Biomedical Physics & Engineering Express

NOTE

Minimum spanning tree and *k*-core decomposition as measure of subject-specific EEG traits

Alessandra Crobe[1], **Matteo Demuru**[2], **Luca Didaci**[2], **Gian Luca Marcialis**[2] **and Matteo Fraschini**[2,3]

[1] Biomedical Engineering PhD Course, University of Cagliari, Piazza D'armi, Cagliari, 09123, Italy
[2] Department of Electrical and Electronic Engineering, University of Cagliari, Piazza D'armi, Cagliari, 09123, Italy
[3] Author to whom any correspondence should be addressed.

E-mail: fraschin@unica.it

RECEIVED
16 July 2015

REVISED
11 December 2015

ACCEPTED FOR PUBLICATION
15 December 2015

PUBLISHED
18 January 2016

Keywords: EEG, minimum spanning tree, networks

Abstract

The identification of subject-specific traits extracted from patterns of brain activity still represents an important challenge. The need to detect distinctive brain features, which is relevant for biometric and brain computer interface systems, has been also emphasized in monitoring the effect of clinical treatments and in evaluating the progression of brain disorders. In this study we propose an approach which aims to investigate the existence of a distinctive functional core (sub-network) using an unbiased reconstruction of network topology. Brain signals from a public and freely available EEG dataset were analysed using a phase synchronization based measure, minimum spanning tree and *k*-core decomposition. The analysis was performed for each classical brain rhythm separately. Highest classification rates from *k*-core decomposition were obtained in the gamma (EER = 0.130, AUC = 0.943) and high beta (EER = 0.172, AUC = 0.905) frequency bands. These results confirm that EEG analysis may represent an effective tool to identify subject-specific characteristics that may be of great impact for several bioengineering applications. However, despite the widespread use of these techniques, critical aspects should be considered when dealing with results from high-frequency scalp EEG.

1. Introduction

The identification of subject-specific traits extracted from patterns of brain activity still represents an important challenge. In contrast to the widely used approach based on identification of functional brain features that allow distinguishing between groups of subjects (generally, patients and healthy controls), the detection of human distinctive traits estimated from brain activity may be of help in several relevant applications.

Recently, the need to detect subject-specific functional brain traits has been emphasized in both clinical [1] and biometric applications [2–4]. In particular, functional connectivity and tools from complex network analysis, which represent a new paradigm in the study of brain organization [5], have suggested that the resting state may reveal network organizations linked with individual cognitive and behavioural trajectories [6, 7]. Moreover, it has been shown [8] that resting-state functional connectivity is associated with individual differences (phenotypic variability) in several domains such as behavioural traits, neurological conditions and response to treatments.

Therefore, subject-specific network features seem to be candidate markers in monitoring the effects of treatments and in evaluating the progression of brain disorders in the era of personalized medicine. Furthermore, biometric applications strongly require the clear definition of distinctive brain activity features. In the last decade, several papers investigated the use of electroencephalographic (EEG) features with the aim to characterize subject-specific brain traits. Even though it has been shown that simple power spectral [2] measures allow us to obtain high identification performances, questions related to subject identification remain to be answered. Recently, functional connectivity [9] and brain network organization measures [10] have been successfully applied in order to investigate human brain distinctiveness, also adding important information on physiological implications of the identified specific traits. Nevertheless, it is well

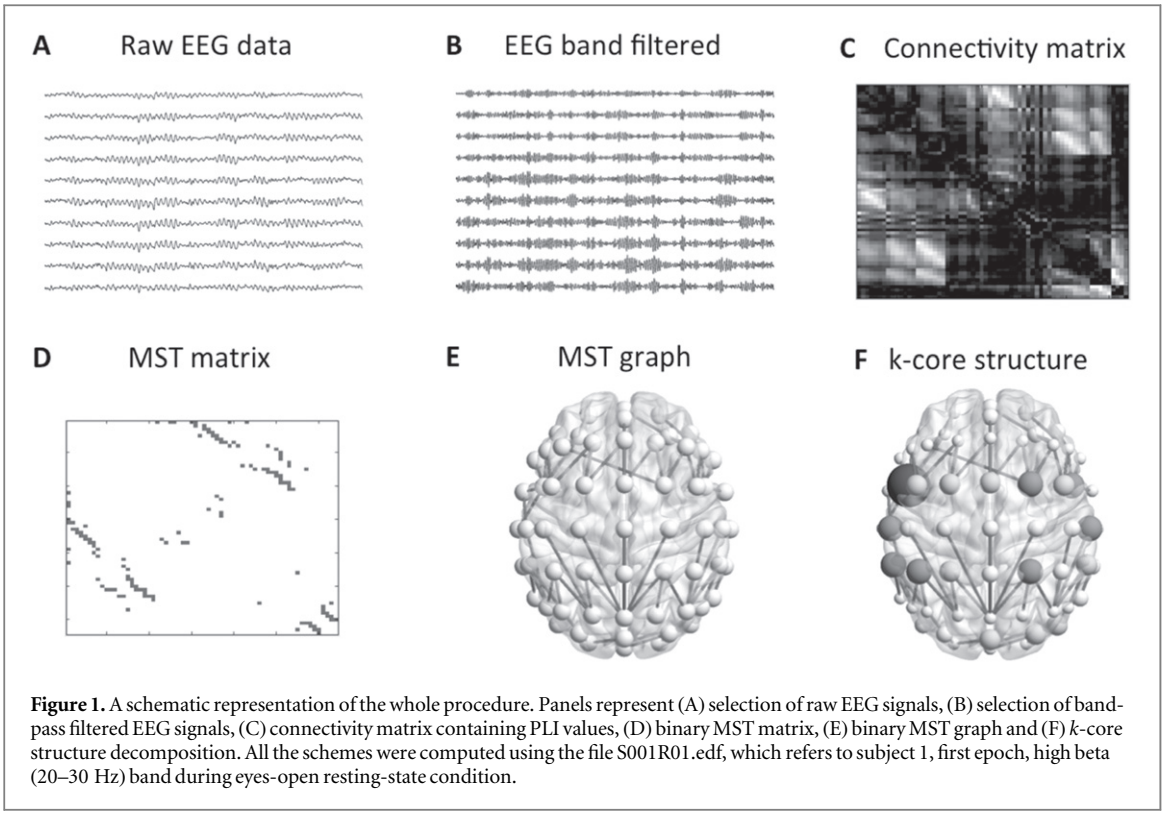

Figure 1. A schematic representation of the whole procedure. Panels represent (A) selection of raw EEG signals, (B) selection of band-pass filtered EEG signals, (C) connectivity matrix containing PLI values, (D) binary MST matrix, (E) binary MST graph and (F) *k*-core structure decomposition. All the schemes were computed using the file S001R01.edf, which refers to subject 1, first epoch, high beta (20–30 Hz) band during eyes-open resting-state condition.

established that comparing different networks is not without difficulties. Since number of nodes and average degree may have a great influence on network parameters, the comparison between measures extracted from empirical networks can yield spurious results [11]. The identification of (arbitrary) thresholds or the use of normalization procedures (based on surrogate data) do not allow for a reliable unbiased comparison. It has been shown that the minimum spanning tree (MST), an acyclic sub-network that connects all nodes, may represent an unbiased method for brain network comparison [12, 13].

Moreover, recent studies have shown that central nodes (hubs) tend to form a densely linked community called rich-club [14–16], which plays an important role in integrating and disseminating information across the entire network. These results suggest that rich-club connections form a central functional core (backbone) responsible for efficient global brain communication.

In this study we propose an approach which aims to investigate the existence of a distinctive central functional core using an unbiased reconstruction of network topology. Brain signals from a public and freely available EEG dataset were analysed using a phase synchronization based measure, namely the phase lag index (PLI) [17], which allows us to estimate statistical interdependences between EEG time series. Successively, the MST was used to filter each network in order to increase comparability between measures extracted from individual networks. Finally, the *k*-core decomposition [18] was applied to disentangle the hierarchical structure of networks and identify the central functional core. We hypothesize that each single subject can be characterized by an distinctive functional core topology.

2. Material and methods

2.1. Data set

A public and freely available high-density (64 channels) EEG dataset [19], consisting of 109 healthy subjects, was used in this study. Raw data can be downloaded from the PhysioNet web site (http://physionet.org/pn4/eegmmidb/). The same dataset has been previously used for brain computer interface [20] and biometric applications [9, 10].

EEG signals, acquired with a sampling rate of 160 Hz and referenced to the average of the ear-lobe electrodes, are organized in several different runs using resting state, motor movement and imaginary tasks. For the subsequent analysis we used two resting-state runs (eyes open and eyes closed), each one lasting 2 min.

2.2. Pre-processing

Five nonoverlapping epochs (12 s long, corresponding to 1920 samples), extracted from the 2 min resting-state conditions, were analysed for each single subject. The raw EEG signals were band-pass filtered (without phase distortion [21]) in the classical frequency bands: delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), low beta (13–20 Hz), high beta (20–30 Hz) and gamma (30–45 Hz).

Table 1. Recognition rates. Recognition rates expressed as EER and AUC for each frequency band and for both eyes-open (EO) and eyes-closed (EC) resting-state conditions. In addition, a comparison with EERs derived from the analysis without the use of MST (No-MST) is also reported.

	MST EO		MST EC		No-MST EO	No-MST EC
Frequency band	EER	AUC	EER	AUC	EER	EER
Delta	0.429	0.605	0.480	0.541	0.371	0.440
Theta	0.445	0.580	0.441	0.585	0.411	0.424
Alpha	0.354	0.693	0.348	0.711	0.339	0.315
Low beta	0.257	0.817	0.240	0.837	0.204	0.229
High beta	0.172	0.905	0.173	0.906	0.114	0.143
Gamma	0.131	0.943	0.130	0.933	0.050	0.072

Successively, for each condition (eyes open and eyes closed), each subject and each frequency band, the following analysis was performed: (i) functional connectivity analysis, (ii) MST filtering and (iii) functional core characterization. A schematic view of the whole procedure is represented in figure 1.

2.3. Functional connectivity analysis

The functional connectivity analysis was performed by computing pair-wise statistical interdependence between EEG time series using the PLI [17, 22]. The PLI, an index of asymmetry of the distribution of instantaneous phase differences between pairs of channels, allows us to address common scalp EEG problems due to signal spreads, linear mixing and active reference [17].

The PLI, which assumes values in the range from 0 (no interaction, or interaction with zero phase lag) to 1 (maximum interaction), can be computed as follows:

$$\mathrm{PLI} = |\langle \mathrm{sign}[\sin(\Delta\phi(t_k))]\rangle| \quad (1)$$

where $\Delta\phi$ is the difference between instantaneous phases, t_k are discrete steps and $\langle\ \rangle$ denotes the average over the time t.

The functional connectivity analysis allows us to represent the functional interaction between brain regions as a weighted connectivity matrix, where the PLI represents the magnitude of the relationship. In order to validate the results, we compared PLI-based features with features computed by using another widely employed phase synchronization measure, namely the phase locking value (PLV) [23]. Before estimating the PLV the time series were orthogonalized by means of linear regression analysis as described by [24, 25] to remove trivial correlations due to field spread or volume conduction (leakage corrected).

2.4. MST filtering

Network analysis has revealed important aspects about the organization and functioning of the complex brain system [5]. However, comparison between measures extracted from empirical networks can yield spurious results [11]. The use of MST has been recently proposed as an unbiased method for brain network comparison [12, 13].

The MST is an acyclic sub-network that connects all nodes minimizing the link weights. In this study MSTs are constructed based on the weighted networks using the Kruskal algorithm [26], starting with the largest link weights.

The weights are sorted in a descending order; the construction of the tree starts using the largest link weight and successively the following largest link weight is added. The procedure, preserving the acyclic condition, continues until all N nodes are connected with $N-1$ links. Finally, the constructed MSTs are binarized (e.g., all weights are assigned a value of one). Furthermore, the construction of the MST is independent of arbitrary thresholds.

2.5. Functional core characterization

It has been shown that the human brain is characterized by the existence of a central core, formed by a densely linked community of the most important nodes (hubs), responsible for efficient global communication. In order to identify the central functional core, the k-core decomposition [18, 27] was applied to the reconstructed binary MSTs.

The k-core, which is the largest sub-network comprising nodes of degree at least k, is computed by recursively peeling off nodes with degree lower than k. The procedure continues until no such nodes remain in the sub-network. For a full description of the algorithm see [18].

The k-core decomposition analysis was performed using the Brain Connectivity Toolbox (which is freely available at the web site https://sites.google.com/site/bctnet/) for MATLAB [28]. Figures were obtained using the Toolbox BrainNet Viewer [29].

2.6. Classification

In order to test the k-core distinctiveness property, a feature vector (consisting of 64 entries), expressing the coreness value of each node within the MST network, was defined for each epoch and subject. The coreness of a node is k if the node belongs to the k-core. For each band and condition, a total of 545 (5 epochs $\times$ 109 subjects) feature vectors were used for the classification procedure.

Pair-wise similarity scores between feature vectors (nodal coreness) were estimated as $1/(1+d)$, where d represents the Euclidean distance. Genuine and impostor scores were successively used to evaluate the equal error rate (EER) and the area under the ROC curve (AUC), which allow us to assess the performance of the proposed approach for each band and condition.

In order to assess the statistical significance of recognition rates a nonparametric permutation testing (using 1000 iterations) was used. During each

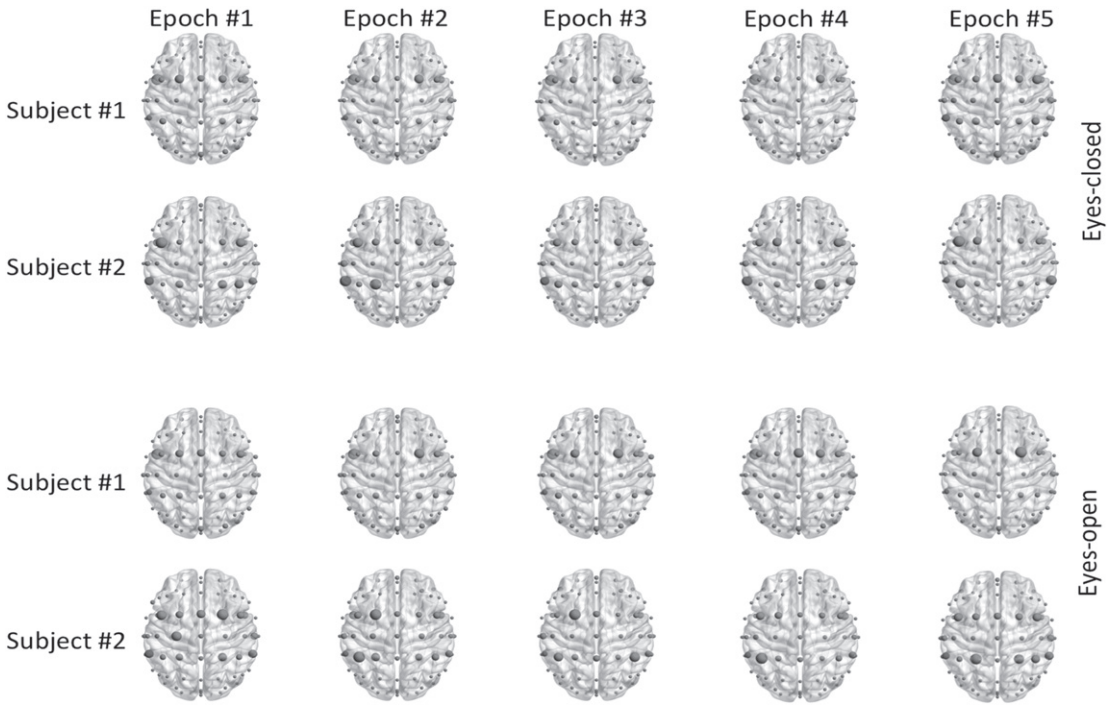

Figure 2. A representation of the *k*-cores for two representative subjects for the gamma frequency band.

single iteration the feature vectors from one subject were randomly assigned to other subjects, the similarity scores were computed and the recognition rate (in terms of EER) evaluated. The *p*-value is given by the percentage of iterations for which the EER obtained is lower than the EER obtained in analysis.

3. Results

The results, which represent recognition rates in terms of EER and AUC, are summarized in table 1. For both eyes-open and eyes-closed resting-state conditions, higher recognition rates were obtained in gamma (EER = 0.131 and AUC = 0.943 in eyes-open condition; EER = 0.130 and AUC = 0.933 in eyes-closed condition) and high beta (EER = 0.172 and AUC = 0.905 in eyes-open condition; EER = 0.173 and AUC = 0.906 in eyes-closed condition) bands. Lower recognition rates were obtained in lower frequency bands. A comparison with the results obtained from *k*-cores extracted without the use of the MST is reported in table 1.

The highest recognition rate (in terms of EER) achieved with the permutation tests was 0.484, thus suggesting a statistical significance ($p = 0.001$, expressed as 1/number of permutations) for all the reported results. In figure 2 we have included a representation of the *k*-cores for two representative subjects in both eyes-closed and eyes-open conditions.

The comparison between PLI- and PLV-based recognition rates (in terms of EER) shows very similar results, thus allowing us to validate the reported results (see table 2).

Table 2. Recognition rates. Recognition rates expressed as EER from PLI- and PLV-based networks, for each frequency band and for both eyes-open and eyes-closed resting-state conditions.

	PLI based		PLV based	
Frequency band	Eyes open	Eyes closed	Eyes open	Eyes closed
Delta	0.429	0.480	0.404	0.464
Theta	0.445	0.441	0.403	0.433
Alpha	0.354	0.348	0.348	0.334
Low beta	0.257	0.240	0.244	0.235
High beta	0.172	0.173	0.156	0.165
Gamma	0.131	0.130	0.112	0.119

4. Discussion

Although the present study is based on a new approach for the detection of individual traits extracted from EEG signals, the reported results confirm, as recently suggested [1, 9, 10], that network analysis may be of help to characterize distinctive functional brain fingerprint.

The proposed approach introduces two relevant aspects that considerably add to the current literature on EEG based subject identification methods. The first aspect is about the identification of a characteristic sub-network, which is responsible for the efficient global brain communication. In contrast to previous methods based on (i) arbitrary selection of a limited

number of channels or (ii) characterization of the whole-brain structure, the proposed approach reflects the role of well known physiological mechanisms. Interestingly, it has been previously shown [18] that the use of *k*-core decomposition allows us to disentangle the hierarchical structure of a network (progressively focusing on its central core), which represents a specific fingerprint of the network. The second important aspect is about the use of the MST. Indeed, even though network analysis has revealed important aspects about the organization and functioning of the complex brain organization [5], it has been highlighted that comparing different networks is strongly hindered by several methodological issues [30–32]. In particular, since the number of nodes and the average degree may have a great influence on network parameters, the comparison between measures extracted from empirical networks can yield spurious results [11]. The MST, proposed as an unbiased method for brain network comparison [12, 13], seems to solve relevant methodological limitations of previous works (e.g. sensitivity to alterations in connection strength). Furthermore, the MST efficiently captures the essential properties of complex networks.

Moreover, with the use of the PLI, an index of phase synchronization that reduces the effects of common problems in scalp EEG, this work tends to limit as much as possible the introduction of methodological biases that hinder the use of network measures in the study of brain organization.

However, the use of these strict and conservative criteria, which simplify the understanding of physiological implications of the findings, may be responsible for the slightly lower reported classification rates if compared with previously published papers [9, 10]. This is also confirmed by the results obtained without introducing the use of the MST (see table 1). Furthermore, the almost identical results obtained from the two conditions (eyes closed and eyes open) are in line with the similar spatial representation of *k*-cores from the same subjects (see figure 2).

Despite the promising results, our study suffers from some limitations that need to be discussed. First, since high-frequency (>20 Hz) scalp EEG components overlap with myogenic activity [33], it is difficult to estimate to what extent the role of high-frequency bands is due to neural sources or to muscle contamination. Second, the used dataset is noisy (which, however represents a more useful approximation for real life applications) and EEG signals are acquired during a single session (which could tend to overestimate the recognition rates).

Therefore, it seems necessary that future works focus their attention to design well controlled experimental setups that can help in to disentangle neural sources from artefact contaminations and to investigate possible changes induced by multiple session recordings. It would also be of great interest to investigate the use of tools that may allow us to filter out muscle activity from scalp EEG (see [33] for a comprehensive review).

5. Conclusions

The proposed approach, based on the detection of distinctive functional fingerprints (core sub-network), may help in elucidating crucial mechanisms related to subject-specific EEG traits. Furthermore, the reported findings may be of great impact for both personalized medicine and bioengineering applications as biometric and BCI systems.

Acknowledgments

A Crobe performed her activity in the framework of the PhD in Biomedical Engineering at the University of Cagliari. M Demuru performed his activity in the framework of the PhD in Electronic Engineering and Computer Science at the University of Cagliari.

References

[1] Miranda-Dominguez O, Mills B D, Carpenter S D, Grant K A, Kroenke C D, Nigg J T *et al* 2014 *Connectotyping : Model Based Fingerprinting of the Functional Connectom PLos Oné* **9** e11048
[2] Del Pozo-Banos M, Alonso J B, Ticay-Rivas J R and Travieso C M 2014 Electroencephalogram subject identification: a review *Expert Syst. Appl.* **41** 6537–54
[3] Campisi P and La Rocca D 2014 Brain waves for automatic biometric based user recognition *IEEE Trans. Inf. Forensics Secur.* **9** 782–800
[4] DelPozo-Banos M, Travieso C M, Weidemann C T and Alonso J B 2015 EEG biometric identification: a thorough exploration of the time–frequency domain *J. Neural Eng.* **12** 056019
[5] Stam C J 2014 Modern network science of neurological disorders *Nat. Rev. Neurosci.* **15** 683–95
[6] Sporns O 2013 Structure and function of complex brain networks *Dialogues Clin. Neurosci.* **15** 247–62
[7] Finn E S, Shen X, Scheinost D, Rosenberg M D, Huang J, Chun M M, Papademetris X and Constable R T 2015 Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity *Nat. Neurosci.* **18** 1664–71
[8] Vaidya C J and Gordon E M 2013 Phenotypic variability in resting-state functional connectivity: current status *Brain Connect.* **3** 99–120
[9] La Rocca D, Campisi P, Vegso B, Cserti P, Kozmann G, Babiloni F and DE Vico Fallani F 2014 Human brain distinctiveness based on EEG spectral coherence connectivity *IEEE Trans. Biomed. Eng.* **9294** 1–7
[10] Fraschini M, Hillebrand A, Demuru M, Didaci L and Marcialis G L 2015 An EEG-based biometric system using eigenvector centrality in resting state brain networks *IEEE Signal Process. Lett.* **22** 666–70
[11] van Wijk B C M, Stam C J and Daffertshofer A 2010 Comparing brain networks of different size and connectivity density using graph theory *PLoS One* **5** e13701
[12] Tewarie P, van Dellen E, Hillebrand A and Stam C J 2015 The minimum spanning tree: an unbiased method for brain network analysis *NeuroImage* **104** 177–88
[13] Stam C J, Tewarie P, Van Dellen E, van Straaten E C W, Hillebrand A and Van Mieghem P 2014 The trees and the forest: characterization of complex brain networks with minimum spanning trees *Int. J. Psychophysiol.* **92** 129–38

[14] Colizza V, Flammini A, Serrano M A and Vespignani A 2006 Detecting rich-club ordering in complex networks *Nat. Phys.* **2** 110–5
[15] van den Heuvel M P, Kahn R S, Goñi J and Sporns O 2012 High-cost, high-capacity backbone for global brain communication *Proc. Natl Acad. Sci. USA* **109** 11372–7
[16] van den Heuvel M P and Sporns O 2011 Rich-club organization of the human connectome *J. Neurosci.* **31** 15775–86
[17] Stam C J, Nolte G and Daffertshofer A 2007 Phase lag index: assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources *Hum. Brain Mapp.* **28** 1178–93
[18] Ignacio Alvarez-hamelin A V J, Dall'Asta L and Barrat A 2005 Large scale networks fingerprinting and visualization using the *k*-core decomposition *Adv. Neural Inf. Process. Syst.* **18** 41–50
[19] Goldberger A L, Amaral L A N, Glass L, Hausdorff J M, Ivanov P C, Mark R G, Mietus J E, Moody G B, Peng C-K and Stanley H E 2000 PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals *Circulation* **101** e215–20
[20] Demuru M, Fara F and Fraschini M 2013 Brain network analysis of EEG functional connectivity during imagery hand movements *J. Integr. Neurosci.* **12** 441–7
[21] Delorme A and Makeig S 2004 EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis *J. Neurosci. Methods* **134** 9–21
[22] Hillebrand A, Barnes G R, Bosboom J L, Berendse H W and Stam C J 2012 Frequency-dependent functional connectivity within resting-state networks: an atlas-based MEG beamformer solution *NeuroImage* **59** 3909–21
[23] Lachaux J P, Rodriguez E, Martinerie J and Varela F J 1999 Measuring phase synchrony in brain signals *Hum. Brain Mapp.* **8** 194–208
[24] Hipp J F, Hawellek D J, Corbetta M, Siegel M and Engel A K 2012 Large-scale cortical correlation structure of spontaneous oscillatory activity *Nat. Neurosci.* **15** 884–90
[25] Brookes M J, Hale J R, Zumer J M, Stevenson C M, Francis S T, Barnes G R, Owen J P, Morris P G and Nagarajan S S 2011 Measuring functional connectivity using MEG: methodology and comparison with fcMRI *NeuroImage* **56** 1082–104
[26] Kruskal J B 1956 On the shortest spanning subtree of a graph and the traveling salesman problem *Proc. of the American Mathematical Society* pp 48–50
[27] Hagmann P, Cammoun L, Gigandet X, Meuli R, Honey C J, Wedeen V J and Sporns O 2008 Mapping the structural core of human cerebral cortex *PLoS Biol.* **6** e159
[28] Rubinov M and Sporns O 2010 Complex network measures of brain connectivity: uses and interpretations *NeuroImage* **52** 1059–69
[29] Xia M, Wang J and He Y 2013 BrainNet Viewer: a network visualization tool for human brain connectomics *PLoS One* **8** e68910
[30] van Diessen E *et al* 2014 Opportunities and methodological challenges in EEG and MEG resting state functional brain network research *Clin. Neurophysiol.* **126** 1468–81
[31] Fornito A, Zalesky A and Breakspear M 2013 Graph analysis of the human connectome: promise, progress, and pitfalls *NeuroImage* **80** 426–44
[32] De Vico Fallani F, Richiardi J, Chavez M and Achard S 2014 Graph analysis of functional brain networks: practical issues in translational neuroscience *Phil. Trans. R. Soc. Lond. B. Biol. Sci.* **369** 20130521
[33] Muthukumaraswamy S D 2013 High-frequency brain activity and muscle artifacts in MEG/EEG: a review and recommendations *Front. Hum. Neurosci.* **7** 138