Bioinformatics 2017-2018
Date: December 29, 2018
Project report

Bioinformatics@Data Science A.Y. 2017-2018

Brain Oscillatory and Network Activity during resting states

Cecilia Martinez Oliva, Vittoria Sessa

Group no. 6

Abstract

In this project we did an analysis of two signals obtained through an Electroencephalography (EEG), one obtained when the subject was in a rest state with her/his eyes open, the other when the subject was in a rest state with her/his eyes closed. The analysis included a spectrum estimation, the creation of several graphs using different metrics and their study, a motifs analysis and a community detection.

Introduction

We analyzed two runs of EEG data (S059R01 and S059R02) taken from PhysioNet's 'EEG Motor Movement/Imagery dataset'[1]. The whole dataset contained data from 109 volunteers of which we analyzed one subject (the fifty-ninth's). The different runs correspond to the subjects performing different motor and imagery tasks, we analyzed the experiments in which the subject had his eyes open (R01) and his eyes closed (R02) while in a resting state. For each run we have 64 channels, each corresponding to an electrode placed on a different area of the scalp as per the international 10-10 system (excluding electrodes Nz, F9, F10, FT9, FT10, A1, A2, TP9, TP10, P9, and P10). During the experiment 160 samples per second were taken for one minute (61 seconds), thus we have a total of 9760 samples. The programming language we used for the analysis is python. The data is provided in the EDF file format, to read this files we used the pyEDFlib library [5] available for python. The analysis comprised the following topics:

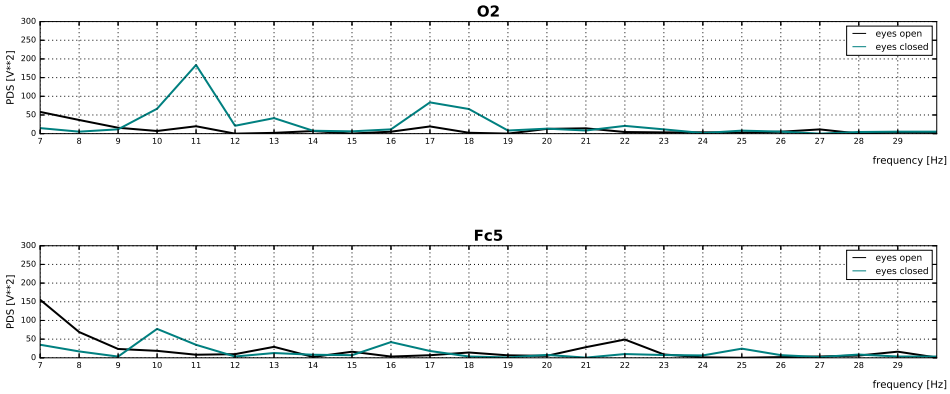
1. spectrum estimation
2. connectivity graphs
3. graph theory indices
4. motif analysis
5. community detection

For the motifs analysis we used the mfinder1.21 software provided by the Uri Alon lab of the Weizmann Institute of Science [12] while the links to python libraries used for the other parts of this project can be found in the references.

1 Spectral Analysis

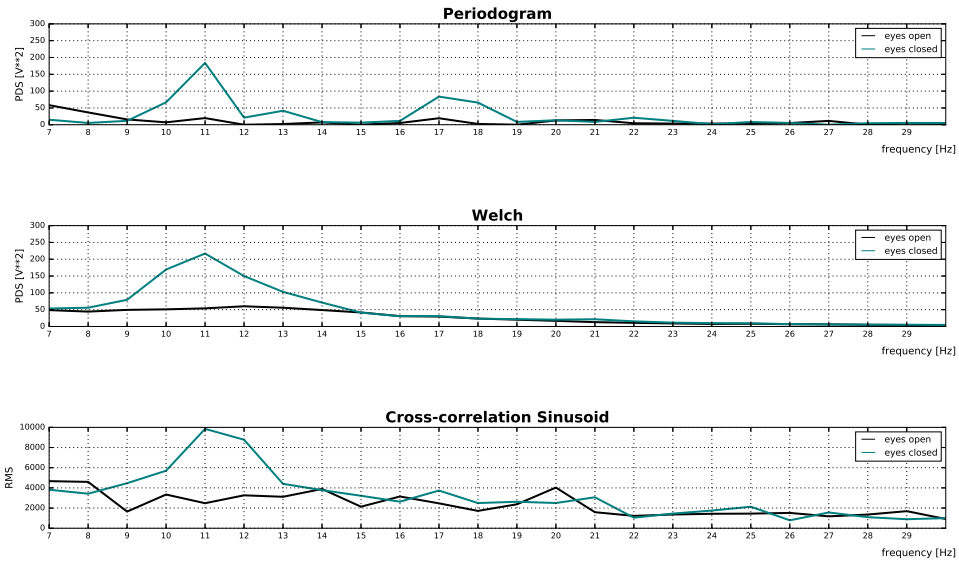
We selected all channels corresponding to the occipital lobe since this is the area of the brain in which the visual cortex is situated and some of the channels of the parietal lobe. The channels are the following: Po7 Po3 Poz Po4 Po8 O1 Oz O2 .
Figure 1 shows the power spectral density (PSD), estimated through the periodogram method, the green line represents the baseline eyes closed experimental run, while the black one represents the baseline eyes open experimental run, the range of the horizontal axis is from 7 to 30Hz since this is the frequency interval of interest (alpha and beta waves). As expected the difference of the alpha waves between eyes open and eyes closed is more visible for the O2 (occipital lobe) channel than for the Fc5 (frontal lobe) channel.

Figure 1: Spectral comparison of channels O2 and Fc5



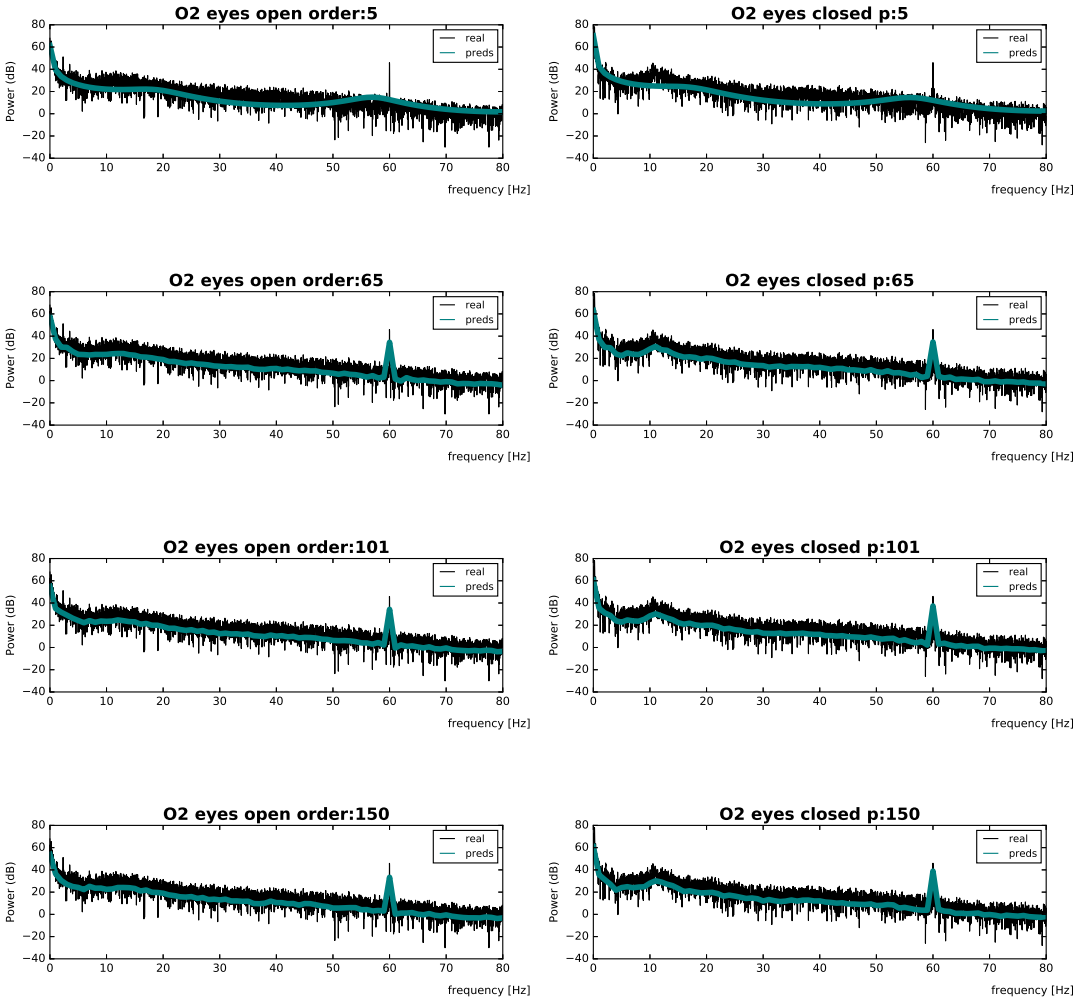
We estimated the power spectral density (PSD) using several methods: the non-parametric methods we used are the periodogram, the Welch method and the cross-correlation sinusoids method, while the parametric method we used is the Yule-Walker method. A comparison of the results obtained using the different non-parametric methods for channel O2 is in figure 2 . To reduce the noise we chose the hamming window . A graphical representation of the spectrum for all selected channels and methods used can be found in the directory bioinf_proj_neuro_group-06.zip (1-correlation.png,1-periodogram.png, 1-welch.png) .

Figure 2: Spectral comparison using different methods



As expected in figure 2 a peak can be seen around the 11Hz frequency (alpha frequencies range) for the eyes closed run. For the cross-correlation sinusoids method a second smaller peak can be seen around 20Hz (beta frequencies range) for the eyes open run, we could not find this second peak with the other methods.
The order parameter p used for the Yule-Walker method was estimated using the Akaike information criterion, the minimum is reached for p = 101 (1-yluearAIC.png). We also tried different values of the parameter (p = 5, 65,101,150) as can be seen in figure 3. It's possible to see that the best results are obtained for higher parameter values and that for p = 5 the spectrum is flat.

Figure 3: Yule Walker for different orders of the parameters



The python library used for all non-parametric methods is SciPy [6] while for Yule-Walker we used Spectrum [7].

2 Connectivity Graph

We estimated the functional brain connectivity among 64 channels using Partial Directed Coherence (PDC) and Direct Transfer Function (DTF). We selected the 11Hz frequency where a peak is visible in the alpha waves range as seen before. We first estimated the best value of the order for the

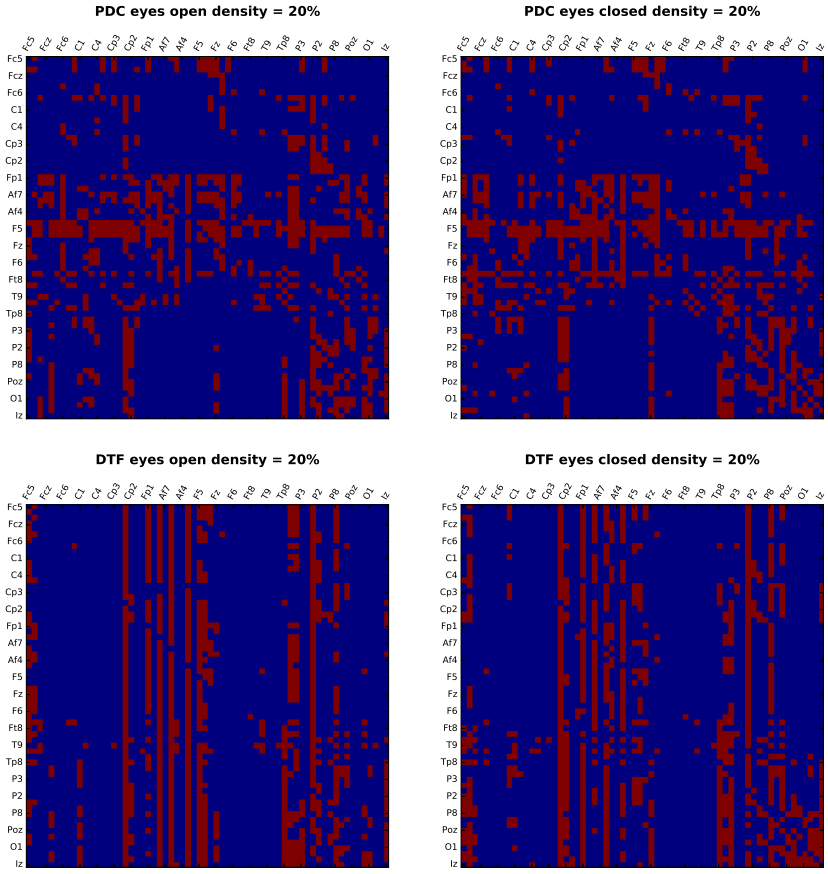
multivariate autoregressive (AR) model using the Akaike information criterion (AIC). The model was created using all 64 channels. The optimum value of the order is 5 for both cases, eyes open and closed. Using the coefficients estimated through the models we computed the PDC (eq. 1) and DTF (eq. 1).

$$\pi_{ij}(f) = \frac{|A_{ij}(f)|^2}{\sum_{m=1}^{L} |A_{mi}(f)|^2} \tag{1}$$

$$\theta_{ij}(f) = \frac{|H_{ij}(f)|^2}{\sum_{m=1}^{L} |H_{mi}(f)|^2} \tag{2}$$

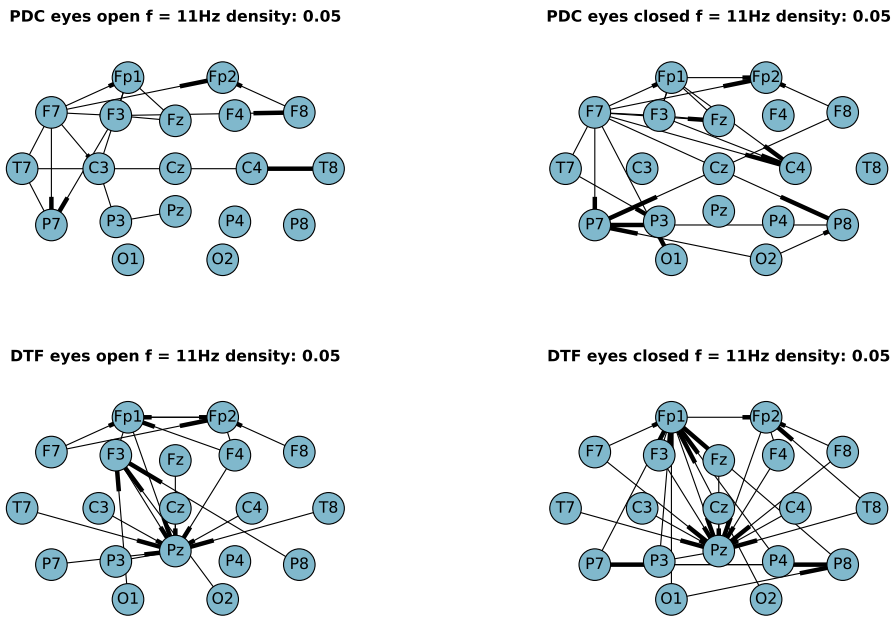
$A_{ij}(f)$ are the coefficients of the AR model and $H_{ij}(f)$ are the elements of the $H(z) = \frac{1}{A(z)}$ matrix. The PDC recontructs the direct interactions between signals while the DTF reconstructs also the indirect ones. We used the connectivipy library [8].
We then created an adjacency matrix with network density equal to 20% by applying a threshold. This matrices can be found in the files Ao_pdc.txt , Ao_dtf.txt (eyes open), Ac_pdc.txt and Ac_dtf.txt (eyes closed) contained in the directory. A graphical representation for the matrices of the pdc and the dtf is in the figures 4.

Figure 4: Graphical representation for the matrices of the pdc and dtf



We also used different thresholds to the obtain adjacency matrices with the following density values: 1%, 5%, 10%, 20%, 30%, 50% , the graphical representations are in the directory
(with the following names 2.3-PDC(eyes_open)_freqs.png, 2.3-PDC(eyes_closed)_freqs.png, 2.3-DTF(eyes_open)_freqs.png, 2.3-DTF(eyes_closed)_freqs.png).
A topographical representation of the networks with 19 channels and density $\leq 5\%$ is in figure 5.

Figure 5: Topographical representation of the networks with 19 channels and density $\leq 5\%$



The same analysis was done with the 20Hz beta rhythm frequency (2.6-20HZ19ch.png, 2.6-PDC_DTF20HZ.png in the directory).

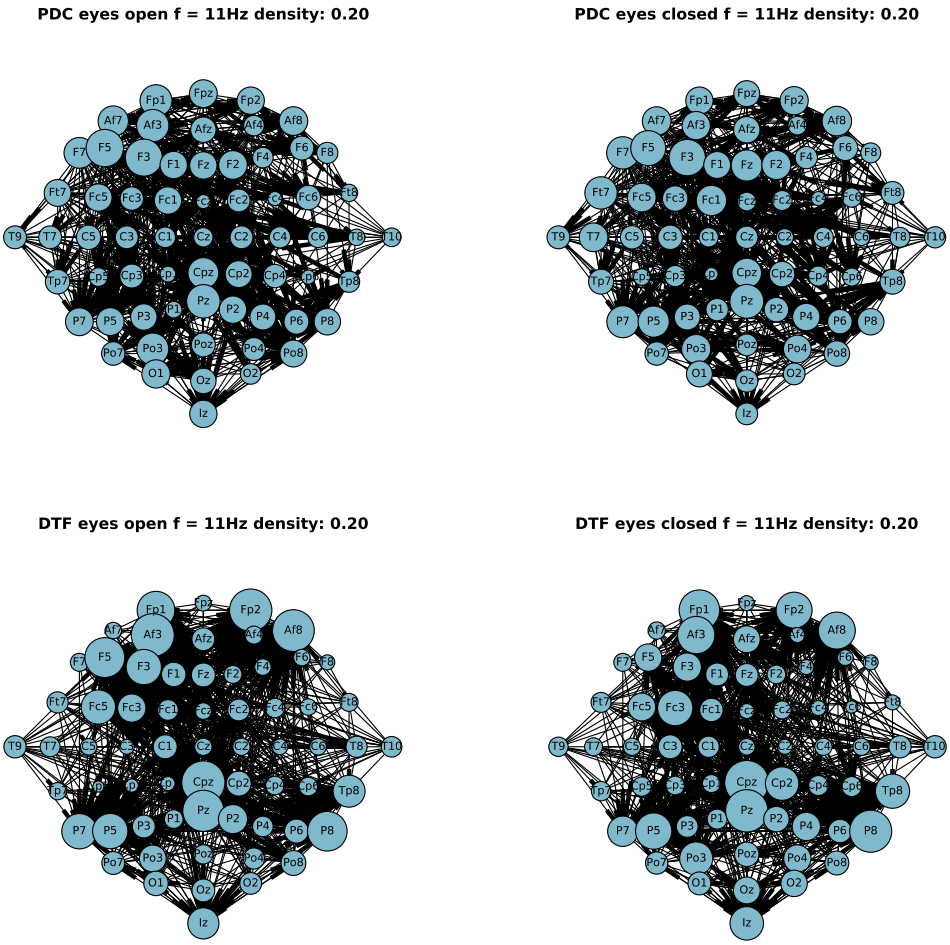
3 Graph Theory Indices

In table 2 global graph indices (clustering coefficient and path length) are shown. In table 1 instead the highest 10 channels according to the degree measure are listed. From this table we see that 10 channels with highest degree are almost the same for both eyes open and eyes closed networks. Bigger changes in the top ten can be seen between the PDC networks and the DTF networks. In general we also see that the DTF has higher degree values compared to the PDC. We also observe that for PDC the out-degree values are higher than the in-degree values while for DTF the opposite is true. A topographical representation of the local indices is in figure 6, the dimensions of the nodes are proportional to the respective degrees. We see that the channels situated between frontal and parietal lobe have in general smaller degree.

Table 1: Local indices of the 10 channels with highest degree value

PDC eyes open				PDC eyes closed				DTF eyes open				DTF eyes closed			
channel	degree	in degree	out degree	channel	degree	in degree	out degree	channel	degree	in degree	out degree	channel	degree	in degree	out degree
F5	56	11	45	F3	53	11	42	Cpz	75	63	12	Cpz	78	63	15
F3	55	11	44	F5	49	9	40	Af3	74	63	11	Pz	72	57	15
Pz	45	35	10	Pz	46	36	10	Fp2	72	61	11	P8	71	51	20
Af3	42	15	27	Ft7	42	1	41	Af8	70	62	8	Fp1	66	57	9
Po3	41	26	15	F7	42	0	42	Pz	68	58	10	Af3	56	48	8
Fp1	41	15	26	P7	41	26	15	F5	64	54	10	Af8	55	46	9
Cpz	38	36	2	P5	39	24	15	P8	63	43	20	P5	53	37	16
F7	38	0	38	Fz	39	29	10	Fp1	57	42	15	Fp2	52	41	11
Af7	37	7	30	Fc1	37	23	14	P5	50	35	15	Fc3	50	39	11
F2	33	22	11	Cpz	35	33	2	P7	50	34	16	Tp8	47	30	17

Figure 6: Topographical representation of the graph's local indices where the dimension of the node is proportional to its degree



3.1 Small-World Networks

A small-world network is characterized, according to Watts and Strogatz, by two main properties: being highly clustered and yet having small characteristic path lengths [9]. This means that even though most nodes are not directly connected to each other, neighbors of neighbors are neighbors of each other themselves. A high number of clusters is a characteristic found in regular lattices, while small path lengths are a characteristic of random graphs.
To identify small-world networks we used two different indices: the small-world coefficient σ (3) and the small-world measurement ω (4) [9].

$$\sigma = \frac{C/C_{rand}}{L/L_{rand}} = \frac{\gamma}{\lambda} \tag{3}$$

$$\omega = \frac{L_{rand}}{L} - \frac{C}{C_{latt}} \tag{4}$$

C and L are respectively the clustering coefficient and the path length of our network while C_{rand} and L_{rand} are the same measures for a random graph which has the same degree distribution as our network. The random network was generated 50 times using the NetworkX library [11], clustering coefficient and average shortest path are obtained as the average of the measures of each iteration. C_{latt} is the clustering coefficient of an equivalent lattice which mantains the same number of nodes and edges. To create this lattice we used the btcpy library [10].
The small-world measurement ω is less susceptible to fluctuations than the small-world coefficient σ since C_{latt} is used instead of C_{rand}.
A network is considered a small-world network if $C \gg C_{rand}$ and $L \approx L_{rand}$ so when $\sigma > 1$. According to ω instead a network is considered a small-world network if $L \approx L_{rand}$ and $C \approx C_{latt}$ which means when ω is close to zero. When ω assumes positive values this means that the graph has more random characteristics while when it assumes negative values the graph is more regular. Results of the small-world coefficient and measurement for our network can be found in table 2. We see that the clustering coefficient is higher and the path length is smaller for the DTF networks than for the PDC networks. However the PDCs have better small-world indices.

Table 2: Global graph indices

	Clustering Coefficient	Path Length	Small World Coefficient	Small World Measurement
PDC eyes closed	0.546581	2.124256	1.364011	-0.055942
PDC eyes open	0.502314	2.104663	1.304200	-0.106823
DTF eyes closed	0.723197	1.090774	0.647261	-0.383796
DTF eyes open	0.776795	0.714534	0.499761	-0.485618

3.2 Comparisons

In table 3 the behaviour of global graph indices in function of network density is depicted. The clustering coefficient increases with the density, while the path length first increases and then decreases for PDC. For the DTF both clustering coefficient and path length increase with the density.

Table 3: Behaviour of global graph indices in function of network density

	PDC eyes open		PDC eyes closed		DTF eyes open		DTF eyes closed	
Density	Clustering Coefficient	Path Length	Clustering Coefficient	Path Length	Clustering Coefficient	Path Length	Clustering Coefficient	Path Length
1.0%	0.132165	0.0171131	0.166531	0.0262897	0.180364	0.031746	0.0833879	0.0386905
5.0%	0.300312	0.481151	0.313945	0.338046	0.765308	0.225694	0.808348	0.256696
10.0%	0.381239	2.77108	0.423923	2.8626	0.788471	0.672867	0.747583	0.979663
20.0%	0.502314	2.10466	0.546581	2.12426	0.776795	0.714534	0.723197	1.09077
30.0%	0.63181	1.96081	0.618336	1.7939	0.768543	0.980407	0.74931	1.14583
50.0%	0.804934	1.53919	0.809789	1.53323	0.831064	1.4308	0.831925	1.30233

In table 4 we compared the global indices for 11Hz and 20Hz networks. The clustering coefficient is higher for 20Hz. The path length for 20Hz frequency is shorter for the PDCs and longer for the DTF networks, compared to the 11Hz networks. The 20Hz networks have better 'small world' values when considering the PDCs while worse when considering the DTFs. In table 5 a comparison for the local indices is depicted. The results for 20 HZ are similar to the ones obtained for 11 Hz. For a topographical comparison of local indices a representation of the 11Hz and 20Hz graphs can be found in the directory (3.6-local_indices20HZ.png).

Table 4: Comparison of global indices for networks obtained for 11Hz and 20Hz frequencies

	11 HZ				20 HZ			
	Clustering Coefficient	Path Length	Small World Coefficient	Small World Measurement	Clustering Coefficient	Path Length	Small World Coefficient	Small World Measurement
PDC eyes open	0.502314	2.10466	1.3042	-0.106823	0.804934	1.53919	1.37237	-0.193606
PDC eyes closed	0.546581	2.12426	1.36401	-0.0559422	0.809789	1.53323	1.37874	-0.194928
DTF eyes open	0.776795	0.714534	0.499761	-0.485618	0.831064	1.4308	1.25146	-0.262191
DTF eyes closed	0.723197	1.09077	0.647261	-0.383796	0.831925	1.30233	1.25067	-0.262173

Table 5: Local indices of the networks obtained 20Hz frequencies

PDC eyes open				PDC eyes closed				DTF eyes open				DTF eyes closed			
channel	degree	in degree	out degree	channel	degree	in degree	out degree	channel	degree	in degree	out degree	channel	degree	in degree	out degree
F5	56	11	45	F3	53	11	42	Cpz	75	63	12	Cpz	78	63	15
F3	55	11	44	F5	49	9	40	Af3	74	63	11	Pz	72	57	15
Pz	45	35	10	Pz	46	36	10	Fp2	72	61	11	P8	71	51	20
Af3	42	15	27	F7	42	0	42	Af8	70	62	8	Fp1	66	57	9
Po3	41	26	15	Ft7	42	1	41	Pz	68	58	10	Af3	56	48	8
Fp1	41	15	26	P7	41	26	15	F5	64	54	10	Af8	55	46	9
Cpz	38	36	2	P5	39	24	15	P8	63	43	20	P5	53	37	16
F7	38	0	38	Fz	39	29	10	Fp1	57	42	15	Fp2	52	41	11
Af7	37	7	30	Fc1	37	23	14	P5	50	35	15	Fc3	50	39	11
Af8	33	17	16	Cpz	35	33	2	P7	50	34	16	Tp8	47	30	17

We also compared our 20% density network with a weighted network. Starting from the PDC and DTF matrices we generated four graphs (eyed open and eyes closed) taking the values of the matrices as the weights of the edges, thus generating almost complete graphs. From table 6 we see that the clustering coefficient is lower for the weighted graph than for the 20% graph, whereas when comparing PDC and DTF for the weighted graph we see that the clustering coefficient is higher for

the PDC. Here the path length can be interpreted as the PDC/DTF average value. In table 7 local indices for the weighed graph are found.

Table 6: Comparison of global indices in a 20% density network and weighted network

	20% density graph		weighted graph	
	Clustering Coefficient	**Path Length**	**Clustering Coefficient**	**Path Length**
PDC eyes open	0.502314	2.10466	0.139434	0.0293599
PDC eyes closed	0.546581	2.12426	0.163047	0.0305121
DTF eyes open	0.776795	0.714534	0.0913334	0.0327027
DTF eyes closed	0.723197	1.09077	0.0840769	0.0300813

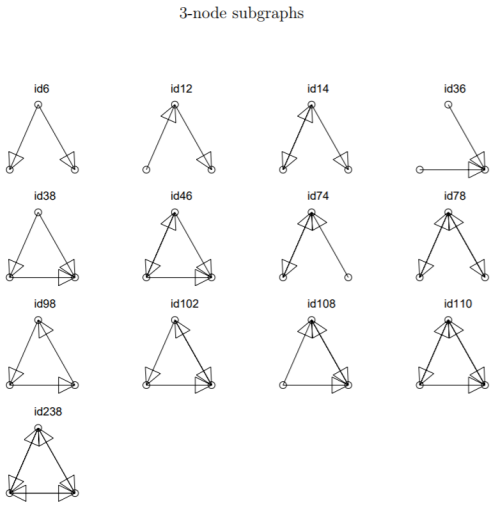
Table 7: Local indices for the weighed graph

PDC eyes open				PDC eyes closed				DTF eyes open				DTF eyes closed			
channel	**degree**	**in degree**	**out degree**	**channel**	**degree**	**in degree**	**out degree**	**channel**	**degree**	**in degree**	**out degree**	**channel**	**degree**	**in degree**	**out degree**
F3	15.5282	5.65479	9.87342	F3	13.513	4.82475	8.68821	Cpz	35.4409	31.2088	4.23211	Cpz	42.4753	38.2106	4.26473
F5	14.8561	5.20601	9.65007	F5	12.682	4.14385	8.53818	Pz	22.2483	18.2033	4.04495	Pz	23.1525	18.7967	4.35576
Af3	12.399	5.80203	6.59701	F7	12.0049	1.07865	10.9263	Af3	21.2813	16.6807	4.60063	Fp1	16.2703	11.6051	4.66523
Pz	11.8849	7.43647	4.44846	Pz	11.858	7.51324	4.34475	Af8	17.1591	13.3641	3.79494	Fp2	14.0959	9.41786	4.678
Fp1	11.5868	5.16579	6.42096	Af8	11.8133	6.3088	5.50446	F5	17.0496	12.0906	4.95905	P5	13.9813	8.61206	5.36927
P7	11.2307	6.59906	4.63161	Af7	11.7299	4.03968	7.69022	Fp2	16.922	12.6282	4.29381	P8	13.9144	8.02437	5.89003
P5	11.1986	6.64934	4.54929	Fc1	11.5504	6.12647	5.42396	Fp1	14.6732	9.61241	5.06075	Tp8	13.4227	8.03364	5.38901
Fz	11.1426	6.51017	4.63239	Af3	11.301	5.08558	6.21545	F3	14.0791	9.22684	4.8523	Af3	13.2404	8.81693	4.42351
P2	11.0032	6.58518	4.41802	Po3	11.0462	6.61714	4.42908	P8	13.3369	7.87	5.4669	Af8	13.0833	8.7218	4.36153
Af8	10.9529	5.73304	5.21989	P4	11.0241	6.69195	4.33215	P7	12.8484	7.37468	5.47376	Po3	12.1019	7.16211	4.93983

4 Motif Analysis

We performed a motifs analysis using mfinder1.21 [12] to investigate the presence of 3-node configurations in the networks. This was done for all four graphs (generated using PDC and DTF and eyes closed and eyes open case). There are 13 possible isomorphic types of 3-node connected subgraphs (figure 7).

Figure 7: 13 possible isomorphic types of 3-node connected subgraphs [12]



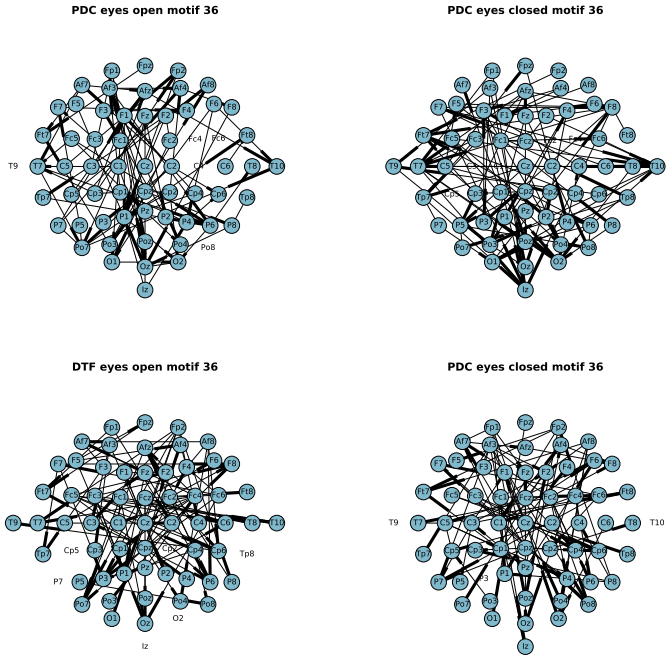
Frequency values and statistical significance, i.e. whether the pattern is a motif or an anti-motif, are in the table 8.

Table 8: Frequency values and statistical significances of the 3-node configurations

	PDC eyes open		PDC eyes closed		DTF eyes open		DTF eyes closed	
id	Frequency	Statistical Significance	Frequency	Statistical Significance	Frequency	Statistical Significance	Frequency	Statistical Significance
6	2974	anti-motif	3119	anti-motif	576	motif	710	-
12	2047	motif	1850	anti-motif	805	motif	1118	-
14	886	anti-motif	763	anti-motif	109	anti-motif	216	anti-motif
36	2353	anti-motif	2092	anti-motif	8708	anti-motif	6538	anti-motif
38	1624	motif	1741	motif	1769	motif	1476	motif
46	451	motif	392	motif	188	motif	241	motif
74	787	anti-motif	843	anti-motif	1496	anti-motif	1795	anti-motif
78	69	anti-motif	66	anti-motif	49	anti-motif	88	anti-motif
98	21	anti-motif	31	anti-motif	12	anti-motif	19	anti-motif
102	89	anti-motif	139	anti-motif	107	-	158	anti-motif
108	363	motif	384	motif	1454	anti-motif	1170	motif
110	158	motif	171	motif	268	-	366	anti-motif
238	53	motif	44	motif	97	motif	116	motif

A topographical representation of the network considering only the motifs with pattern A→B←C is found in the directory. In figure 8 a topographical representation of the network depicting only 100 edges, taken at random, can be found, the edges seem to be oriented outwards.

Figure 8: Topographical representation of the network depicting only 100 edges considering only the motifs with pattern A→B←C



We chose the channel Po3 in the parieto-occipital scalp region, a representation of the motifs which involve it is in table 9.

Table 9: Frequency values and statistical significances of the 3-node configurations for the Po3 channel

	PDC eyes open	PDC eyes closed	DTF eyes open	DTF eyes closed
id	frequency	frequency	frequency	frequency
6	128	98	28	57
12	194	99	94	74
14	43	37	2	26
36	171	84	63	66
38	150	112	111	99
46	51	29	25	21
74	96	77	62	167
78	6	5	1	17
98	5	5	2	2
102	7	16	14	19
108	34	35	31	99
110	11	24	12	43
238	4	2	3	10

Finally we performed the motif analysis considering 4-node motifs. With 4-nodes we can have 199 isomorphic types of connected subgraphs. The table depicting the results is in the directory. It is likely that the motifs represent simpler patterns starting from which the more complex network is formed.

5 Community Detection

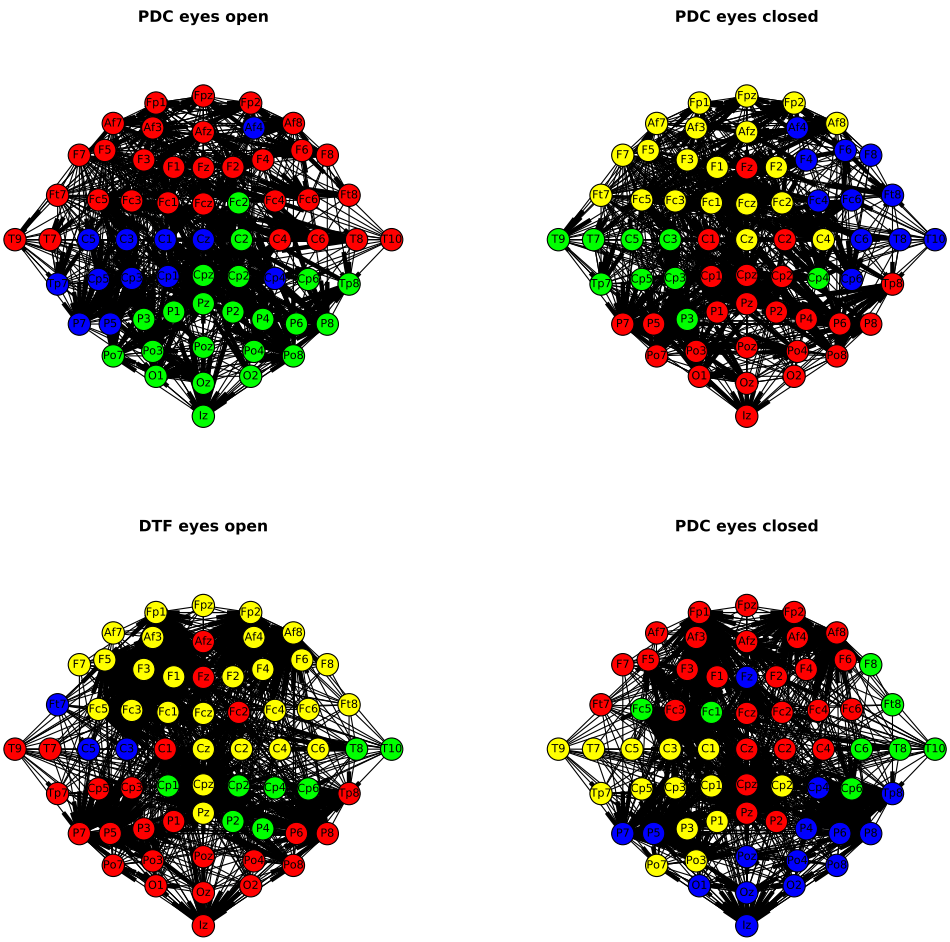
To detect communities in our networks we used the Leicht& Newman algorithm implemented by the community_newman python library [13]. This method takes into account information about the direction contained in the edges. In this algorithm the goodness of partition is evaluated using the following quality score:

$$Q = \frac{1}{2m} \sum_{ij} (A_{ij} - P_{ij}) \delta(C_i, C_j) \tag{5}$$

where
A is the adjacency matrix
m is the total number of links in the network
$P_{ij} = \frac{k_i^{out} k_j^{in}}{m}$ is the expected number of edges between nodes i and j in the null model
$\delta(C_i, C_j) = 1$ only if i and j belong to the same community ($C_i = C_j$)

The maximum value of Q is not always indicative of a good partition.
The community structure in both rest conditions obtained using this algorithm can be seen in figure 9.

Figure 9: Community structure in both rest conditions obtained using Leicht&Newman algorithm



6 List of tasks

Table 10: List of tasks chosen for the project

Task	Class
1.1	mandatory
1.2	B
1.3	C
2.1	mandatory
2.2	A
2.3	A
2.5	C
2.6	B
3.1	mandatory
3.2	D
3.3	B
3.4	C
3.5	B
3.6	B
3.7	C
4.1	mandatory
4.2	C
4.3	C
4.4	E
5.1	mandatory
5.2	B

References

[1] **PhysioNet EEG Motor Movement/Imagery dataset:** https://physionet.org/pn4/eegmmidb/

[2] **PhysioNet:** Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 101(23):e215-e220 [Circulation Electronic Pages; http://circ.ahajournals.org/cgi/content/full/101/23/e215]; 2000 (June 13).

[3] http://www.bci2000.org/

[4] http://www.ncbi.nlm.nih.gov/pubmed/15188875
Schalk, G., McFarland, D.J., Hinterberger, T., Birbaumer, N., Wolpaw, J.R. BCI2000: A General-Purpose Brain-Computer Interface (BCI) System. IEEE Transactions on Biomedical Engineering 51(6):1034-1043, 2004. [In 2008, this paper received the Best Paper Award from IEEE TBME.]

[5] **pyEDFlib:** http://pyedflib.readthedocs.io/en/latest/

[6] **SciPy:** https://www.scipy.org/

[7] **spectrum:** https://pypi.python.org/pypi/spectrum/0.7.3

[8] **connectivipy:** https://https://github.com/dokato/connectivipy

[9] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3604768/
Telesford QK, Joyce KE, Hayasaka S, Burdette JH, Laurienti PJ. The ubiquity of small-world networks. Brain Connect. 2011;1(5):367–375. doi: 10.1089/brain.2011.0038.

[10] **bctpy:** https://github.com/aestrivex/bctpy

[11] **NetworkX:** https://networkx.github.io/

[12] **mfinder1.21:** http://www.weizmann.ac.il/mcb/UriAlon/research/network-motifs

[13] **community_newman:**https://github.com/zhiyzuo/python-modularity-maximization