

# SAPIENZA UNIVERSITY OF ROME

BIOINFORMATICS & NETWORK MEDICINE

PROJECT 2

---

## Brain network study during resting states

---

January 16, 2021

*Group no. 2:*

Giammarco D'ALESSANDRO 1753102

Luca GIOFFRÈ 1763652

Narges SHARGI 1852342

### Abstract

In this report we describe how we conducted the study of the a brain network during resting phase with eyes open and eyes closed. We used the EEG data of subject S003 from the dataset of PhysioNet [1] [2] (reachable here [3]) to perform functional connectivity analysis.

The analysis is divided into four steps; connectivity estimation and network creation, network analysis through local and global indices, motif analysis and community detection. We used Python to conduct these tests.

The table below (Tab. 1) indicates which tasks we carried out.

Task	Class
<b>1.1</b>	mandatory
<b>1.2</b>	A
<b>1.3</b>	A
<b>1.4</b>	D
<b>1.5</b>	C
<b>1.6</b>	B
<b>2.1</b>	mandatory
<b>2.2</b>	D
<b>2.3</b>	B
<b>2.4</b>	C
Task	Class
<b>2.5</b>	B
<b>2.6</b>	B
<b>2.7</b>	C
<b>3.1</b>	mandatory
<b>3.2</b>	C
<b>3.3</b>	C
<b>3.4</b>	E
<b>4.1</b>	mandatory
<b>4.2</b>	B
<b>4.3</b>	C

Table 1: Chosen tasks for the project.

# 1 Connectivity graph

In this project we analyze the brain connectivity of subject S003 in two states: resting state with eyes open (denoted in the dataset as R01) and resting state with eyes closed (denoted in the dataset as R02). Since we are interested in wakeful rest, we chose to focus on the  $\alpha$ -band rhythm. [1]

We performed a test to understand whether we should focus on the entire  $\alpha$ -band or only on one frequency of the  $\alpha$ -band. We computed the mean matrix and the variance matrix for both states and both estimators and we found out that the variance values are two order of magnitude less than the values of the mean matrices and so they are not significant; we show an example in Fig. 1. We chose then to use only one frequency, 10Hz.

We chose the **Partial Directed Coherence** [4] estimator to compute the functional brain connectivity, since it is more accurate in representing direct influence of one signal on one another and we wanted to study directed networks. However, we computed the same matrix also with the **Directed Transfer Function** [5] to show the differences. The EEG data are in the standard EDF format [6] and we loaded them using the **pyedflib** [7] library for Python.

We computed the **PDC** adjacency matrix and then we applied a threshold to obtain a binary adjacency matrix with density equal to 20% (Fig. 2). We did the same using the **DTF** estimator (Fig. 4). To do so, we used the *connectivity* [8] library for Python. We also computed the adjacency matrices with the following densities: 1%, 5%, 10%, 20%, 30%, 50%. Results can be seen in Fig. 5. Threshold values are specified in the code. Moreover with PDC we investigated connectivity in a 19 channel configuration (Fig.??), for both states, using the resampling procedure to filter the values via *connectivity* tool. We produced the two connectivity matrices (Fig.8) and the related connectivity plot showing the brain connectivity value for each couple of channels 9.

In addition to the graphical representation of the matrices, we also present a topographical representation of the network. We propose three representation, where nodes are colored depending on their in-degree, out-degree and degree. In this way, it is possible to understand the most influent and influenced channels (Fig. 6). We used the *NetworkX* [9] library for Python to handle the graph representation.

Finally, we computed the adjacency matrix using PDC also for the frequency of 25Hz, which belongs to the  $\beta$ -band (Fig. 7). The frequencies in this band are related to active thinking and some muscle contractions. We were expecting a quite different matrix wrt to the 10Hz version, however the two results are similar.

# 2 Graph theory indices

In this section analyze the network resulting from the binary PDC adjacency matrix with density of 20% in terms of graph indices. Local results can be seen in Fig. 10, while the global results are reported below.

The local indices (in-degree, out-degree, degree) can be used to identify the most connected nodes in the graph, which in our case are channels that exploit the function of network hubs. They are important because they play a key role in supporting integrated brain functions. A topographical representation of the three local indices can be seen in 6: we can see that the main difference between state R01 and state R02 is that in the first (eyes open) the frontal part of the scalp is more active, while in the second (eyes closed), the central and occipital parts are more connected. One of the channel in which this behaviour is more evident is channel F6, which has the maximum degree in state R01 and one of the lowest in state R02.

Regarding the global indices, we used the average clustering coefficient and the characteristic path length. The former is a measure about the tendency of the nodes to form highly connected groups, while the latter measures the efficiency in the network communication.

Brain network can also be studied using a small-world index; small-world network (such as the Watts-Strogatz network [10]) are a halfway between regular and random network, because they are characterized by a small average path length (typical of random networks) and by a high clustering coefficient (typical of regular networks). In this way the network ensures a fast transmission of information.

We used the following definition of *Small-Worldness* [11]; a network  $G$  is defined a *small-world network* if

$$S = \frac{C_G/C_{rand}}{L_G/L_{rand}} > 1$$

where  $C_G$  is the average clustering coefficient of  $G$ ,  $C_{rand}$  is the average clustering coefficient for a set of random networks,  $L_G$  is the average path length of  $G$  and  $L_{rand}$  is the average path length for a set of random networks.

Estimator	State	Average Clustering Coefficient	Average Path Length	Small-Worldness
PDC	R01	0.322694	1.915178	0.504817
PDC	R02	0.320938	1.807539	0.528159
DTF	R01	0.482523	0.707093	3.001706
DTF	R02	0.558414	0.481151	3.005174
random	-	0.399983	1.600244	-

Table 2: Comparison between the global indices and the small-worldness of the networks resulting PDC and DTF estimators in both states. The last row contains the averaged values computed for a set of random Erdős-Rényi networks; these values are used in the computation of the small-worldness index.

For the PDC estimator there is no significant differences between state R01 and R02; all three values are comparable. They tend to vary more with the DTF estimator, especially the average path length. The results suggest that PDC networks tend to be less efficiently connected but more clusterized compared to the DTF networks. We were expecting a high score for the small-worldness index, but the results are low; this could be due to the fact that we needed to test on more type of random networks (we tested 500 random generated networks using the Erdős-Rényi algorithm [12]) and/or because the networks we are analyzing are relatively small and, as reported in [11], that the small-world property is not robustly achieved for small networks.

However, these results were obtained by testing only one network density (20%); to investigate more these properties we computed the same indices for different density values (Fig. 4). We can see that the scale-free property of these type of network is present, because scaling the nodes degree does not give many information about the topology of the network (both the average path length and the small-worldness do not follow a linear curve). Moreover, the small-worldness tends to increase with the density of the graph after the threshold of 20% density.

In order to understand better the previous findings about the 25Hz PDC graph, we computed its local (Fig. 12) and global indices.

Frequency	State	Average Clustering Coefficient	Average Path Length
25Hz	R01	0.323690	2.166421
25Hz	R02	0.315331	2.021812

Table 3: Comparison between the global indices for the network using frequency of 10Hz and 25Hz.

We can see that the results are consistent with what we found before. The apparently strange results

may be due to the fact that we are analyzing a single subject for which we have few data available in these conditions.

Finally, instead of a binarized adjacency matrix we computed these indices also for a weighted adjacency matrix, which results in a weighted directed graph; local indices are shown in Fig. 11, while the global indices are the following table (Tab. 4);

State	Average Weighted Clustering Coefficient	Average Weighted Path Length
R01	0.108107	0.281965
R02	0.110925	0.252326

Table 4: Global indices for the weighted network in both states.

### 3 Motif analysis

In this section we perform motif analysis to investigate the presence of 3-node configuration in the connectivity network (obtained from PDC binary adjacency matrix, 20% density). In order to do so we computed motif frequency (the occurrences of each specific pattern in the network), and determine their statistical significance, calculating the z-score of each pattern [13].

Motif analysis can be useful to understand which are the building blocks of a network; in fact it is likely that, as the network become more complex, already existing simpler sub-networks are largely preserved and extended, while it is less likely to observe newly formed complex structures. We analyzed the 3-node pattern frequencies in both experiment states (R01, R02), as shown in Fig.13. We observe that in both cases pattern 1,2,3 and 5 are the most represented, while the others (the most connected ones) are less present. We also see that in both experiments runs the 3-node configurations are similarly distributed, as the two frequency fingerprints have peak values in the same nodes (e.g. node 33 and 49 in Fig. 14).

Even though motif frequency gives some interesting insights over network structure, it is necessary to take into account their statistical significance in order to understand which motif are the real building blocks. Thus we computed the z-score, a statistical significance measure, for each of the 3-nodes pattern as follows:

$$Z = \frac{f_{orig}(G_k) - \langle f_{rand}(G_k) \rangle}{\sigma_{rand}}$$

where  $f_{orig}(G_k)$  is the frequency of a 3-nodes sub-graph  $G_k$  in the real graph  $G$ ;  $\langle f_{rand}(G_k) \rangle$  is the average frequency of sub-graph  $G_k$  in an ensemble of random graphs generated with the same degree sequence of  $G$  (i.e. a similar configuration); and  $\sigma_{rand}$  is the standard deviation of the frequency over the random graphs [14]. This score serves as a measure of how much the network configuration obtained from the connectivity data is characterized by the data themselves, and how likely is to observe the same pattern frequency in random generated structures. We first computed the average frequencies in the random network ensemble, and compared them to the real ones (Fig.16 *top panels*), observing that in all cases, except for pattern 7, real network frequencies are higher than the random average. Then we analyzed the z-score w.r.t. the *minimum-deviation* thresholds for each configuration (Fig.16 *bottom panels*), observing that the highest values belongs to the less represented patterns 8,9,10,11,12 and 13 that are less frequent than the others in the real graph, but significantly more present than in random graphs. Finally we implemented to definition of network motif: *over representation* and *minimum deviation*. According to the former all patterns, except 2,3 and 7 (R01), 2 and 7 (R02), can be considered motifs; whilst according to the latter, 4,8,9,10,11,12,13 (R01), and 1,5,6,8,9,10,11,12,13 (R02) configurations are network motif. We also analyzed frequency of motif 4 ( $A \rightarrow B \leftarrow C$ ) (Fig.15), and the frequencies for channel Po4 (Fig.17) in both states.

## 4 Community detection

The last step in our analysis is to investigate the communities present in the brain connectivity network. A brain network can be divided into communities, subset of densely connected nodes that can possibly be interpreted as responsible for specific brain functions.

In fact, the modular structure promotes mechanism of both integration (the ability to quickly combine specialized information coming from different modules) and segregation (the ability for specialized processes to occur within densely connected group). We used two algorithms based on two different approaches; the Louvain algorithm [15] and the Infomap algorithm [16].

The Louvain algorithm tries to optimize the modularity quality function of Newman's Algorithm [17] [18] and it consists of two phases repeated iteratively.

During the first, each node is considered as a distinct community and is moved in the neighbour's community that maximizes the gain in modularity; the process terminates when no individual move can improve the modularity.

During the second phase, each community is treated as a node (super-node) and the edge's weights are computed accordingly. The two phases are repeated until there are no more changes to do.

The Infomap algorithm is based on the principles of the information theory and it identifies communities according to the information flow in the network; instead of modularity, the algorithm tries to optimize the minimum description length [19].

It is important to note that the number of communities and their structure which result from the first algorithm may depend on the order in which the nodes are analyzed while the second does not have this limitation.

The results of both algorithm are shown in Fig. 18. We can see how the difference in the two approaches reflects in the final results; the Louvain algorithm identifies four to five different communities, that can be roughly mapped onto the four different brain region, while the Infomap algorithm seems to be lacking the resolution for distinguish small communities and in both cases identifies one big community and another one or two small communities.

However, in both cases, the connectivity network in state R02 can be divided in a number of communities greater than the number of communities in state R01, and so an increase in segregation, which may be interpreted as a higher number of processing being held during state R02.

We can also note how the detected communities can be quite 'sparse' (with respect to the location of the electrodes on the scalp). This suggests that interaction and cooperation between different brain regions is possibly also without a direct link and it is no surprise, since we are studying the functional connectivity of the brain.

For example, in the *top-left corner* of Fig. 18 we can see how the 'green' community is spread across temporal, frontal and parietal lobe; the same happens, even more intensely, with the 'red' community in the *bottom-left corner* of the figure.

It is more difficult to identify this behaviour in the Infomap communities (*right side*), since the red community is predominant, but they are not necessarily wrong results.

Assuming that the Infomap algorithm is less sensible, we can give the following interpretation to its results; since in a state of wakeful rest we can assume that the cerebral activity is generally low, the recorded signals are not intense enough for the algorithm to clearly distinguish different communities.

As our last step, we computed the Jaccard similarity among communities in the two partitions, in both states, to see quantitatively how much the algorithms output is different. The results are presented in Table 5, only non-zero values are taking into consideration; a complete mapping of the two community sets is not possible, since the number of communities is not the same, but we highlighted the mapping with the highest scores.

<b>State</b>	<b>Louvain Community</b>	<b>Infomap Community</b>	<b>Jaccard similarity</b>
R01	<b>RED</b>	<b>RED</b>	<b>37.73%</b>
R01	ORANGE	RED	35.84%
R01	GREEN	RED	14.51%
R01	<b>GREEN</b>	<b>PURPLE</b>	<b>45.00%</b>
R01	PURPLE	RED	9.09%
R01	PURPLE	PURPLE	12.50%
<b>State</b>	<b>Louvain Community</b>	<b>Infomap Community</b>	<b>Jaccard similarity</b>
R02	RED	RED	15.78%
R02	<b>RED</b>	<b>YELLOW</b>	<b>47.36%</b>
R02	<b>ORANGE</b>	<b>RED</b>	<b>35.41%</b>
R02	YELLOW	RED	24.48%
R02	YELLOW	YELLOW	4.54%
R02	GREEN	RED	5.66%
R02	<b>GREEN</b>	<b>PURPLE</b>	<b>55.55%</b>
R02	PURPLE	RED	14.28%
R02	PURPLE	PURPLE	7.69%

Table 5: Jaccard similarities for the communities found with the two algorithms.

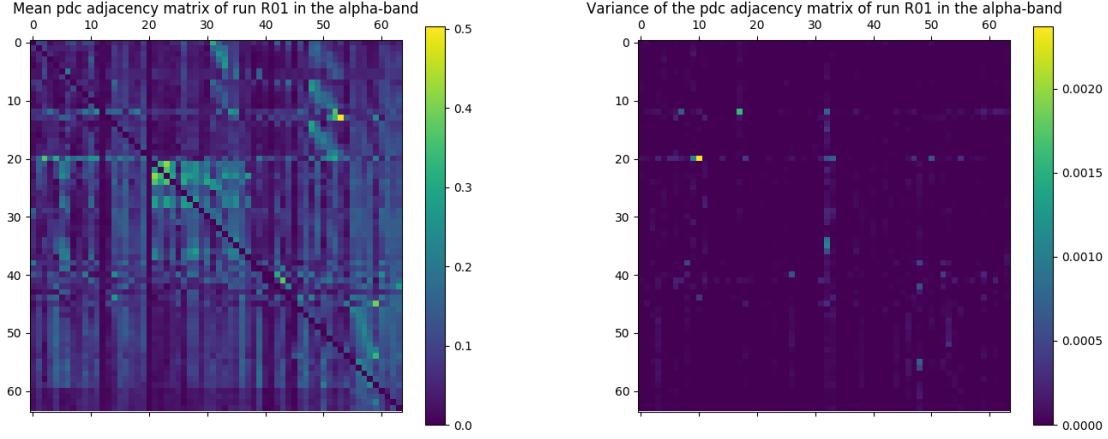


Figure 1: Mean matrix of all PDC matrices in state R01 in the  $\alpha$ -band (*left*) and its variance (*right*). The maximum value for the mean matrix is 0.5, while the maximum for the variance matrix is around 0.0025. Since the latter is two order of magnitude smaller than the former, we used one specific frequency (10Hz) instead of the entire  $\alpha$ -band.

0	Fc5..	16	Cp1..	32	F1..	48	P3..
1	Fc3..	17	Cpz..	33	Fz..	49	P1..
2	Fc1..	18	Cp2..	34	F2..	50	Pz..
3	Fcz..	19	Cp4..	35	F4..	51	P2..
4	Fc2..	20	Cp6..	36	F6..	52	P4..
5	Fc4..	21	Fp1..	37	F8..	53	P6..
6	Fc6..	22	Fpz..	38	Ft7..	54	P8..
7	C5..	23	Fp2..	39	Ft8..	55	Po7..
8	C3..	24	Af7..	40	T7..	56	Po3..
9	C1..	25	Af3..	41	T8..	57	Poz..
10	Cz..	26	Afz..	42	T9..	58	Po4..
11	C2..	27	Af4..	43	T10..	59	Po8..
12	C4..	28	Af8..	44	Tp7..	60	O1..
13	C6..	29	F7..	45	Tp8..	61	Oz..
14	Cp5..	30	F5..	46	P7..	62	O2..
15	Cp3..	31	F3..	47	P5..	63	Iz..

Table 6: Index to channel mapping.

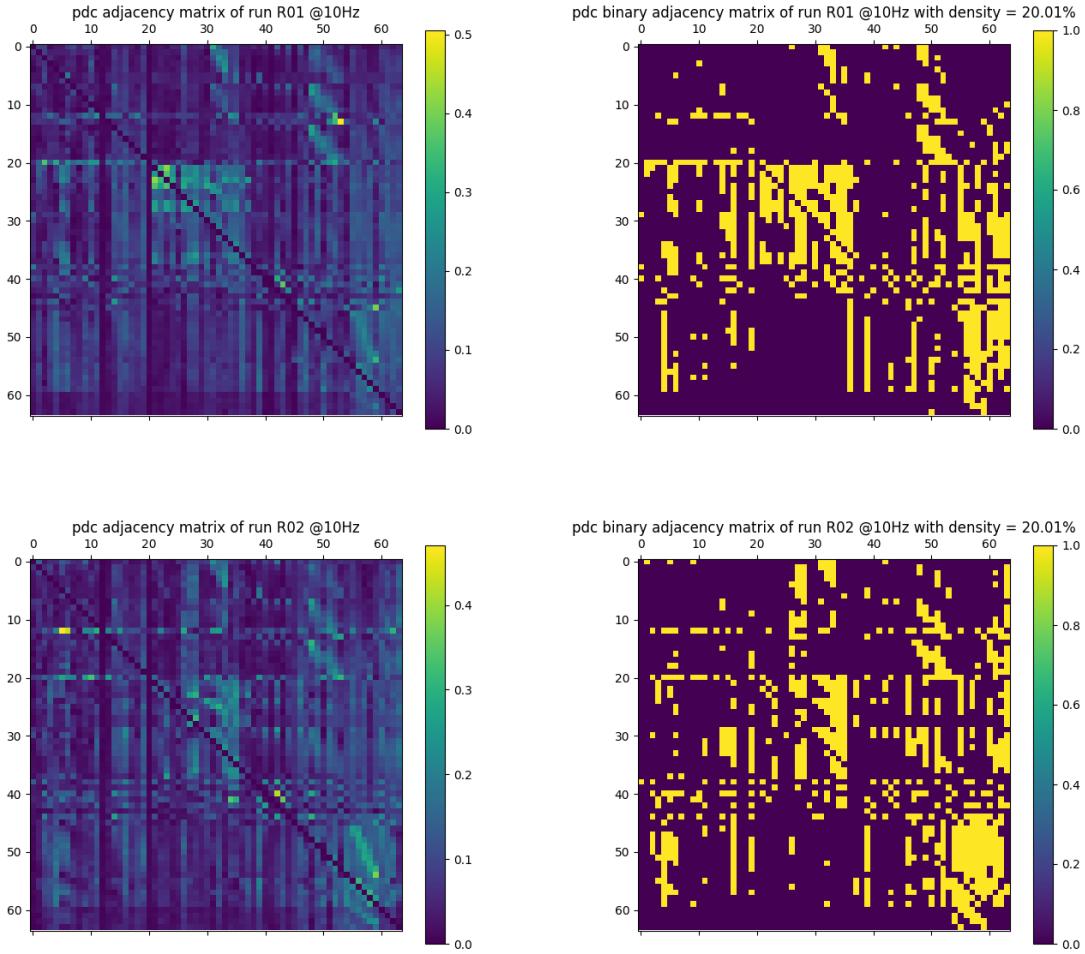


Figure 2: PDC adjacency matrices without thresholds in R01 (*top-left*) and in R02 (*bottom-left*) and their binary version (*right side*) for the frequency of 10Hz.

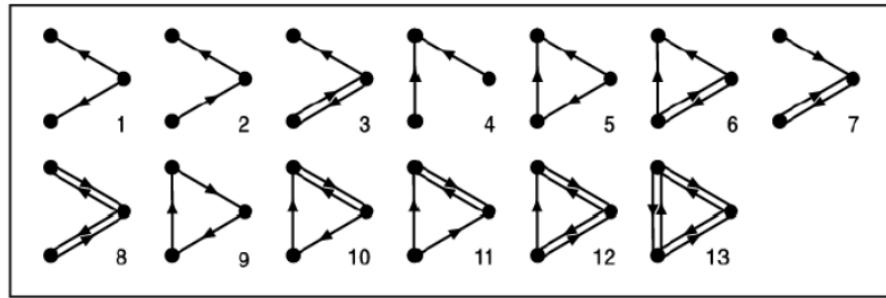


Figure 3: IDs of all possible 3-nodes motif configurations, where all nodes are connected (at least two edges). The IDs are the same as in [14].

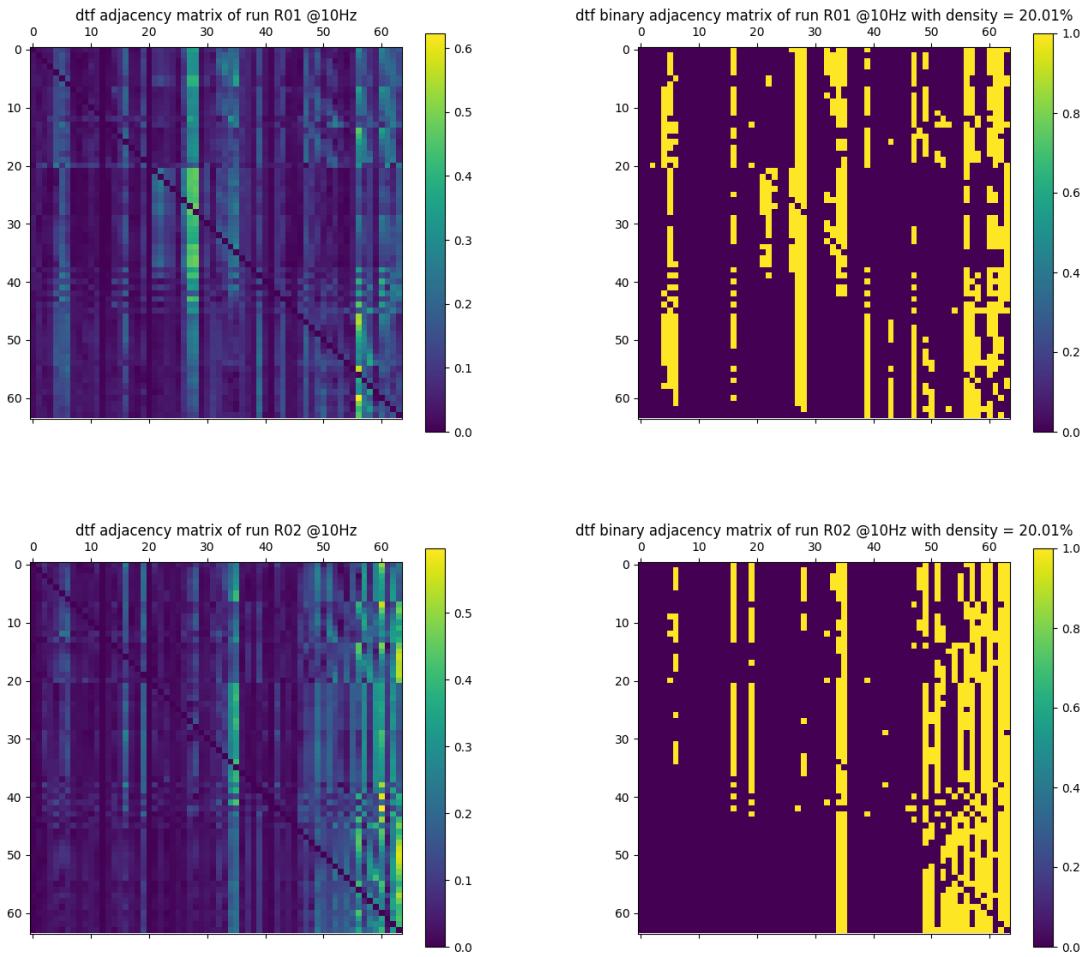


Figure 4: DTF adjacency matrices without thresholds in R01 (*top-left*) and in R02 (*bottom-left*) and their binary version (*right side*) for the frequency of 10Hz.

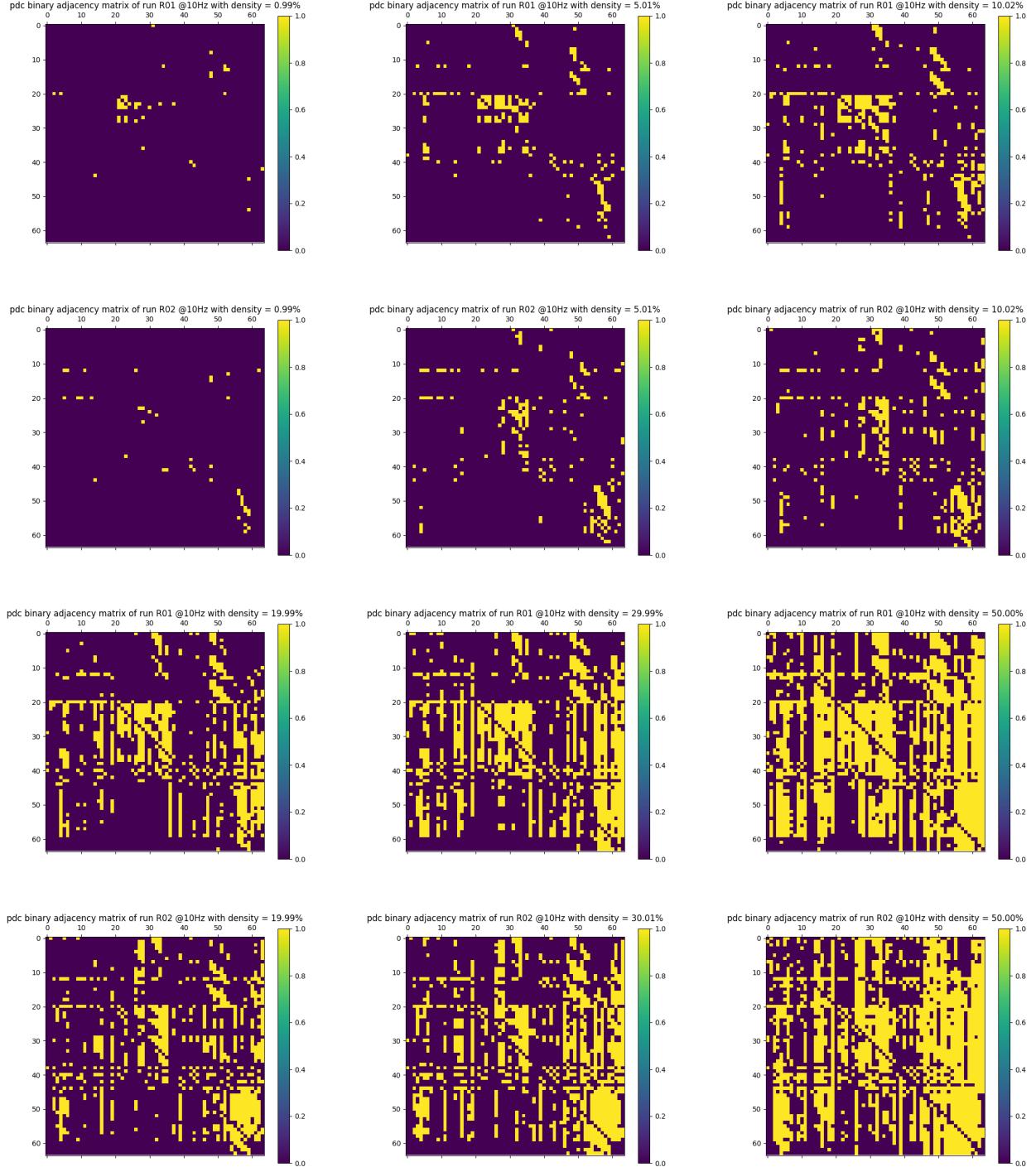


Figure 5: From *top-left* corner moving horizontally; PDC binary matrix with densities 1%, 5%, 10% (first row is state R01, second row is state R02), 20%, 30%, 50% (third row is state R01, fourth row is state R02)

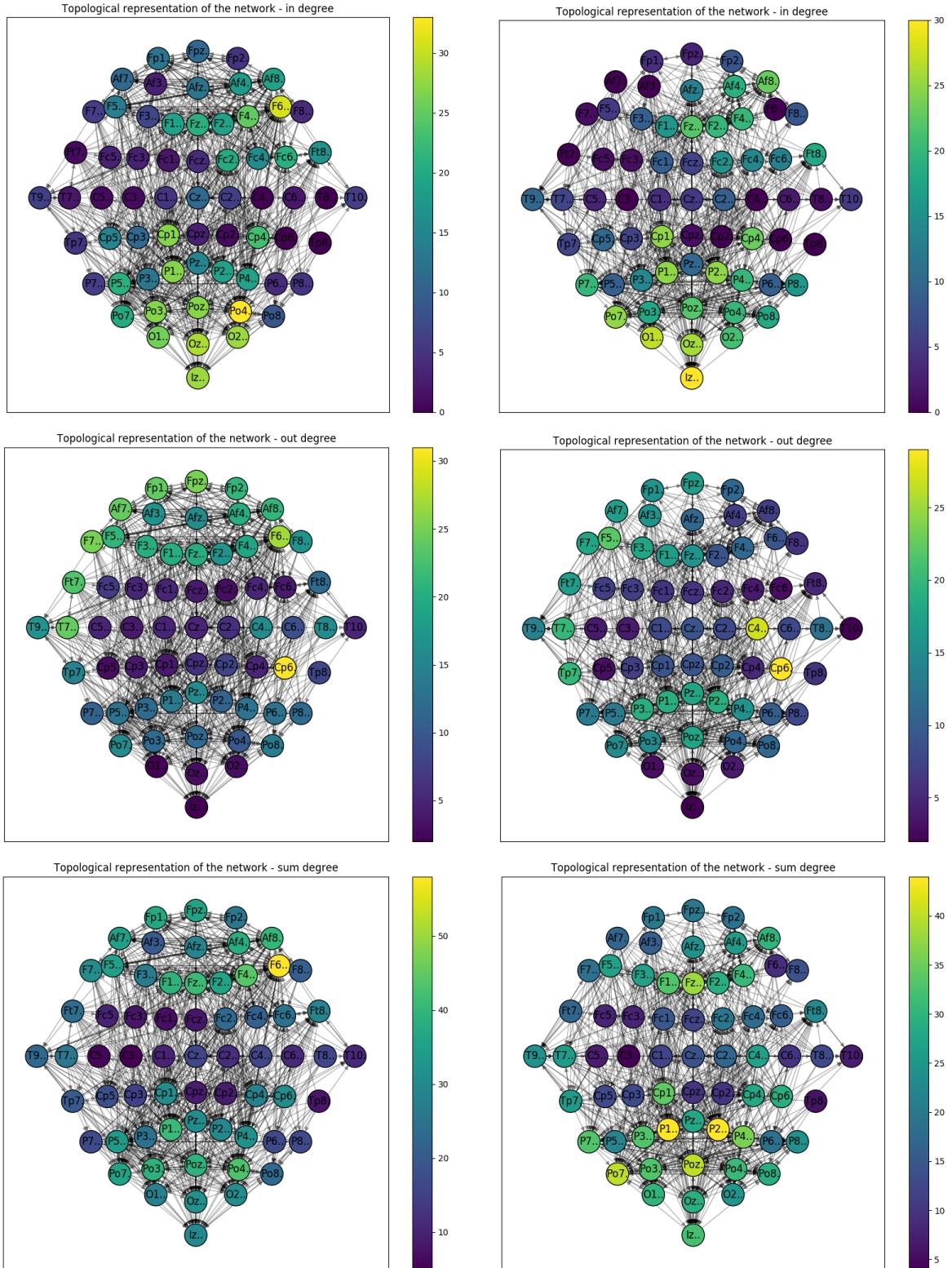


Figure 6: Topographical representation in R01 (*left column*) and in R02 (*right column*). Nodes are colored depending on their in-degree (*first row*), on their out-degree (*second row*) and on their degree (*third row*).

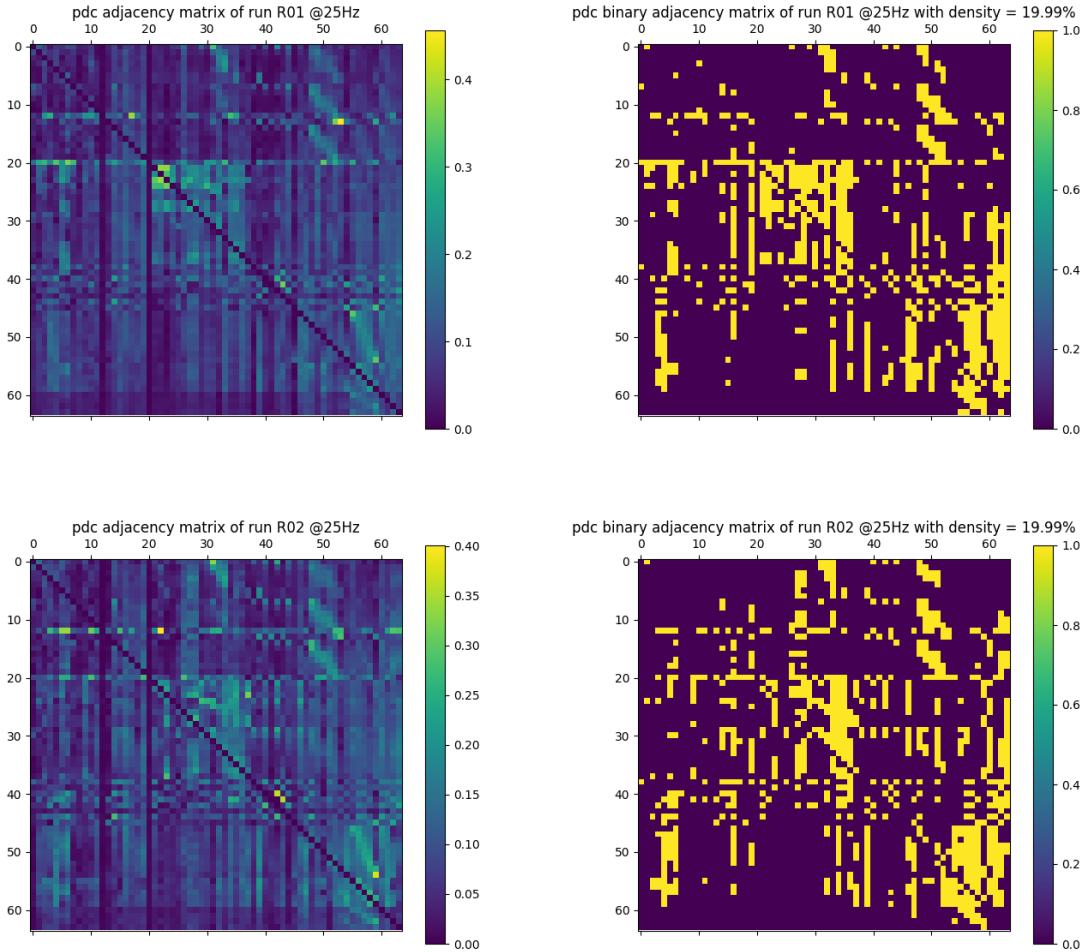


Figure 7: PDC adjacency matrices without thresholds in R01 (*top-left*) and in R02 (*bottom-left*) and their binary version (*right side*) for the frequency of 25Hz.

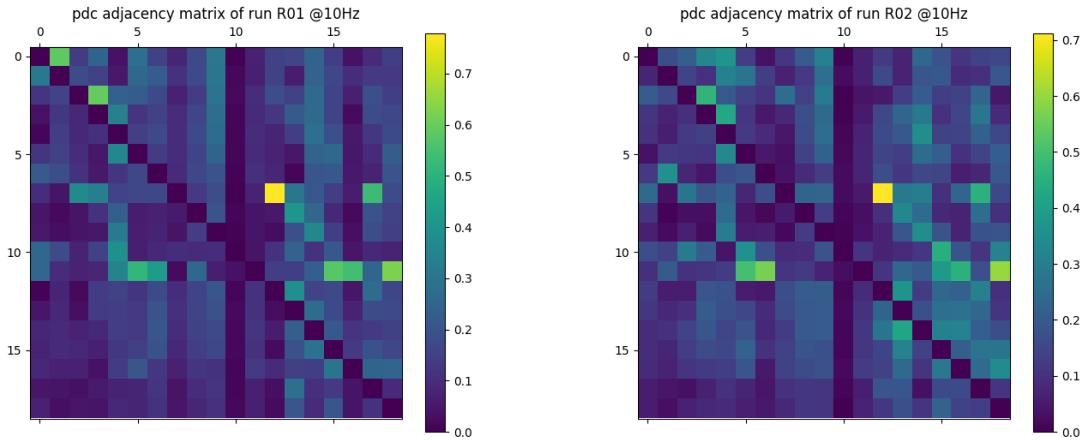


Figure 8: Connectivity matrix obtained with 19 channel configuration in state R01 (*left*), and in state R02 (*right*).

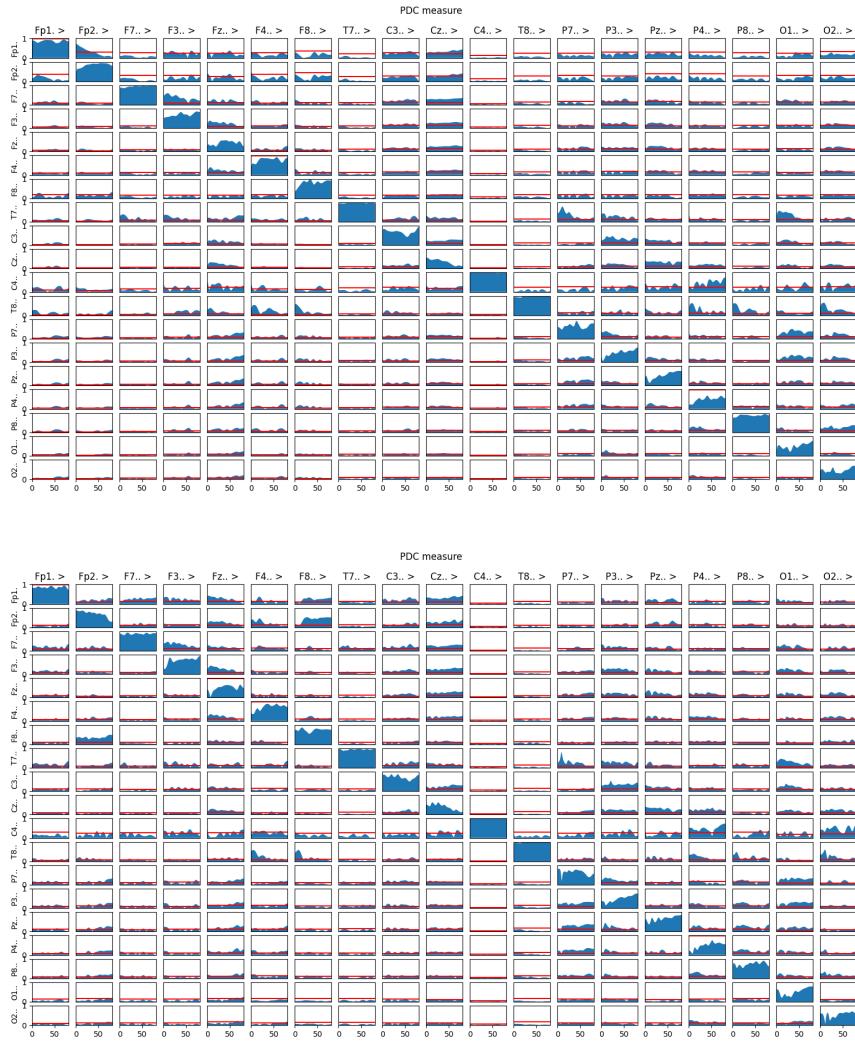


Figure 9: Connectivity plot obtained with 19 channel configuration in state R01 (*left*), and in state R02 (*right*).

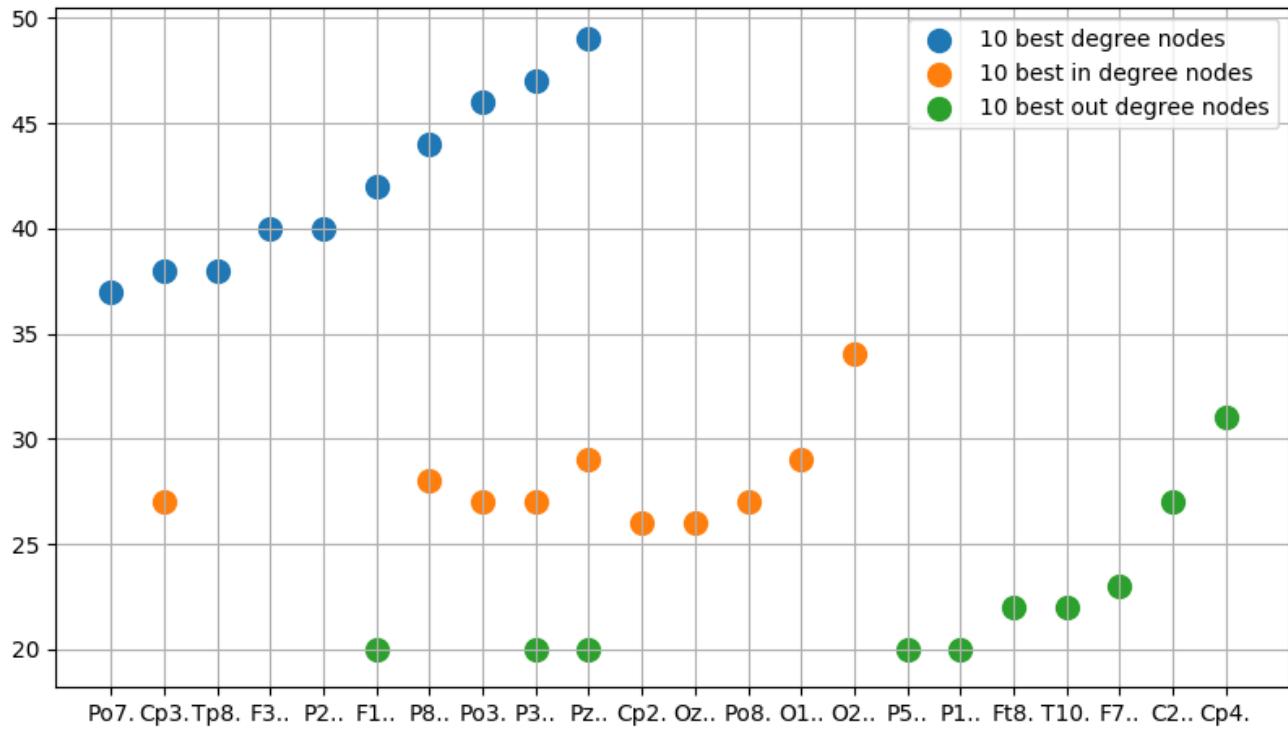
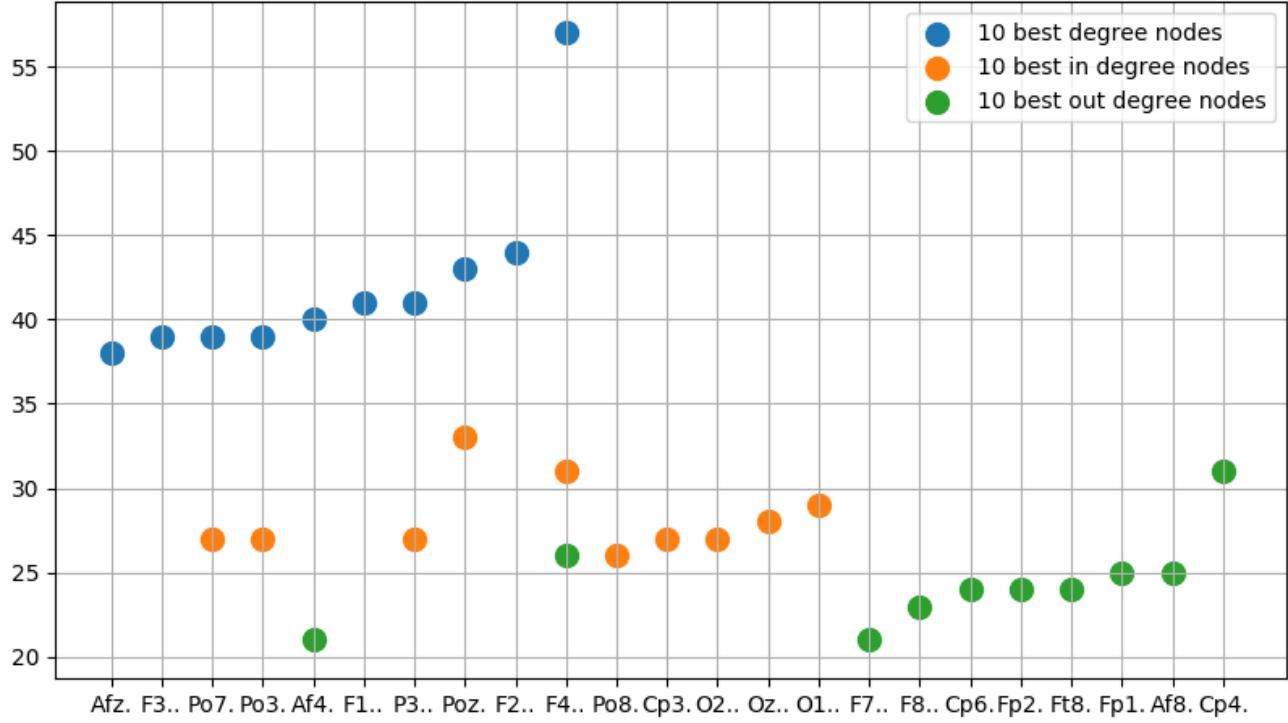


Figure 10: 10 highest channels ranked for node degree (*left*), in-degree (*center*) and out-degree (*right*) in state R01 (*first row*) and in state R02 (*second row*).

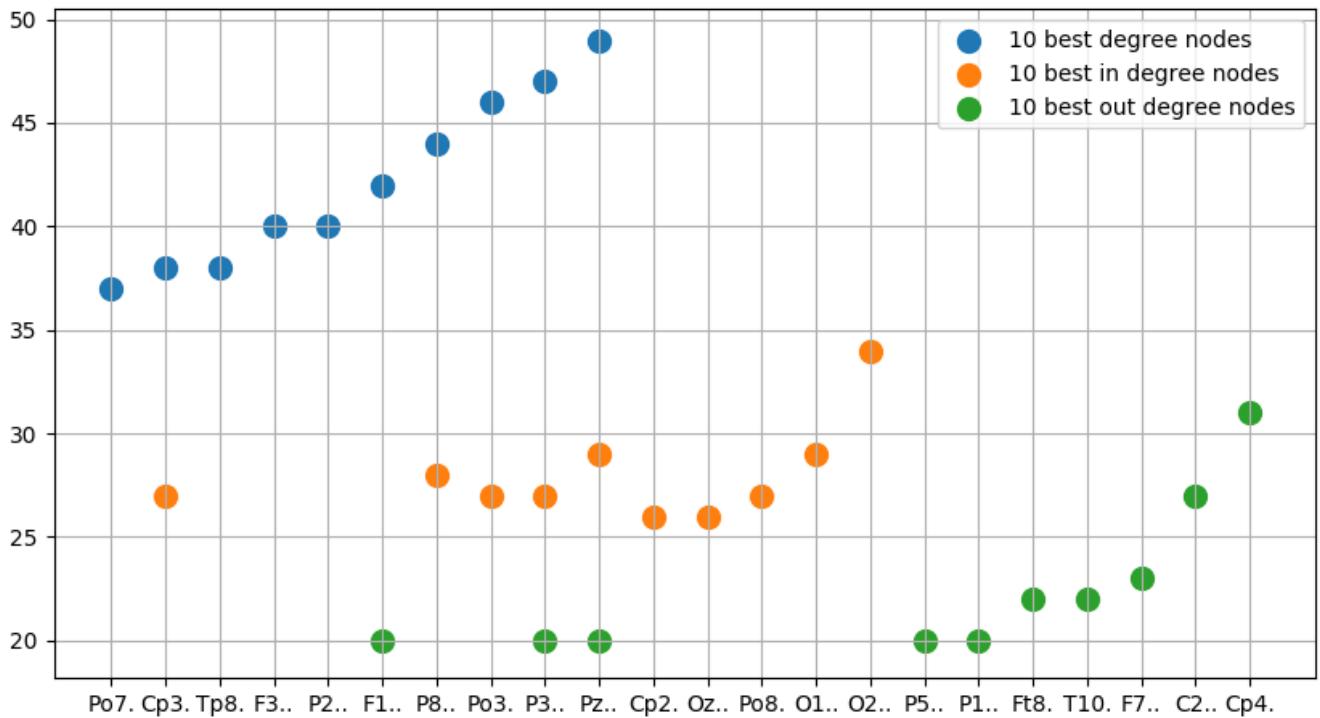
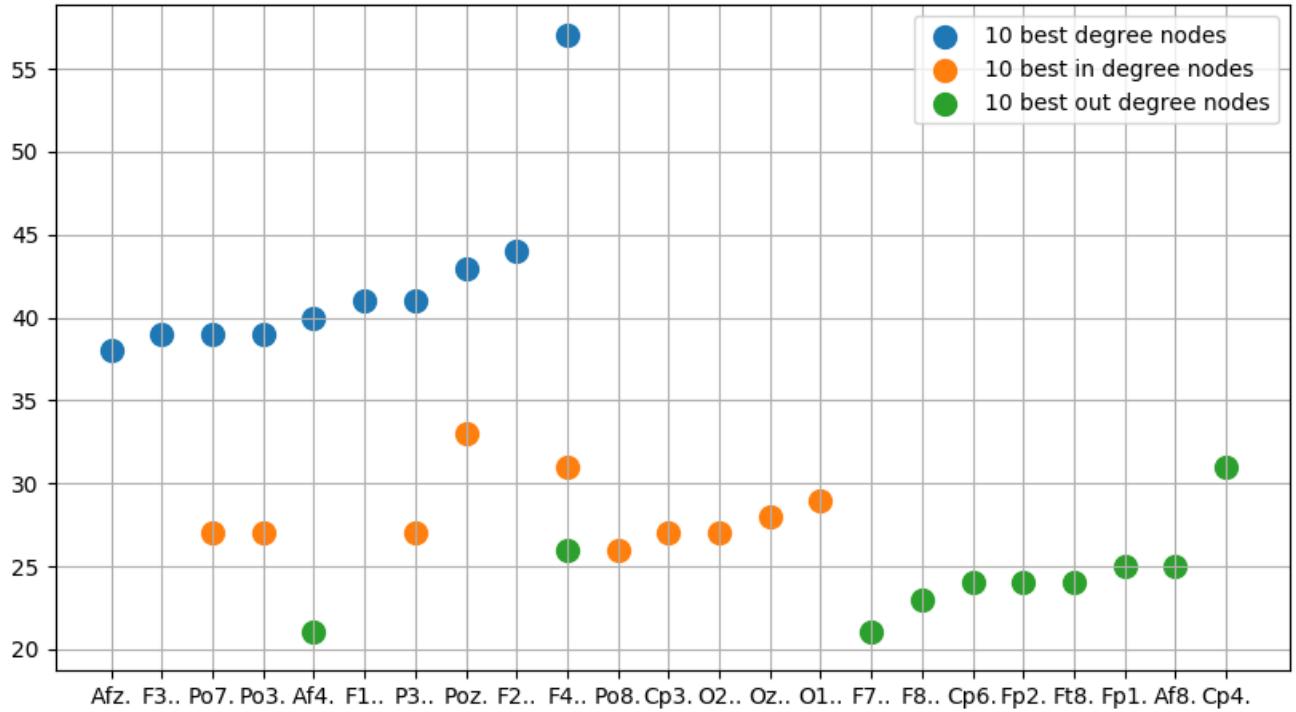


Figure 11: 10 highest channels ranked for node degree (blue), in-degree (orange) and out-degree (green) in state R01 (*top*) and in state R02 (*bottom*) for the weighted version of the network.

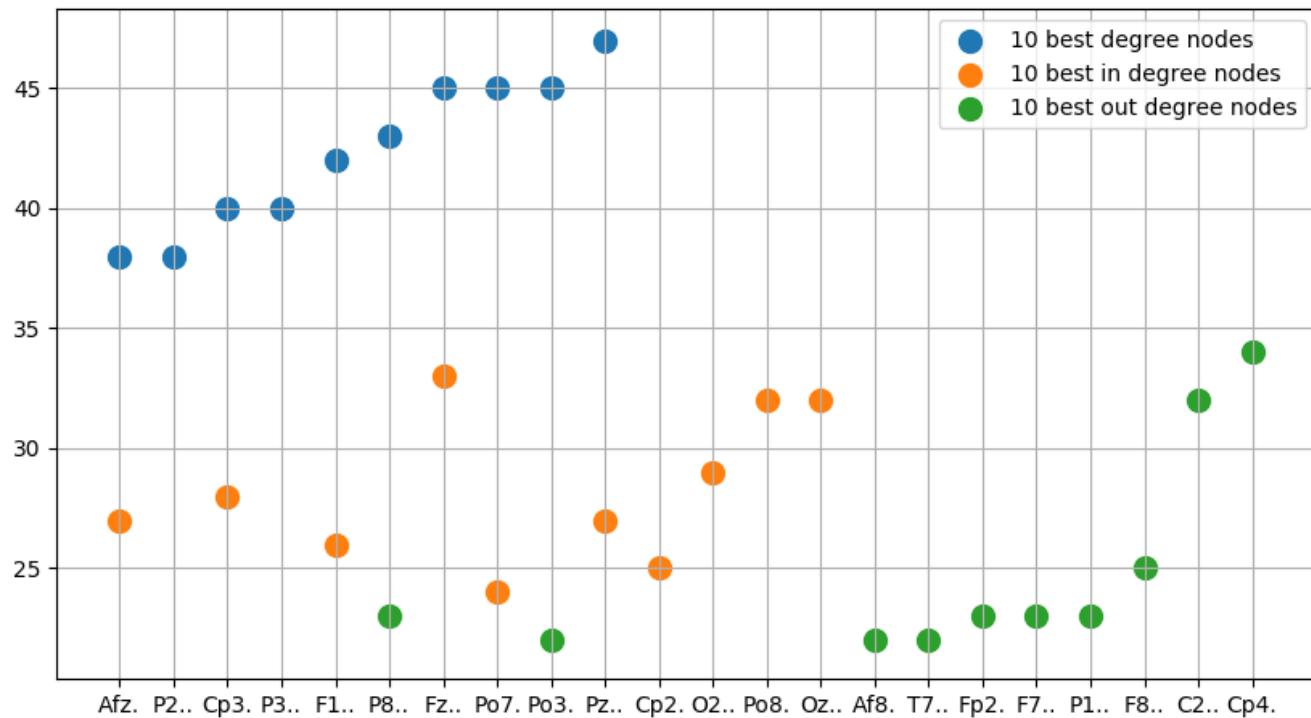
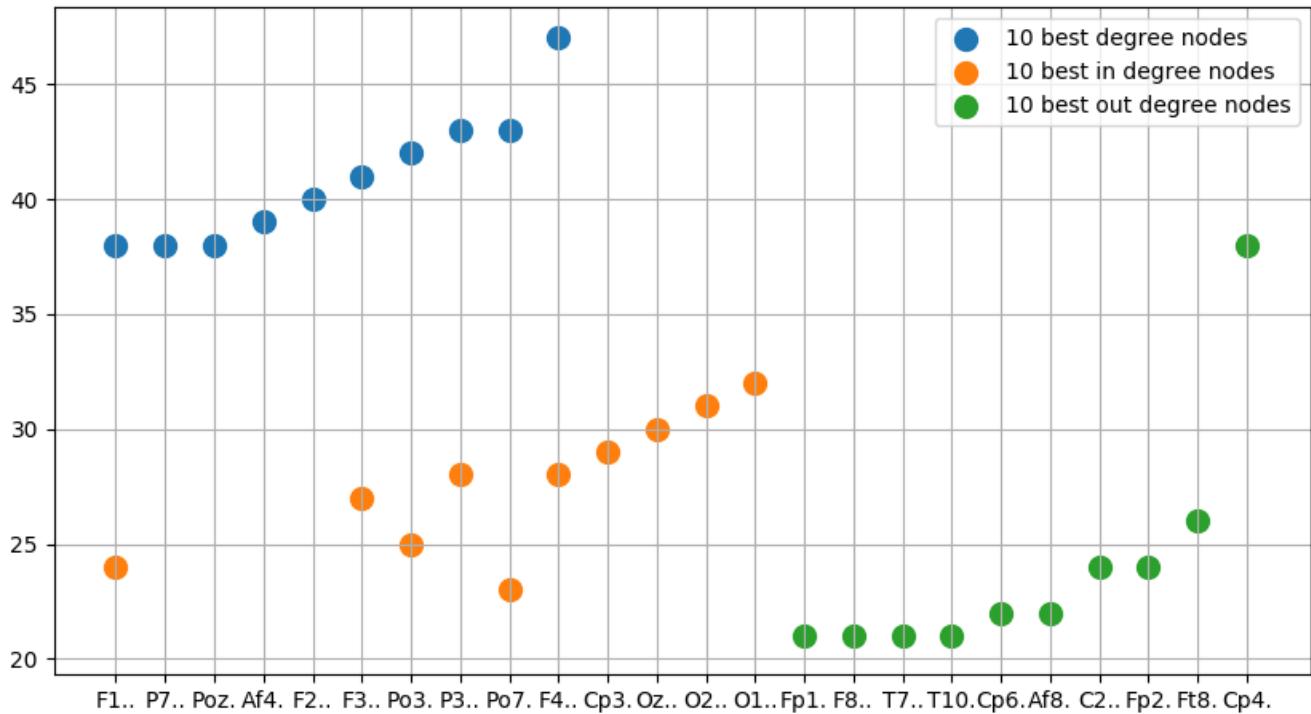
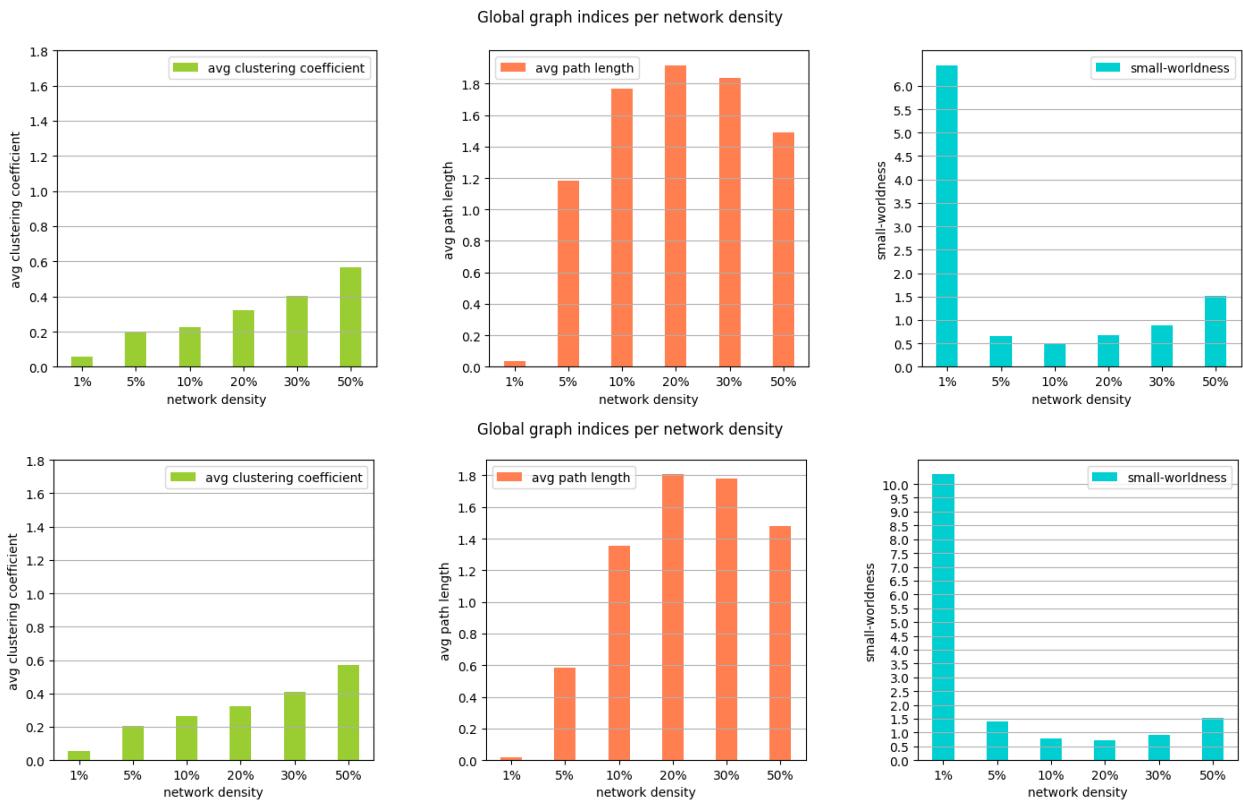


Figure 12: 10 highest channels ranked for node degree (blue), in-degree (orange) and out-degree (green) in state R01 (*top*) and in state R02 (*bottom*) for the 25Hz version of the network.



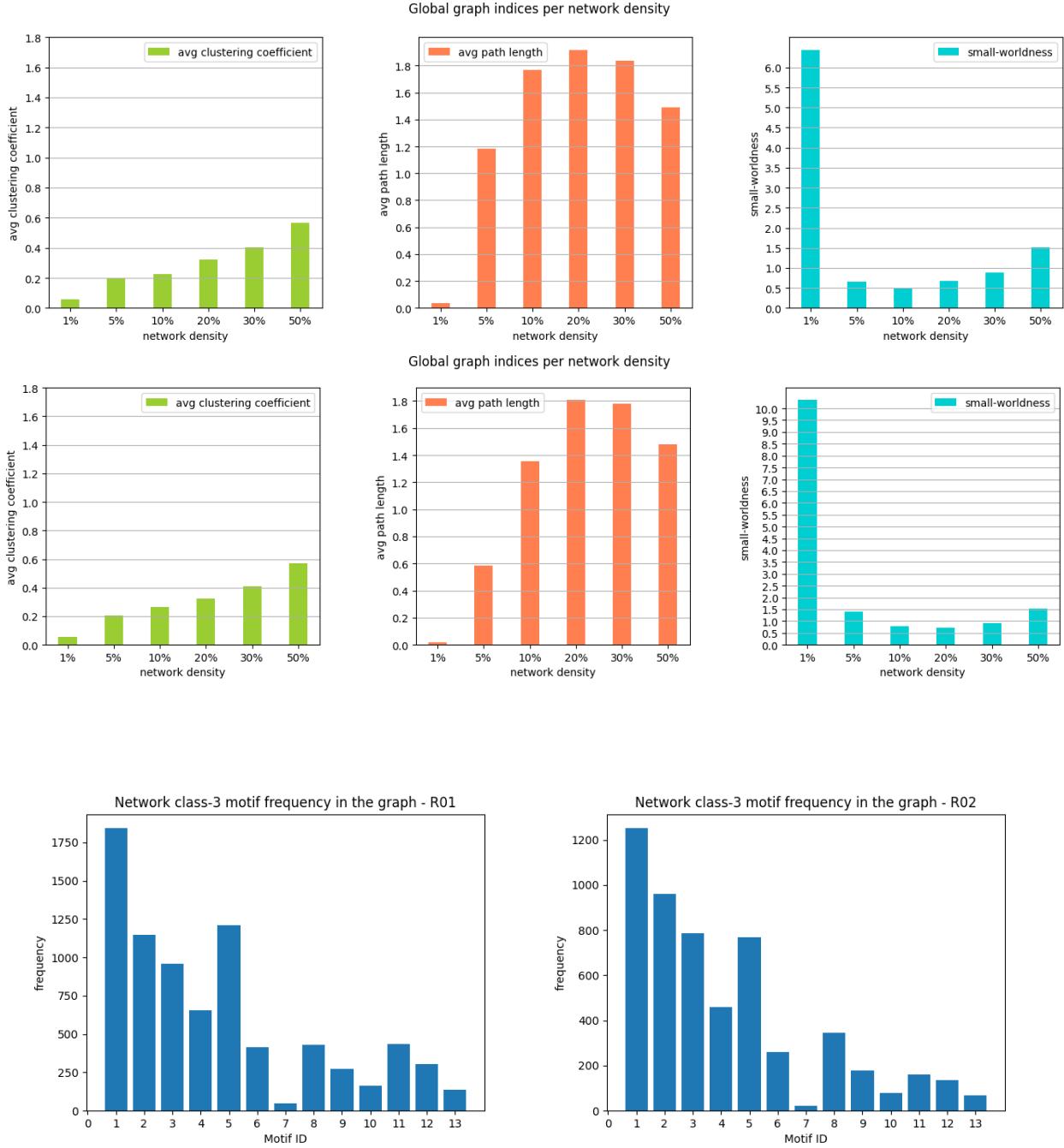


Figure 13: Class-3 motif frequency in the connectivity graph in state R01 (*left*), and in state R02 (*right*).

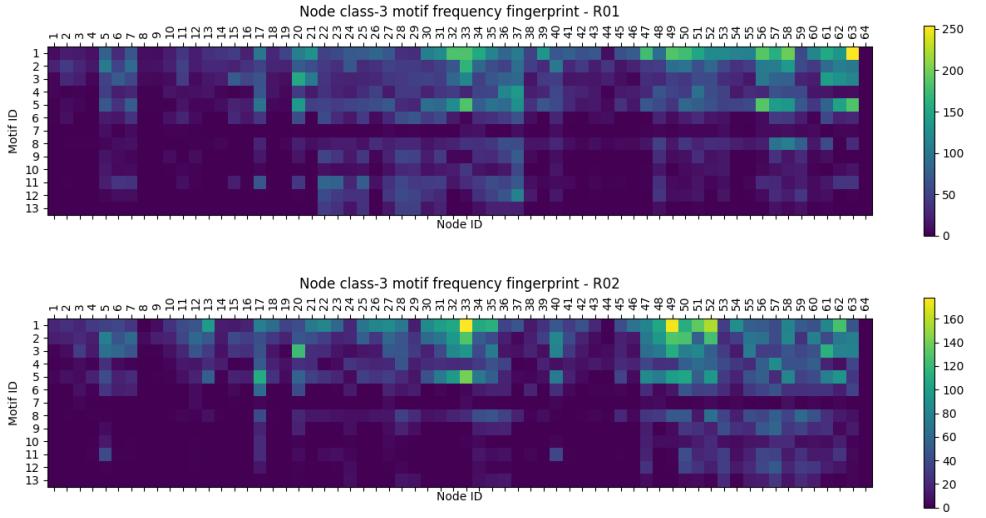


Figure 14: Frequency of patterns involving each node (channel) in state R01 (*top*), and in state R02 (*bottom*).

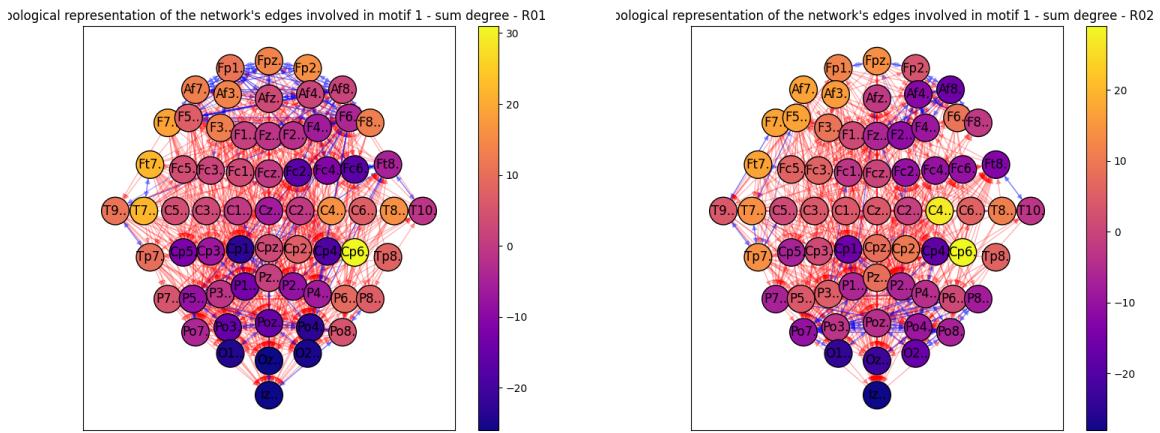


Figure 15: Topographical plot of connections involving pattern  $A \rightarrow B \leftarrow C$ , scaled on node degree, in state R01 (*left*), and in state R02 (*right*).

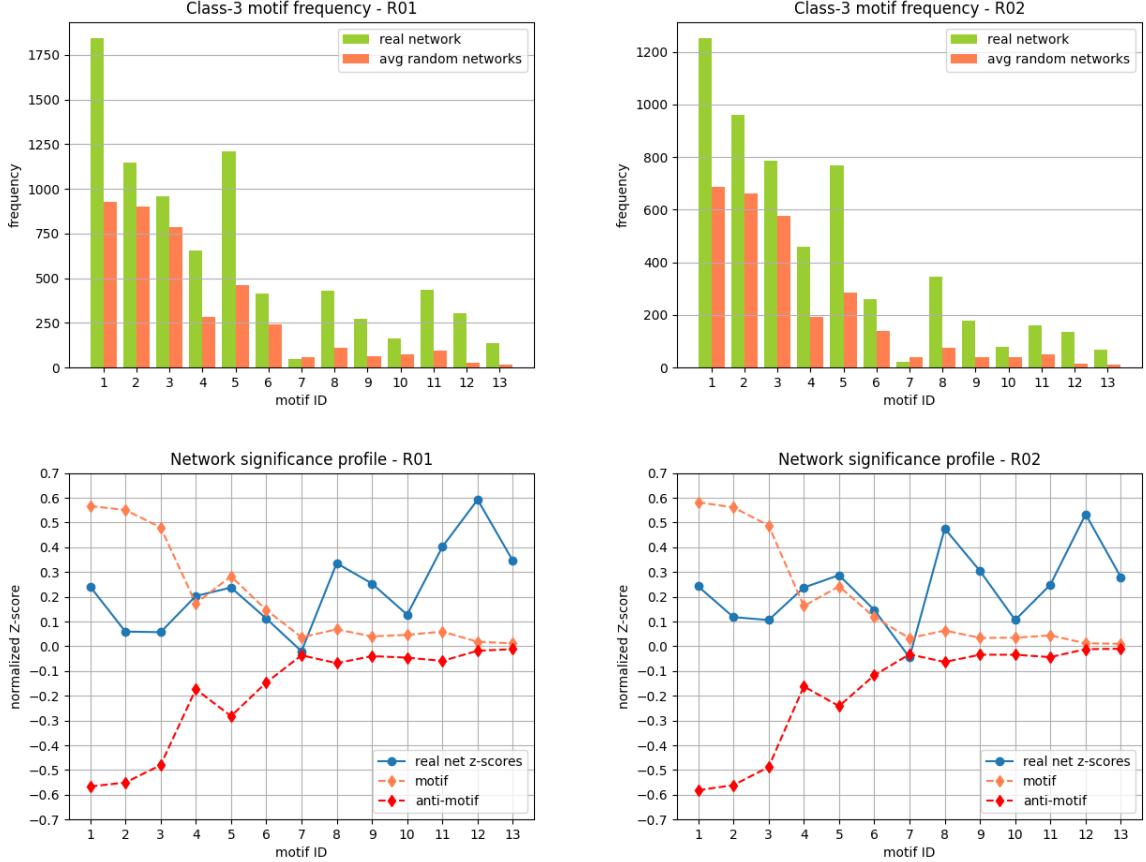


Figure 16: Class-3 motif frequencies in the real network, compared to the average frequencies in the random networks tested for statistical significance; state R01 (*top left*), and in state R02 (*top right*). Below them we show the significance profile of the connectivity network, i.e. the z-scores values for each of the 3-node motif; in state R01 (*bottom left*), and in state R02 (*bottom right*).

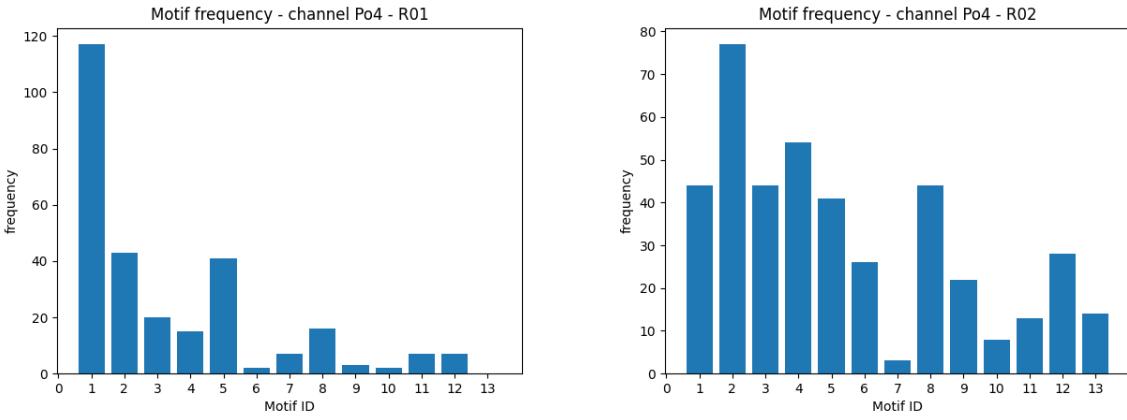


Figure 17: Frequency of patterns involving channel Po4 in state R01 (*left*), and in state R02 (*right*).

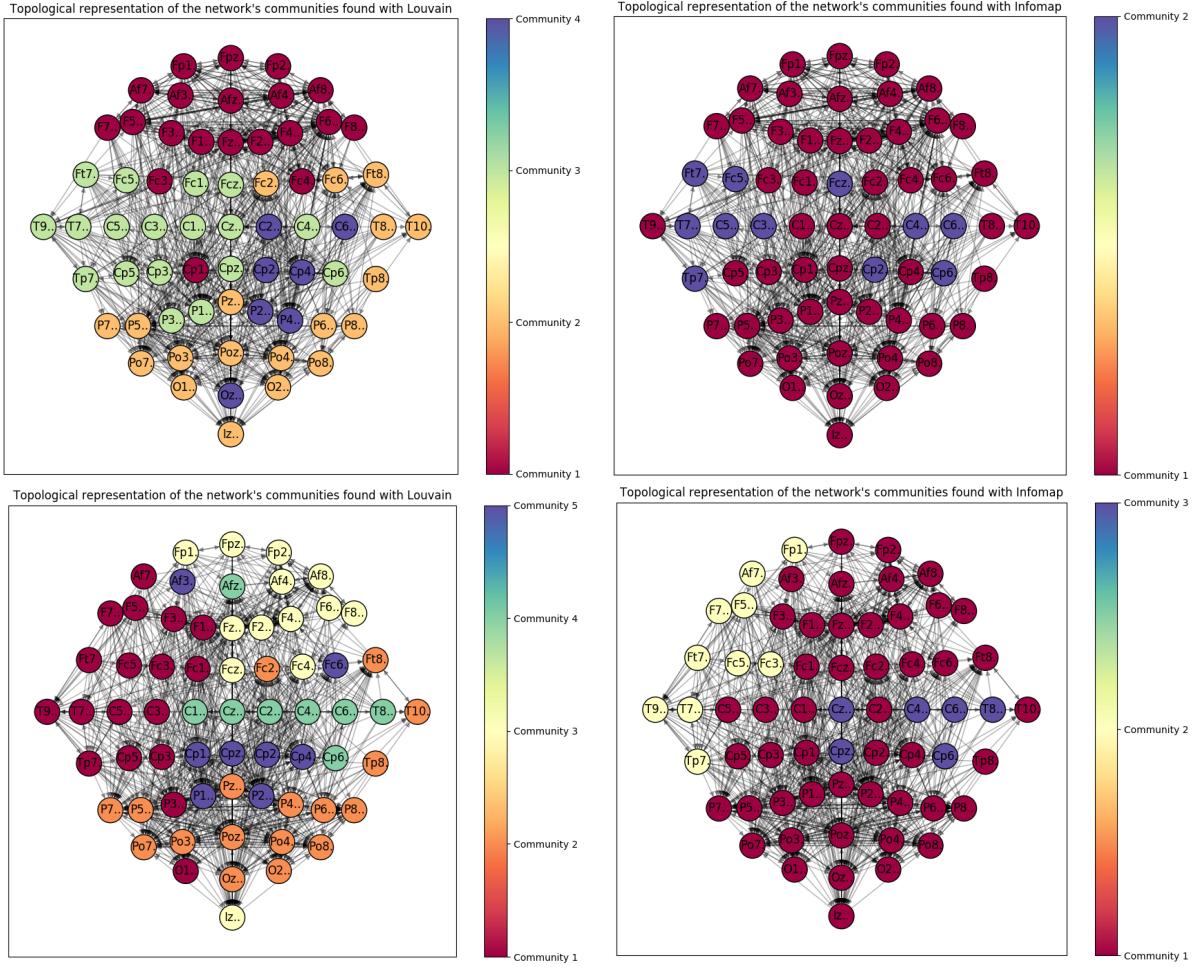


Figure 18: Topographical view of the communities found using the Louvain algorithm (*left side*) and the Infomap algorithm (*right side*) in state R01 (*top*) and R02 (*bottom*)

## References

- [1] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, “Bci2000: a general-purpose brain-computer interface (bci) system,” *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1034–1043, 2004.
- [2] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, “PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals,” *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000 (June 13), circulation Electronic Pages: <http://circ.ahajournals.org/content/101/23/e215.full> PMID:1085218; doi: 10.1161/01.CIR.101.23.e215.
- [3] G. L. H. J. e. a. Goldberger A., Amaral L., “Eeg motor movement/imagery dataset.” [Online]. Available: <https://physionet.org/content/eegmmidb/1.0.0/>
- [4] L. Baccala and K. Sameshima, “Partial directed coherence: A new concept in neural structure determination,” *Biological Cybernetics*, vol. 84, pp. 463–474, 05 2001.
- [5] M. Kamiński and K. J. Blinowska, “A new method of the description of the information flow in brain structures,” *Biological Cybernetics*, vol. 65, p. 203–210, 1991.
- [6] D. Alvarez-Estevez, “European data format.” [Online]. Available: <https://www.edfplus.info/index.html>
- [7] H. Nahrstaedt, “Pyedflib - edf/bdf toolbox in python.” [Online]. Available: <https://pyedflib.readthedocs.io/en/latest/>
- [8] M. K. Dominik Krzemiński, “Connectivipy - python connectivity module.” [Online]. Available: <https://connectivipy.readthedocs.io/en/latest/index.html>
- [9] A. A. Hagberg, D. A. Schult, and P. J. Swart, “Exploring network structure, dynamics, and function using networkx,” in *Proceedings of the 7th Python in Science Conference*, G. Varoquaux, T. Vaught, and J. Millman, Eds., Pasadena, CA USA, 2008, pp. 11 – 15.
- [10] S. S. H. Watts, Duncan J, “Collective dynamics of ‘small-world’ networks,” *Nature*, pp. 440–442, 06 1998. [Online]. Available: <https://doi.org/10.1038/30918>
- [11] G. K. Humphries Mark D., “Network ‘small-world-ness’: a quantitative method for determining canonical network equivalence,” *PLoS One*, vol. 3, no. 4, 04 2008.
- [12] R. A. Erdős, P., “On random graphs. i,” *Publicationes Mathematicae.*, vol. 6, p. 290–297, 1959.
- [13] R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenstat, M. Sheffer, and U. Alon, “Superfamilies of evolved and designed networks,” *Science*, vol. 303, no. 5663, pp. 1538–1542, 2004. [Online]. Available: <https://science.sciencemag.org/content/303/5663/1538>
- [14] O. Sporns and R. Kötter, “Motifs in brain networks,” *PLOS Biology*, vol. 2, no. 11, 10 2004. [Online]. Available: <https://doi.org/10.1371/journal.pbio.0020369>
- [15] V. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, “Fast unfolding of communities in large networks,” *Journal of Statistical Mechanics Theory and Experiment*, vol. 2008, 04 2008.

- [16] M. Rosvall, D. Axelsson, and C. T. Bergstrom, “The map equation,” *The European Physical Journal Special Topics*, vol. 178, no. 1, p. 13–23, Nov 2009. [Online]. Available: <http://dx.doi.org/10.1140/epjst/e2010-01179-1>
- [17] A. Clauset, M. E. J. Newman, and C. Moore, “Finding community structure in very large networks,” *Physical Review E*, vol. 70, no. 6, Dec 2004. [Online]. Available: <http://dx.doi.org/10.1103/PhysRevE.70.066111>
- [18] E. A. Leicht and M. E. J. Newman, “Community structure in directed networks,” *Phys. Rev. Lett.*, vol. 100, p. 118703, Mar 2008. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.100.118703>
- [19] “Minimum description length.” [Online]. Available: [https://en.wikipedia.org/wiki/Minimum\\_description\\_length](https://en.wikipedia.org/wiki/Minimum_description_length)