Bioinformatics 2018-2019
Date: 13 January 2019
Project report

Bioinformatics@Data Science A.Y. 2018-2019

Neuroscience Application

Pochiraju Venkata Naga Sai Krishna Abhinay (1819771)

Rachuri Mani Niharika (1819748)

Group no. 7

Abstract

In this project we analyze brain signals recorded by EEG for two eye-states: Eyes-open and Eyes-closed conditions. This EEG data is recorded from 64 electrodes with the subject at rest in eyes-open and eyes-closed conditions respectively. The analysis includes Estimating the connectivity of the 64 channels and representing the networks, computing the graph theory indices, performing motif analysis, and community detection. We perform these analyses on two datasets and report their comparison.

Introduction

The main goal of this project is to understand brain functionality during resting states. We use the EEG data from PhysioNet, "EEG Motor Movement/Imagery dataset" at https://physionet.org/physiobank/database/eegmmidb/ . We perform various analysis about and related to: Connectivity graphs, Graph Theory Indices, Motif Analyses, and Community Detection. There are 4 mandatory tasks for the four analyses, and several optional analyses. The list of all the tasks that we have performed are shown in Table 0.

Dataset

The data is from PhysioNet: EEG Motor Movement/Imagery dataset. This overall dataset consists of over 1500 one- and two-minute EEG recordings acquired from 109 subjects, each containing 14 runs (files) of acquisition, out of which, we work on the first two experimental runs: R01 (recorded during eyes-open resting state), and R02 (recorded during eyes-closed resting state). Our group was assigned the subject S064. So, the two files we are using are S064R01 and S064R02. The files are provided in EDF format (European Data Format), which includes metadata, among which the sampling frequency and the channel labels. We use the python library 'pyedflib' to read the edf files into python. To understand the format of the dataset we have converted the edf files to csv format and understand it. Both the files have the same format.

Connectivity graph

1.1 (Mandatory) Estimating functional brain connectivity using Direct Transfer Function (DTF)

Out of the given MVAR estimators, DTF – Direct Transfer Function was used to estimate the functional brain connectivity. The python module *'connectivipy'* was used to perform this analysis. The inbuilt function *'order_akaike()'* was used to find the best fit model using Vieira-Morf algorithm, and then fit this best fit parameters by using the Yule Walker algorithm to fit the mvar model (the plots for this best fit are shown in figure 1). And then, using the *'DTF()'* to calculate the Direct Transfer Function from the data. Then a threshold is applied so that the resulting binary connectivity matrices have network density equal to 20%. The graphical representation of the binary adjacency matrices for both the cases (eyes-open and eyes-closed) are shown in figure 2.

1.2 (Class A) Performing the same task as above for different density values

Threshold is applied so that the resulting binary connectivity matrices have density values of 1%, 5%, 10%, 20%, 30% and 50%. The binary adjacency matrices for all these density values are shown in figure 3.

1.5 (Class C) Topological representation of the networks

The channel locations were provided by the instructor to represent the networks. But there was problem with the values provided, so wrote a small script in R which uses the library *'eegkit',* which has the coordinates for 87 channels. We write these coordinates into a csv file from R (the script for this process in is the file: *eegCoordfromR.R*). Then using some basic excel functions like index and match, we get the coordinates for our 64 channels and save it in a text file (channel_locations.txt). Depending on these coordinates, we make the topographical representation of the networks, which are shown in figure 4.

Graph theory indices

2.1 (Mandatory) Computing Binary Global and Local graph indices

We compute the binary global indices – Clustering Coefficient and Path Length; and local graph indices - Degree, in-degree, out-degree for our networks. For the global indices, we use the inbuilt functions: *average_clustering() and average_shortest_path_length()* from the network library. In case of local indices we wrote a small function which returns the Degree, In-degree and out-degree of the graph passed. The list of highest 10 channels for local indices are shown in table 1. The Average Clustering Coefficient for Eyes Open Graph is :0.7370509093949917 and Average Shortest Path Length for Eyes Open Graph is :1.6364087301587302. The Average Clustering Coefficient for Eyes Closed Graph is :0.7160179866083559 and the Average Shortest Path Length for Eyes Closed Graph is :1.6369047619047619.

2.2 (Class D) Small-Worldness Index

A small-world network is defined to be a network where the typical distance L between two randomly chosen nodes grows proportionally to the logarithm of the number of nodes N in the network

$$L \propto \log N$$

First, we check the conditions to call a network as a small-world network, ($\lambda \approx 1$ and $\gamma > 1$). These two conditions were satisfied. Then we used the R library *qgraph* to calculate the smallworldness index.

To perform this, we used the python library *rpy2* to import the R library and calculate the Index. The results for this analysis are as follows:
transitivity = 0.498335; transitivity_random = 0.197021
APL = 1.616667; APL_random = 1.913198
And finally, the Small-worldness Index for our network was calculated to be 2.993280

2.4 (Class C) Behavior of Global Graph Indices in function of network density

We analyze the behavior of Global Graph indices (both Clustering Coefficient and Path Length) and compare the results for both the cases eyes-open and eyes-closed.
For Clustering Coefficient, the behavior is plotted and shown in Figure 5. We can observe that the clustering coefficient is increasing as the density increases.
In case of Path Length, the behavior is plotted and shown in Figure 6. The average path length is decreasing as the density increases.

Motif Analysis

3.1 (Mandatory) Motif Analysis – 3 Node configurations

We perform Motif Analysis to investigate the presence of 3 node subgraphs in the networks for both eyes-open and eyes-closed cases. The graphs obtained in the previous question for density 20% was used to perform this. For this task, we utilize the software *mfinder*. First, the edges of the respective graphs for eyes-open and eyes-closed are written in a text file which will be the input for mfinder. Then, we use the library *subprocess* to run the .exe file of mfinder with the corresponding parameters and the input file. The output from mfinder will result in text file, which we copy into a csv file and then again import into python. We format this data and store it in a pandas dataframe, which are shown in tables 2 and 3 for eyes-open and eyes-closed cases respectively. The parameters that we used in this case are: *[mfinder.exe,input txt file, -s (motif size), size, -r, number of random networks to generate, number, -f, output filename]*

3.2 (Class C) Motif Analysis – with pattern A -> B <- C and topological representation

For this task, we are supposed to consider only the configuration A -> B <- C only. This configuration is described in the mfinder manual (motif dictionary) as id36 under 3-node subgraphs. So, the process is similar to the previous part, but we add an extra parameter here, i.e., -ospmem (which outputs members of specific subgraph only). The topological representation of the networks considering only the connections involved in this configuration are shown in figure 7.

3.3 (Class C) Motif Analysis for a channel in parietooccipital scalp region

We found a research paper (https://www.researchgate.net/figure/Schematic-display-of-EEG-electrode-positions-For-statistical-analyses-ERS-ERD-was_fig2_233539681) mentioning channels in parietooccipital region as (PO; PO7, PO3, O1). So, we considered the channel **Po3**. The numeric id for Po3 is 56. First, we execute the same command as before considering only the channel Po3 and we get the output from mfinder – OUT and Members. Then read the Members file and add the edges for the networkx graph. The graphs are plotted and shown in figure 8 for both the cases. The total number of nodes and edges for eyes-open case were observed to be 49 and 473 respectively, where as for the eyes-closed case they were 42 and 333 respectively.

3.4 (Class E) Motif Analysis – 4 Node configurations

We perform Motif Analysis to investigate the presence of 4 node subgraphs in the networks for both eyes-open and eyes-closed cases. We perform the same process as of 3.1 with a slight change in the passed parameters, where we change the motif size from 3 to 4. The results are tabulated and shown in tables 4 and 5 for eyes-open and eyes-closed cases respectively.

Community Detection

4.1 (Mandatory) Community Detection – Louvain clustering

The python library community was used to perform the best-partition algorithm for Louvain Clustering. In the case of eyes-open, the number of communities detected by the algorithm were 4 with 21, 16, 8, and 19 elements in each community, where as in the case of eyes-closed the number of communities were 3 with 15, 27, and 22 elements in each community respectively. The results have also been saved as csv files with community number and the case respectively.

4.3 (Class C) Community Detection – Comparison of various methods

Using the library *igraph* in python, various algorithms for community detection were analyzed and compared. The results are shown in table 6. The various algorithms applied are:

- Community infomap
- Label propagation
- Leading eigen vector
- Spinglass
- Community multilevel

These algorithms have been applied for eyes-open and eyes-closed cases and the comparisons are analyzed.

TABLES:

Table 0. List of Tasks chosen

S. No	Question	Class
1	1.1	Mandatory
2	1.3	A
3	1.5	C
4	2.1	Mandatory
5	2.2	D
6	2.4	C
7	3.1	Mandatory
8	3.2	C
9	3.3	C
10	3.4	E
11	4.1	Mandatory
12	4.3	C

Table 1. Top 10 channels – eyes open (left) and eyes closed (right)

node	in-degree	out-degree	Degree
F3	63	11	74
F5	62	12	74
Cp4	56	13	69
Fz	56	12	68
F1	55	12	67
Cpz	54	12	66
Cp3	48	12	60
Fc2	40	13	53
Po3	43	10	53
Cp1	32	14	46

node	in-degree	out-degree	degree
F5	55	12	67
Poz	54	12	66
Cpz	48	17	65
F3	56	9	65
Pz	48	13	61
O1	56	4	60
Fz	47	7	54
Cz	38	15	53
Iz	36	12	48
P1	32	16	48

Table 2. Motif Analysis – eyes open – 3 nodes

MOTIF ID	Fre-quency	Interval	z score	p - value	lower-Limit	upper-Limit	Statistical Significance
6	492	632.8+-28.6	-4.92	1	604.2	661.4	['ANTI-MOTIF']
12	799	825.6+-27.7	-0.96	0.855	797.9	853.3	['NA']
14	50	72.8+-10.2	-2.24	0.99	62.6	83	['ANTI-MOTIF']
36	8816	8911.8+-35.4	-2.7	0.995	8876.4	8947.2	['ANTI-MOTIF']
38	1711	1645.8+-33.9	1.92	0.015	1611.9	1679.7	['MOTIF']
46	80	49.3+-5.0	6.16	0	44.3	54.3	['MOTIF']
74	1411	1523.8+-33.6	-3.36	1	1490.2	1557.4	['ANTI-MOTIF']
78	32	78.4+-8.2	-5.69	1	70.2	86.6	['ANTI-MOTIF']
98	14	18.6+-5.0	-0.94	0.845	13.6	23.6	['NA']
102	96	120.7+-7.9	-3.13	1	112.8	128.6	['ANTI-MOTIF']
108	1628	1552.3+-19.2	3.94	0	1533.1	1571.5	['MOTIF']
110	211	224.9+-11.2	-1.24	0.92	213.7	236.1	['ANTI-MOTIF']
238	101	80.9+-4.0	5.06	0	76.9	84.9	['MOTIF']

Table 3. Motif Analysis – eyes closed – 3 nodes

MOTIF ID	Fre-quency	Interval	z score	p - value	lower-Limit	upper-Limit	Statistical Significance
6	512	658.4+-32.2	-4.54	1	626.2	690.6	['ANTI-MOTIF']
12	1177	1189.8+-28.9	-0.44	0.675	1160.9	1218.7	['NA']
14	98	144.3+-14.0	-3.3	1	130.3	158.3	['ANTI-MOTIF']
36	7797	7923.8+-41.1	-3.08	1	7882.7	7964.9	['ANTI-MOTIF']
38	2038	1939.2+-36.7	2.69	0	1902.5	1975.9	['MOTIF']
46	148	120.0+-9.0	3.12	0.005	111	129	['MOTIF']
74	1519	1604.7+-31.8	-2.7	1	1572.9	1636.5	['ANTI-MOTIF']
78	59	91.7+-8.7	-3.76	1	83	100.4	['ANTI-MOTIF']
98	17	39.0+-6.2	-3.57	1	32.8	45.2	['ANTI-MOTIF']
102	138	158.0+-11.0	-1.82	0.965	147	169	['ANTI-MOTIF']
108	1135	1087.3+-18.1	2.63	0.005	1069.2	1105.4	['MOTIF']
110	226	215.7+-11.7	0.88	0.175	204	227.4	['NA']
238	67	59.5+-3.8	1.95	0.035	55.7	63.3	['MOTIF']

Table 4. Motif Analysis – eyes open – 4 nodes (first 10 rows)

MOTIF ID	Fre-quency	Interval	z score	p - value	lower-Limit	upper-Limit	Statistical Significance
14	169	251.9+-29.0	-2.86	1	222.9	280.9	['ANTI-MOTIF']
28	135	354.9+-60.5	-3.63	1	294.4	415.4	['ANTI-MOTIF']
30	4	19.4+-7.1	-2.18	0.995	12.3	26.5	['ANTI-MOTIF']
74	912	979.4+-88.5	-0.76	0.77	890.9	1067.9	['NA']
76	10724	12849.1+-477.9	-4.45	1	12371.2	13327	['ANTI-MOTIF']
78	816	889.3+-56.8	-1.29	0.92	832.5	946.1	['ANTI-MOTIF']
90	9	17.1+-7.2	-1.13	0.88	9.9	24.3	['ANTI-MOTIF']
92	377	643.1+-81.3	-3.27	0.995	561.8	724.4	['ANTI-MOTIF']
94	27	20.3+-6.5	1.03	0.195	13.8	26.8	['MOTIF']
204	1467	2364.4+-358.3	-2.5	1	2006.1	2722.7	['ANTI-MOTIF']

Table 5. Motif Analysis – eyes closed – 4 nodes (first 10 rows)

MOTIF ID	Fre-quency	Interval	z score	p - value	lower-Limit	upper-Limit	Statistical Significance
14	121	207.8+-28.2	-3.08	1	179.6	236	['ANTI-MOTIF']
28	166	508.0+-69.7	-4.9	1	438.3	577.7	['ANTI-MOTIF']
30	7	30.5+-10.1	-2.33	0.995	20.4	40.6	['ANTI-MOTIF']
74	833	1052.3+-84.2	-2.6	0.99	968.1	1136.5	['ANTI-MOTIF']
76	10054	12493.8+-552.0	-4.42	1	11941.8	13045.8	['ANTI-MOTIF']
78	693	877.9+-70.1	-2.64	0.99	807.8	948	['ANTI-MOTIF']
90	5	26.8+-9.0	-2.42	0.995	17.8	35.8	['ANTI-MOTIF']
92	503	811.5+-100.3	-3.08	1	711.2	911.8	['ANTI-MOTIF']
94	28	42.6+-10.5	-1.39	0.93	32.1	53.1	['ANTI-MOTIF']
204	1567	2372.0+-451.0	-1.79	0.97	1921	2823	['ANTI-MOTIF']

Table 6. Various algorithms comparison

Dataset	Algorithm	Num of Communities
Eyes – open	Louvain Clustering	4
Eyes – open	Community Infomap	1
Eyes – open	Label propagation	29
Eyes – open	Leading Eigen vector	3
Eyes – open	Spinglass	3
Eyes – open	Community Multilevel	3
Eyes – closed	Louvain Clustering	3
Eyes – closed	Community Infomap	1
Eyes – closed	Label propagation	24
Eyes – closed	Leading Eigen vector	4
Eyes – closed	Spinglass	3
Eyes – closed	Community Multilevel	4

FIGURES:

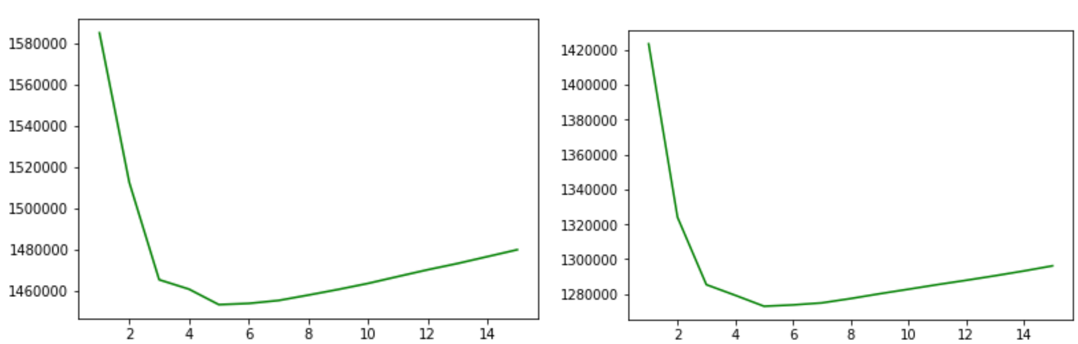


Figure 1. Best Fit for eyes-open (left) and eyes-closed (right)

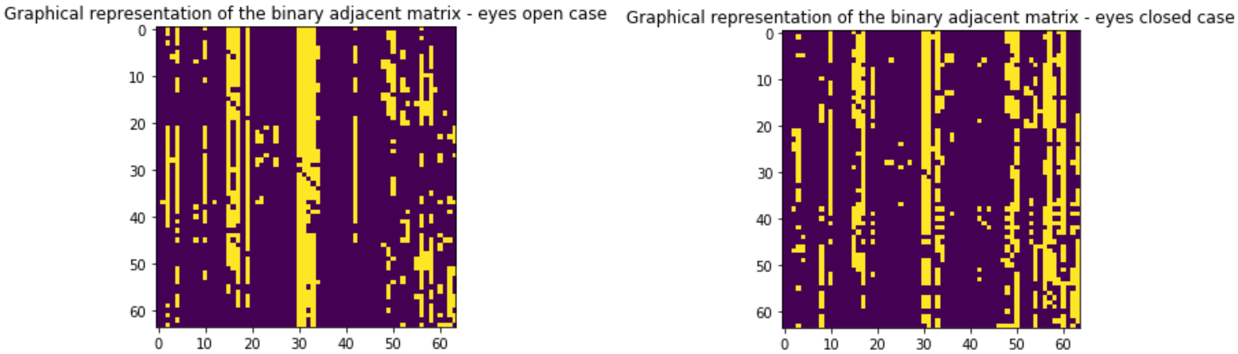


Figure 2. Binary adjacency matrices for eyes-open (left) and eyes-closed (right)

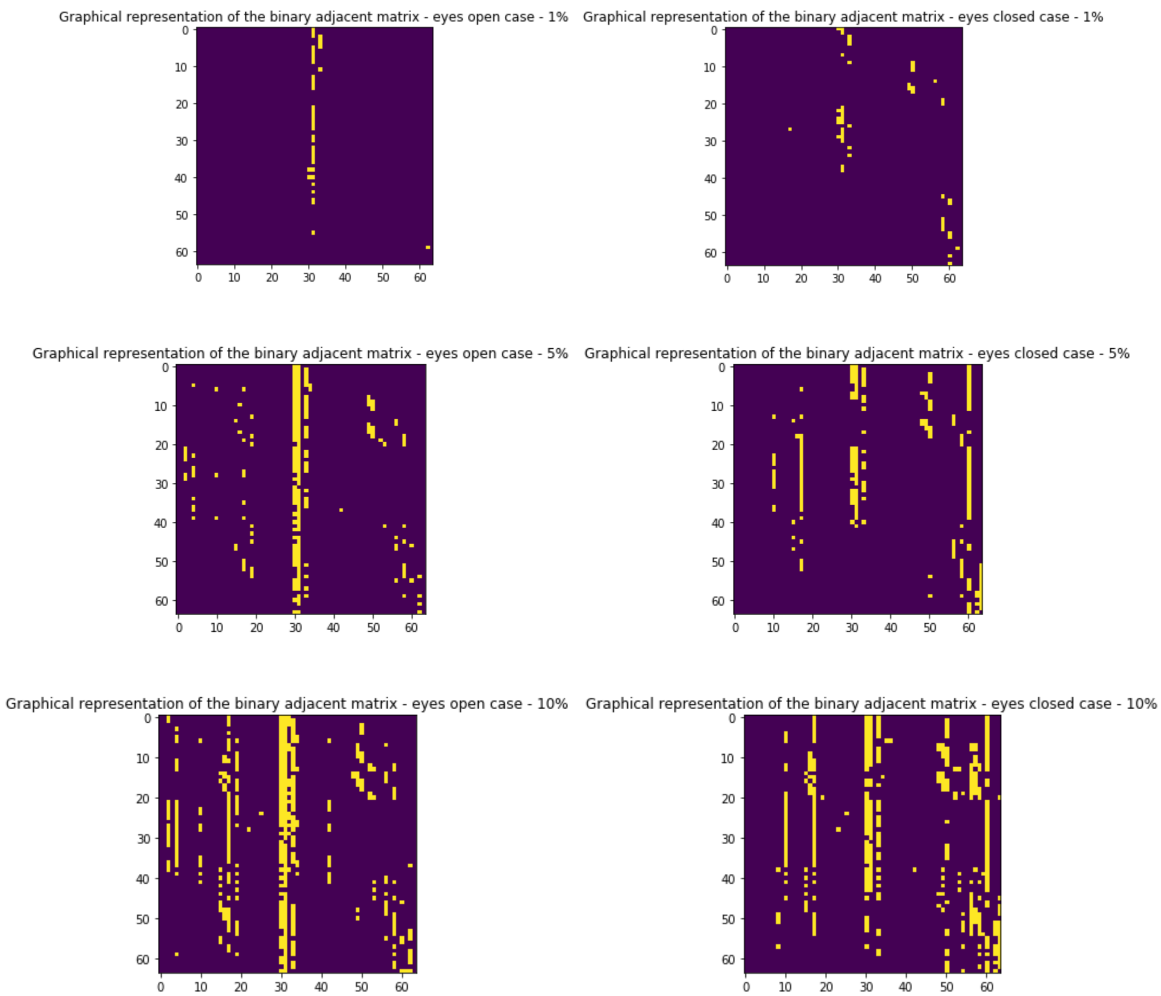


Figure 3a. Binary Adjacency matrices for various densities: 1%, 5%, 10%

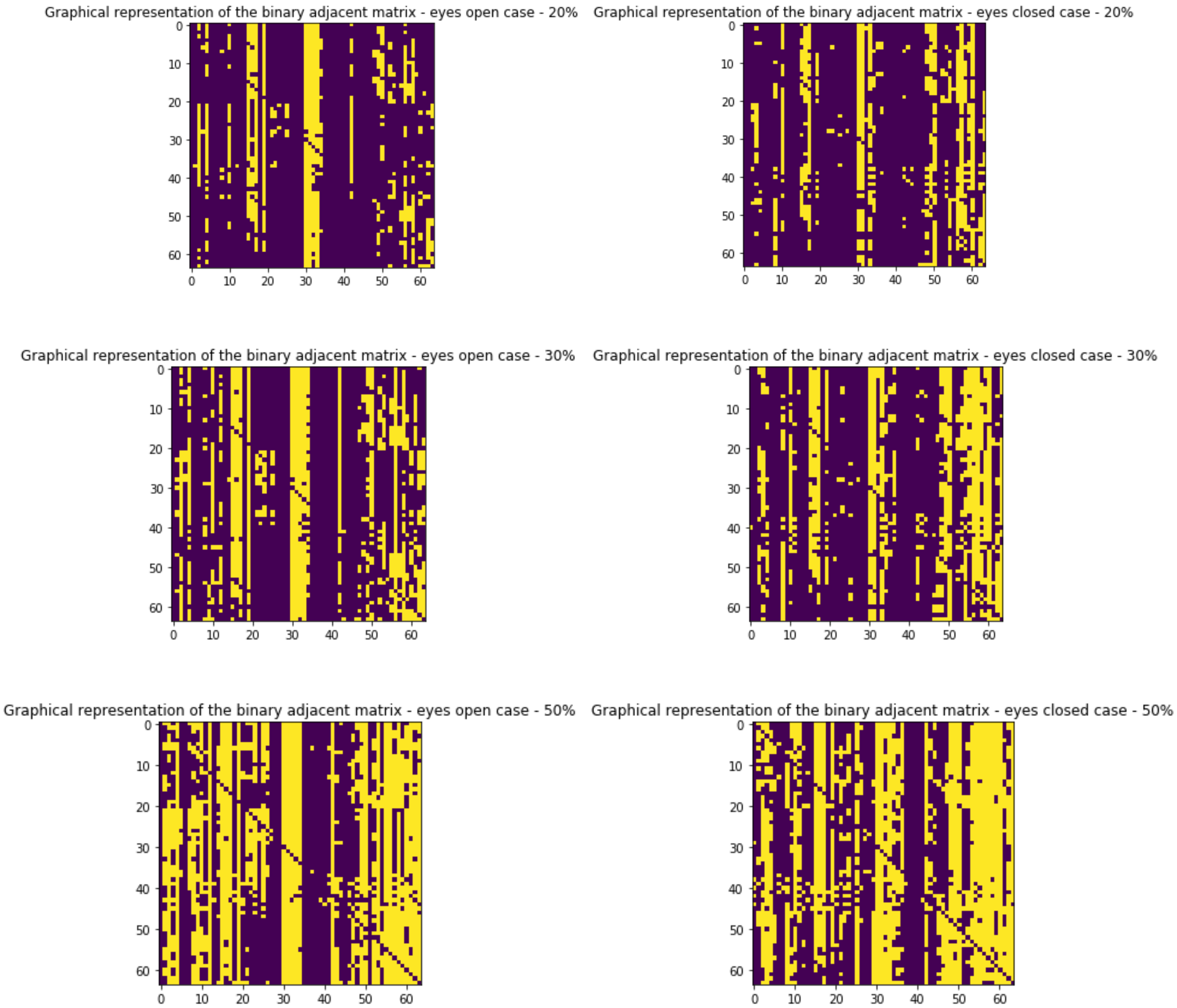


Figure 3b. Binary Adjacency matrices for various densities: 20%, 30%, 50%

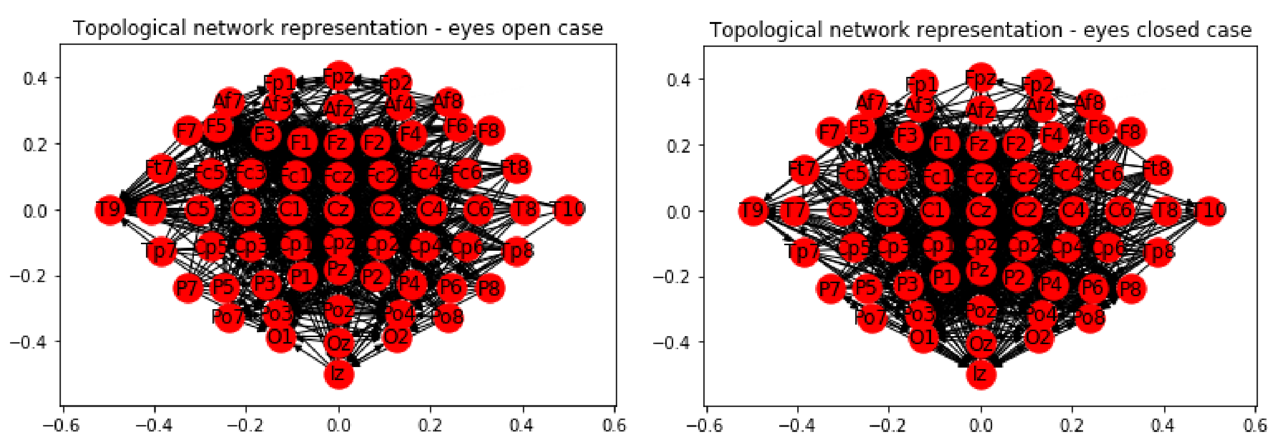


Figure 4. Topological Network representation for eyes – open (left) and eyes – closed (right)

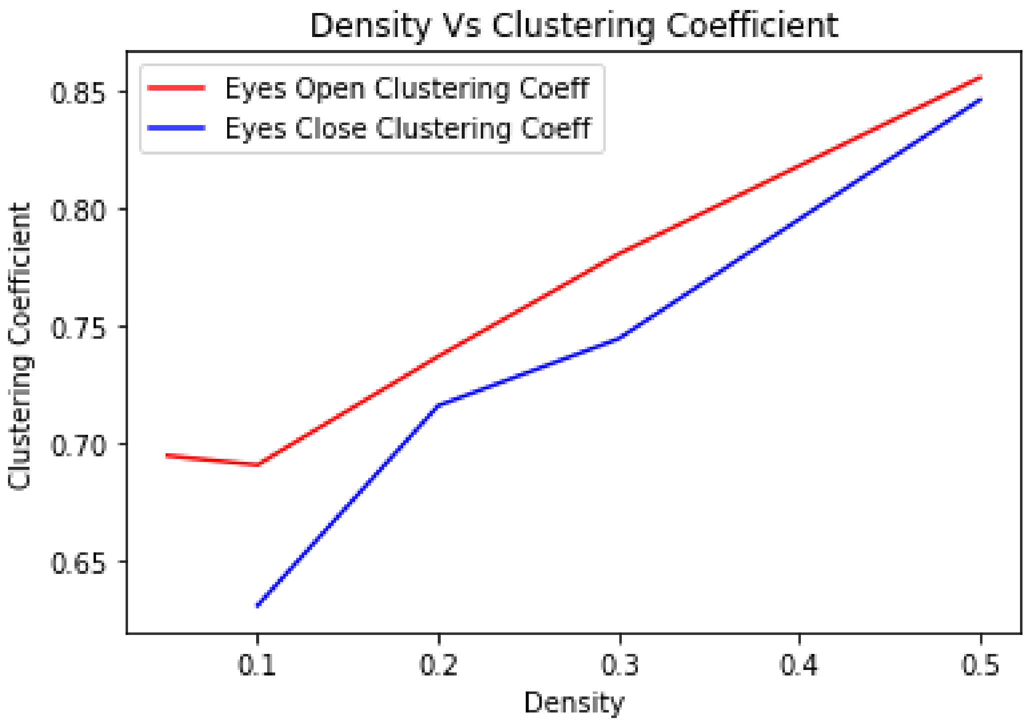


Figure 5. Clustering Coefficient vs Density

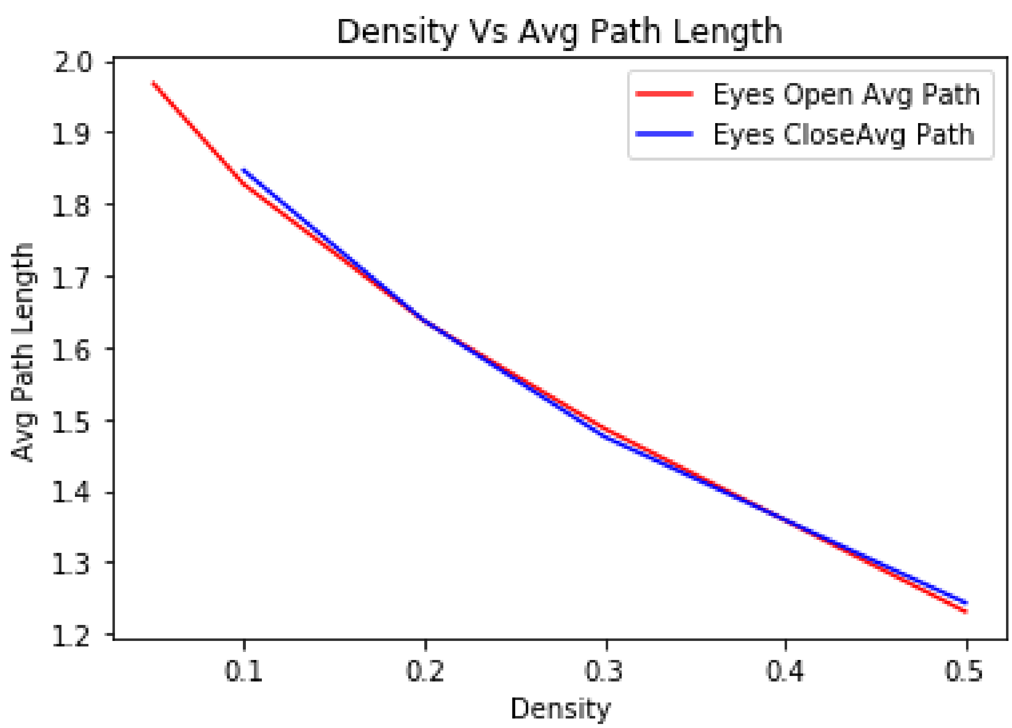
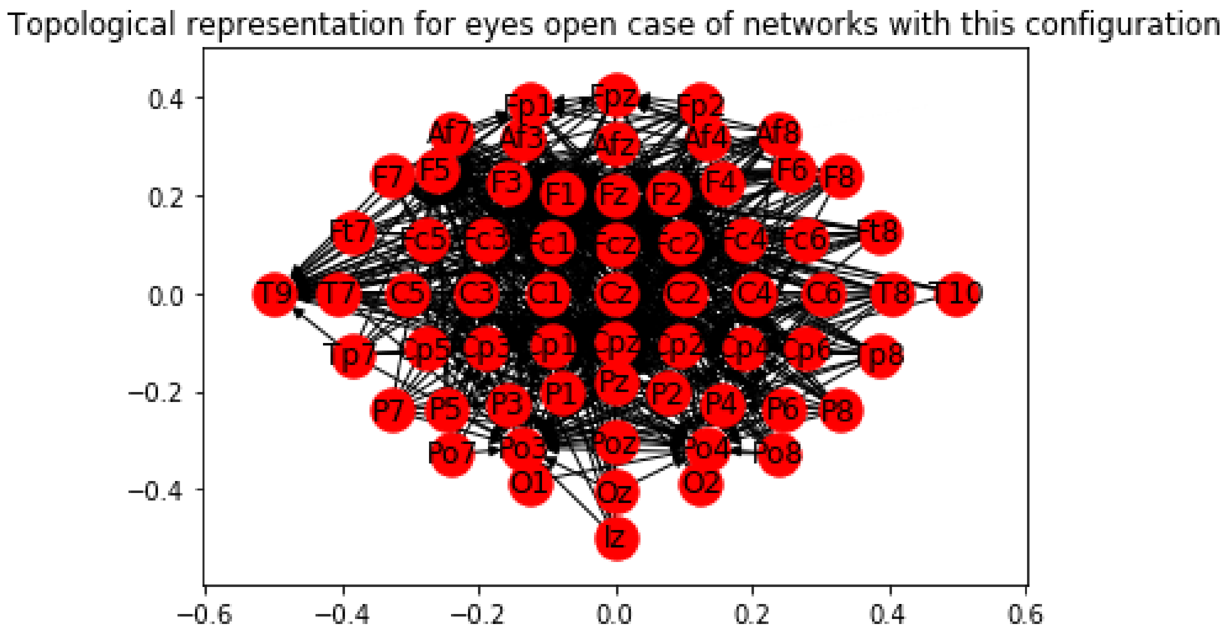


Figure 6. Path Length vs Density



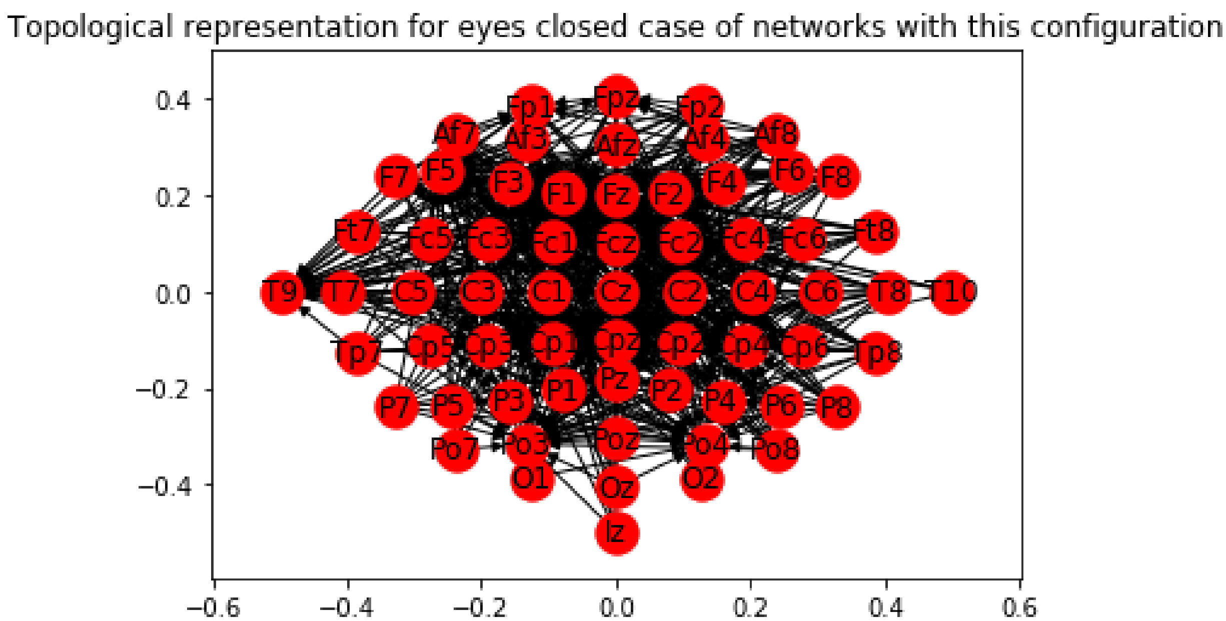
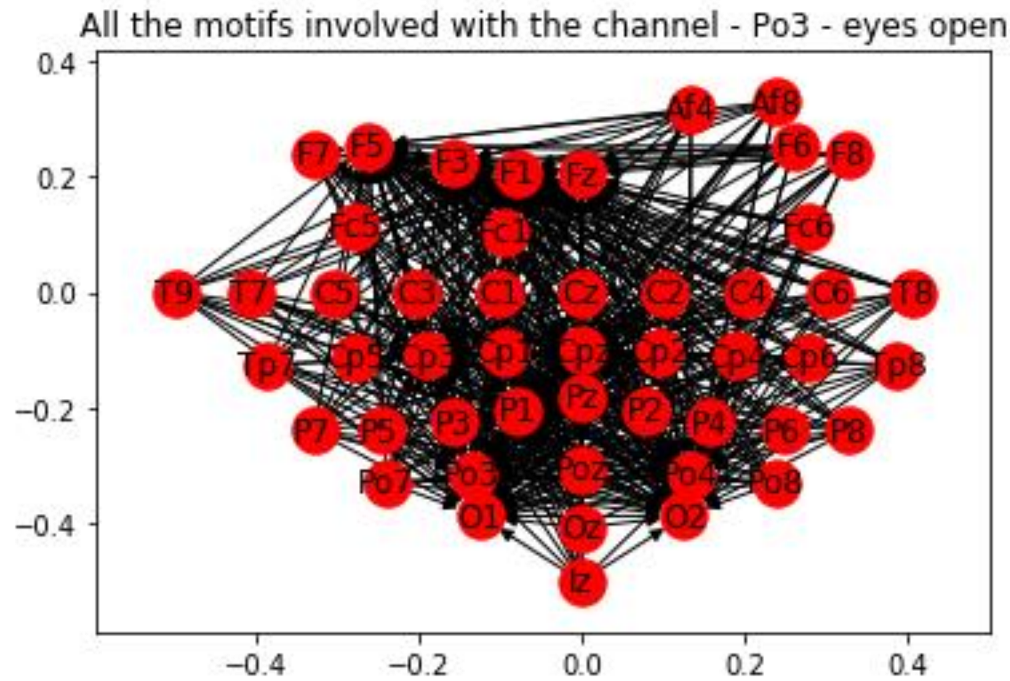


Figure 7. Topological representation of networks with the configuration A -> B <- C for eyes-open (top) and eyes-closed (bottom)



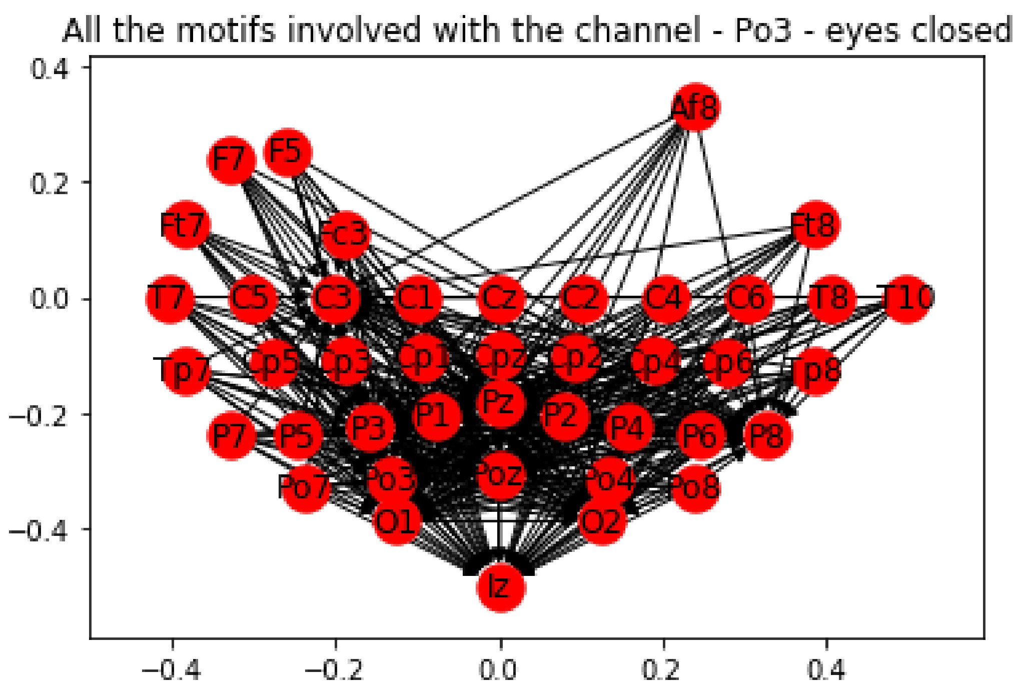


Figure 8.Motif Analysis - Topological representation of the networks involved only with the channel (Po3) in parietooccipital region for eyes-open (top) and eyes-closed (bottom)

References

Epskamp, S., Costantini, G., Haslbeck, J., Cramer, A. O., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2018, April 25). Smallworldness. Retrieved from https://rdrr.io/cran/qgraph/man/smallworldness.html

Introduction to rpy2¶. (n.d.). Retrieved from https://rpy2.readthedocs.io/en/version_2.8.x/introduction.html

Small-world network. (2018, November 17). Retrieved from https://en.wikipedia.org/wiki/Small-world_network

Yildmar, Ü Y. (n.d.). How to test statistically whether my network (graph) is a "small-world" network or not? Retrieved from https://stats.stackexchange.com/a/211047

Examples¶. (n.d.). Retrieved from https://connectivipy.readthedocs.io/en/latest/tutorial.html

Toppi, J., Fallani, F. D., Vecchiato, G., Maglione, A. G., Cincotti, F., Mattia, D., . . . Astolfi, L. (2012). How the Statistical Validation of Functional Connectivity Patterns Can Prevent Erroneous Definition of Small-World Properties of a Brain Connectivity Network. *Computational and Mathematical Methods in Medicine,2012*, 1-13. doi:10.1155/2012/130985

Methods for the estimation of the cortical activity and connectivity during cognitive tasks in humans(Unpublished master's thesis). Thesis / Dissertation ETD. (2013).

Brain connectivity estimators. (2018, December 22). Retrieved from https://en.wikipedia.org/wiki/Brain_connectivity_estimators

Schematic Display of EEG electrode positions for Statistical Analyses. Retrieved from https://www.researchgate.net/figure/Schematic-display-of-EEG-electrode-positions-For-statistical-analyses-ERS-ERD-was_fig2_233539681