A Comparison of the Cluster-Span Threshold and the Union of Shortest Paths as objective thresholds of EEG functional connectivity networks from Beta activity in Alzheimer's disease

A Comparison of the Cluster-Span Threshold and the Union of Shortest Paths as objective thresholds of EEG functional connectivity networks from Beta activity in Alzheimer's disease

Keith Smith[1,2], *student member, IEEE,* Daniel Abasolo[3], *member, IEEE* & Javier Escudero[1], *member, IEEE*

***Abstract*— The Cluster-Span Threshold (CST) is a recently introduced unbiased threshold for functional connectivity networks. This binarisation technique offers a natural trade-off of sparsity and density of information by balancing the ratio of closed to open triples in the network topology. Here we present findings comparing it with the Union of Shortest Paths (USP), another recently proposed objective method. We analyse standard network metrics of binarised networks for sensitivity to clinical Alzheimer's disease in the Beta band of Electroencephalogram activity. We find that the CST outperforms the USP, as well as subjective thresholds based on fixing the network density, as a sensitive threshold for distinguishing differences in the functional connectivity between Alzheimer's disease patients and control. This study provides the first evidence of the usefulness of the CST for clinical research purposes.**

I. Introduction

Network science applies graph theory to real world problems in order to understand the interdependencies of complex systems [1]. It is widely used in the analysis of functional brain recordings where brain regions show statistical dependencies in a highly complex topological manner, bringing many different functions into one seamless experience for the subject [2][3][4].

In Alzheimer's disease (AD) it is held that the breakdown of these interactions, through the aggregation of toxic plaques, leads to noticeable functional differences which can be captured by the Electroencephalogram (EEG) using topological information of the network created from pairwise connectivity information of the EEG channels [5]. However, this process results in a complete weighted network where an edge between every two nodes in the network is weighted as the output of the corresponding connectivity measure, which is likely to contain many spurious low weights between largely independent channels [3].

To counter this problem, binarisation of the edges is proposed to offer an equalised form of the important information in the network whilst rejecting many of the spurious connections. This also allows for the use of binary metrics which are simpler to use than weighted metrics [6]. However, the lack of objective methods for binarisation results in arbitrary choices made in studies leading to different and sometimes conflicting results [7].

Novel solutions for network binarisation are thus forthcoming to solve this problem. Recently, Smith et al. [8] proposed the Cluster-Span Threshold (CST) based on selecting the threshold of the network that results in a balance of open to closed triples in the graph, ensuring a trade-off of sparsity and density of information. It was compared favourably with the Minimum Spanning Tree (MST), another binarisation solution which had previously been implemented successfully in a number of studies [9][10]. The MST results in a very sparse network, which gives robustness to fluctuations in the edge weights but it was noted that this also results in a poverty of connectivity information which appeared to be insensitive to more subtle cognitive differences [8]. Further, in general, the MST requires the rejection of some very strong edges because their inclusion would create cycles in the network and thus disqualify it as a tree.

The Union of Shortest Paths (USP) is another approach which overcomes this issue of edge rejection by focusing on the union of edges constituting the shortest paths between all pairs of nodes in the network [11]. It ensures connectedness (a path necessarily exists from every node to every other) and allows for higher connection density than the MST as it accepts cycles. Thus it initially seems a promising binarisation method.

In this study we compare the CST with the USP and proportional thresholds on clinical data of Alzheimer's disease patients against healthy control to contrast these techniques and assess their viability in the clinical setting.

II. Materials

A. Subjects

The EEG recordings used here were taken from 12 AD patients and 11 healthy control subjects. The patients -5 men and 7 women; age = 72.8 $\pm$ 8.0 years, mean $\pm$ standard deviation (SD)- were recruited from the Alzheimers Patients Relatives Association of Valladolid (AFAVA). They all fulfilled the criteria for probable AD. EEG activity was recorded at the University Hospital of Valladolid (Spain) after the patients had undergone clinical evaluation including clinical history, neurological and physical examinations, brain scans

*This work was partially supported by the Engineering and Physical Sciences Research Council.

[1]Keith Smith and Javier Escudero are with the Institute for Digital Communications, School of Engineering, University of Edinburgh, King's Buildings, West Mains Road, Edinburgh, UK, EH9 3FB. k.smith@ed.ac.uk, javier.escudero@ed.ac.uk

[2]Keith Smith is with the Alzheimer Scotland Dementia Research Centre, School of Philosophy, Psychology and Language Sciences, University of Edinburgh 7 George Square, Edinburgh, EH8 9JZ.

[3]Dr Daniel Abasolo is with the Centre for Biomedical Engineering, Department of Mechanical Engineering Sciences, Faculty of Engineering and Physical Sciences, University of Surrey, Guildford, GU2 7XH, UK.

and a Mini Mental State Examination (MMSE) to assess their cognitive ability [12].

The mean MMSE score for the AD group was 13.3 $\pm$ 5.6 points (mean $\pm$ SD), indicating moderate disease on average, however five of the patients had a score below 12 points, indicating severe dementia. Two of the patients were undergoing lorapezam treatment which may enhance beta activity with therapeutic doses, although no prominent rapid rhythms were observed in the visual inspection of their EEG recordings. No other patients were taking medication reported to influence EEG activity.

The 11 age-matched, elderly subjects who made up the control group- 7 men and 4 women; 72.8 $\pm$ 6.1 years $\pm$ SD- did not have any past or present mental disorder. They all scored 30 on the MMSE. The local ethics committee approved the study and control subjects and all caregivers of the patients gave their informed consent for participation [13].

B. EEG Recordings

The EEG recordings were made using Profile Study Room 2.3.411 EEG equipment (Oxford Instruments) at electrodes F3, F4, F7, F8, Fp1, Fp2, T3, T4, T5, T6, C3, C4, P3, P4, O1, O2, Fz, Cz and Pz, in accordance with the international 10-20 system. More than 5 minutes of data were recorded for each subject. Subjects were asked to remain in a relaxed awakened state with eyes closed while the recording process was taking place. Sampling was performed at 256 Hz, with 12 bit A-to-D precision.

Recordings were visually inspected by a specialist physician who selected epochs with minimal artefactual activity of 5s (1280 points) from the data for further analysis. The average number of these epochs per electrode per subject was 28.8$\pm$ 15.5 (mean $\pm$ SD).

The data were then written to ASCII files and the selected epochs were digitally filtered with a bandpass filter from 0.5-40 Hz before analysis [13].

III. Methods

A. Pre-Processing

We closely followed the procedure undergone in the introductory study of the CST [8]. All network pre-processing was performed using FieldTrip [14]. For each epoch, the 16 channels were first re-referenced to the grand average. The muli-taper method was applied using a Hanning window where the 5s recordings provided a 0.2 Hz frequency resolution. We then applied the debiased weighted phase-lag index [15] to the β (12.2-32 Hz) frequencies and averaged over frequency to obtain one connectivity matrix per trial for β. These connectivity matrices were further averaged over epochs [16] to obtain one matrix per person. Only one trial was available for one of the healthy control subjects and only 2 trials for another. These data sets are thus too small for reliable connectivity information and had to be disregarded for this study. In the end, we have a group of 12 AD patients and 9 healthy control subjects.

B. Network Binarisation

Before outlining the binarisation techniques we wish to compare, it is necessary to cover the graph theory on which their foundations are based. Let $G = (\mathcal{V}, \mathcal{E}, \mathbf{A})$ be an undirected graph, where $\mathcal{V}$ is the vertex set with $|\mathcal{V}| = n$, $\mathcal{E}$ is the edge set consisting of ordered pairs of elements from $\mathcal{V}$ with $|\mathcal{E}| = 2m$, and $\mathbf{A}$ is a weighted adjacency matrix such that the ith row, j column entry of $\mathbf{A}$, A_{ij}, is the weight of edge $(i, j) \in \mathcal{E}$. We then look for methods to transform the weighted graphs obtained from connectivity measures into corresponding simple graphs, which are unweighted, undirected and with no multiple edges or self-loops [1]. For an example of how the networks look like after implementing the binarisation techniques, see Fig.2.

1) CST: The CST computes an extensive list of proportional thresholds and chooses the one for which the proportion of open triples to closed triples is balanced. This balance occurs exactly where the global clustering coefficient, C_{glob}, obtains the value 0.5, i.e. when

$$C_{\text{glob}} = \sum_i \frac{(\mathbf{A}^3)_{ii}}{\sum_j T_{ij}} = 0.5,$$

where

$$\mathbf{T} = \mathbf{A}^2 - \mathbf{D}_{\mathbf{A}^2}$$

is the matrix of triples between nodes so that T_{ij} is the number of triples starting at node i and ending at node j. $\mathbf{D}_{\mathbf{A}^2}$ here is the matrix consisting of the diagonal entries of $\mathbf{A}^2$ on the diagonal and zeros elsewhere, thus $\mathbf{T}$ is just $\mathbf{A}^2$ with diagonal entries set to 0.

In application the ratio of open to closed triples is unlikely to balance exactly because graphs are discrete objects. Therefore, as in [8], we take the CST as the proportional threshold at which the closest value of C_{glob} is 0.5 for any given network.

2) USP: The union of shortest paths uses Dijkstra's algorithm [17] for finding the shortest path between two nodes and then takes the unweighted edges constituting those paths for all nodes as the network [11]. This guarantees connectedness of the network since all nodes must be connected to all others.

However, connectivity information is inverse to distance since, for connectivity, the stronger the connection is the higher is the weight, whereas with distance the stronger the connection (i.e. the shorter the distance) the smaller the weight. Before finding the USP one must first transform the weights accordingly [11]. Further, if the weights are too similiar, the shortest paths between two nodes will mostly consist of the edge adjacent to both nodes, which would result in an almost complete graph where the only relevant information of the network would be related to the weakest connections, which are already seen to be spurious. Therefore, to attempt to avoid these issues we take the negative log of the weights divided by a suitably chosen

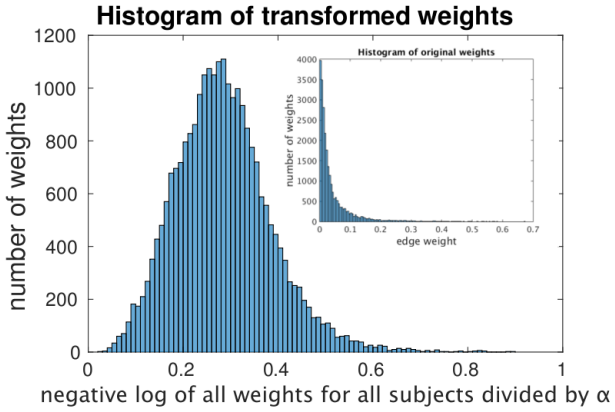

Fig. 1. Histogram of the transformed edge weights, $\hat{w}_{ij}$, for all subjects in the study. This shows a roughly normal distribution which has much wider spread than the power law distribution of the original weights (inset).

upper bound. That is, we use weights $\hat{w}_{ij}$ where

$$\hat{w}_{ij} = \begin{cases} 0, & i = j \\ \dfrac{-log(w_{ij})}{\alpha}, & i \neq j, \end{cases}$$

and compute the shortest paths between these transformed weights. To find α we simply computed the negative log of all weights and looked for the next integer number greater than the maximum value, which was 14 in this instance. A histogram of the transformed weights in this study is shown in Fig.1, which shows how the weights with a roughly power law distribution (inset Fig.1) have been transformed into a more spread, roughly normal distribution.

3) Proportional thresholds: To compare these binarisation methods with a widely utilised standard, we choose arbitrary proportional thresholds at 20%, 30% and 40%, similarly as in [8]. This simply keeps the given percentage number of strongest weights and uses those as the unweighted edges for the network.

C. Network Metrics

We computed the local clustering coefficient, C_{loc}, and the degree variance, V, for the CST, 20%, 30% and 40% threshold binarised networks and USPs.

C_{loc} is different to C_{glob} in that it computes the mean of the clustering coefficients centred at every node [1]. What this means is that, for a node, i, one computes the percentage of triples centred at i which are triangles and averages this value over all nodes. This is different from just the percentage of triples which are triangles as in C_{glob}, thus one can use this as a measure for CST networks and, indeed, this difference could extract interesting differences in the underlying data. C_{loc} is one of the most widely used network measures and has been utilised in many AD studies [7].

The degree variance, V, is the variance of the degrees of the graph, where the degree of node i is the number of adjacent connections to i. This measures the topological heterogeneity of the degrees [18] and is thus relevant to the scale-freeness of network topology [19].

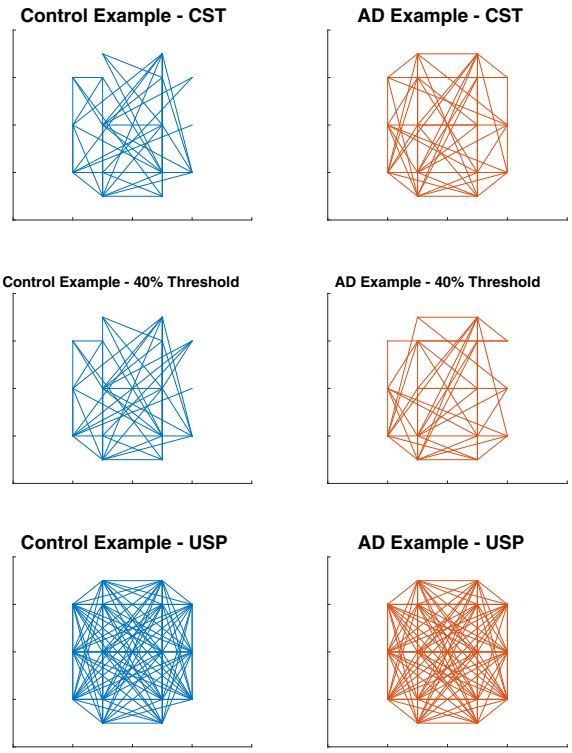

Fig. 2. Basic network representation of EEG activity of one healthy control subject, left, and one Alzheimer's disease patient, right, for Cluster-Span Threshold (CST), 40 % proportional threshold and Union of Shortest Paths (USP). Nodes correspond to a top view of the rough layout of electrodes on the 10-20 system- top to bottom is front to back of the head.

We computed these measures using the Brain Connectivity Toolbox [20] and simple computations in MATLAB.

IV. Results and Discussion

The average connection density for the CST networks in this analysis was $45.63 \pm 5.23\%$ (mean $\pm$ standard deviation). This mean breaks down into $47.50 \pm 3.48\%$ for Alzheimer's patients and $43.15 \pm 6.29\%$ for healthy control. This indicates that the Alzheimer's networks appear to be more random than healthy control networks since random networks have connection density $= C_{\text{glob}} = 0.5$ [1]. This is in agreement with the findings in [6]. For the USP networks the average connection density was $94.40 \pm 4.35\%$. This is still very high. However, it must be stressed that methods to counter this in the weight transformation should not involve variable parameters as this would introduce a similar arbitrary parameter that the USP is introduced to avoid. This breaks down into $94.58 \pm 4.37\%$ for Alzheimer's patients and $94.17 \pm 4.56\%$ for healthy control.

Results of student t-tests for AD vs. Control for the network metrics of the different binarisation approaches are shown in Table I. The box plots of these results are found in Fig.3. We see that the CST shows significance with C_{loc} alongside the 40% threshold at the standard $p < 0.05$ level. No other significant differences are found, particularly the USP appears far the least sensitive for finding differences in topology between AD and control. However, the degree

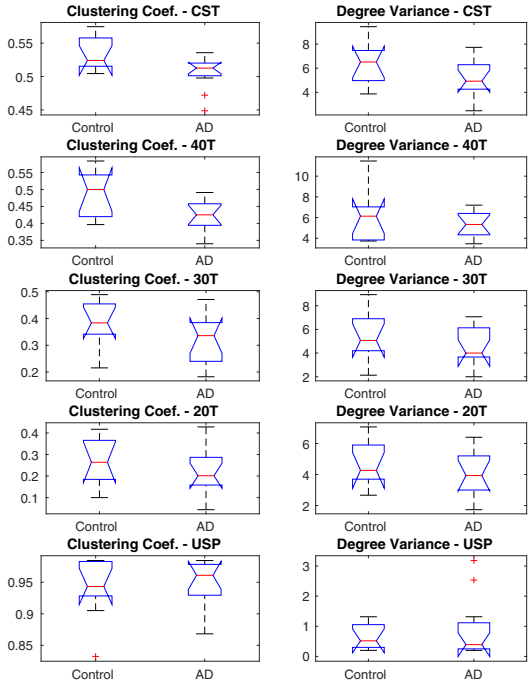

Fig. 3. Box plots of metric values: AD vs. healthy Control for local clustering coefficient and degree variance as indicated. The arbitrary proportional thresholds are denoted by '#T' for the #% threshold.

variance for the CST does appear to exhibit some sensitivity ($p = 0.0899$) in contrast with V for the other approaches.

For the USP, the connection density is too large as it stands. Particularly, having such large densities means that the topology being studied by the USP relates to the weakest connections in the network rather than the strongest ones. Thus, though based on a theoretically sound idea, we advise that it is not useful for functional connectivity where the weights of the edges tend to have similar magnitudes.

The box plots show that local network clustering appears to be lower for AD than for controls at higher densities of network binarisation of EEG β activity. This may correspond to less organised brain connectivity.

TABLE I

THE p-VALUES FROM TOPOLOGICAL METRICS FOR t-TESTS OF AD PATIENTS VS. HEALTHY CONTROL.

Metric	CST	40%	30%	20%	USP
C	0.0260	0.0282	0.1519	0.2892	0.7203
V	0.0899	0.3052	0.4175	0.2649	0.5388

V. CONCLUSIONS

We presented a comparison of novel network binarisation techniques to negate the problem of biased subjective thresholding in brain network research on clinical EEG data of Alzheimer's patients vs healthy control. We found further evidence that the CST provides a sensitive framework for functional brain network analysis, this time providing clinical evidence that resting-state functional connectivity of beta in Alzheimer's diseased brains is more random than in healthy brain activity. The USP was found to not be as appropriate in this setting from both the theory and in application. In the future we aim to provide a comprehensive evaluation of the CST using both simulated and real brain networks. We will provide an extensive comparison with the USP, MST, arbitrary thresholds (including higher density thresholds) and also with weighted metrics to ascertain in which scenarios each of these may be more powerful and appropriate.

REFERENCES

[1] M.E.J. Newman, "Networks", *Oxford University Press*, Oxford, 2010.

[2] E. Bullmore, O. Sporns, "Complex brain networks: graph theoretical analysis of structural and functional systems", *Nature Neuroscience*, 10:186-198, 2009.

[3] F. Fallani, J. Richiardi, M. Chavez, S. Achard, "Graph analysis of functional brain networks: practical issues in translational neuroscience", *Phil Trans. R Soc. B*, 369(1653): 20130521, 2014.

[4] D. Papo, M. Zanin, J.A. Pineda-Pardo, S. Boccaletti, J.M. Buldú, "Functional brain networks: great expectations, hard times, and the big leap forward", *Phil. Trans. R Soc. B*, 369(1653): 20130525, 2014.

[5] C.J. Stam, B.F. Jones, G. Nolte, M. Breakspear, P. Scheltens, "Small World Networks and Functional Connectivity in Alzheimers Disease", *Cerebral Cortex*, 17:92-99, doi:10.1093/cercor/bhj127, 2007.

[6] C.J. Stam, "Modern network science of neurological disorders", *Nature Reviews Neuroscience*, 15: 683695, doi:10.1038/nrn3801, 2014.

[7] B. Tijms et al., "Alzheimer's disease: connecting findings from graph theoretical studies of brain networks ", *Neurobiology of Aging*, 34: 2023-2036, 2013.

[8] K. Smith, H. Azami, M.A. Parra, J.M. Starr & J. Escudero, " Cluster-Span Threshold: An unbiased threshold for binarising weighted complete networks in functional connectivity analysis", *IEEE Engineering in Medicine and Biology Society Conference*, 2840-2843, 2015.

[9] P. Tewarie et al., "The minimum spanning tree: an unbiased method for brain network analysis ", *Neuroimage*, 104: 177-188, 2015.

[10] C.J. Stam, P.Tewarie, E. Van Dellen, E.C.W.vanStraaten, A. Hillebrand, P. Van Mieghem, "The Trees and the Forest: Characterization of complex brain networks with minimum spanning trees", *International Journal of Psychophysiology*, 92: 129-138,2014.

[11] J. Meier, P. Tewarie & P. Van Mieghem, "The Union of Shortest Path Trees of Functional Brain Networks", *Brain Connectivity*, 5(9): 575-581. doi:10.1089/brain.2014.0330, Nov. 2015.

[12] M.F. Folstein, S.E. Folstein and P.R. McHugh, "Mini-mental state: a practical method for grading the cognitive state of patients for the clinician", *J. Psychiatr. Res.*, 12: 189198, 1975.

[13] J. Escudero, D. Abásolo, R. Hornero, P. Espino & M. López, "Analysis of electroencephalograms in Alzheimer's disease patients with multiscale entropy", *Physiological Measurement*, 27(11): 1091-1106, Sept. 2006.

[14] R. Oostenveld, P. Fries, E. Maris and J-M. Schoffelen, "FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data", *Computational Intelligence and Neuroscience*, Volume 2011, 156869, 9 pages, 2011.

[15] M. Vinck, R. Oostenveld, M. van Wingerden, F. Battaglia, C.M.A. Pennartz, "An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias", *NeuroImage*, 55:1548-1565, 2011.

[16] E. van Diessen, T. Numan, E. van Dellen, A.W. van der Kooi, M. Boersma, D. Hofman, R. van Lutterveld, B.W. van Dijk, E.C.W. van Straaten, A. Hillebrand, C.J. Stam, "Opportunities and methodological challenges in EEG and MEG resting state functional brain network research", *Clinical Neurophysiology*, doi:10.1016/j.clinph.2014.11.018, 2014.

[17] E.W. Dijkstra, "A note on two problems in connexion with graphs", *Numerische Mathematik*, 1(1):269-271, 1959.

[18] T.A.B. Snijders, "The degree variance: an index of graph heterogeneity", *Social Networks*, 3(3): 163-174, 1981.

[19] V.M. Eguluz, D.R. Chialvo, G.A. Cecchi, M. Baliki, A.V. Apkarian, "Scale-Free Brain Functional Networks", *Phys. Rev. Lett*, 9: 018102, Jan. 2005.

[20] M. Rubinov, O. Sporns, "Complex network measures of brain connectivity: Uses and interpretations", *NeuroImage*, 52:1059-1069, 2010.