Bioinformatics - Project 1
Neuroscience application
Brain network study during resting states

Peikova Kateryna, Sofianos Theodoros, Zavadskaya Katsiaryna
Group 10

January 2, 2020

Abstract

The aim of the project is to analyze two datasets of EEG data. The datasets contain EEG data recorded from 64 electrodes with subject S071 being in two resting states: (i) eyes-open and (ii) eyes-closed. In order to complete project, number of tasks were performed using Python programming language. Full list of chosen task can be seen in table 1.

1 Connectivity

1.1 Overview

In the first part of the project, we investigated and analysed the functional brain connectivity of the subject S071. We used data from the "EEG Motor Movement/Imagery Dataset". For the project, we analysed two-minute Electroencephalography (EEG) recordings for the eyes-open (EO) and eyes-closed (EC) states. We used two different Multivariate Autoregressive Models (MVAR) estimation methods as well as experimented with various thresholds to obtain a full view of how connectivity can be estimated. To support our results we provide a topographical representation of obtained brain networks and heatmaps of adjacency matrices.

1.2 Estimation

To build connectivity graphs we used MVAR based on two estimators Partial Directed Coherence (PDC) and Direct Transfer Function (DTF). We noticed, that even though these two methods differ, the resulting binary adjacency matrices remained very similar (Figure 2). To set the order of models we performed the estimation of optimal order on the EEG recordings. Result of PDC or DTF is a matrix with significance values that correspond to connection strength between each pair of nodes in the subject brain. To obtain connectivity graphs we applied different thresholds to transform matrices to binary adjacency matrices.

1.3 Threshold adjustment

The choice of the threshold influences how dense the network is. We manually picked thresholds that would lead to required densities of $1\%, 5\%, 10\%, 20\%, 30\%$ and 50%. We noticed that higher threshold results in a lower density. The significance threshold values are presented on the Table 2. We noticed that higher threshold results in a lower density. Moreover, EC recording requires smaller threshold values to achieve the same density compared to EO state. Based on this, we assume that the eyes-open state expresses connectivity more, while there is less confidence in obtained connections in eyes-closed case.

1.4 Node subset

Additionally, we analysed the given subset of 19 channels and repeated connectivity estimation using PDC. We used a resampling method with $p<5\%$ to filter out non-significant connections. Significant values were transformed into binary adjacency matrix using threshold 0.121 for EO and 0.09 for EC, that leads to 20% density of the network. We give results for both recordings in a topographical representation of the network in Cartesian coordinates on Figure 1. We noticed that EO network has higher connection degree for the pre-frontal and frontal area, while for EC network connectivity distributed between pre-frontal, frontal and parietal areas.

2 Graph theory indices

We calculated global and local network indices to represent network structure. These graph properties allow comparison of networks. In this work, we used four indices. Clustering coefficient and Average path length are global indices while Indegree and Outdegree are local indices.

2.1 Global indices

The clustering coefficients of the eyes-open run for PDC and DTF are 0.33 and 0.26 respectively. The average path length indices are 0.23 and 0.21. Regarding eyes-closed recordings, values of global indices do not differ a lot between PDC and DTF.

We observed a small difference in global indices between methods of connectivity estimation. Furthermore, with a comparison of two runs, we noticed that average path length differs a lot between eyes-open and eyes-closed states.

2.2 Local indices

Regarding local indices, EO run and EC run vary a lot. Based on these results we assume that there are different regions important for connectivity in EO and EC states. In other words, when eyes are closed different parts of the brain are active compared to the eyes-open state.

Table 3 What is more, sets that represent top connective nodes in EO and EC do not overlap.

2.3 Weighted indices

Additionally, we calculated weighted versions of global indices for both recordings (Table 4). We treated the significance of a connection between channels as the weight of an edge. As EC connectivity estimation resulted in lower significance values we observe lower average path length compared to EO state. The reserve effect is noticeable in clustering coefficient values. Though the difference in average path lengths is not significant, clustering coefficient values show that eyes-closed network tends to have more distinct clusters than eyes-open.

3 Motif analysis

For this part, the mfinder software tool was used to track down the 3-node and 4-node motifs and anti-motifs. The threshold values to determine if a motif is statistically significant were picked to be: $Z - score > 2$, $mfactor > 1$, $uniqueness \geq 4$, propositional threshold $D = 0.1$ and p value is ignored, since the number of random created graphs is relatively small(100). Practically, all the values are the default ones,recommended by Milo et al (Science, 2002). The only parameter changed is the mfactor, from 1.1 to 1. For the variable mfactor, there is not explanation on the mfinder documentation as to what this parameter is. Perhaps it is the propotional threshold D, to define the minimum difference between f_{original} and f_{random}. For its default value, we cant get any significant motifs. However, for values from 0 to 1, we get exactly the same motifs, so most probably it is not the variable D. We decided to set its value as 1. Also, since the software of mfinder is able to find only motifs and not anti-motifs, the output of the software will be again used as input in Python, in order to apply the formula for anti-motifs recognition. f_{random}-$f_{\text{original}} > D \times f_{\text{random}}$

3.1 Overview

For the first dataset (eyes-open), the motif found that exceeds the threshold values is the one with id 6, with Z-score 5.56 and Uniqueness 17. For anti-motifs, we found no signficant ones, for D = 0.1 For the second dataset (eyes-closed), the motif with id 6 is also exceeding the threshold values, with Z-score=8.48 and Uniqueness=19. Again, we found no statistical significant anti-motifs. We can claim that the same motif, with id6 ,is appearing in both datasets more times than the average appearance in the 100 random graphs and it is statistically significant in both cases, in terms of Z-score and Uniqueness. In the second dataset, the motif is even 'stronger', since it has quite higher Z-score and number of uniqueness than the first dataset.

3.2 Topographic representation of configuration

The motif we are looking for is the motif with id number 36 in the mfinder Dictionary. By observing the graphs in the appendix, we see that the graph of dataset 1 includes 58 channels and the second graph includes all the 64 channels. By looking more closely, we see that the channels not included in the graph of the 1st dataset are the most frontal ones, in the topographic representation. More concretely, the frontal channels Fp1,Fpz,Fp2, Af7 and Af8 exist only in the second graph, plus a channel from the Parietal lobe, P7. (Figure 3)

3.3 Motifs involved in channel Po7

The channel selected for this task is the Po7. By looking at the table in the appendix, we notice that the channel chosen is involved in more times in motifs in dataset 1 rather than in dataset 2. We also observe that for 7 motifs(all those with id number greater than 46), the channel in both datasets has zero involvement. The biggest difference between the two datasets in a specific motif occurs for motif id 6, where in dataset 1, the channel Po7 appears 117 times and only 79 times in dataset 2. (Table 5)

3.4 4-nodes motifs and anti-motifs

For the first dataset, the motif that exceeds the threshold values is the one with id 14, with Z-score =2.63 and Uniqueness=13.For the second dataset, we find two motifs, id 14 with Z-score= 4.48 and Uniqueness=12 and motif id 204, with Z-score=2.55 and U=12. The first thing that stands out is that dataset 2 has many more anti-motifs, compared to dataset 1. We also see that motif id 14 is statistically significant in both datasets, with higher Z-score in dataset 2. Common anti-motifs are those with id 280,344,904,906. Since dataset 2 has too many anti-motifs, we can see that the majority of anti-motifs of dataset 1 are also in 2 (4/7) (Table 6)

4 Community detection

4.1 Overview

The aim of this section is to investigate how EEG information recorded from 64 electrodes flows between different brain regions. In order to do that we are representing brain as a graph, where brain areas are nodes and signals are edges which are connecting nodes. In order to perform this task there are two main approaches: modularity-based and information theory-based.

4.2 Modularity-based approach

To determine the composition of communities using modularity-based approach, the Louvain algorithm was used. The Louvain algorithm is based on modularity maximization. The algorithm consists two steps. In the first step, the algorithm iterates over the nodes in the graph and assigns each node to a community if the assignment will lead to an increase in modularity. In the second step, the algorithm creates super-nodes out of the clusters found in the first step. This process repeats iteratively.

4.3 Information theory-based approach

In order to find optimal communities with information theory-based approach the InfoMap algorithm was used. The Infomap algorithm is based on the principles of information theory. Infomap characterizes the problem of finding the optimal clustering of a graph as the problem of finding a description of minimum information of a random walk on the graph. The algorithm maximizes an objective function called the Minimum Description Length.

4.4 Results

Both of these approaches have a quality function and then search the space of graph partitions to find the partition that optimizes that quality function. The difference between them is the quality function: InfoMap focuses on the information needed to compress the movement of the random walker, while Modularity defines modules based on edge density.

The number of optimal communities found by two methods are presented in the table 7. The visualisation of clustering into communities is shown on figures 5 and 6 for modularity and information theory-based partition, respectively; where each community is drown in different colour.

Since the thresholds applied to signals for both eyes states are relatively small, Infomap method finds only 1 optimal community, while Louvain's approach determines 5 communities in each of states.

References

[1] Scott Emmons, Stephen Kobourov, Mike Gallant, and Katy Börner, 2016. Analysis of Network Clustering Algorithms and Cluster Quality Metrics at Scale. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4938516/

[2] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, U. Alon,2002, Network Motifs: Simple Building Blocks of Complex Networks. https://www.ncbi.nlm.nih.gov/pubmed/12399590

A Tables and figures

Task	Class
1.1.	mandatory
1.2.	A
1.3.	A
1.4.	D
1.5.	C
2.1.	mandatory
2.3.	B
2.4.	C
2.7.	C
3.1.	mandatory
3.2.	C
3.3.	C
3.4.	E
4.1.	mandatory
4.2.	B
4.3.	C

Table 1: Tasks chosen for the project

Density, %	Threshold (eyes open)	Threshold (eyes closed)
1	0.268	0.15
5	0.202	0.12
10	0.159	0.105
20	0.121	0.09
30	0.1	0.08
50	0.076	0.07

Table 2: Threshold values for network density for open eyes and closed eyes runs.

Channel	Indegree	Channel	Outdegree
T9	28	Af7	63
P6	25	Af8	63
P8	24	Fp1	62
P3	23	Fpz	62
T10	21	Fp2	62
Po4	21	P7	53
Cp5	20	Af3	47
Cp6	20	Af4	47
Tp8	20	Afz	41
P1	19	Cp3	40

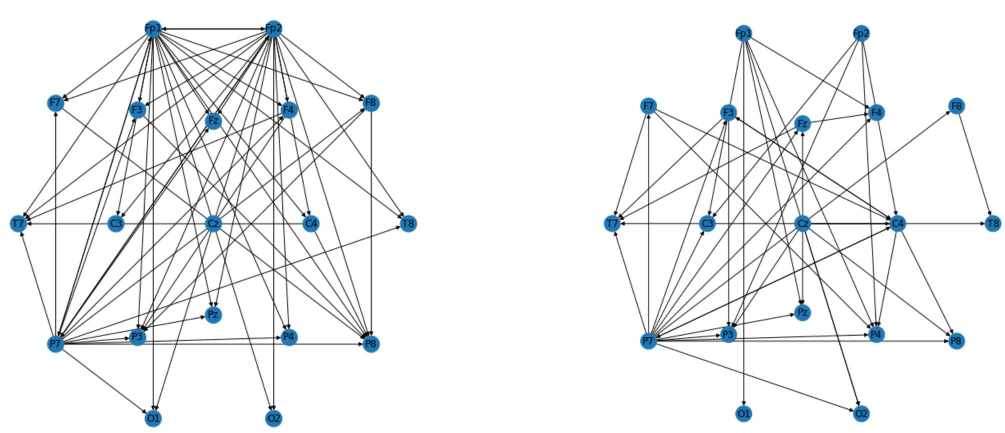
Table 3: Local indices

Global index	Eyes-open	Eyes-closed
Clustering coefficient	0.2093	0.3111
Path length	0.0895	0.0740

Table 4: Weighted global indices values

Dataset	Motif id	Frequency
1	6	117
1	12	0
1	14	0
1	36	72
1	38	81
1	46	10
1	74	0
1	78	0
1	98	0
1	102	0
1	108	0
1	110	0
2	6	79
2	12	0
2	14	5
2	36	68
2	38	78
2	46	1
2	74	0
2	78	0
2	98	0
2	102	0
2	108	0
2	110	0
2	238	0

Table 5: Existance of channel Po7 in motifs

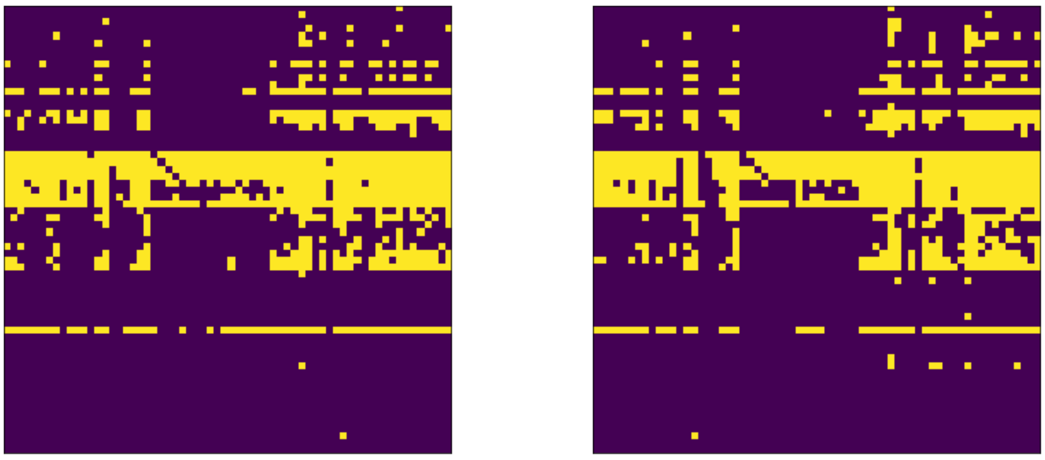


(a) Eyes-open recording network of 19 nodes (b) Eyes-closed recording network of 19 nodes

Figure 1

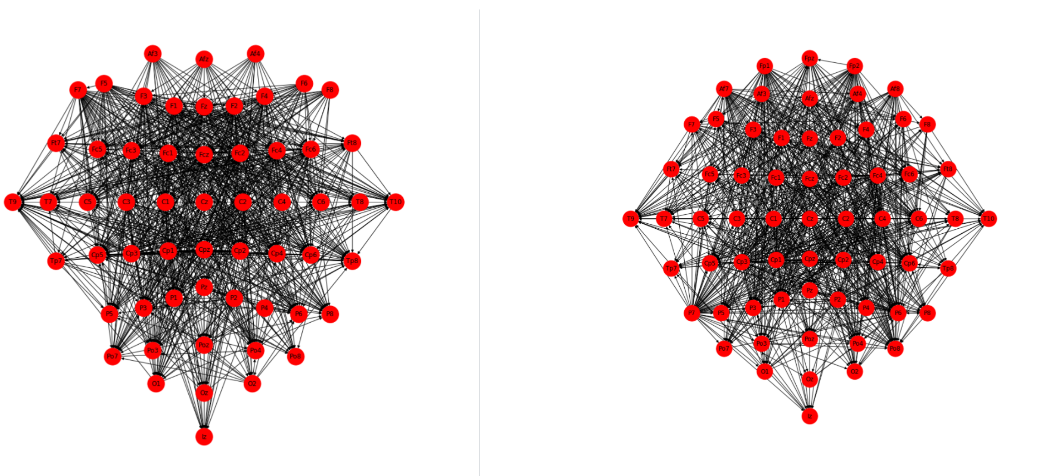
Dataset	**Motif id**
1	74
1	280
1	344
1	392
1	408
1	904
1	906
2	30
2	280
2	332
2	334
2	348
2	350
2	396
2	398
2	414
2	460
2	476
2	478
2	904
2	906
2	922
2	2202
2	4370
2	4422
2	4428
2	4430
2	4434
2	4556
2	4572
2	4574
2	4686
2	4698
2	4700
2	4702
2	4740
2	4748
2	4750
2	4764
2	4766
2	4812
2	4814
2	4830
2	4946
2	4998
2	5004
2	5006
2	5012
2	5016
2	5020
2	5058
2	5064
2	5068
2	5076
2	5080
2	5082
2	6350
2	6550
2	6554

Table 6: 4-nodes antimotifs



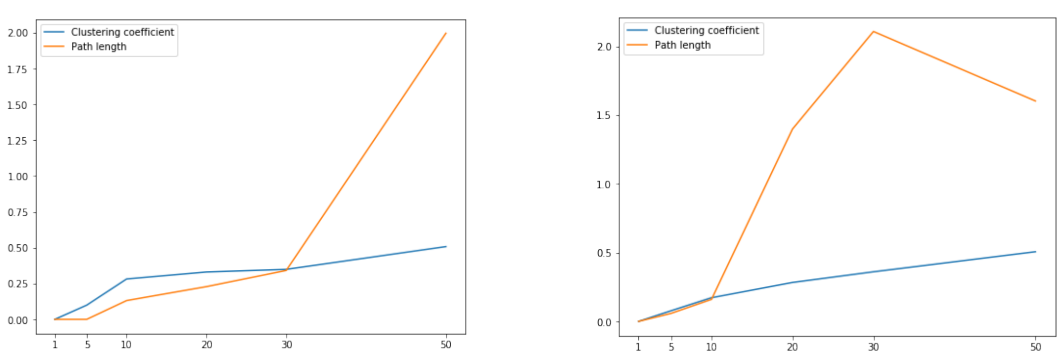
(a) PDC estimator (b) DTF estimator

Figure 2: Binary adjacency matrix with density 20% from PDC and DTF estimators of eyes-open recording.



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

Figure 3: Topographical representation of nodes for part 3.2

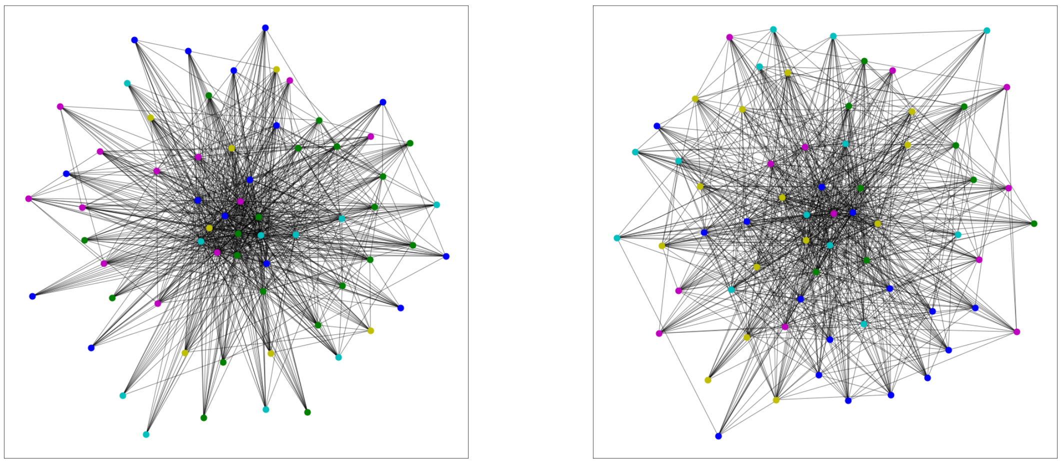


(a) Eyes-open state (b) Eyes-closed state

Figure 4: Global indices dependency on network density.

Method	Eyes opened	Eyes closed
Louvain	4	5
Infomap	1	1

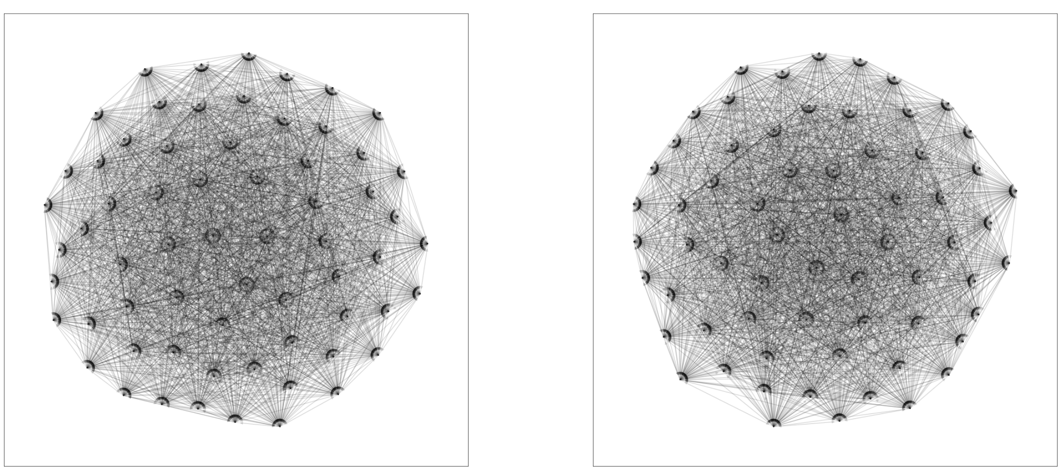
Table 7: Number of optimal communities



(a) Louvain communities eyes-opened

(b) Louvain communities eyes-closed

Figure 5: Louvain communities



(a) Infomap communities eyes-opened

(b) Infomap communities eyes-closed

Figure 6: Infomap communities