Bioinformatics Report HW3 Group 5

Brain Oscillatory and Network Activity during resting states

Giorgio Giannone[1], Livia Lombardi[1]

INTRODUCTION

All body activities and actions are timed and coordinated by the nervous system. The nervous system includes a network of neurons that transmit signal throughout the body. Nervous system is divided in Central Nervous System (CNS) and the Peripheral Nervous System (PNS). Spinal cord and brain form the CNS. CNS activity comprehends electro physiological, metabolic and electrochemical phenomena. Neurons communicate with each other via electrical pulses that causes the release of chemical neurotransmitter. When two neurons joint each other is called synapse. A widespread method to monitor and measure the synaptic potential is the EEG, that stands for electroencephalogram, which records excitatory and inhibitory post-synaptic potentials. By placing a set of electrodes on the scalp in a specific way - which depend on the area of the skull (left, right, central sides) EEG tracks brain wave patterns. The aim of this project is to analyze two EEG datasets performed using the BCI2000 system and coming from two different methodology records:

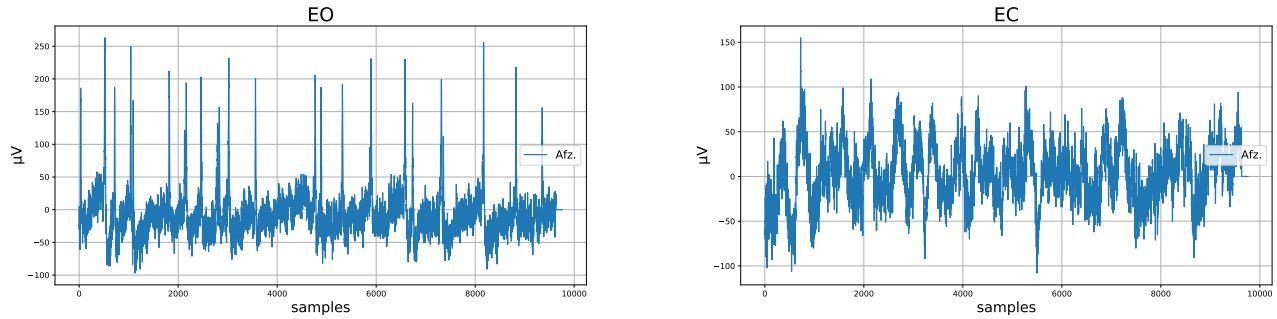
- The first group of records is extracted from subjects who took the experiment with closed eyes, during a resting state.
- The second group of records is extracted from subjects who took the experiment with opened eyes, during a resting state.

SPECTRAL ANALYSIS

An EEG signal can be analyzed studying the sinusoids that form the signal, so thanks to the connection with the reference signal, a complex signal can be decomposed in simpler signals originally embedded in it: this process is called cross-correlation.

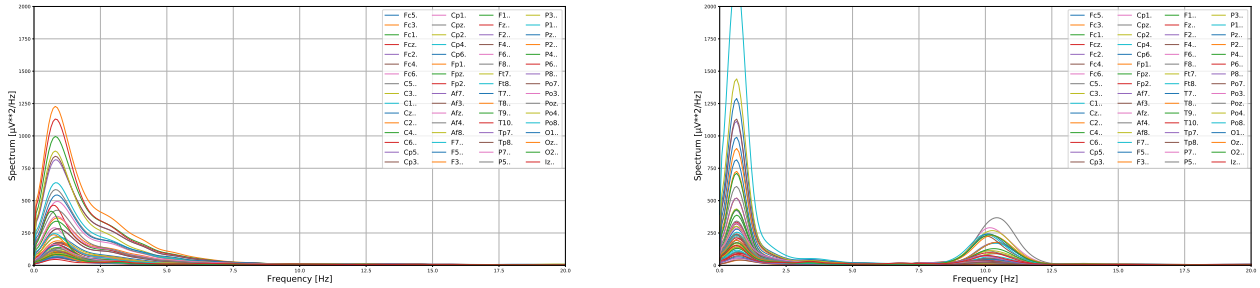
1.1

The Welch's method [3] computes an estimate of the Power Spectral Density (PSD) dividing the data into overlapping segments, computing a modified periodogram for each segment and averaging the periodograms; we consider segments of 256 samples points, with an overlap of 0.5 and an hamming



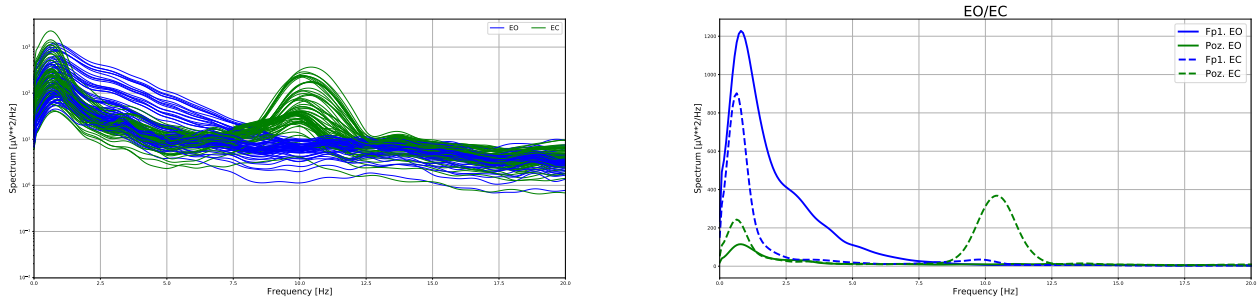
(a) eyes-open (b) eyes-closed

Fig. 1: Signals channels Afz two rest condition



(a) eyes-open (b) eyes-closed

Fig. 2: PSDs for all channels two rest condition



(a) log psd (b) psd Fp1 and Poz channels

Fig. 3: PSDs all channels and relevant channels

window. The result is shown in Fig.3a, where a clear pick is visible around 10 Hz with closed eyes: the alpha waves (cerebral activity between 8-13 Hz) are the reason of this activity because they are related with the closing eyes state. Given this result we decided to use a data driven strategy: we selected the two most active channels for low (<5Hz) and high (>5Hz) frequency regime. The final choice for channels is one in the frontal lobe (Fp1) and one in the parietal-occipital lobe (Poz); when we need to choose a particular frequency we consider 10Hz (3b).

1.2

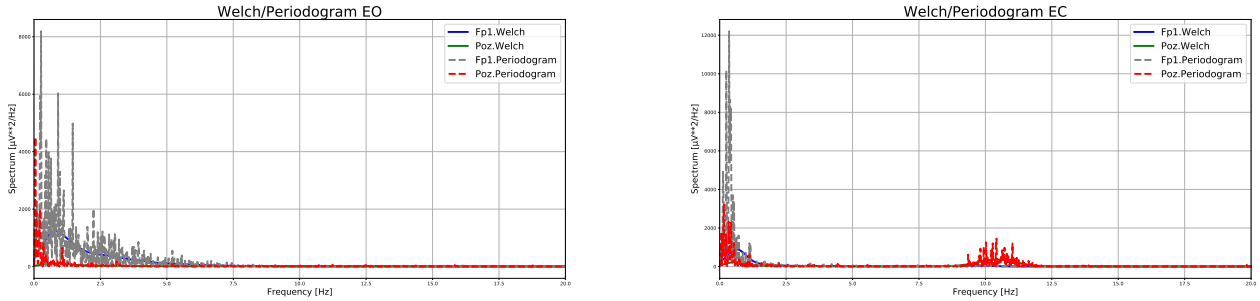
In Fig.4 we compared the averaged Welch's method with the periodogram method: the Welch's method is an average of PSDs and for this reason smoother respect the classical periodogram method. In general the two approaches return the same information in different ways.

1.3

In Fig. 5 we computed the PSD with different parameters for the two rest states: in particular we tried a range of value for the window's overlap and for the fft's length. The results are similar and the biggest difference is in the shape of the curves.

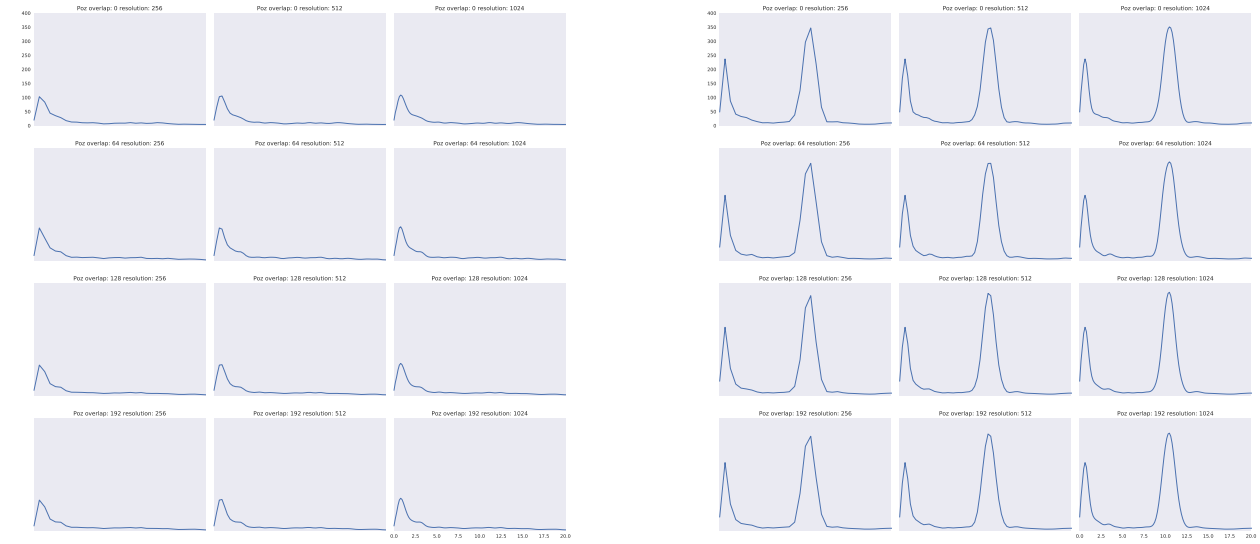
1.4

In Fig. 6 the frequency is fixed to 10Hz and the values of every channel for the two rest states (0-open eyes / 1-closed eyes) are plotted: there is practically no activity for open-eyes; there is an increasing in activity for the closed-eyes condition.



(a) eyes-open (b) eyes-closed

Fig. 4: comparison PSD methods



(a) eyes-open (b) eyes-closed

Fig. 5: PSDs different parameters

1.5

We tried two different approaches to solve this problem: the first approach is an asymptotic analysis; the second one is a statistical test. For the asymptotic analysis (Fig.7) we assume a white noise and we verify a non normal distribution for the error sample (false negative rate 3%); for the statistical test we perform a Mann-Whitney rank test (this test checks if two samples are from the same distribution without normal hypothesis): we obtain a pi-value of 4.2 e-28 and we reject the null hypothesis. So we conclude that the two samples are from different distribution and that the difference is caused by some activity and not noise.

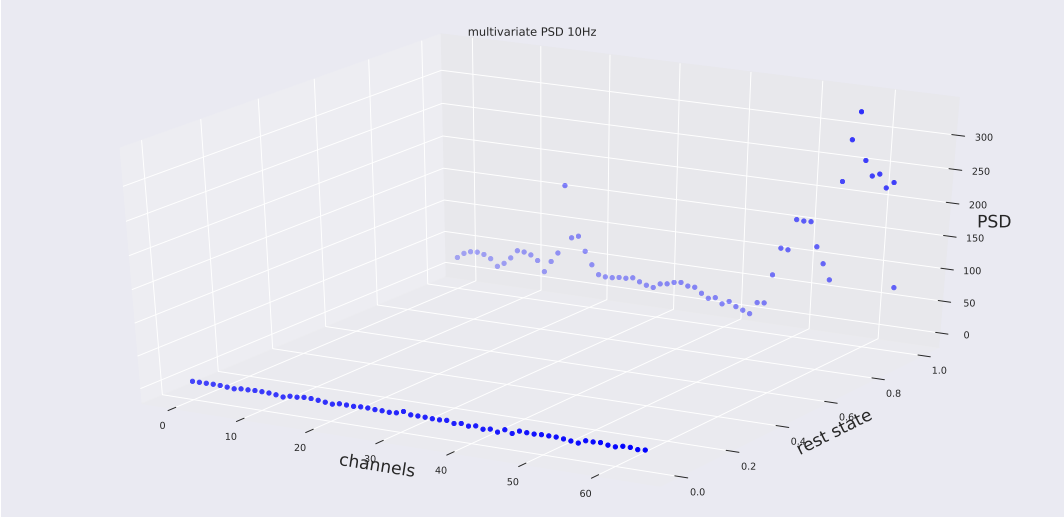


Fig. 6: multivariate PSD 10Hz channels-rest condition

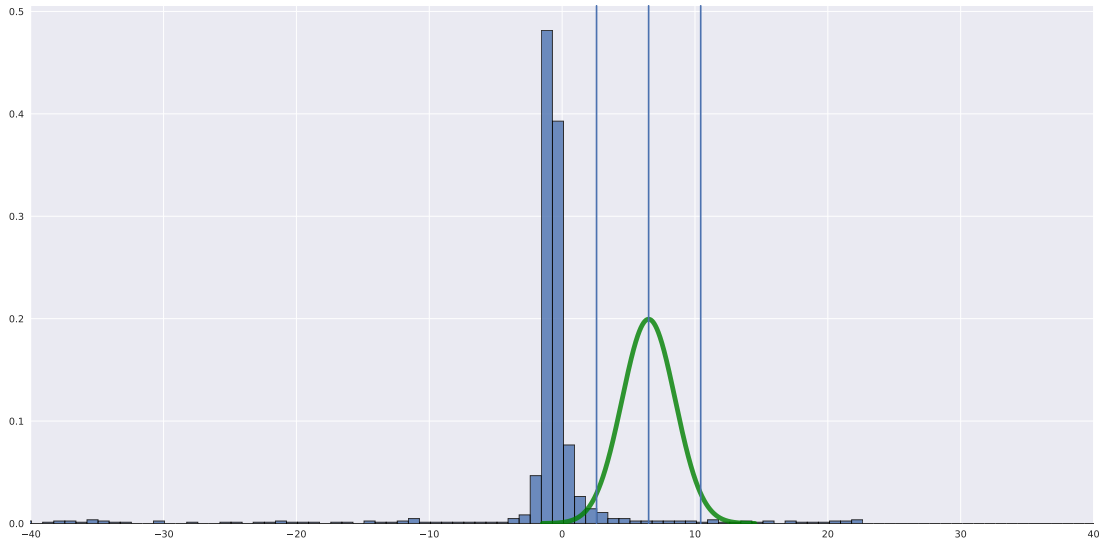


Fig. 7: asymptotic analysis

CONNECTIVITY

From an EEG output can be performed a brain connectivity analysis: how cerebral areas cooperate for execute some tasks, thanks to different regional activations scanned in different times. A general time series contains information about the process that generated it. A method used to analyze this process is called Autoregressive Model (AR). The idea of AR method is modeling the current value of a variable

through a weighted linear sum of its previous values. The natural extensions of the AR is the Multivariate Autoregressive Model (MVAR) which generalize the AR approach to multiple time series [16].

2.1-2.2

Applying a MVAR model, the Direct Transfer Function (DTF) [1] (it measures the overall effect of one signal on another, considering both direct and indirect paths) and the Partial Directed Coherence (PDC) [2] are computed; we used cross-validation with the akaike information criterion as metric to choose the best model order (Fig. 8) for the two rest states.

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

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

Selecting the matrix connectivity at frequency 10Hz (Fig. 9, 10) we built a connectivity graph with density around 20% for the two estimators and the two rest states (Fig. 12, Fig. 13).

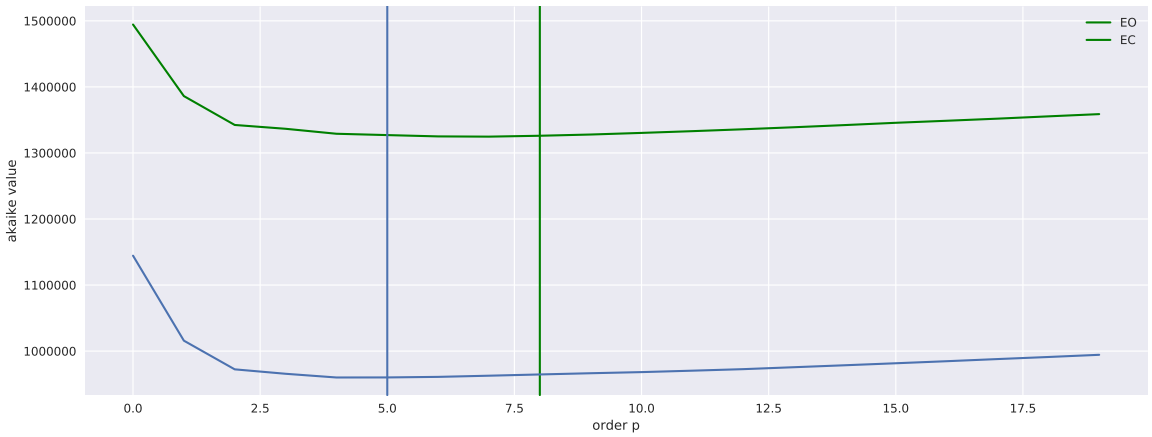


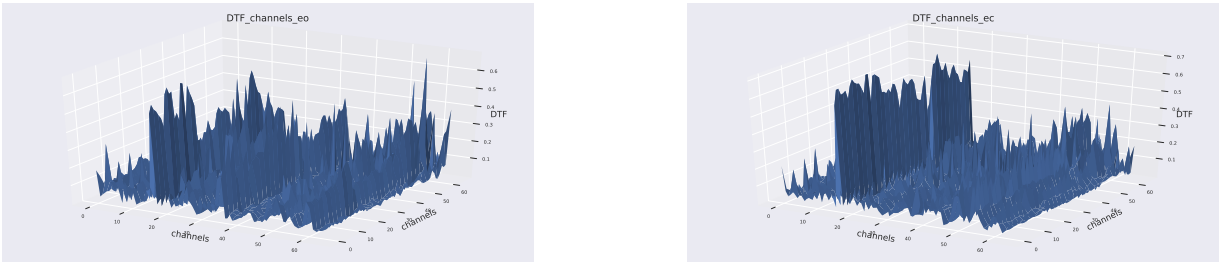
Fig. 8: akaike for model selection DTF/PDC

2.3

we performed the same procedure with the DTF for different density thresholds (Fig. 14, Fig. 15, Fig. 16). Decreasing the density, the most connected nodes and the most active zones are clearly visible.

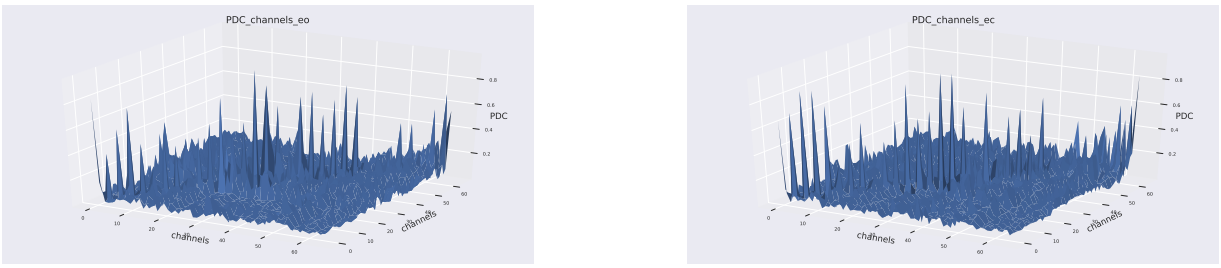
2.4-2.5

we selected 19 channels among 64, we computed a new DTF, and using a re-sampling technique we removed not statistically significant connection for these new networks (Fig. 17). Finally we plotted a graphical representation for the closed-eyes and open-eyes conditions.



(a) eyes-open (b) eyes-closed

Fig. 9: DTF vs channels 10 Hz



(a) eyes-open (b) eyes-closed

Fig. 10: PDC vs channels 10 Hz

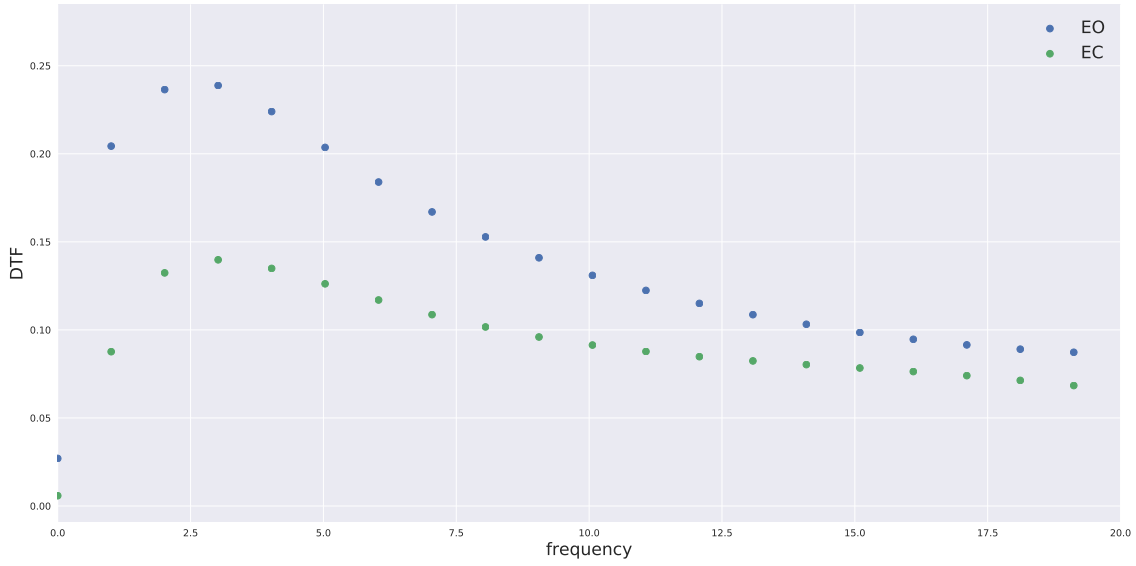
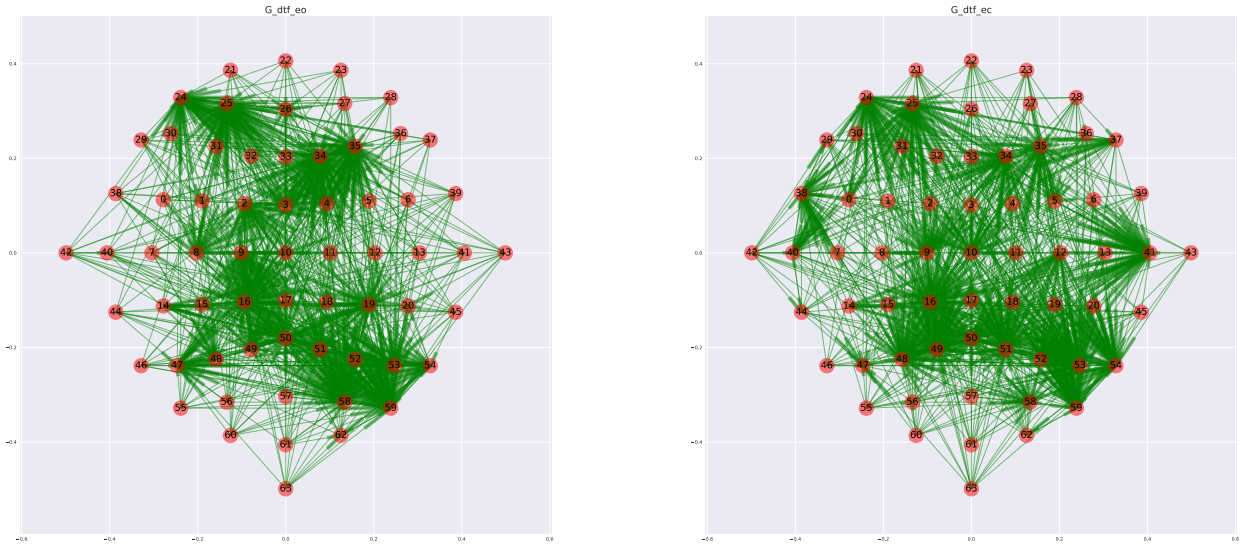
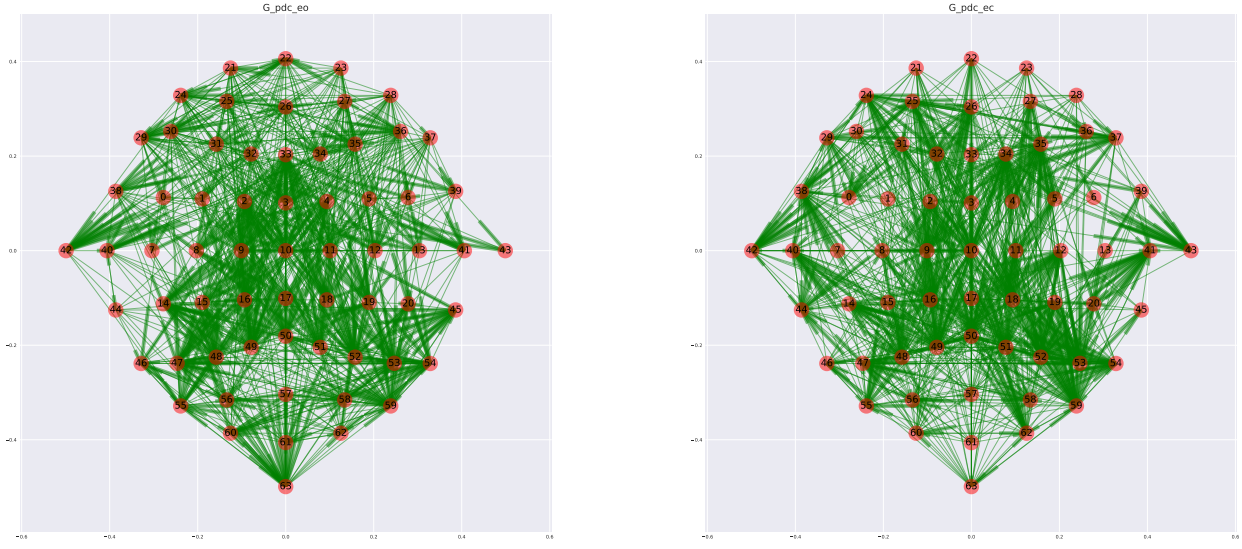


Fig. 11: DTF channels(1,2) vs frequency



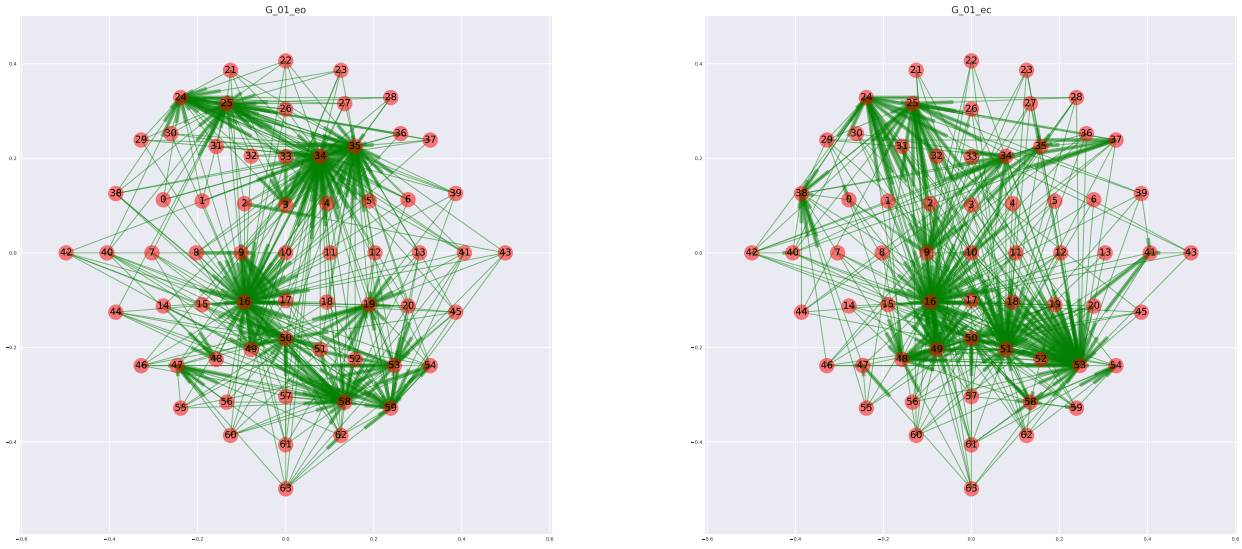
(a) eyes-open (b) eyes-closed

Fig. 12: connectivity graph DTF density 0.2 10 Hz



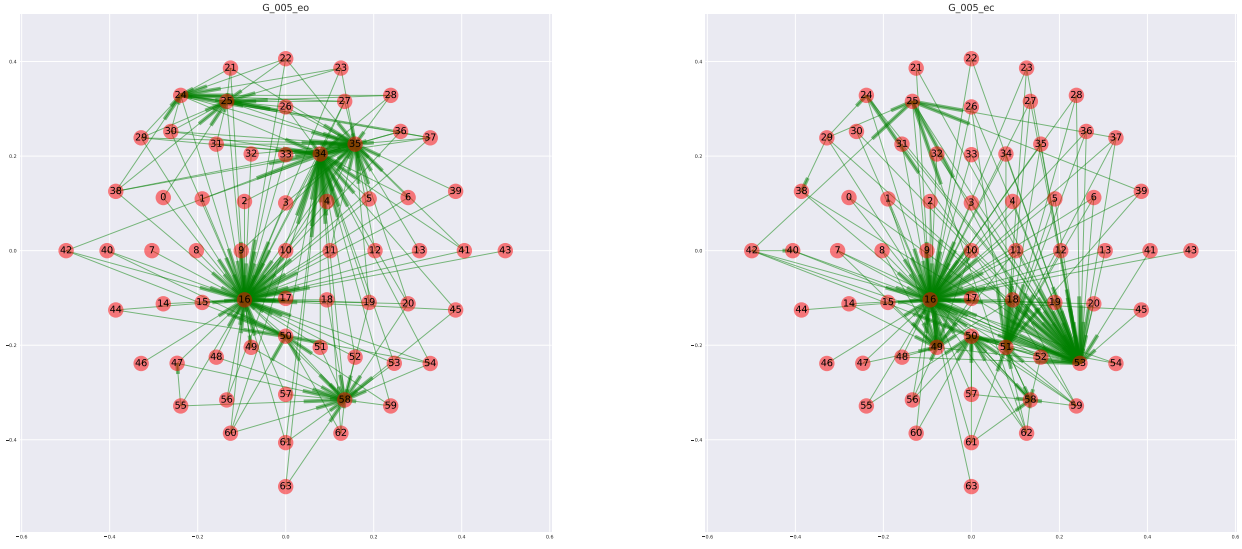
(a) eyes-open (b) eyes-closed

Fig. 13: connectivity graph PDC density 0.2 10 Hz



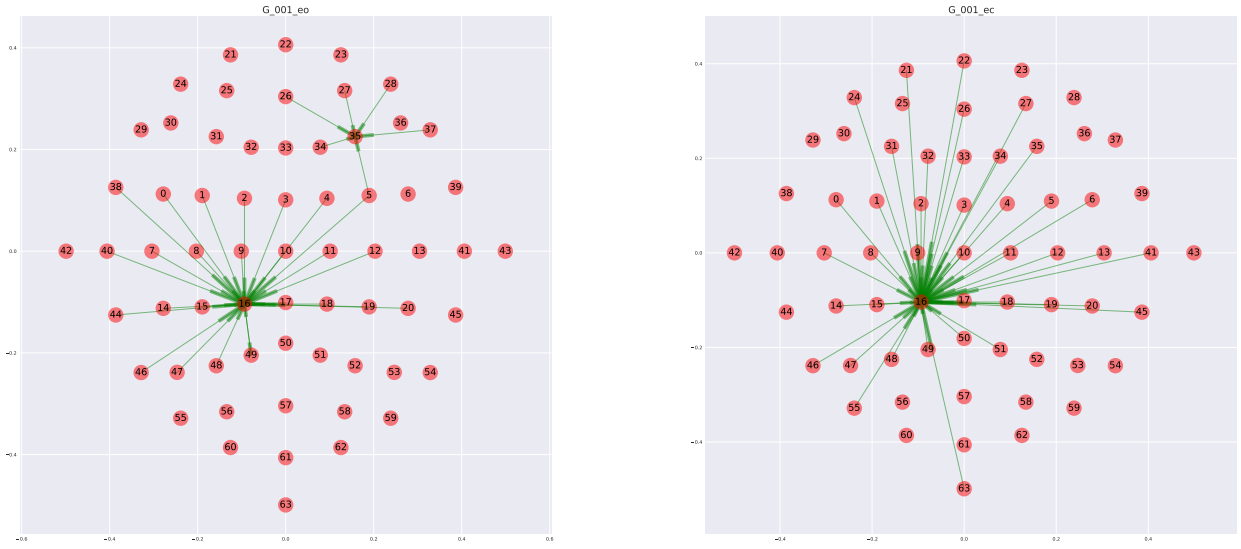
(a) eyes-open (b) eyes-closed

Fig. 14: connectivity graph DTF 10 Hz density 0.1



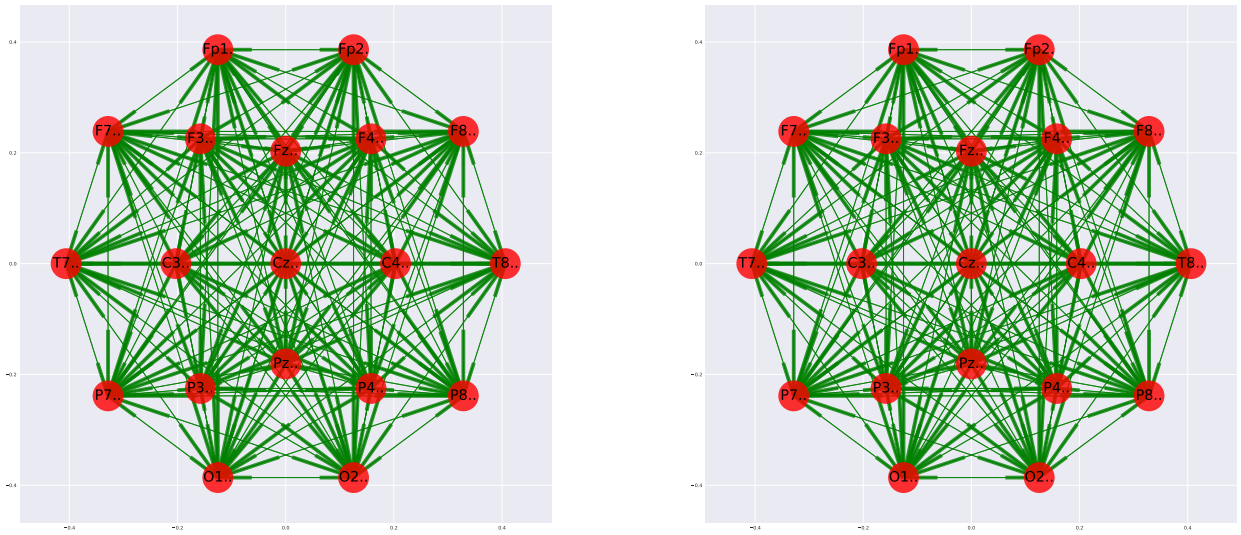
(a) eyes-open (b) eyes-closed

Fig. 15: connectivity graph DTF 10 Hz density 0.05



(a) eyes-open (b) eyes-closed

Fig. 16: connectivity graph DTF 10 Hz density 0.01



(a) eyes-open (b) eyes-closed

Fig. 17: sub-sample connectivity graph DTF 10 Hz

GRAPH THEORY

Brain connectivity means information flows between different brain regions. Undertaking a graph analysis of brain connectivity can lead to a complete description of the structural organization of the brain. Starting from the standard graph theory, graphs as objects made up of nodes and edges can be mirrored respectively as brain areas and functional connections. In this prospective the graph theory can be used to understand the brain network organization, quantify brain connectivity properties and look for markers. Looking at networks in neuroscience, can show small-world networks. A small-world network in mathematical graph defines networks in which most nodes are not neighbors among themselves but at the same time most nodes are reachable from every other node by a small number of steps [11] [13]. This kind of organization let a rapid integration of information between different specialized brain areas even when they are distant.

3.1-3.3

Global indexes (for the networks with density 0.2) are summarized in table I for the two estimator and the two rest states. In particular the average shortest path and the average clustering coefficient are computed as global measures; and node degree, in-degree and out-degree for the local measures. The 20

TABLE I: Global indices

	average clustering coefficient	average shortest path
dtf ec	0.76	1.66
dtf eo	0.78	1.66
pdc ec	0.48	1.73
pdc eo	0.53	1.67

channels with highest degree are reported (using the DTF) in tables II (left:EO; right:EC).

TABLE II: table local indices E0-EC

	degree	in-degree	out-degree
Cp1.	77	64	13
F2..	76	63	13
Po4.	75	64	11
Po8.	74	59	15
F4..	70	58	12
Af3.	67	55	12
Pz..	60	44	16
Cp4.	58	42	16
Af7.	54	42	12
P6..	52	39	13
P2..	43	27	16
P5..	42	27	15
P8..	37	20	17
Fcz.	34	23	11
C1..	34	23	11
P1..	33	21	12
Fc1.	31	22	9
C3..	30	19	11
Fc2.	28	15	13
P3..	27	12	15

	degree	in-degree	out-degree
P6..	75	64	11
Cp1.	72	64	8
P3..	65	53	12
P2..	64	54	10
P1..	63	56	7
P8..	60	44	16
Po8.	58	43	15
T8..	56	43	13
Af3.	50	34	16
Pz..	47	36	11
Af7.	45	30	15
F2..	44	31	13
F4..	44	31	13
Ft7.	43	26	17
C1..	38	29	9
Po4.	33	22	11
Cz..	31	23	8
T7..	29	13	16
P5..	27	12	15
P4..	25	14	11

3.2

This graph configuration is a typical small-world organization (a graph where most nodes are not neighbors of one another, but the neighbors of any given node are likely to be neighbors of each other and most nodes can be reached from every other node by a small number of hops or steps). In formula we can use the small-coefficient [12]

$$\sigma = \frac{\frac{C}{C_r}}{\frac{L}{L_r}}$$

and define a small-word configuration a network where $\sigma > 1$ (Table III)

TABLE III: small-word

	C	C_r	L	L_r	σ
EO	0.78	0.20	1.66	1.85	1.26
EC	0.76	0.20	1.66	1.86	1.21

3.4

The behavior of the global indexes is analyzed in Fig. 18 varying the density graph (0.5, 0.2, 0.1, 0.05, 0.01)

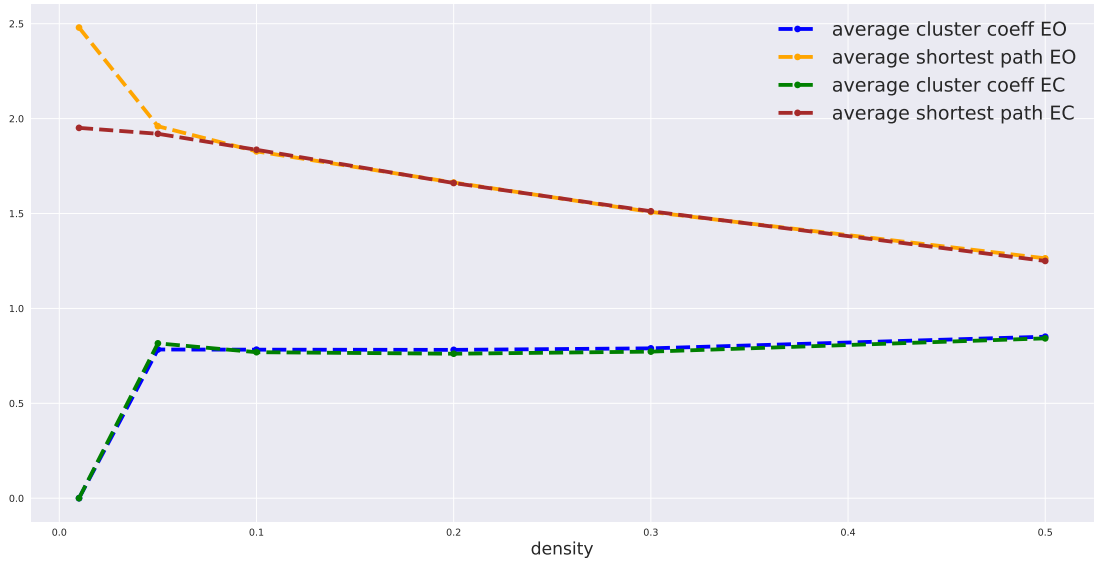
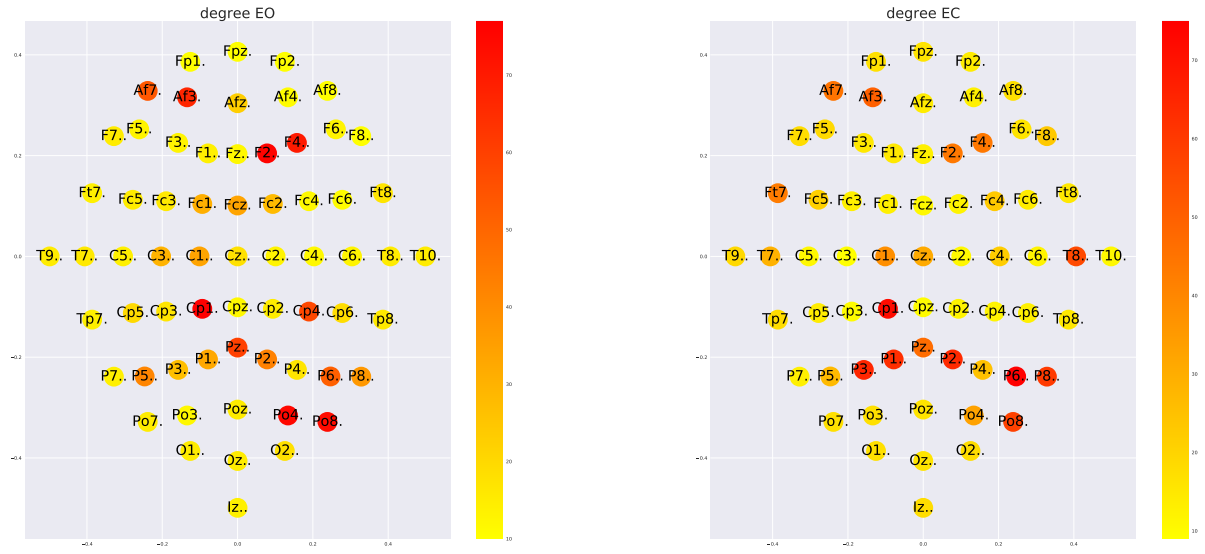


Fig. 18: global indices vs density

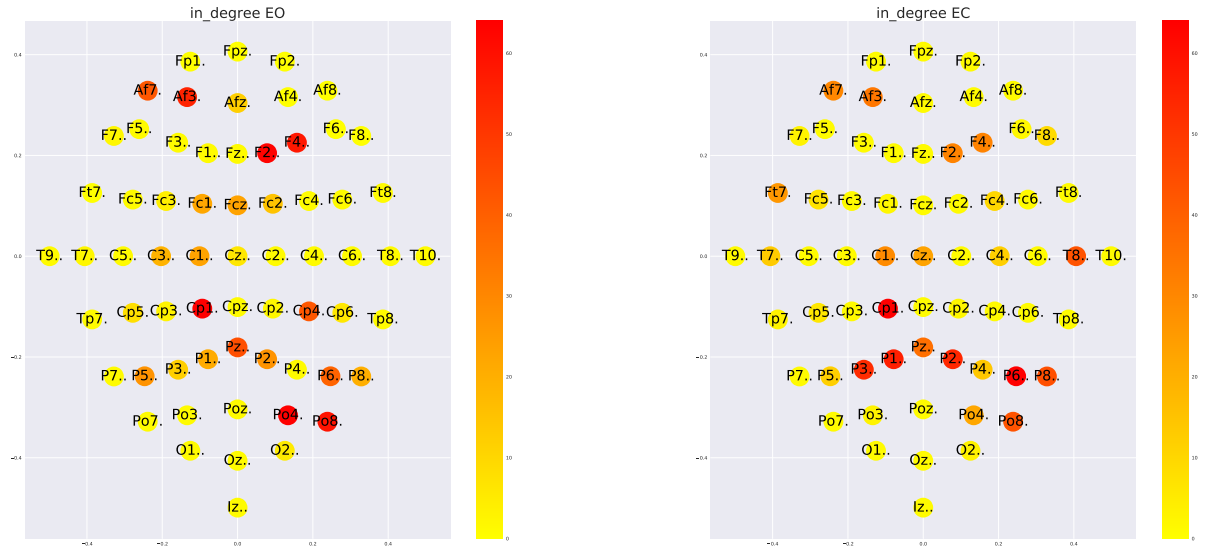
3.5

Topological visualization of the local indexes (degree, in-degree, out-degree) for the two rest states are plotted in Fig.19, Fig.20, Fig.21. It is possible to see an increase in degree in the bottom part of the images with eyes-closed.



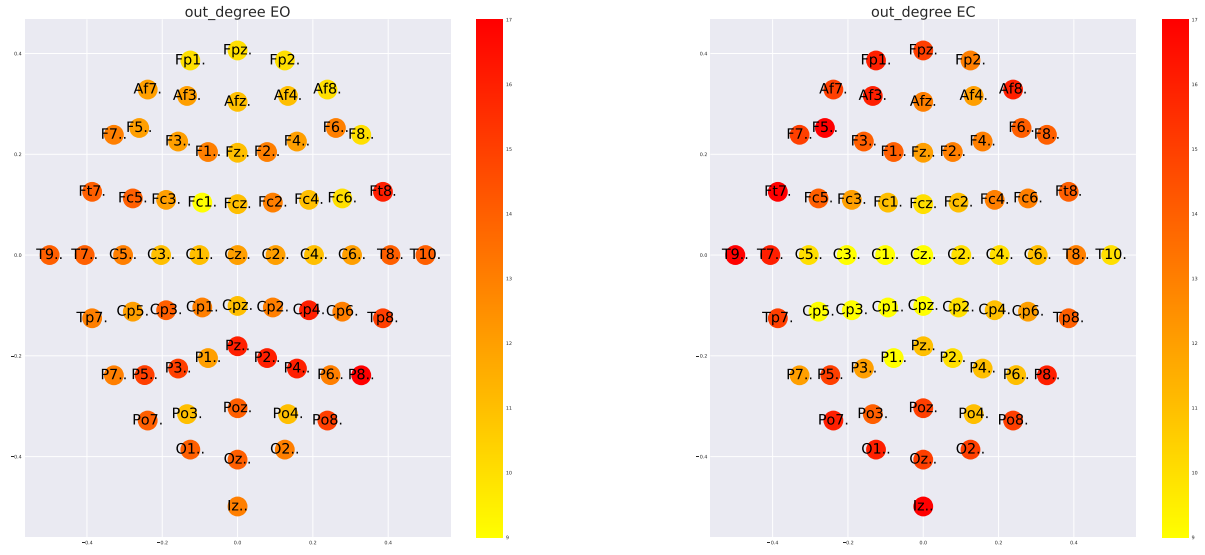
(a) eyes-open (b) eyes-closed

Fig. 19: nodes degree



(a) eyes-open (b) eyes-closed

Fig. 20: nodes in-degree



(a) eyes-open (b) eyes-closed

Fig. 21: nodes out-degree

MOTIFS

A network in Biology, is useful to represent interactions. Interactions can be in proteins, DNA, genes, neurons etc. and these groups can have specific characterization identified by redundant sequences. In a sequence, what we call motif is a subsequence that have a functional significance across the network. A motif has an import since means that the subsequence appears with a high frequency in the graph. So, a motif is a pattern of inter-connection that is occurring for a significantly number of times. Conduct a motif analysis means counting how many times a pattern is occurring in a network, when subnetworks are over and/or under represented important information is gathered. Motifs show that behind a complex network exist simpler networks which play a very important role.

4.1

We investigate under-represented and over-represented third order motifs using the mfinder [7] tool: the program compute a large number of random graph with same density and size of the original graph; after that compares the average number of motifs in the original graph and the random graph to perform a statistical validation test. In our case we are particularly interested in the motif type 36: this motif is under-represented in both the connectivity graphs. Analyzing the table IV we see that the motif configuration 38, 74 and 108 are differently represented in the two rest states.

4.2

Visualization of the networks obtained considering only the motif36 (Fig. 22)

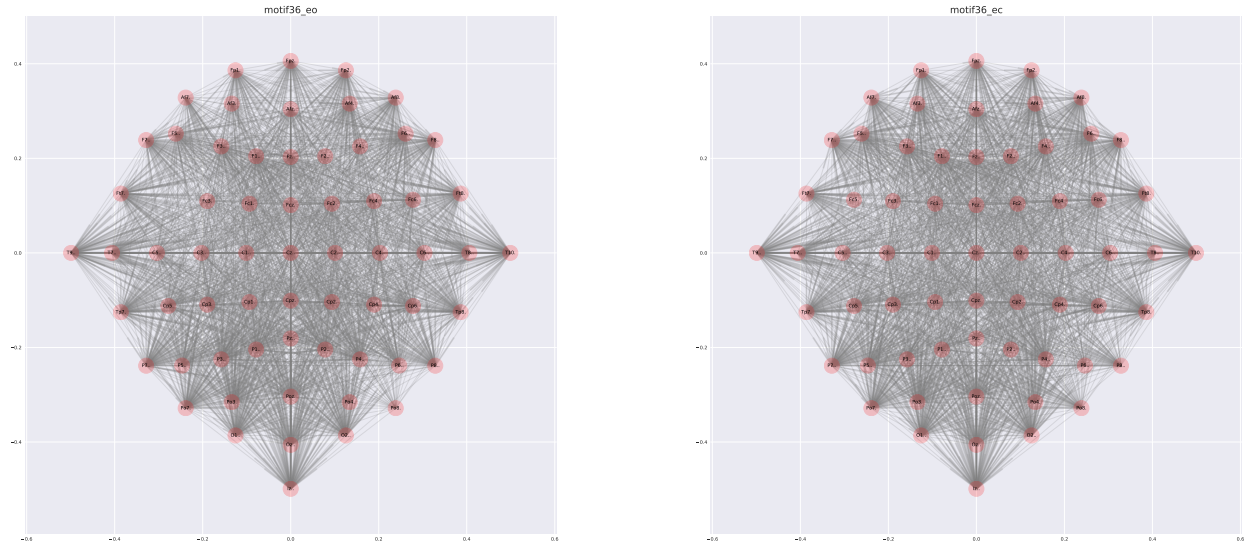
4.4

Same analysis with forth order motifs and sample of the results in tables V

TABLE IV: motif3 EO-EC

	frequency	statistical significance
6	196	under-represented
12	565	normal-represented
14	46	under-represented
36	8251	under-represented
38	1256	normal-represented
46	88	over-represented
74	1746	under-represented
78	33	under-represented
98	6	normal-represented
102	108	under-represented
108	1479	over-represented
110	240	normal-represented
238	128	over-represented

	frequency	statistical significance
6	434	under-represented
12	718	normal-represented
14	60	under-represented
36	7974	under-represented
38	1715	over-represented
46	131	over-represented
74	1462	normal-represented
78	53	under-represented
98	10	normal-represented
102	81	under-represented
108	1098	normal-represented
110	181	normal-represented
238	65	over-represented



(a) eyes-open

(b) eyes-closed

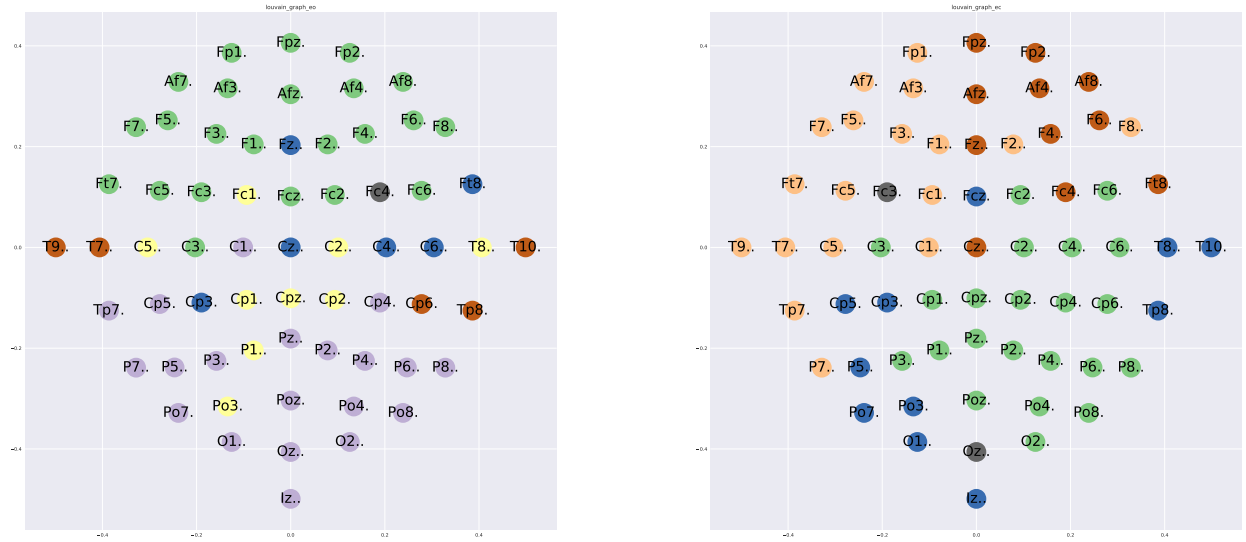
Fig. 22: motif36

COMMUNITY DETECTION

Finally we perform a multi-level cluster analysis: we used a modularity-based vs information theory-based method.

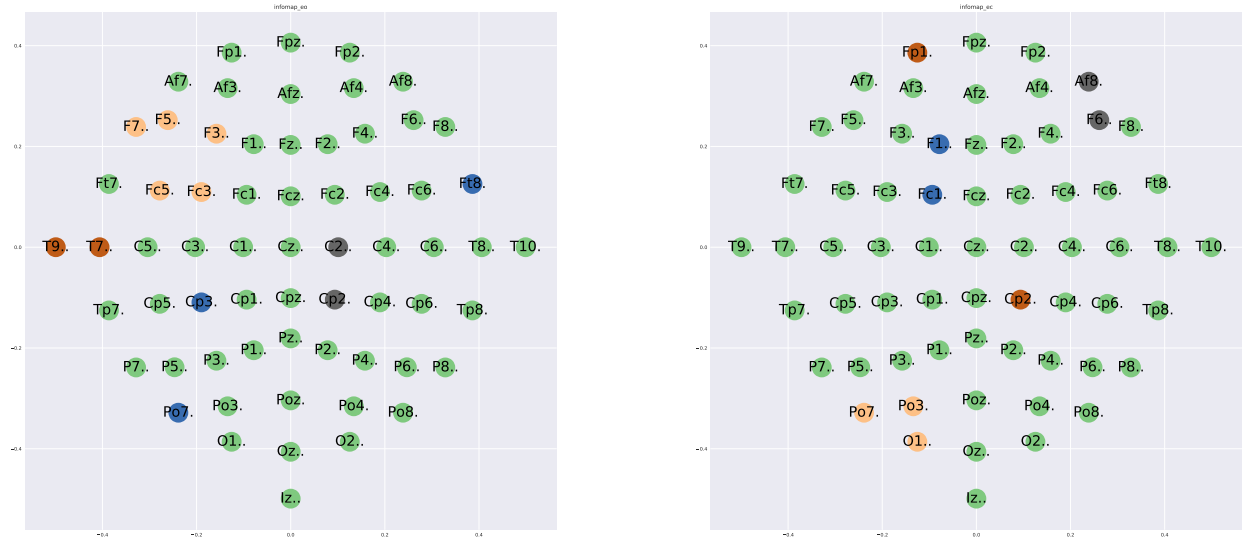
5.1-5.2-5.3

The first method is the Louvain algorithms [8] [14]: The algorithm finds high modularity partitions of large networks in short time and that unfolds a complete hierarchical community structure for the network, thereby giving access to different resolutions of community detection. The second method that we apply is the Map Equation [10] method: the algorithm exploits local interactions considering a flow of information inside the network. Topological graphical representations with the results for the two methods and the two rest states are plotted in Fig. 23 and Fig. 24



(a) eyes-open (b) eyes-closed

Fig. 23: louvain



(a) eyes-open (b) eyes-closed

Fig. 24: infomap

CONCLUSION

In this work we developed a complete pipeline to analyze a signal; starting at low level with the spectral analysis, we abstracted, considering connectivity and network properties. We obtained useful informations and insights on cerebral activity, and performed different typologies of clustering and visualizations. Finally, we extracted post-processed data to quantify our work. In Table VI there is the list of tasks completed. All the code to generate tables and plots is into the S050.ipynb notebook.

TABLE V: sample motif4 EO-EC

	frequency	statistical significance
14	45	under-represented
28	105	under-represented
30	6	under-represented
74	296	under-represented
76	4105	under-represented
78	145	under-represented
90	2	under-represented
92	182	under-represented
94	15	normal-represented
204	466	under-represented
206	47	under-represented
222	7	over-represented
280	2513	normal-represented
282	208	under-represented
286	1	under-represented
328	376	normal-represented
330	4	normal-represented
332	219	under-represented
334	5	under-represented
344	241	normal-represented
346	7	under-represented
348	278	normal-represented
350	16	normal-represented
390	28	over-represented
392	6370	over-represented
394	269	over-represented
396	458	over-represented
398	11	normal-represented
404	367	under-represented
406	7	under-represented
408	3716	over-represented
410	374	over-represented
412	195	under-represented
414	11	normal-represented
454	7	normal-represented
456	125	under-represented
458	3	normal-represented
460	140	under-represented
462	4	under-represented
468	61	under-represented
470	3	under-represented
472	62	under-represented
474	13	normal-represented
476	98	under-represented
478	13	normal-represented
856	6	normal-represented
858	15	normal-represented
862	14	normal-represented
904	106	normal-represented
906	193	over-represented
908	4	under-represented
910	7	normal-represented
922	223	over-represented
924	8	normal-represented
926	22	over-represented
972	7	normal-represented
974	4	normal-represented
990	6	normal-represented
2184	85650	under-represented
2186	25201	normal-represented
2190	415	under-represented
2202	999	over-represented
2204	967	under-represented
2206	50	normal-represented
2252	8408	normal-represented
2254	710	normal-represented
2270	29	over-represented
2458	1164	normal-represented
2462	21	under-represented

	frequency	statistical significance
14	106	under-represented
28	412	over-represented
30	21	under-represented
74	648	normal-represented
76	8665	under-represented
78	607	normal-represented
90	16	normal-represented
92	764	over-represented
94	81	over-represented
204	1075	under-represented
206	320	over-represented
222	52	over-represented
280	3389	normal-represented
282	332	under-represented
286	7	under-represented
328	422	normal-represented
330	14	normal-represented
332	481	under-represented
334	43	normal-represented
344	617	over-represented
346	18	normal-represented
348	446	normal-represented
350	30	normal-represented
390	24	normal-represented
392	6908	under-represented
394	327	normal-represented
396	283	under-represented
398	7	under-represented
404	808	over-represented
406	54	normal-represented
408	6012	normal-represented
410	528	over-represented
412	152	under-represented
414	11	under-represented
454	30	normal-represented
456	195	under-represented
458	3	under-represented
460	161	under-represented
462	11	normal-represented
468	278	under-represented
470	45	under-represented
472	311	normal-represented
474	22	over-represented
476	215	under-represented
478	25	normal-represented
856	49	over-represented
858	51	over-represented
862	18	under-represented
904	156	under-represented
906	384	over-represented
908	7	under-represented
910	36	over-represented
922	254	over-represented
924	9	normal-represented
926	2	under-represented
972	2	under-represented
974	2	under-represented
990	22	over-represented
2184	75485	under-represented
2186	29300	normal-represented
2190	623	under-represented
2202	1489	over-represented
2204	1111	under-represented
2206	66	under-represented
2252	12279	under-represented
2254	1876	over-represented
2270	140	over-represented
2458	1991	over-represented
2462	54	under-represented

TABLE VI: Tasks Completed

Tasks	Class
1.1	Mandatory
1.2	B
1.3	C
1.4	D
1.5	E
2.1	Mandatory
2.2	A
2.3	A
2.4	D
2.5	C
3.1	Mandatory
3.2	D
3.3	B
3.4	C
3.5	B
3.7	C
4.1	Mandatory
4.2	C
4.4	E
5.1	Mandatory
5.2	B
5.4	C

REFERENCES

[1] Kaminski, M. J., and Katarzyna J. Blinowska. "A new method of the description of the information flow in the brain structures." Biological cybernetics 65.3 (1991): 203-210.

[2] Baccal, Luiz A., and Koichi Sameshima. "Partial directed coherence: a new concept in neural structure determination." Biological cybernetics 84.6 (2001): 463-474.

[3] P. Welch, The use of the fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms, IEEE Trans. Audio Electroacoust. vol. 15, pp. 70-73, 1967.

[4] M.S. Bartlett, Periodogram Analysis and Continuous Spectra, Biometrika, vol. 37, pp. 1-16, 1950.

[5] Wiener, Norbert. "Nonlinear prediction and dynamics." Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability. Vol. 3. Berkeley, University of California Press, 1956.

[6] Milo, Ron, et al. "Network motifs: simple building blocks of complex networks." Science 298.5594 (2002): 824-827.

[7] Nadav Kashtan, Shalev Itzkovitz, Ron Milo, Uri Alon 2002-2005.

[8] Blondel, Vincent D., et al. "Fast unfolding of communities in large networks." Journal of statistical mechanics: theory and experiment 2008.10 (2008): P10008.

[9] Bohlin, Ludvig, et al. "Community detection and visualization of networks with the map equation framework." Measuring Scholarly Impact. Springer International Publishing, 2014. 3-34.

[10] Rosvall, Martin, and Carl T. Bergstrom. "Maps of random walks on complex networks reveal community structure." Proceedings of the National Academy of Sciences 105.4 (2008): 1118-1123.

[11] Watts, Duncan J., and Steven H. Strogatz. "Collective dynamics of small-worldnetworks." nature 393.6684 (1998): 440-442.

[12] Humphries, Mark D., Kevin Gurney, and Tony J. Prescott. "The brainstem reticular formation is a small-world, not scale-free, network." Proceedings of the Royal Society of London B: Biological Sciences 273.1585 (2006): 503-511.

[13] Muldoon, Sarah Feldt, Eric W. Bridgeford, and Danielle S. Bassett. "Small-world propensity and weighted brain networks." Scientific reports 6 (2016): 22057.

[14] Pujol, Josep M., Javier Bjar, and Jordi Delgado. "Clustering algorithm for determining community structure in large networks." Physical Review E 74.1 (2006): 016107.

[15] www.met.reading.ac.uk/~sws02hs/teaching/Fourier/Fourier_2_student.pdf

[16] citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.380.4017&rep=rep1&type=pdf