

PAPER • OPEN ACCESS

The effect of density thresholding on the EEG network construction

To cite this article: Ilya Zakharov *et al* 2021 *J. Phys.: Conf. Ser.* **1727** 012009

View the [article online](#) for updates and enhancements.

You may also like

- [A comparison of methods for fully automatic segmentation of tumors and involved nodes in PET/CT of head and neck cancers](#)
Aurora Rosvoll Groendahl, Ingerid Skjei Knudtsen, Bao Ngoc Huynh et al.
- [Optimization of wavelet- and curvelet-based denoising algorithms by multivariate SURE and GCV](#)
R Mortezaejad and A Gholami
- [Systematic analysis of wavelet denoising methods for neural signal processing](#)
Giulia Baldazzi, Giuliana Solinas, Jaume Del Valle et al.



ECS
The
Electrochemical
Society
Advancing solid state &
electrochemical science & technology

DISCOVER
how sustainability
intersects with
electrochemistry & solid
state science research

The effect of density thresholding on the EEG network construction

Ilya Zakharov^{1,2}, Timofey Adamovich^{1,2}, Anna Tabueva^{1,2}, Victoria Ismatullina^{1,2} and Sergey Malykh^{1,2}

¹ Psychological Institute of Russian Academy of Education, Behavioural Developmental Genetics Lab, 125009, Moscow, Russia

² Russian Academy of Education, Interdisciplinary Centre in Educational Sciences, 119121, Moscow, Russia

e-mail: iliazaharov@gmail.com

Abstract. The procedure of thresholding for graph construction is one of the common steps in calculating networks of brain connections. However, this procedure can lead to incomparable results from different studies. In the present study we aim to test the effect of thresholding or algorithmic reduction of the number of connected nodes on the construction of a set of widely used connectivity graph metrics derived from EEG data. 164 people took part in our study. Participants were recruited via social networks. EEG was recorded during resting state. At the beginning of the procedure each participant was asked to relax and not to think about anything. Source reconstruction was performed using standard source localization pipeline from MNE-package. Desikan-Killiany Atlas was used for cortical parcellation with 34 ROI per hemisphere. Synchronization was estimated with weighted phase lag index in 4–30 Hz frequency range for eyes closed and eyes open separately. We have found that All metrics except average participation coefficient vary monotonously as a function of density level (moreover, we have found, that for Cluster Coefficient, more than 95% and for the Characteristic Path Length ~50% of the variance is related to thresholding cut-off). The different data-driven approaches to the network construction leads to significant changes in the group-level graph metrics and can eliminate the variance in the data that can be crucial for individual differences studies.

1. Introduction

The human brain is an interconnected network of communicating regions. One way to understand this network comes from mathematical graph-theory which can be used to describe complex structures and dynamics within networks [1]. Recently, this approach has been used to study the neurobiological correlates of psychopathology [2], cognitive processes and personality characteristics (e.g. temperament or creativity; [3, 4]). It has been shown that graph-theory derived brain network characteristics are stable and unique for an individual [5].

Network neuroscience is new and developing area, and one of its challenges is considerable variability in the connectivity calculation routines, which is not always detailed and explicitly motivated in studies. When constructing the graph of functional connections within the brain one starts from creating the adjacency matrix of the given relationship between any pair of sources of the recorded brain activity. Technically, the functional connectivity graph is a complete graph, because in



a real data it is virtually impossible to get a zero correlation between a pair of signals. However, both from computational and anatomical perspectives complete graphs within the brain don't make sense. To go from complete graph to more meaningful incomplete graph the thresholding procedure is used. The thresholding is basically choosing a value according to which the connection would be regarded as present (if it is higher than this value) or absent (if it is low).

This procedure is supposed to reduce spurious connections [6]. However, while thresholding is a good way to reduce the possibility of too much influence of spurious connections, it leads to confounds related to arbitrary choice of the calculation routines. According to the recent International Federation of Clinical Neurophysiology (IFCN) report, for M/EEG studies "there is no clear consensus on the statistical thresholds for the confirmation of a significant effect at the scalp or source level" [7]. The effect of the choosing different thresholding procedures for the calculation of the EEG-derived connectivity graph metrics has never been directly assessed. In the present study we aim to test the effect of thresholding or algorithmical reduction of the number of connected nodes on the construction of a set of widely used connectivity graph metrics derived from EEG data.

2. Materials and Methods

2.1. Participants

164 people took part in our study. Participants were recruited via social networks without any monetary incentives. The exclusion criteria for participating was any report of psychiatric or neurological disorder or head trauma. Participants age ranged from 17 to 34 ($M=21.7$, $SD=3.36$) with 30% female.

2.2. Experimental procedure and data acquisition

EEG was recorded during resting state. At the beginning of the procedure each participant was asked to relax and not to think about anything. The recording started with closed eyes, every 2 minutes experimenter asked to change the state ("Close your eyes, please" and "Open your eyes, please"). In total 10 minutes was recorded, with 3 eyes closed intervals and 2 eyes open intervals.

EEG was recorded with 64-channel Brain Products ActiChamp system with 500 Hz sampling frequency with Cz as the reference electrode. EEG preprocessing steps included downsampling to 256 Hz, re-reference to common reference, filtering from 0.1 Hz to 30 Hz, manual removal of artifacts and noisy channels, except blinks and eyes movement artifacts for which Independent Component Analysis (ICA) was used. Topographic Interpolation was performed on excluded noisy channels if needed.

2.3. Source Reconstruction and Synchronization Measures

Source reconstruction was performed using standard source localization pipeline from MNE-package. First, source space with 1026 sources per hemisphere was created. Second, BEM (boundary-element model) method was used to create three-layer model of the hemispheres. Three layers: inner skull, outer skull and outer skin. The conductivity of layers was standard for MNE package (0.3, 0.006, 0.3 for three layers accordingly). MNE exploit anatomical information from FreeSurfer standard model head (Colin27). Third, the forward operator was constructed basin on the source space and BEM model. Fourth, the individual inverse operator was created for every participant with individual noise covariance matrix. Source reconstruction of each individual was performed with the appropriate inverse operator using dSPM method. Desikan-Killiany Atlas [8] was used for cortical parcellation with 34 ROI per hemisphere. Synchronization was estimated with weighted phase lag index (wPLI, [9]) in 4-30 Hz frequency range for eyes closed and eyes open separately.

2.4. Graph analysis

2.4.1. Proportional approach to thresholding. On a basis of each connectivity matrix, we created a weighted graph for which relative threshold were applied. Threshold was calculated as quantile of synchronization strength distribution for every participant individually with values below threshold set to 0, representing lack of connection. Threshold values started from 0.1 (90% of the connections included) to 0.9 (10% of the strongest connections included) with 0.1 step.

2.4.2. Data-driven approach to thresholding. An important problem of the thresholding procedure is that in a number of cases it leads to constructing the disconnected sub-graphs (graphs, that are not connected by at least one edge). Mathematically, disconnected graphs cannot be described by any single metric. In the present study we developed new thresholding approach, based on choosing the graph density level so that there were no disconnected nodes within the graph.

Another way to construct the graph is based on the minimum spanning tree (MST) algorithms. In the present study we have chosen two version of this algorithm. The classic MST [10, 11] and the orthogonal MST developed by G. Dimitriadis [12]. We also add intra-individual means of the synchronization as the thresholds to compare it with the data-driven graphs. We also add intra-individual means of the synchronization as the thresholds to compare it with the data-driven graphs.

The following measures were chosen to describe the structure of a graph:

- Characteristic path length
- Clustering coefficient
- Small-world Index
- Participation coefficient

3. Results

3.1. Proportional approach to thresholding problem

The figures 1 demonstrate changes in the mean and standard deviations for 4 graph-based connectivity metrics separately for wPLI and ImCoh as a synchronisation measure and for eyes open/closed conditions. Thresholds are presented from left to right in order of descending density, from full graph (density 1) to 90th quantile (density 0.1), followed by the mean density as the threshold and for the minimally connected graph (graph with minimum possible density without disconnected nodes).

At the next step of our analysis we wanted to estimate the proportion of the variance in the calculation of graph metrics as the result of the thresholding procedure. Linear regression analysis was performed in order to investigate the dependence of measures from the thresholding value. At this step the measures were recalculated: we used 1%-step instead of 10% step to have a smoother scale.

Table 1. Linear regression coefficients and adjusted R Squared value for all measures.

	Intercept	Density value (Coef)	Adj.R.Squared
CharacteristicPL	0.929989	-0.82389	0.499992
ClustCoef	0.192732	0.800123	0.976311
Participation_coef	0.565608	0.083153	0.065674

*Calculation of SWI with step 1% is computationally too heavy, and 10 points are insufficient for the linear regression, so, SWI was removed on this step.

The regression analysis showed that for CPL and Cluster Coefficient metrics the density-based thresholding explained a large amount of the overall variance in the data. The participation coefficient was shown to be much more stable across various thresholding values.

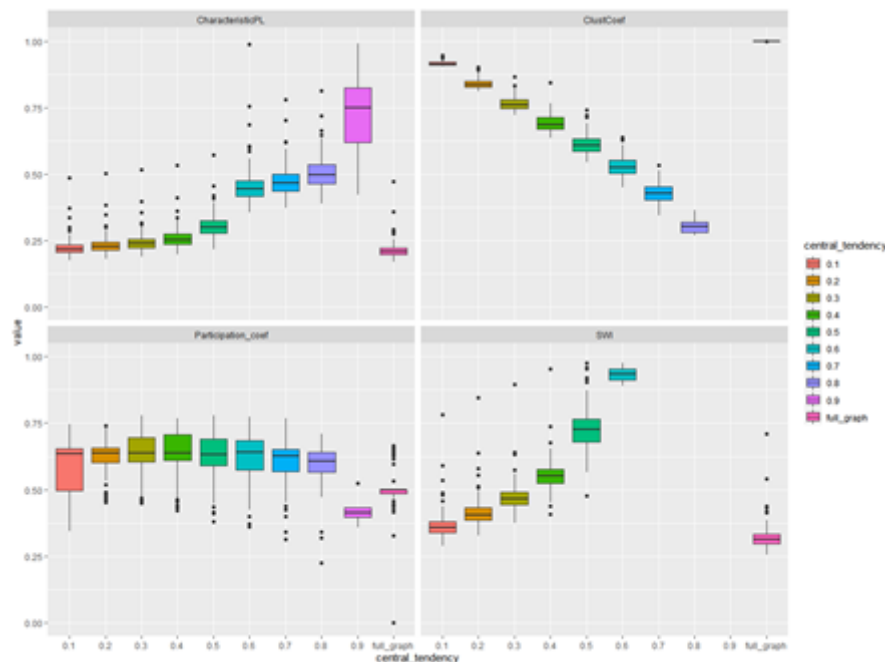


Figure 1. The effect of proportional thresholding on the EEG graph metrics

3.2. Data-driven approach to thresholding problem

To avoid the subjectivity in the choosing the threshold, the data-driven approach to spurious nodes elimination has been proposed. In the present study we compared graph metrics based on the networks with two subjective thresholds with intra-individual mean and median density as the cut-offs, the classic MST-based network, and the orthogonal MST-based network. We have also

The next step of our analysis was comparing the different data-driven approaches to graph construction. The results of the comparison can be seen in Fig. 2

The one-way ANOVA model showed that the thresholding factor was significant for all the graph metrics. It can be also seen, that choosing the classic MST algorithm for network construction led to elimination of any individual differences in our sample.

4. Discussion

In the present study we analyzed the dependence of the graph functional connectivity metrics on the widely-used density-based thresholding procedure. All metrics except average participation coefficient vary monotonously as a function of density level (moreover, we have found, that for Cluster Coefficient, more than 95% and for the Characteristic Path Length ~50% of the variance is related to thresholding cut-off).

The different data-driven approaches to the network construction leads to significant changes in the group-level graph metrics and can eliminate the variance in the data that can be crucial for individual differences studies. Overall, according to our results the way the researcher solve the thresholding issue heavily affects the outcome of the connectivity analysis. Our results add to the previous results by Garrison and colleagues [13] with fMRI and Civier and colleagues [14] with DTI, showing that in EEG studies the thresholding should be used with considerable cautious.

Our results are in line with the results of the broader study regarding the analytic choices in the neuroimaging studies in general [15]. The combined results from 70 independent research teams have

showed that an existing level of analytical flexibility in modern neuroscience can lead to substantial variability of the scientific conclusions. So, what can be done, if there is no strong arguments that can be made for the best choice of the analysis routine?

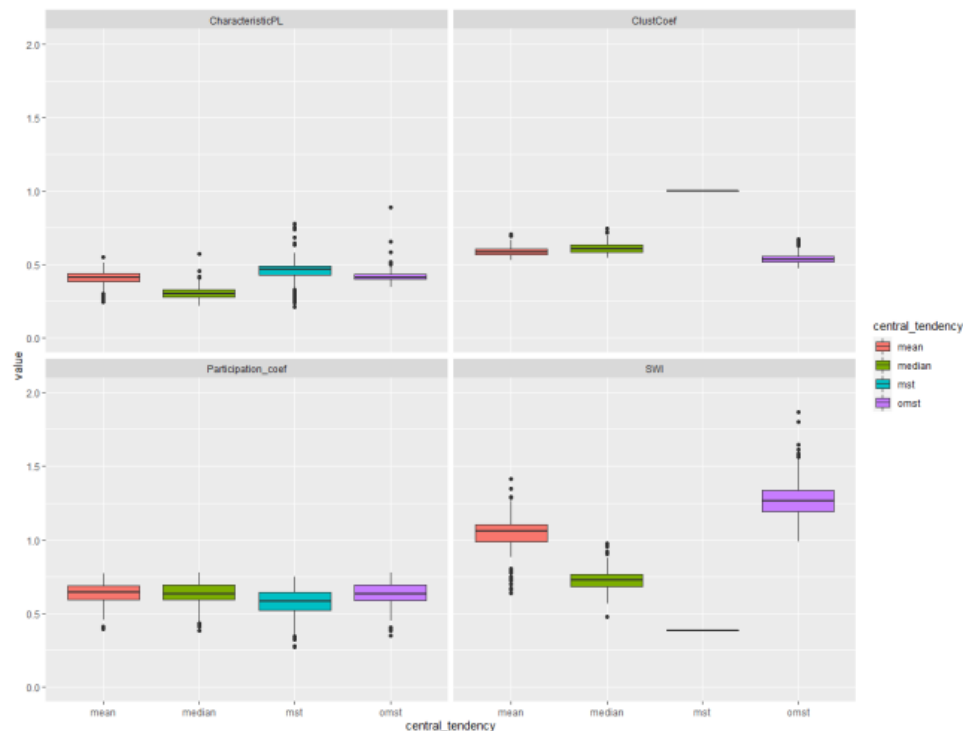


Figure 2. The comparison of proportional and data driven thresholding on the EEG graph metrics

First, sharing the raw or preprocessed data should be incentivized as the common practice between researchers. The individual researchers and the scientific community benefits of the open science approach to sharing data have been discussed extensively [16]. Second, the degrees of freedom in the analysis can be minimized with the adoption of pre-registration of the new studies or the re-analysis of the previously recorded data [17].

Finally, the multiverse approach to the analysis of the complex datasets can be accepted [18]. The multiverse approach suggests that multiple analytic pipelines can be applied to the datasets and the total conclusion about the effects of interest can be settled only after accounting for the differences in the results of the different analytic choices. One way to increase the reproducibility and the integrity of this analysis can be achieved by blinding the analysts to the hypothesis of interest. In general, deflating the multiverse involves developing a better and more complete theorizing of the constructs of interest and improving their measurement.

5. Conclusion

To sum up, in the present study we have shown that in agreement to the broader problem of the analytic choices in neuroimaging studies, the analysis of the brain connectivity can be highly dependent on choosing the thresholds for connectivity matrices construction with no best way to choose between different analytic options. Potential approaches that could be used to mitigate issues related to analytical variability involves new practices of data sharing, preregistration and multiverse approach to analysis.

References

- [1] Avena-Koenigsberger A, Misić B and Sporns O 2017 Communication dynamics in complex brain networks *Nature Reviews. Neuroscience* **19** (1) 17–33 <https://doi.org/10.1038/nrn.2017.149>
- [2] van den Heuvel M P and Sporns O 2019 A cross-disorder connectome landscape of brain dysconnectivity *Nature Reviews Neuroscience* **20** (7) 435–46 <https://doi.org/10.1038/s41583-019-0177-6>
- [3] Markett S, Weber B, Voigt G, Montag C, Felten A, Elger C and Reuter M 2013 Intrinsic connectivity networks and personality: The temperament dimension harm avoidance moderates functional connectivity in the resting brain *Neuroscience* **240** 98–105 <https://doi.org/10.1016/j.neuroscience.2013.02.056>
- [4] Beaty R E, Kenett Y N, Christensen A P et al 2018 Robust prediction of individual creative ability from brain functional connectivity *Proceedings of the National Academy of Sciences* **115** (5) 1087–92 <https://doi.org/10.1073/pnas.1713532115>
- [5] 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 *Nature Neuroscience* **18** (11) 1664–71 <https://doi.org/10.1038/nn.4135>
- [6] Rubinov M and Sporns O 2010 Complex network measures of brain connectivity: Uses and interpretations *NeuroImage* **52** (3) 1059–69 <https://doi.org/10.1016/j.neuroimage.2009.10.003>
- [7] Babiloni C, Barry R J, Başar E, Blinowska K J et al 2020 International Federation of Clinical Neurophysiology (IFCN) – EEG research workgroup: Recommendations on frequency and topographic analysis of resting state EEG rhythms. Part 1: Applications in clinical research studies *Clinical Neurophysiology* **131** (1) 285–307 <https://doi.org/10.1016/j.clinph.2019.06.234>
- [8] Desikan R S, Ségonne F, Fischl B et al 2006 An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest *NeuroImage* **31** (3) 968–80 <https://doi.org/10.1016/j.neuroimage.2006.01.021>
- [9] Vinck M, Oostenveld R, van Wingerden M, Battaglia F and Pennartz C M A 2011 An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias *NeuroImage* **55** (4) 1548–65 <https://doi.org/10.1016/j.neuroimage.2011.01.055>
- [10] Meier J, Tewarie P and Van Mieghem P 2015 The Union of Shortest Path Trees of Functional Brain Networks *Brain Connectivity* **5** (9) 575–81 <https://doi.org/10.1089/brain.2014.0330>
- [11] 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 <https://doi.org/10.1016/j.neuroimage.2014.10.015>
- [12] Dimitriadis S I, Antonakakis M, Simos P, Fletcher J M and Papanicolaou A C 2017 Data-Driven Topological Filtering Based on Orthogonal Minimal Spanning Trees: Application to Multigroup Magnetoencephalography Resting-State Connectivity *Brain Connectivity* **7** (10) 661–70 <https://doi.org/10.1089/brain.2017.0512>
- [13] Garrison K A, Scheinost D, Finn E S, Shen X and Constable R T 2015 The (in)stability of functional brain network measures across thresholds *NeuroImage* **118** 651–61 <https://doi.org/10.1016/j.neuroimage.2015.05.046>
- [14] Civier O, Smith R E, Yeh C-H, Connelly A and Calamante F 2019 Is removal of weak connections necessary for graph-theoretical analysis of dense weighted structural connectomes from diffusion MRI? *NeuroImage* **194** 68–81 <https://doi.org/10.1016/j.neuroimage.2019.02.039>
- [15] Botvinik-Nezer R, Holzmeister F, Camerer C F et al 2020 Variability in the analysis of a single neuroimaging dataset by many teams *Nature* **582** (7810) 84–88 <https://doi.org/10.1038/s41586-020-2314-9>

- [16] Munafò M R, Nosek B A, Bishop D V M et al 2017 A manifesto for reproducible science *Nature Human Behaviour* **1** (0021) <https://doi.org/10.1038/s41562-016-0021>
- [17] Lakens D 2019 *The Value of Preregistration for Psychological Science: A Conceptual Analysis* <https://doi.org/10.31234/osf.io/jbh4w>
- [18] Steegen S, Tuerlinckx F, Gelman A and Vanpaemel W 2016 Increasing Transparency Through a Multiverse Analysis *Perspectives on Psychological Science* **11** (5) 702–12 <https://doi.org/10.1177/1745691616658637>