graph based on average estimated partial coherence over frequencies, and associated p-values are found. For each edge these are used in a multiple hypothesis test across individuals and the proportion rejecting the hypothesis of no edge is used to construct a connectivity group graph. This study provides a framework for integrating results on multiple individuals into a single overall connectivity structure.

KEYWORDS
brain functional connectivity, EEG time series, graphical model, multivariable power spectra, schizophrenia, spectral matrix clustering

1. Introduction

A very important area of modern neuroscience is that of graph analysis of complex brain networks. As pointed out in the seminal paper [5], "functional brain networks are thought to provide the physiological basis for information processing and mental representations." In neuroscience, functional connectivity [5, 21] denotes the symmetrical statistical association or dependency between elements of the system, and graphs may be used to describe this connectivity. Many neurological and psychiatric disorders, including Alzheimer's disease and schizophrenia, may be described as dys-connectivity syndromes [4, 5, 11, 24]. In this paper we consider time series methodology for building a single, representative, brain functional connectivity graph from a group of individuals with the same medical diagnosis (positive syndrome schizophrenia). We use electroencephalogram (EEG) measurements representing cortical activity as potential, measured over the scalp of each patient.

The EEG dataset analyzed here is discussed in more detail in Section 2. It was acquired from a group of $n = 15$ patients. There were ten sensor locations on the scalp

Corresponding author. Email: a.walden@imperial.ac.uk

of each individual. The data is modelled by a stationary vector-valued time series, $\{\boldsymbol{X}_t\}$, of dimension $p = 10$, for each individual. For any individual, if all connections were present, the complete graph shown in Fig. 1 is obtained.

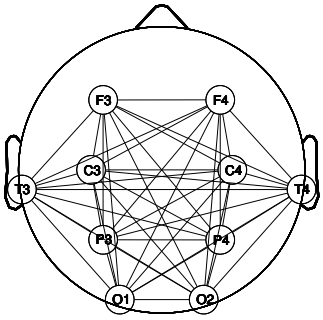

Figure 1. Schematic of 10 scalp sensor locations for recording EEG time series with complete graph.

The problem of clinical heterogeneity over the patient's group is a well recognized and difficult problem. Some differences may be resolved by combining different sorts of measurements on the same patients [5, p. 195]. Our approach is quite different: we seek to identify 'outlier' individuals, and remove them from further analyses. Using the estimated spectral matrices for the individuals in a group, we develop a spectral matrix clustering method using a similarity matrix constructed from spectral matrix distances and the resulting normalized graph Laplacian. This incorporates an approach for choosing the number of clusters, and allows us to perform clustering using a traditional clustering method, such as k-means clustering, which assumes the number of clusters to be known *a priori*. We refer to our novel approach as 'spectral matrix clustering via the normalized graph Laplacian.' Using this approach we identify two outlier individuals.

This use of graphs for outlier identification via clustering, is subsidiary to our main aim, graph representations for functional connectivity. For each individual we need to determine if an edge in the graph (a connection between two nodes) exists. Using frequency domain analysis, the absence of an edge is indicated by zero partial coherencies for all frequencies in the frequency band. In practice the partial coherencies for any individual will have to be estimated, and thus will include sampling variability. Estimated partial coherencies will never be exactly zero, so that hypothesis testing is required to test an edge to see if it should be declared to be missing. Unfortunately, the partial coherencies to be zero-tested for every frequency in the frequency band are correlated due to smoothing; one approach for this situations would be to follow Dahlhaus [6] and use a test based simply on the maximum of the estimated partial coherencies over the frequency band, but the exact asymptotic null distribution of this test statistic is not known. Alternatively, [22, 23] employed Holm's stepdown procedure [14], which, while suitable for dependent tests, is very conservative. A third aproach using p-value combiners (e.g., [32]) was discussed in [30]; here independent tests are assumed so that other unsatisfactory adjustments are required.

Instead, we use a test statistic to test for edge inclusion which is formed as an integrated — over frequencies — functional of the estimated spectral matrix. The effects of tapering and smoothing are fully built into the statistical analysis. Since the asymptotic distribution of the functional is known, p-values can be associated to the particular graph edge for all individuals in the group, allowing us to find the proportion

of individuals whose associated statistic rejects the hypothesis of no edge. Using this information, a group-specific connectivity graph can be readily constructed.

1.1. Graphs and time series

A graph $G = (V, E)$ consists of vertices V and edges E, where $E \subset \{(j,k) \in V \times V : j \neq k\}$. (Note: We only consider simple graphs where there are no loops from a vertex to itself, nor multiple edges between two vertices.) In order to represent $\{\boldsymbol{X}_t\}$ the vertices of the graph correspond to the p individual series $\{X_{j,t}\}$, so $V = \{1, \ldots, p\}$. Edges $(j,k) \in E$ for which both $(j,k) \in E$ and $(k,j) \in E$ are called undirected edges. An undirected graph is one with only undirected edges and is the type considered here. For a p-vector-valued process there are $p(p-1)/2$ unordered pairs of vertices for its graph. Thus there are $2^{p(p-1)/2}$ possible distinct graph structures. So for our data set there are 45 possible connections between the series (edges to the graph) and 2^{45} possible graph structures.

In a time series graph, an edge between two vertices (or series j and k, say), corresponds to a *direct* connection between the measured series at the two locations. As we shall see in Section 3, the absence of an edge is indicated by a corresponding zero partial coherence for all frequencies in the frequency band for the two series. Partial coherence is denoted $\gamma^2_{jk\bullet(\backslash jk)}(f)$, where $\{\backslash jk\} = \{1 \leq i \leq p : i \neq j, k\}$, and the $_{\bullet(\backslash jk)}$ terminology indicates that the linear effects of all the other series have been removed from series j and k. The assessment of the interaction between series j and k thus discounts the indirect effects of the other series. Partial coherence has been used elsewhere in neuroscience (e.g., [28, 29]), but not in the same way as here. Partial coherence is determined from the spectral matrix. We write $\boldsymbol{X}_t = [X_{1,t}, \ldots, X_{p,t}]^T \in \mathbb{R}^p$, $t \in \mathbb{Z}$, and ^{T} denotes transposition. Without loss of generality $\{\boldsymbol{X}_t\}$ may be taken to have a mean of zero. Then the matrix-valued sequence $\boldsymbol{s}_\tau = \mathrm{cov}\{\boldsymbol{X}_{t+\tau}, \boldsymbol{X}_t^T\} = E\{\boldsymbol{X}_{t+\tau}\boldsymbol{X}_t^T\}, \tau \in \mathbb{Z}$, is the (matrix) autocovariance sequence, and, assuming it exists, its Fourier transform is the spectral matrix

$$\boldsymbol{S}(f) \stackrel{\mathrm{def}}{=} \Delta_{\mathrm{t}} \sum_{\tau=-\infty}^{\infty} \boldsymbol{s}_\tau \, \mathrm{e}^{-\mathrm{i}2\pi f \tau \Delta_{\mathrm{t}}}, \quad |f| \leq f_{\mathcal{N}}, \tag{1}$$

where $f_{\mathcal{N}} = 1/(2\Delta_{\mathrm{t}})$ is the Nyquist frequency and Δ_{t} is the sampling interval. In practice we work with some estimate $\hat{\boldsymbol{S}}(f)$ of $\boldsymbol{S}(f)$.

1.2. Contributions

(1) We firstly low-pass filter and downsample the data to eliminate the contaminating alpha rhythm around 10Hz, leaving the so-called delta frequency band for further analysis. For individual i, we derive a test statistic for edge ℓ in the connectivity graph, based on the average partial coherence over just the delta frequency band. By drawing on [7, 8] the asymptotic null distribution of this statistic can be deduced and hence a corresponding p-value found. The use of the average partial coherence statistic means that multiple hypothesis testing across frequencies is not required.

(2) For building a representative graph for the group, with regard to each possible edge in the graph, the associated p-values for each individual in the group are used in a multiple hypothesis test *across individuals* to find the proportion of

values of the statistic rejecting the hypothesis of no edge. The edge is included in the group graph when this proportion exceeds a given threshold.

(3) To identify (and remove) any outlying cases in the group, we use novel clustering techniques based on the n estimated spectral matrices $\hat{\boldsymbol{S}}(f)$ for the group: (i) We calculate Kullback-Leibler disparities and Riemannian distances between the spectral matrices. Both measures integrate across frequencies, and an analysis of these measures suggests to define 'distance' matrices based on (a) the square root of the Kullback-Leibler disparity measure and (b) the Riemannian distance. (ii) The two distance matrices $\boldsymbol{D}$ are used to define two similarity matrices $\boldsymbol{W}$, each of which defines a weighted adjacency matrix from which the normalized graph Laplacian is constructed. The eigenvalues and eigenvectors of this latter matrix enable the determination of the *number* and structure of the clusters. As a result of our clustering, two individuals are eliminated from the study.

Remark 1. The use of the graph Laplacian for clustering is discussed in [33] where it is called 'spectral clustering' since it uses the eigenvalues of the normalized Laplacian matrix. It is a general method which merely requires a weighted adjacency matrix. In our context this adjacency matrix is determined from a similarity function involving the distances between (power) 'spectral' matrices. Confusingly, this second use of 'spectral' is quite different to that in 'spectral' clustering. We refer to our novel approach as 'spectral matrix clustering via the normalized graph Laplacian.' □

1.3. Development of paper

Section 2 explains the nature of the motivating data for 15 patients with a diagnosis of positive syndrome schizophrenia, the necessary preprocessing steps and their implications for subsequent analyses. Section 3 looks at the spectral analysis methodology, hypothesis testing machinery and test statistic, used in determining the edges for the group graph. This is predicated on a homogeneous — in a spectral matrix sense — group of patients. So, in Sections 4 and 5, we explain how we selected a homogeneous group of 13 patients from the original 15. The spectral matrix data for these 13 is then analyzed using the techniques of Section 3. Results and Conclusions are presented in Section 6.

2. Background on the data

2.1. Data details

The data was acquired from $n = 15$ male forensic patients with a diagnosis of positive syndrome schizophrenia (symptoms such as delusions, hallucinations) from the Serbsky Institute in Moscow. All subjects gave written informed consent for the investigation. Ethical approval came from the local Moscow ethics committee, in compliance with national legislation and the Declaration of Helsinki; see [23] for further discussion of this clinical dataset.

EEG was recorded using an EEG mapper from 10 scalp sites (F3, F4, C3, C4, T3, T4, P3, P4, O1 and O2) referenced to linked ears, using a bandpass filter of $0.5 - 45$Hz while the patients were resting with eyes closed. The electrode positions are shown in Fig. 1. A sample interval of $\Delta_t = 0.01$s was used when recording the data, so the

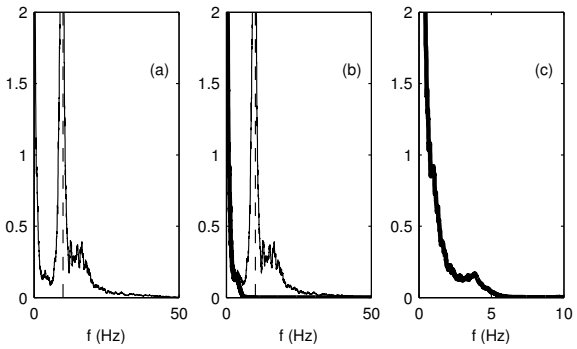

Figure 2. Spectra for series 7 (O1) of EEG time series. (a) as recorded, (b) after low-pass filtering (heavy line) (c) after low-pass filtering and down-sampling by 5. The vertical dashed line marks the frequency of 10Hz.

Nyquist frequency is $f_{\mathcal{N}} = 1/(2\Delta_{\rm t}) = 50$Hz.

2.2. Data processing

The estimated spectrum for one channel of the data, (corresponding to location O1), and for one individual, is shown in Fig. 2(a). An enormous peak/line is seen in the spectrum around 10Hz due to alpha waves, neural oscillations in the range of 7.5–12.5 Hz arising from synchronous and coherent electrical activity of thalamic pacemaker cells in humans. They predominantly originate from the occipital lobe during wakeful relaxation with closed eyes. (O1 is an occipital location). These waves must be removed from the data before processing because (i) they will cause huge side-lobe leakage, and (ii) they will be highly coherent and cause connections purely due to their presence. The data was thus filtered by a Butterworth low-pass filter with cut-off 4.6Hz. The spectrum of the filtered data is shown in Fig. 2(b) by the heavy line; as a result of the filtering there is now no power above about 5Hz so that the majority of the power spectrum is null. Next the data was downsampled by a factor of 5, i.e., we keep only every fifth data point. After downsampling there is, for each patient, a total of $N = 612$ points for each of the $p = 10$ series. The new sample interval is $\Delta_{\rm t} = 0.05$s so the new Nyquist frequency is 10Hz. After this data preprocessing, the spectral matrices over $[-10, -5] \cup [5, 10]$Hz are close to null and matrix inverses are erroneous in this frequency band. There is no aliasing because the spectrum is already zero beyond 5Hz. The spectrum of the filtered and downsampled data is shown in Fig. 2(c).

There is consistent evidence of abnormal cerebral oscillations in schizophrenia. Patients typically show unusually high EEG power in the low frequencies [12]. As a result (see also [16, 23]) we shall focus our analysis on the so-called 'delta frequency band' $f \in [0.5, 4]$Hz. So, we define a restricted frequency interval

$$\mathcal{R} \stackrel{\rm def}{=} [f_{\rm a}, f_{\rm b}] \subseteq [0, f_{\mathcal{N}}] = [0, 1/(2\Delta_{\rm t})], \tag{2}$$

with $f_{\rm a} = 0.5$Hz and $f_{\rm b} = 4$Hz with $\Delta_{\rm t} = 0.05$s.

Detailed prescreening had been carried out to eliminate time periods with missing data channels and other recording problems. Since the patients were resting with eyes closed, and no task was performed, stationarity for the utilised series segments is a reasonable assumption and was not contraindicated.

3. Determining the edges for the connectivity group graph

3.1. Further background to time series graphs

Define $X_j = \{X_{j,t} : t \in \mathbb{Z}\}$ and $X_{(\backslash jk)} = \{X_i : i \in \{\backslash jk\}\}$. For a time series graph, edges between the vertices represent partial correlation between two series, so that there is no connection between nodes j and k if and only if X_j and X_k are partially uncorrelated given $X_{(\backslash jk)}$, [6].

Using filtering, the linear effects of $X_{(\backslash jk)}$ are removed from X_j to obtain the jth 'residual' series $\{\nu_{j,t}\}$. The kth residual series is defined likewise. The sequence $s_{\nu_j\nu_k,\tau} = \mathrm{cov}\{\nu_{j,t+\tau}, \nu_{k,t}\}, \tau \in \mathbb{Z}$, is the partial cross-covariance sequence and, if it is everywhere zero, the two residual series are partially uncorrelated and then we write $X_j \perp\!\!\!\perp X_k | X_{(\backslash jk)}$. With $\iff$ denoting 'if and only if,' we then have $(j,k) \notin E \iff X_j \perp\!\!\!\perp X_k | X_{(\backslash jk)}$.The resulting graph G is called a partial correlation graph.

The Fourier transform of the partial cross-covariance sequence is the partial cross-spectral density function, denoted $S_{jk\bullet(\backslash jk)}(f)$. The partial coherence is defined as

$$\gamma^2_{jk\bullet(\backslash jk)}(f) = \frac{|S_{jk\bullet(\backslash jk)}(f)|^2}{S_{jj\bullet(\backslash jk)}(f) S_{kk\bullet(\backslash jk)}(f)}, \quad |f| \le f_{\mathcal{N}}.$$

It is the squared partial correlation between series j and k at frequency f. Notice

$$\begin{aligned} (j,k) \notin E &\iff \mathrm{cov}\{\nu_{j,t+\tau}, \nu_{k,t}\} = 0, \ \forall \tau \in \mathbb{Z} \iff S_{jk\bullet(\backslash jk)}(f) = 0, \quad |f| \le f_{\mathcal{N}} \\ &\iff \gamma^2_{jk\bullet(\backslash jk)}(f) = 0, \quad |f| \le f_{\mathcal{N}}. \end{aligned} \tag{3}$$

In other words, a missing edge corresponds to the partial coherence being zero over the frequency range. Assuming $\boldsymbol{S}(f)$ is full rank, the partial coherence can be expressed as, (e.g., [6]),

$$\gamma^2_{jk\bullet(\backslash jk)}(f) = \frac{|S^{jk}(f)|^2}{S^{jj}(f) S^{kk}(f)}, \tag{4}$$

where $S^{jk}(f)$, is the (j,k)th element of $\boldsymbol{S}^{-1}$. Hence testing $(j,k) \notin E$ can be done using the null hypothesis

$$H_0 : \gamma^2_{jk\bullet(\backslash jk)}(f) = 0 \text{ for } |f| \le f_{\mathcal{N}}. \tag{5}$$

Estimated $\hat{\gamma}^2_{jk\bullet(\backslash jk)}(f)$ values are found by substituting $\hat{\boldsymbol{S}}^{-1}(f)$ into (4).

Testing (5) was considered by Eichler [8, p. 993]. With a modification to incorporate the sample interval, we define the test statistic

$$T_N \stackrel{\mathrm{def}}{=} \Delta_{\mathrm{t}} \int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} \hat{\gamma}^2_{jk\bullet(\backslash jk)}(f) \mathrm{d}f. \tag{6}$$

T_N is the average estimated partial coherence, and measures the deviation from the null hypothesis.

Remark 2. In our application, connectivity between two series may be due to a flow of information between two or more spatially separate regions of the brain. Also, a more

distant source may affect several electrodes simultaneously, through the phenomenon of 'volume conduction.' Our use of *partial* coherence makes no assumptions about the nature of cerebral sources and markedly reduces the impact of volume conduction by only considering *direct* connections. (It is shown in [18] how ordinary — rather than partial — covariance between two variables may be decomposed into a sum of path weights for all paths connecting the two variables in the graph, only one of which is the direct path.) Solving the inverse problem for source localisation in the brain requires strong assumptions about the nature of the sources to be localised and these assumptions would invariably affect the estimates of connectivity [23].

3.2. Restricted frequency method

We concentrate attention on the low-frequency 'delta band' and so are interested in the partial coherence for $f \in \mathcal{R}$, rather than for $|f| \leq f_{\mathcal{N}}$. For us, testing $(j,k) \notin E$ is equivalent to testing the null hypothesis

$$H_0 : \gamma^2_{jk\bullet(\backslash jk)}(f) = 0, \quad f \in \mathcal{R}. \tag{7}$$

Using (2), let

$$T_{N,\mathcal{R}}(j,k) \stackrel{\text{def}}{=} \frac{1}{f_\text{b} - f_\text{a}} \int_{f_\text{a}}^{f_\text{b}} \hat{\gamma}^2_{jk\bullet(\backslash jk)}(f)\mathrm{d}f. \tag{8}$$

$T_{N,\mathcal{R}}$ is the average value of the estimator of partial coherence over the restricted domain $\mathcal{R}$. Assuming N is large, to calculate (8) we shall use

$$T'_{N,\mathcal{R}}(j,k) = \frac{1}{|R|} \sum_{l \in R} \hat{\gamma}^2_{jk\bullet(\backslash jk)}\left(f_l\right), \tag{9}$$

(where $|R|$ is the number of terms in the set R), because

$$\begin{aligned} T'_{N,\mathcal{R}}(j,k) &= \frac{N\Delta_\text{t}}{|R|N\Delta_\text{t}} \sum_{l \in R} \hat{\gamma}^2_{jk\bullet(\backslash jk)}\left(f_l\right) \approx \frac{N\Delta_\text{t}}{|R|} \int_{f_\text{a}}^{f_\text{b}} \hat{\gamma}^2_{jk\bullet(\backslash jk)}(f)\mathrm{d}f \\ &\approx \frac{1}{f_\text{b} - f_\text{a}} \int_{f_\text{a}}^{f_\text{b}} \hat{\gamma}^2_{jk\bullet(\backslash jk)}(f)\mathrm{d}f = T_{N,\mathcal{R}}(j,k), \end{aligned}$$

with $f_\text{b} - f_\text{a} \approx |R|/(N\Delta_\text{t})$. Henceforth we do not distinguish between $T_{N,\mathcal{R}}(j,k)$ and $T'_{N,\mathcal{R}}(j,k)$ since the integral approximation should be very accurate. Under H_0, (7), the partial coherence estimator has the same distribution at any frequency in $\mathcal{R}$, so the average in (9) is that of identically distributed, but correlated, terms.

3.3. Spectrum estimation

The statistics $T_{N,\mathcal{R}}(j,k)$ in (8) are integrated functionals of the spectral matrix estimator $\hat{\boldsymbol{S}}(f)$. Eichler [8] studied the asymptotic distribution (as $N \to \infty$) of such statistics when the spectral estimation combines tapering and spectral window (ker-

nel) smoothing. Such an estimator is of the form

$$\hat{\boldsymbol{S}}(f) \stackrel{\text{def}}{=} \int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} V_m(f-\phi)\hat{\boldsymbol{S}}^{(\mathrm{D})}(f)\,\mathrm{d}\phi, \tag{10}$$

where $V_m(\cdot)$ is the window. The direct spectral estimator, or tapered periodogram, $\hat{\boldsymbol{S}}^{(\mathrm{D})}(f)$, is constructed as follows.

3.3.1. Direct spectral estimator

Given a length-N sample $\boldsymbol{X}_0, \ldots, \boldsymbol{X}_{N-1}$, form $h_t\boldsymbol{X}_t$ where $\{h_t\}, t = 0, \ldots, N-1$, is a taper with $\sum_{t=0}^{N-1} h_t^2 = 1$, and compute $\boldsymbol{J}(f) = \Delta_{\mathrm{t}}^{1/2} \sum_{t=0}^{N-1} h_t \boldsymbol{X}_t \mathrm{e}^{-\mathrm{i}2\pi f t \Delta_{\mathrm{t}}}$. The direct spectral estimator is $\hat{\boldsymbol{S}}^{(\mathrm{D})}(f) \stackrel{\text{def}}{=} \boldsymbol{J}(f)\boldsymbol{J}^H(f)$.

In [8, Assumption 3.3] conditions are stated which must be satisfied by the taper, spectral window and the sequence of smoothing bandwidths as $N \to \infty$, in order for the stated asymptotic results on the statistic to hold. Some standard spectral windows (Daniell, Bartlett) are excluded, but the Bartlett-Priestley window does satisfy the requirements, as does the $100q\%$ cosine taper [26, p. 209] (which we use subsequently). Further, stated mixing condition on $\{\boldsymbol{X}_t\}$ must hold.

3.3.2. Bartlett-Priestley window

The Bartlett-Priestley window (also known as the quadratic or Epanechnikov window) has the form [27, p. 447] for $m \geq 1$,

$$V_m(f) \stackrel{\text{def}}{=} \begin{cases} \frac{3m\Delta_{\mathrm{t}}}{2}\left[1 - (2fm\Delta_{\mathrm{t}})^2\right], & |f| \leq 1/(2m\Delta_{\mathrm{t}}); \\ 0, & 1/(2m\Delta_{\mathrm{t}}) < |f| \leq f_{\mathcal{N}}, \end{cases}$$

and for $|f| > f_{\mathcal{N}}$ by $2f_{\mathcal{N}}$ periodic extension. Then $V_m(f) = m\Delta_{\mathrm{t}} K(m\Delta_{\mathrm{t}} f)$, for $|f| \leq f_{\mathcal{N}}$, where

$$K(f) = \begin{cases} \frac{3}{2}\left[1 - (2f)^2\right], & |f| \leq 1/2; \\ 0, & |f| > 1/2, \end{cases} \tag{11}$$

and for $|f| > f_{\mathcal{N}}$, by periodic extension.

3.3.3. Smoothed Direct Spectral Estimator

Now consider the smoothed direct spectral estimator (10). Then

$$\begin{aligned} \hat{\boldsymbol{S}}(f) &= \sum_{k=-((N/2)-1)}^{N/2} \int_{(k-1)/(N/\Delta_{\mathrm{t}})}^{k/(N\Delta_{\mathrm{t}})} V_m(f-\phi)\hat{\boldsymbol{S}}^{(\mathrm{D})}(\phi)\,\mathrm{d}\phi \\ &\approx \frac{1}{N\Delta_{\mathrm{t}}} \sum_{k=-((N/2)-1)}^{N/2} V_m\left(f - f_k\right)\hat{\boldsymbol{S}}^{(\mathrm{D})}\left(f_k\right). \end{aligned}$$

At the Fourier frequencies, $\hat{\boldsymbol{S}}(f_j)$ is given by

$$\frac{1}{N\Delta_{\mathrm{t}}} \sum_{k=-((N/2)-1)}^{N/2} V_m(f_j - f_k)\, \hat{\boldsymbol{S}}^{(\mathrm{D})}(f_k) = \frac{1}{N\Delta_{\mathrm{t}}} \sum_{l=(j+(N/2)-1)}^{j-(N/2)} V_m(f_l)\, \hat{\boldsymbol{S}}^{(\mathrm{D})}(f_j - f_l)\,.$$

Both $V_m(f)$ and $\hat{\boldsymbol{S}}^{(\mathrm{D})}(f)$ have periodicity $2f_{\mathcal{N}}$, so since $V_m(f)$ is only non-zero for $|f| \leq 1/(2m\Delta_{\mathrm{t}})$ within $|f| \leq f_{\mathcal{N}}$, we have the requirement $|l| \leq N/(2m) \stackrel{\mathrm{def}}{=} M$, and we obtain

$$\begin{aligned}\hat{\boldsymbol{S}}(f_j) &= \frac{1}{N\Delta_{\mathrm{t}}} \sum_{l=-M}^{M} V_m(f_l)\, \hat{\boldsymbol{S}}^{(\mathrm{D})}(f_j - f_l) = \sum_{l=-M}^{M} \frac{1}{2M} K\left(\tfrac{l}{2M}\right) \hat{\boldsymbol{S}}^{(\mathrm{D})}(f_j - f_l) \\ &= \sum_{l=-M}^{M} g_l \hat{\boldsymbol{S}}^{(\mathrm{D})}(f_j - f_l)\,, \end{aligned} \tag{12}$$

where $g_l = \frac{1}{2M} K\left(\frac{l}{2M}\right)$, $l = -M, \ldots, M$, and the estimator (12) is just a discretely-averaged tapered spectral estimator.

3.3.4. Bandwidth

In terms of finite-N estimators, let $B_N \stackrel{\mathrm{def}}{=} 1/(m\Delta_{\mathrm{t}}) = 2M/(N\Delta_{\mathrm{t}}) = 2M\Delta_{\mathrm{f}}$, where Δ_{f} is the frequency spacing. This is an approximation to the frequency width across which the direct spectrum estimator is being smoothed, so B_N is a crude bandwidth measure. However, B_N does not properly take account of the effect of tapering. The bandwidth, $B_{\mathcal{U}}$, of the spectral window, $\mathcal{U}(f)$, which takes account of both tapering and smoothing, can be expressed through its autocorrelation width [34], and is defined by $1/B_{\mathcal{U}} = \Delta_{\mathrm{t}} \sum_{\tau=-(N-1)}^{N-1} w_{m,\tau}^2 \left(\sum_{t=0}^{N-|\tau|-1} h_{t+|\tau|} h_t\right)^2$. Here $\{w_{m,\tau}\}$ is the lag window corresponding to the Bartlett-Priestley window $V_m(f)$,

$$w_{m,\tau} = \begin{cases} 1, & \tau = 0; \\ \frac{3m^2}{\pi^2\tau^2}\left[\frac{\sin(\pi\tau/m)}{\pi\tau/m} - \cos(\pi\tau/m)\right], & 1 \leq |\tau| \leq N-1; \\ 0, & |\tau| \geq N. \end{cases}$$

Remark 3. For our dataset $N = 612$. The choice of $q = 0.2$ (20%) cosine taper along with a choice for m determines a value $B_{\mathcal{U}}$. Use of $N/(2m) = M$, enables us to express $B_{\mathcal{U}}$ as a function of M. $M = 9$ gives $B_{\mathcal{U}} \approx 0.5$Hz, virtually the same as that used in the spectral estimation for the clustering scheme (Section 5). □

3.4. Distribution of normalized test statistic

Under the null hypothesis

$$Q_{N,\mathcal{R}}(j,k) = \frac{NT_{N,\mathcal{R}}(j,k) - \mu}{\sigma} \tag{13}$$

is asymptotically distributed as a standard Gaussian, denoted $\mathcal{N}(0,1)$, where

$$\mu \stackrel{\text{def}}{=} mC_h\, C_{k,2} \quad \text{and} \quad \sigma \stackrel{\text{def}}{=} C_h\, [2mC_{k,4}/|\mathcal{R}|]^{1/2}. \tag{14}$$

Here $|\mathcal{R}| \stackrel{\text{def}}{=} (f_{\rm b} - f_{\rm a})/f_{\mathcal{N}} = (f_{\rm b} - f_{\rm a})2\Delta_{\rm t}$. C_h is the variance inflation caused by the normalized taper, (traditionally calculated as $N\sum_{t=0}^{N-1} h_t^4$), and the $C_{k,i}$ are integrals of powers of the lag window generator: $C_{k,i} = \int_{-\infty}^{\infty} k^i(u)\,\mathrm{d}u$, where $k(u)$ is the inverse Fourier transform of $K(f)$ in (11), i.e., [27, p. 447],

$$k(u) = \frac{3}{(\pi u)^2}\left[\frac{\sin \pi u}{\pi u} - \cos \pi u\right], \qquad u \in \mathbb{R}. \tag{15}$$

For the Bartlett-Priestley design we find that $C_{k,2} = 1.2000$ and $C_{k,4} = 0.8676$ with $C_h = 1.1159$ for $q = 0.2$. The forms of (13) and (14) follow by applying results in [7, p. 143] to our set-up.

We now write (13) in a computationally convenient form.

$$Q_{N,\mathcal{R}}(j,k) = \frac{\frac{N}{|R|}\sum_{l\in R}\hat{\gamma}^2_{jk\bullet(\backslash jk)}\,(f_l) - \frac{N}{2M}C_h\, C_{k,2}}{C_h\left[\frac{N}{M}C_{k,4}/\left(\frac{2|R|}{N}\right)\right]^{1/2}}, \tag{16}$$

where we have used that $m = N/(2M)$ and $(f_{\rm b} - f_{\rm a})2\Delta_{\rm t} \approx |R|/(N\Delta_{\rm t}) \times 2\Delta_{\rm t} = 2|R|/N$. Then (16) simplifies to

$$Q_{N,\mathcal{R}}(j,k) = \frac{\frac{1}{|R|}\sum_{l\in R}\hat{\gamma}^2_{jk\bullet(\backslash jk)}\,(f_l) - \frac{1}{2M}C_h\, C_{k,2}}{C_h\,[C_{k,4}/\,(2|R|M)]^{1/2}}. \tag{17}$$

3.5. Multiple hypothesis testing

After the analyses in section 5, we will have identified the n individuals deemed to be forming the homogeneous group for graph construction. Suppose we relabel the $L = p(p-1)/2$ edges $(1,2),\ldots,(1,p),(2,3),\ldots,(p-1,p)$ as $l = 1,\ldots,L$. Also let $l_i,\ i = 1,\ldots,n$, denote edge l, (corresponding to some (j,k)), for individual i, We can write the test statistic as $Q_{N,\mathcal{R}}(l_i)$ for individual i.

Then we test the null hypotheses

$$H_{0i} : \text{edge } l_i \text{ is missing},\ i = 1,\ldots,n. \tag{18}$$

We can do this using the false discovery rate (FDR) of [2], the expected value of the proportion of rejections that are incorrect. Let $P_{(1)} < \cdots < P_{(n)}$ denote the ordered p-values, found from $Q_{N,\mathcal{R}}(l_i), i = 1,\ldots,n$. For a given α, define $Y = \max\{\ell : P_{(\ell)} < \ell\alpha/n\}$. Then we reject the null hypotheses H_{0i} for which $P_i \le P_{(Y)}$; this ensures FDR $\le \alpha$. The proportion of values of the test statistics in the group which indicate the connection l is present is $\theta = v/n$, where v is the number of rejected hypotheses. A graph can be constructed [30] by the rule that edge l is included if $\theta > \theta_0$, where θ_0 is some threshold. The resulting estimated graph for the positive syndrome group is a function of α and θ_0, which we denote by $G(\alpha,\theta_0)$. To keep the FDR low, we took $\alpha = 0.01$.

3.6. Improving poor conditioning: diagonal up-weighting

The estimator $\hat{\boldsymbol{S}}(f_j)$ in (12) must be non-singular so that the spectal matrix can be inverted for partial coherence estimation. If tapering were absent, non-singularity is guaranteed if $2M+1 \geq p$, e.g., [9, p. 3007]. The situation is more complicated when, as here, tapering is used, as the components in the sum in (12) are then correlated. The bandwidth of $\hat{\boldsymbol{S}}(f_j)$, for $M = 9$ and the 20% cosine taper, approximately matches that of a multitaper estimator employing 15 tapers; we know such an equivalent estimator will be non-singular since $15 > p$ with $p = 10$. The excess degrees of freedom is not large however, so we would expect that the $\hat{\boldsymbol{S}}(f_j)$ terms would benefit from regularization which can be achieved by applying some diagonal up-weighting to the $\hat{\boldsymbol{S}}(f_j)$ estimates. Details of such a procedure, along with discussion of the choice of upweighting parameter, are given in [22, p. 378]; here we employ an upweighting parameter of half of one percent.

3.7. Parallelizability

In many applications the numbers of electrodes and subjects will be much larger than the 10 and 15, respectively, encountered with the data of Section 2. To get to the point of using the multiple hypothesis testing of Section 3.5, we need to complete the following steps for each individual:

(1) Compute the smoothed direct spectral estimator $\hat{\boldsymbol{S}}(f)$, and its inverse, over the delta frequency band.
(2) Calculate all the necessary test statistics $Q_{N,\mathcal{R}}(j,k)$ in (17).

Parallelization can be used to greatly speed up the calculation of the test statistics in step 2. There is no dependency between the calculation of each of the test statistics. On a multicore CPU a test statistic can be assigned to each core, and upon completion the next statistic needing calculation is assigned. The overall timing will be close to linear in the reciprocal of the number of cores used, as shown in [35] for a situation where the statistics are different to ours, but again have no dependency. So the algorithm is parallelizable with potential huge benefits.

Once all the test statistics have been computed, the extra time taken to select the graphical model via the multiple hypothesis testing of Section 3.5, is negligible.

3.8. General comments

In our graph analysis we firstly made use of partial coherence (rather than coherence) to ensure only direct connections were considered. Secondly, we used an integrated functional of these partial coherencies, covering all frequencies in the delta band of interest, with the normalized test statistic allowing for the tapering and smoothing choices applied. Finally, we paid attention to numerical conditioning of the spectral matrices, since their inversion is a key step. As a result of these 'optimized' building blocks, we are able to have high confidence in the results of the multiple hypothesis testing.

4. Determining homogeneous groups for analysis. I: disparities and distances

4.1. Disparity and distance measures

We denote n positive definite spectral matrices by $\boldsymbol{S}_i(f)$, $i = 1, \ldots, n$, and write $\boldsymbol{S}_i(f) \in \mathcal{H}^+(p)$, the set of $p \times p$ positive definite Hermitian matrices. We now consider how to measure differences between such matrices, taking into account all frequencies. We denote a disparity measure by $\delta(\boldsymbol{S}_1, \boldsymbol{S}_2)$ and a distance measure by $d(\boldsymbol{S}_1, \boldsymbol{S}_2)$; note that f is missing here because of the aggregation over frequencies. A properly defined distance must satisfy all the properties of a metric:

- positivity $d(\boldsymbol{S}_1, \boldsymbol{S}_2) \geq 0$; $\quad d(\boldsymbol{S}_1, \boldsymbol{S}_2) = 0$ iff $\boldsymbol{S}_1 = \boldsymbol{S}_2$
- symmetry $d(\boldsymbol{S}_1, \boldsymbol{S}_2) = d(\boldsymbol{S}_2, \boldsymbol{S}_1)$
- triangle inequality $d(\boldsymbol{S}_1, \boldsymbol{S}_3) \leq d(\boldsymbol{S}_1, \boldsymbol{S}_2) + d(\boldsymbol{S}_2, \boldsymbol{S}_3)$

For a disparity measure, only the first two properties necessarily hold [20, p. 333].

We shall make use of the squared Frobenius norm, $\|\boldsymbol{X}\|_{\text{Fr}}^2 \stackrel{\text{def}}{=} \text{tr}\{\boldsymbol{X}\boldsymbol{X}^H\}$, where tr denotes trace and ^{H} denotes Hermitian (complex-conjugate) transpose.

4.2. Kullback-Leibler disparity

A disparity measure for comparing, across frequencies, the spectral matrices, $\boldsymbol{S}_1(f)$ and $\boldsymbol{S}_2(f)$, of two vector-valued time series of dimension p, was given by [19, 20]. Denoted $\delta_{\text{KL}}(\boldsymbol{S}_1, \boldsymbol{S}_2)$, it takes the form

$$\int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} \tfrac{1}{2} \left[\text{tr}\{\boldsymbol{S}_1(f)\boldsymbol{S}_2^{-1}(f)\} + \text{tr}\{\boldsymbol{S}_2(f)\boldsymbol{S}_1^{-1}(f)\} - 2p \right] \, \text{d}f, \tag{19}$$

and is symmetric in its arguments. It is derived via the Kullback-Leibler information measure for discriminating between two multivariate normal densities with different covariance matrices.

Let $\{\lambda_i(f)\}_{i=1}^p$, be the eigenvalues of $\boldsymbol{S}_1(f)\boldsymbol{S}_2^{-1}(f)$. Now $\boldsymbol{S}_2^{1/2}(f)[\boldsymbol{S}_2^{-1/2}(f)\boldsymbol{S}_1(f)\boldsymbol{S}_2^{-1/2}(f)]\boldsymbol{S}_2^{-1/2}(f) = \boldsymbol{S}_1(f)\boldsymbol{S}_2^{-1}(f)$, so consequently, $\boldsymbol{S}_2^{-1/2}(f)\boldsymbol{S}_1(f)\boldsymbol{S}_2^{-1/2}(f)$ and $\boldsymbol{S}_1(f)\boldsymbol{S}_2^{-1}(f)$ share the same eigenvalues (by similarity transform). But from its form , $\boldsymbol{S}_2^{-1/2}(f)\boldsymbol{S}_1(f)\boldsymbol{S}_2^{-1/2}(f)$ is positive definite (and Hermitian) and hence the eigenvalues of $\boldsymbol{S}_1(f)\boldsymbol{S}_2^{-1}(f)$ are all positive. Further, (19) can be written as

$$\int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} \sum_{i=1}^{p} \tfrac{1}{2} \left[\lambda_i^{1/2}(f) - \lambda_i^{-1/2}(f) \right]^2 \, \text{d}f, \tag{20}$$

a form which will prove particularly useful.

Remark 4. The integrand in (19) may be rewritten as [17, p. 1726])

$$\tfrac{1}{2} \|\boldsymbol{S}_1^{-1/2}(f)\boldsymbol{S}_2^{1/2}(f) - \boldsymbol{S}_1^{1/2}(f)\boldsymbol{S}_2^{-1/2}(f)\|_{\text{Fr}}^2. \tag{21}$$

$\|\boldsymbol{C} - \boldsymbol{B}\|_{\text{Fr}}^2$ defines the squared Euclidean distance between $\boldsymbol{C}$ and $\boldsymbol{B}$ so that in this

sense the integrand behaves like a squared distance (between $\boldsymbol{S}_1^{1/2}(f)\boldsymbol{S}_2^{-1/2}(f)$ and $\boldsymbol{S}_1^{-1/2}(f)\boldsymbol{S}_2^{1/2}(f)$); the same conclusion was reached in [25]. □

4.3. Riemannian distance

Now $\boldsymbol{S}(f) \in \mathcal{H}^+(p)$, and the space $\mathcal{H}^+(p)$ is a differentiable manifold. A manifold equipped with a Riemannian metric is a Riemannian manifold.

Given a suitable inner product for elements in the tangent space that varies differentiably along the manifold, a Riemannian distance is induced between any two members $\boldsymbol{S}_1(f)$ and $\boldsymbol{S}_2(f)$ of $\mathcal{H}^+(p)$. The minimum length curve connecting two points on the manifold is called the geodesic. The Riemannian distance between the points is given by the length of this curve.

One Riemannian distance, which is invariant to coordinate transformations, is of the same form as Rao's distance between two p-variate normal distributions with zero means but different covariance matrices. This Riemannian distance, $d_{\mathrm{R}}(\boldsymbol{S}_1(f), \boldsymbol{S}_2(f))$, is defined by (e.g., [10], [17, p. 1730]),

$$\begin{aligned} d_{\mathrm{R}}^2(\boldsymbol{S}_1(f), \boldsymbol{S}_2(f)) &= \mathrm{tr}\{\log^2(\boldsymbol{S}_1(f)^{-1/2}\boldsymbol{S}_2(f)\boldsymbol{S}_1^{-1/2}(f))\} \\ &= \|\log(\boldsymbol{S}_1^{-1/2}(f)\boldsymbol{S}_2(f)\boldsymbol{S}_1^{-1/2}(f))\|_{\mathrm{Fr}}^2. \end{aligned}$$

Now $\boldsymbol{S}(f) = \boldsymbol{U}(f)\boldsymbol{M}(f)\boldsymbol{U}^H(f)$ where $\boldsymbol{M}(f) = \mathrm{diag}(\mu_1(\boldsymbol{S}(f)), \ldots, \mu_p(\boldsymbol{S}(f)))$ is the diagonal matrix having entries which are the ordered eigenvalues of $\boldsymbol{S}(f)$, and $\boldsymbol{U}(f) \in \mathbb{C}^{p\times p}$ is the unitary matrix with ith column the eigenvector corresponding to $\mu_i(\boldsymbol{S}(f)) > 0$. The matrix logarithm $\log \boldsymbol{S}(f)$ is then [3, p. 429]

$$\boldsymbol{U}(f)\,\mathrm{diag}(\log\mu_1(\boldsymbol{S}(f)), \ldots, \log\mu_p(\boldsymbol{S}(f)))\,\boldsymbol{U}^H(f). \tag{22}$$

$\log \boldsymbol{S}(f)$ is also Hermitian. In addition to being a disparity measure, the Riemannian distance satisfies the triangle inequality, and so is a true distance. It was further pointed out in [10] that $d_{\mathrm{R}}^2(\boldsymbol{S}_1(f), \boldsymbol{S}_2(f))$ can be rewritten as $d_{\mathrm{R}}^2(\boldsymbol{S}_1(f), \boldsymbol{S}_2(f)) = \sum_{i=1}^p \log^2\lambda_i(f)$ where again we denote the eigenvalues of $\boldsymbol{S}_1(f)\boldsymbol{S}_2^{-1}(f)$ by $\{\lambda_i(f)\}_{i=1}^p$. To see this note that $\log^2 \boldsymbol{S}(f)$ is

$$\boldsymbol{U}(f)\,\mathrm{diag}(\log^2\mu_1(\boldsymbol{S}(f)), \ldots, \log^2\mu_p(\boldsymbol{S}(f)))\,\boldsymbol{U}^H(f).$$

Now $d_{\mathrm{R}}^2(\boldsymbol{S}_1(f), \boldsymbol{S}_2(f)) = d_{\mathrm{R}}^2(\boldsymbol{S}_2(f), \boldsymbol{S}_1(f))$ with the latter given by

$$\begin{aligned} \mathrm{tr}\{\log^2(\boldsymbol{S}_2(f)^{-1/2}\boldsymbol{S}_1(f)\boldsymbol{S}_2^{-1/2}(f))\} &= \sum_{i=1}^p \log^2\mu_i(\boldsymbol{S}_2(f)^{-1/2}\boldsymbol{S}_1(f)\boldsymbol{S}_2^{-1/2}(f))) \\ &= \sum_{i=1}^p \log^2\lambda_i(\boldsymbol{S}_1(f)\boldsymbol{S}_2^{-1}(f)), \end{aligned}$$

as required. (Except for a factor of one half, this is the expression for Rao's distance.) To compare power spectral matrices parameterized by frequency f, we can use the

geodesic distance along the geodesic path connecting $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$, given by [17, p. 1731],

$$d_{\mathrm{R}}(\boldsymbol{S}_1, \boldsymbol{S}_2) = \left[\int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} d_{\mathrm{R}}^2(\boldsymbol{S}_1(f), \boldsymbol{S}_2(f)) \mathrm{d}f \right]^{1/2}, \tag{23}$$

which involves integration over frequencies. Equivalently, this is

$$\left[\int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} \| \log(\boldsymbol{S}_1(f)^{-1/2} \boldsymbol{S}_2(f) \boldsymbol{S}_1^{-1/2}(f)) \|_{\mathrm{Fr}}^2 \, \mathrm{d}f \right]^{1/2} \tag{24}$$

$$= \left[\int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} \sum_{i=1}^{p} \log^2 \lambda_i(f) \, \mathrm{d}f \right]^{1/2}. \tag{25}$$

Remark 5. We see immediately that $d_{\mathrm{R}}(\boldsymbol{S}_1, \boldsymbol{S}_2)$ in (25) involves taking the square root of the integral, whereas $\delta_{\mathrm{KL}}(\boldsymbol{S}_1, \boldsymbol{S}_2)$ given in (19) does not. □

4.4. Comparing the KL disparity and Riemannian distance

We start by comparing the elements of the inner sums of (20) and (25), namely

$$\tfrac{1}{2} \left[\lambda_i^{1/2}(f) - \lambda_i^{-1/2}(f) \right]^2 \quad \text{and} \quad \log^2 \lambda_i(f). \tag{26}$$

Fig. 3 shows these functional forms in terms of λ. We see that for eigenvalues between about 1/20 and 20 the two quantities are quite similar, but as λ gets much smaller than 1/20 or much larger than 20 the differences quickly become quite marked. The eigenvalues in question are those of $\boldsymbol{S}_1(f)\boldsymbol{S}_2^{-1}(f)$; if these two matrices were equal then this quantity would be $\boldsymbol{I}_p$ and all eigenvalues would be unity. $\boldsymbol{S}_1(f)\boldsymbol{S}_2^{-1}(f)$ is in general not Hermitian (unless the two matrices commute), but as we have seen the product does at least have positive eigenvalues.

We know that the condition number κ, of $\boldsymbol{S}_1(f)\boldsymbol{S}_2^{-1}(f)$ satisfies [15, p. 337] $\kappa(\boldsymbol{S}_1(f)\boldsymbol{S}_2^{-1}(f)) \leq \kappa(\boldsymbol{S}_1(f))\kappa(\boldsymbol{S}_2^{-1}(f)) = \kappa(\boldsymbol{S}_1(f))\kappa(\boldsymbol{S}_2(f))$ Since $\boldsymbol{S}_1(f) \in \mathcal{H}^+(p)$, its condition number is the ratio of largest to smallest eigenvalue of the matrix, and likewise for $\boldsymbol{S}_2(f)$. So if $\boldsymbol{S}_1(f)$ and $\boldsymbol{S}_2(f)$ are well-conditioned, i.e., κ is small in each case, $\boldsymbol{S}_1(f)\boldsymbol{S}_2^{-1}(f)$ should be quite well conditioned, and we don't expect the matrix product to have very large or very small eigenvalues. Hence the terms in (26) should be similar. This suggests that it would be reasonable to take the square root of $\delta_{\mathrm{KL}}(\boldsymbol{S}_1, \boldsymbol{S}_2)$, so that $\delta_{\mathrm{KL}}^{1/2}(\boldsymbol{S}_1, \boldsymbol{S}_2)$ and $d_{\mathrm{R}}(\boldsymbol{S}_1, \boldsymbol{S}_2)$ are generally comparable quantities.

Remark 6. We know from Remark 4 that the *integrand* in (19) looks like a squared distance. The same is true for the integrand in $d_{\mathrm{R}}(\boldsymbol{S}_1, \boldsymbol{S}_2)$ in (23). Moreover, from the discussion above we expect the integrands to be somewhat similar. Yet, the integral in (23) is subsequently square-rooted, while that in (19) is not. Consequently, as suggested above, we will indeed take the square root of the disparity measure, so that $\delta_{\mathrm{KL}}^{1/2}(\boldsymbol{S}_1, \boldsymbol{S}_2)$ and $d_{\mathrm{R}}(\boldsymbol{S}_1, \boldsymbol{S}_2)$ are comparable quantities. □

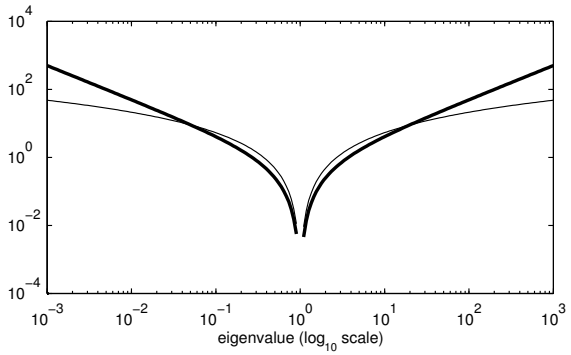

Figure 3. The heavy line shows $\frac{1}{2}\left[\lambda^{1/2} - \lambda^{-1/2}\right]^2$ and the thinner line shows $\log^2 \lambda$, as λ varies from 10^{-3} to 10^3, excluding $\lambda = 1$. Each axis has a $\log_{10}$ scale.

4.5. Distances in practice

As discussed in Section 2.2 the matrix-valued spectra are estimated at $N = 612$ discretized frequencies $f_l = l/(N\Delta_t)$ for $|f_l| \leq 10$Hz, but we are interested in the frequency interval $\mathcal{R}$ defined in (2). So, let $R = \{l \in \mathbb{Z} : f_l \in \mathcal{R}\}$. The disparity/distance between spectral matrices is found by summing over R, i.e.,

$$\tilde{\delta}^{1/2}_{\mathrm{KL}}(\boldsymbol{S}_1, \boldsymbol{S}_2) \stackrel{\mathrm{def}}{=} \left[\Delta_f \sum_{l \in R} \sum_{i=1}^{p} \tfrac{1}{2}\left[\lambda_{il}^{1/2} - \lambda_{il}^{-1/2}\right]^2\right]^{1/2}, \tag{27}$$

and

$$\tilde{d}_{\mathrm{R}}(\boldsymbol{S}_1, \boldsymbol{S}_2) \stackrel{\mathrm{def}}{=} \left[\Delta_f \sum_{l \in R} \sum_{i=1}^{p} \log^2 \lambda_{il}\right]^{1/2}, \tag{28}$$

where λ_{il} is the ith eigenvalue of $\boldsymbol{S}_1(f_l)\boldsymbol{S}_2(f_l)^{-1}$ and $\Delta_f = 1/(N\Delta t)$ is the frequency interval length. We now define symmetric 'distance' matrices $\boldsymbol{D}_{\mathrm{KL}}$ and $\boldsymbol{D}_{\mathrm{R}}$ where, for $1 \leq i < j \leq n$, $(\boldsymbol{D}_{\mathrm{KL}})_{ij} = \tilde{\delta}^{1/2}_{\mathrm{KL}}(\boldsymbol{S}_i, \boldsymbol{S}_j)$ and $(\boldsymbol{D}_{\mathrm{R}})_{ij} = \tilde{d}_{\mathrm{R}}(\boldsymbol{S}_i, \boldsymbol{S}_j)$. These will be utilised in the next section on clustering.

5. Determining homogeneous groups for analysis. II: clustering

Our graphical model for brain connectivity will be constructed from the estimated spectral matrices $\hat{\boldsymbol{S}}_i(f), f \in \mathcal{R}$. Multitaper spectral matrix estimation [26, Chapter 7] was carried out using 15 sine tapers, for which the bandwidth of the corresponding spectral window is about 0.52Hz; see [34]. The resulting estimated spectral matrices were then used in the clustering, discussed next. Note that the similarity graph discussed in this section is purely constructed for the purposes of clustering individuals, and is not related to the main connectivity group graph which we wish to construct.

5.1. Similarity graph and graph Laplacian

The (i, j)th entry, W_{ij}, of the weighted adjacency matrix, $\boldsymbol{W}$, of a 'similarity graph' G is determined by the similarity between the observations i and j, where $0 \leq W_{ij} \leq 1$, and a larger value indicates a stronger similarity. G is undirected, so $\boldsymbol{W}^T = \boldsymbol{W}$. There are several ways to construct a similarity graph from a given distance matrix. (i) The ϵ-neighbourhood graph: two vertices i, j are connected if their distance is smaller than a certain threshold ϵ; then $W_{ij} = 1$, otherwise, $W_{ij} = 0$. In this case, weighted edges are not necessary, because all edges are roughly of the same scale (i.e., ϵ). (ii) ℓ-nearest neighbour graphs: for every vertex i, i is connected to j, if j is among the ℓ-nearest neighbours of i. Because G is undirected, the direction of edges is ignored. (iii) The fully connected graph: G is completely connected. In both the ℓ-nearest neighbour graph and the fully connected graph, W_{ij} is decided by a similarity function which needs to be defined.

Similarity determines the strength of linkage between vertices in a similarity graph. However, there is no rigorous definition of similarity. One possible choice is the Gaussian similarity function (e.g., [33]). For general distance or disparity measures, with a given 'distance' matrix $(\boldsymbol{D})_{ij} = d_{ij}$, the similarity function can be defined as $W_{ij} = \exp(-d_{ij}^2/(2\sigma^2))$. The limiting cases $\sigma^2 \to \infty$ and $\sigma^2 = 0$ correspond to one and n clusters, respectively. So the choice of σ^2 is influential and must be carefully chosen. The effect of σ^2 on $\boldsymbol{W}$ is similar to ϵ in the ϵ-neighbourhood graph [33].

For a vertex $i \in V$, the degree of i is $\nu_i = \sum_{j=1}^n W_{ij}$. The degree matrix $\mathcal{D} = \mathrm{diag}(\nu_1, \ldots, \nu_n)$ is a diagonal matrix. The graph Laplacian matrix is defined as $\boldsymbol{L} \stackrel{\text{def}}{=} \mathcal{D} - \boldsymbol{W}$. The normalized graph Laplacian matrix is defined as $\boldsymbol{L}_{rw} \stackrel{\text{def}}{=} \boldsymbol{I}_n - \mathcal{D}^{-1}\boldsymbol{W}$, where $\boldsymbol{I}_n$ is the $n \times n$ identity matrix.

The following properties of eigenvalues and eigenvectors of the normalized graph Laplacian are important for clustering [33, Proposition 4]:

- $\boldsymbol{L}_{rw}$ is positive semi-definite and has n non-negative real-valued eigenvalues $0 = \lambda_1 \leq \cdots \leq \lambda_n$.
- The multiplicity k of the eigenvalue 0 of $\boldsymbol{L}_{rw}$ equals the number of connected components $U_1, \ldots, U_k$ in the graph. (A subset U of V is called a connected component if it is connected and there is no path connecting U and $V \backslash U$.)
- The eigenspace of 0 is spanned by the vectors $\mathbf{1}_{U_i}$ of those components.

Here $\mathbf{1}_{U_i} = (y_1, \ldots, y_n)^T$, where $y_j = 1$ if $v_j \in U_i$ and zero otherwise.

Hence (i) the number of clusters can be determined by the multiplicity of the eigenvalue 0, and (ii) the membership of each vertex is clearly defined by eigenvectors.

The clustering algorithm based on the normalized graph Laplacian $\boldsymbol{L}_{rw}$ and Gaussian similarity, is implemented as follows:

(1) Given $\boldsymbol{W} \in \mathbb{R}^{n \times n}$, the eigenvalues $0 = \lambda_1 \leq \cdots \leq \lambda_n$ and eigenvectors $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n\}$ of the normalized Laplacian $\boldsymbol{L}_{rw}$ are found.
(2) The number of clusters, k, is chosen from the pattern of the eigenvalues: $\lambda_2, \ldots, \lambda_k$ are very close to 0, but λ_{k+1} is relatively large. (Note that λ_1 is always 0.)
(3) The eigenvector matrix $[\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k] \in \mathbb{R}^{n \times k}$ defines n points $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ in $\mathbb{R}^k$. Cluster the $\boldsymbol{x}_i$'s using the k-means algorithm into k clusters.

See [31, 33] for full discussion of, and justifications for, this clustering algorithm.

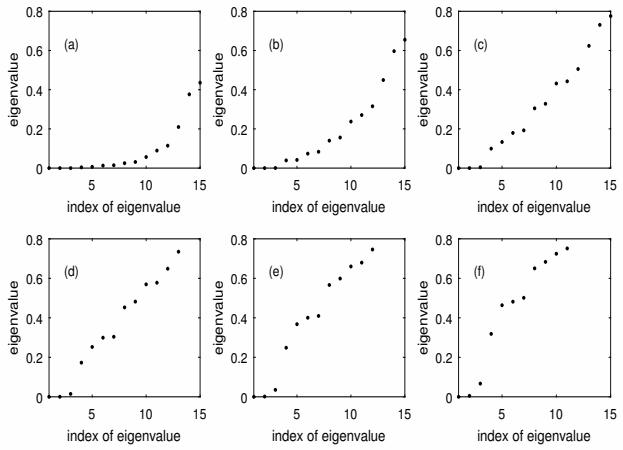

Figure 4. For $\boldsymbol{D}_{\mathrm{KL}}$: eigenvalues of $\boldsymbol{L}_{rw}$ for (a) $\sigma^2 = 40$, (b) $\sigma^2 = 60$, (c) $\sigma^2 = 80$, (d) $\sigma^2 = 100$, (e) $\sigma^2 = 120$, and (f) $\sigma^2 = 140$.

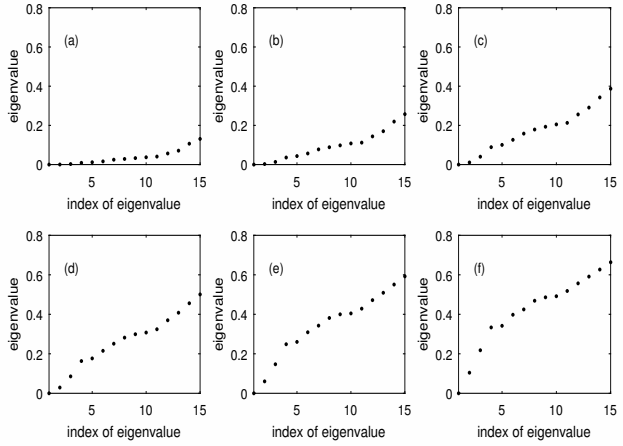

Figure 5. For $\boldsymbol{D}_{\mathrm{R}}$: eigenvalues of $\boldsymbol{L}_{rw}$ for (a) $\sigma^2 = 15$, (b) $\sigma^2 = 20$, (c) $\sigma^2 = 25$, (d) $\sigma^2 = 30$, (e) $\sigma^2 = 35$, and (f) $\sigma^2 = 40$.

5.2. Results

In this study there were 15 patients, so that sparsity of the weighted adjacency matrix $\boldsymbol{W}$ is not a main concern. Hence we adopt a fully-connected similarity graph and concentrate our attention on the choice of σ in the Gaussian similarity function.

Using $\boldsymbol{D}_{\mathrm{KL}}$, the eigenvalues of $\boldsymbol{L}_{rw}$ for $\sigma^2 \in \{20, 40, 60, 80, 100, 120, 140\}$ are shown in Fig. 4. We see a persistent characteristic, namely that the first three eigenvalues are close to zero for plots (a)-(d). In plot (d) the gap between λ_3 and λ_4 is large. This suggests to take $k = 3$ along with $\sigma^2 = 100$. The corresponding results using $\boldsymbol{D}_{\mathrm{R}}$ for $\sigma^2 \in \{15, 20, 25, 30, 35, 40\}$ are given in Fig. 5. Fig. 5(b) shows similar behaviour to Fig. 4(b), but generally it is more difficult to choose k.

Remark 7. Although the clustering algorithm nominally requires the first k eigenvectors, the first eigenvector is a constant vector $\mathbf{1}$ and does not provide any information to the clustering. So we need only include $[\boldsymbol{u}_2\ \boldsymbol{u}_3]$. □

So $k = 3$ was deemed the best choice and we treat each row of $[\boldsymbol{u}_2\ \boldsymbol{u}_3]$ as a point in $\mathbb{R}^2$, as seen in Fig. 6, which indicates the number of clusters is 3, giving the par-

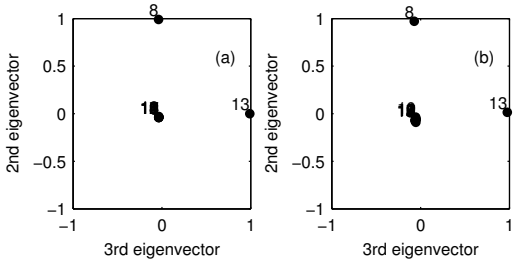

Figure 6. Eigenvectors corresponding to λ_2 and λ_3 of $\boldsymbol{L}_{rw}$. The distances between estimated spectral matrices are measured via $\boldsymbol{D}_{\mathrm{KL}}$ in (a), and by $\boldsymbol{D}_{\mathrm{R}}$ in (b).

tition $\{\{1,\ldots,7,9,\ldots,12,14,15\},\{8\},\{13\}\}$. Patients 8 and 13 appear as 'outliers' compared to the group of 13 remaining individuals.

6. Results and Conclusions

Following the clustering in Section 5, individuals 8 and 13, identified as outliers, are removed from the original set of 15 individuals, leaving us with a homogeneous group of $n = 13$ individuals. We then seek to construct a representative connectivity graph, i.e., a group graph, for these 13 individuals. To do this we use the methodology of Section 3. It was explained in Section 3.5 how group connectivities may be defined in terms of a threshold θ_0. Example connectivities are shown in Fig. 7 for $\theta_0 = 0.4, 0.5, 0.6$ and 0.69, corresponding to 6 or more, 7 or more, 8 or more, 9 or more, individuals out of the 13, respectively. The connection F3-O2 and symmetric connections C3-P3/C4-P4 are clearly quite persistent. There are no connections which are present for 9 or more individuals. Connection F3-O2 is particularly interesting since relatively long physical distances between connected regions, (compatible with inefficient axonal wiring), have already been associated with schizophrenia [1].

The building of a representative connectivity group graph is not an easy task. Firstly quite sophisticated clustering methods are required to eliminate outlying individuals. Our novel approach of spectral matrix clustering via the normalized graph Laplacian greatly aids clustering since it identifies the number of clusters, which then enables use of standard k-means clustering. Having 'cleaned' the data, it is then necessary to consider possible edges in a graphical model. In a time series setting even this is difficult, since conventional hypothesis testing involves testing over the whole frequency range of interest [30]; use of an integrated functional eliminates this complication. Moreover, the test statistic has a simple distribution, so that p-values are easily found, which can be used in multiple hypothesis testing over *individuals*; this provides the theshold parameter θ_0 which can be varied to easily view persistent connections over the group.

Apart from these considerations, it is necessary to check that chosen spectral smoothing parameters and taper choices correspond to bandwidths consistent with the overall delta frequency band. Further, it is necessary that the spectral matrices are invertible, which also constrains parameter choices.

This analytical effort is worthwhile since the production of group analyses from EEG in schizophrenia studies, (and elsewhere), is an interesting, and important prac-

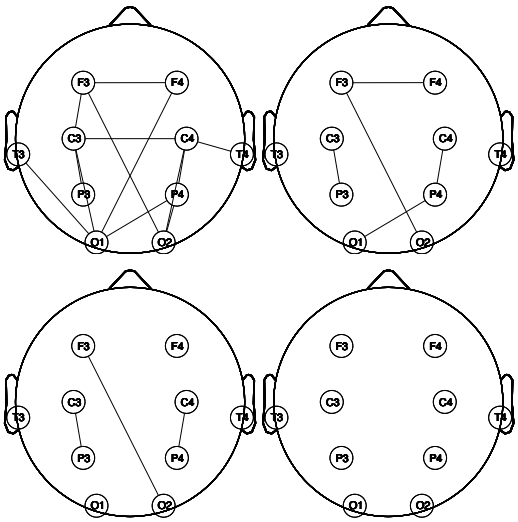

Figure 7. Derived connectivity for a threshold of $\theta_0 = 0.4$ (top left), 0.5 (top right), 0.6 (lower left) and 0.69 (lower right).

tical task [13, 23], which informs clinical practitioners and neuroscientists about the syndrome under study. We believe our novel approach provides a framework for integrating results on multiple individuals into a single overall connectivity structure, and is worthy of further study on other datasets.

Acknowledgement

Linjie Zhuang was supported by a President's PhD Scholarship from Imperial College London. The authors thank the associate editor and anonymous referees for their helpful comments and suggestions.

References

[1] D. S. Bassett, E. Bullmore, B. A. Verchinski, V. S. Mattay, D. R. Weinberger and A. Meyer-Lindenberg, *Hierarchical organization of human cortical networks in health and schizophrenia,* J. Neuroscience 28 (2008), pp. 9239–9248.

[2] Y. Benjamini and Y. Hochberg, *Controlling the false discovery rate: a practical and powerful approach to multiple testing,* J. Roy. Stat. Soc. B 57 (1995), pp. 289-300

[3] D. S. Bernstein, *Matrix Mathematics*. Princeton, NJ: Princeton University Press, 2005.

[4] E. Bullmore and P. Fletcher, *The eye's mind: brain mapping and psychiatry,* British J. Psychiatry 182 (2003), pp. 381–384.

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

[6] R. Dahlhaus, *Graphical interaction models for multivariate time series*, Metrika 51 (2000), pp. 157-172.

[7] M. Eichler, *A frequency-domain test for non-correlation between stationary time series*, Metrika 65 (2007), pp. 133–157.

[8] M. Eichler, *Testing nonparametric and semiparametric hypotheses in vector stationary processes* J. Multivar. Anal. 99 (2008), pp. 968–1009.

[9] M. Fiecas, H. Ombao, C. Linkletter, W. Thompson & J. Sanes, *Functional connectivity: shrinkage estimation and randomization test*, NeuroImage 49 (2010), pp. 3005–3014.

[10] W. Förstner and B. Moonen, *A metric for covariance matrices*, in *Quo vadis geodesia ... Festschrift for Erik W. Grafarend on the occasion of his 60th birthday,* Eds: F. Krumm and V. S. Schwarze, Tech. Report, Dept. of Geodesy and Geoinformatics, Stuttgart University, 1999.

[11] K. J. Friston and C. D. Frith, *Schizophrenia: a disconnection syndrome?* Clin. Neurosci. 3 (1995), pp. 89–97.

[12] S. Galderisi, A. Mucci, U. Volpe & N. Boutros, *Evidence-based medicine and electrophysiology in schizophrenia*, Clin. EEG Neurosci. 40 (2009), pp. 62–77.

[13] K. R. Henshall, A. E. Sergejew, G. Rance, C. M. McKay and D. L. Copolov, *Interhemisheric EEG coherence is reduced in auditory cortical regions in schizophrenia patients with auditory hallucinations*, Int. J. Psychophysiology 89 (2013), pp. 63–71.

[14] S. Holm, *A simple sequentially rejective multiple test procedure*, Scand. J. Stat. 6 (1979), pp. 65–70.

[15] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge UK: Cambridge University Press, 1985.

[16] T. Itoh, T. Sumiyoshi, Y. Higuchi, M. Suzuki & Y. Kawasaki, *LORETA analysis of three-dimensional distribution of delta band activity in schizophrenia: relation to negative symptoms?*, Neurosci. Res. 70 (2011), pp. 442–448.

[17] X. Jiang, L. Ning and T. T. Georgiou, *Distances and Riemannian metrics for multivariate spectral densities,* IEEE Trans. Autom. Control 57 (2012), pp. 1723–1735.

[18] B. Jones and M. West, *Covariance decomposition in undirected Gaussian graphical models*, Biometrika 92 (2005) pp. 779–86.

[19] D. Kazakos and P. Papantoni-Kazakos, *Spectral distance measures between Gaussian processes*, IEEE Trans. Autom. Control 25 (1980), pp. 950-59.

[20] Y. Kakizawa, R. H. Shumway and M. Taniguchi, *Discrimination and clustering for multivariate time series,* J. of the American Statistical Association 93 (1998), pp. 328–340.

[21] E. W. Lang, A. M. Tomé, I. R. Keck, J. M. Górriz-Sáez and C. G. Puntonet, *Brain connectivity analysis: a short survey,* Comput. Intell. and Neurosci. vol. 2012, ID: 412512, 21 pages, 2012.

[22] T. Medkour, A. T. Walden and A. Burgess, *Graphical modelling for brain connectivity via partial coherence,* J. Neurosci. Meth. 180 (2009), pp. 374–383.

[23] T. Medkour, A. T. Walden, A. P. Burgess & V. B. Strelets, *Brain connectivity in positive and negative syndrome schizophrenia,* Neuroscience 169 (2010), pp. 1779–88.

[24] S. Micheloyannis, E. Pachou, C. J. Stam, M. Breakspear, P. Bitsios, M. Vourkas, S. Erimaki & M. Zervakis, *Small-world networks and disturbed functional connectivity in schizophrenia,* Schizophrenia Research 87 (2006), pp. 60–66.

[25] M. Moakher and P. G. Batchelor, *Symmetric positive-definite matrices: From geometry to applications and visualization,* in *Visualization and Image Processing of Tensor Fields,* Eds: J. Weickert and H. Hagen, chapter 17, pp. 285–298, New

York: Springer, 2005.
[26] D. B. Percival and A. T. Walden, *Spectral Analysis for Physical Applications.* Cambridge: Cambridge University Press, 1993.
[27] M. B. Priestley, *Spectral Analysis and Time Series*, New York: Academic Press, 1981.
[28] J. R. Rosenberg, D. M. Halliday, P. Breeze and B. A. Conway, *Identification of patterns of neuronal connectivity — partial spectra, partial coherence, and neuronal interactions,* J. Neurosci. Meth., 83 (1998), pp. 57–72.
[29] R. Salvador, J. Suckling, C. Schwarzbauer and E. Bullmore, *Undirected graphs of frequency-dependent functional connectivity in whole brain networks,* Phil. Trans. R. Soc. Lond., B, 360 (2005), pp. 937-46.
[30] D. Schneider-Luftman, *p-value combiners for graphical modelling of EEG data in the frequency domain,* J. Neurosci. Meth. 271 (2016), pp. 92–106.
[31] J. Shi and J. Malik, *Normalized cuts and image segmentation,* IEEE Trans. Pattern Anal. Mach. Intell., 22 (2000),pp. 888–905.
[32] R. J. Simes, *An improved Bonferroni procedure for multiple tests of significance?*, Biometrika 73 (1986), pp. 751–754.
[33] U. von Luxburg, *A tutorial on spectral clustering,* Statistics and Computing 17 (2007), pp. 395–416.
[34] A. T. Walden, E. J. McCoy, and D. B. Percival, *The effective bandwidth of a multitaper spectral estimator,* Biometrika 82 (1995), pp. 201–214.
[35] R. J. Wolstenholme and A. T. Walden, *An efficient modelling approach to graphical modeling of time series,* IEEE Transactions on Signal Processing 63, pp. 3266–3276.