BIOINFORMATICS Class – PROJECT part 3

Neuroscience application

Brain Oscillatory and Network Activity during resting states

N.B. Make sure you read and understand all sections of this document. Comprehension of the instructions is part of the assignment. If you are unsure about any part, do contact the instructors to receive clarifications.

Aim

Goal of this project is to analyze two datasets of EEG data and to prepare a report on their comparison.

EEG data are recorded from 64 electrodes with subject at rest in (i) eyes-open and (ii) eyes-closed conditions, respectively.

Analyses will span the following topics:

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

Details on mandatory and optional analyses are provided below.

Dataset

The EEG data are available from PhysioNet, "EEG Motor Movement/Imagery Dataset" [[1]]. The whole dataset contains data acquired from 109 subjects, each containing 14 runs (files) of acquisition. Only the first two runs (SxxxR01 and SxxxR02) are relevant for this project: R01 is recorded during eyes-open (EO) resting state; R02 is recorded during eyes-closed (EC) resting state.

Select the subject according to Table 1, depending on the group you belong to (review groups' composition in this shared spreadsheet [[2]]).

Data is provided in EDF files (European Data Format). This format includes metadata, among which the sampling frequency and the channel labels. Several tools to read this format are freely available. For instance, a Python package can be found on GitHub [[3]], a Matlab mfile on MatlabCentral [[4]], an R package on CRAN (untested) [[5]]. As last option, if you cannot find a suitable tool for the programming language you chose you can use the PhysiobankATM utility [[6]] to export data in text (CSV) format.

Tasks

The analysis will be organized in tasks. At least one task per topic ("mandatory analysis") must be carried out. Optional analyses are associated with a *difficulty class*, ranging from 'A' (easiest) to 'E' (most difficult). <u>At least 4 optional analyses of class 'C' or higher must be carried out in the project</u>.

The list of mandatory and optional tasks follows:

---

[1] https://physionet.org/physiobank/database/eegmmidb/ On this web page you also find a thorough description of the dataset

[2] https://drive.google.com/open?id=1hCbXQs36MPil28t50e2YEH6xPRtk--Ztj9hIVX6VVn8

[3] https://github.com/holgern/pyedflib/blob/master/demo/readEDFFile.py

[4] https://it.mathworks.com/matlabcentral/fileexchange/31900-edfread

[5] https://cran.r-project.org/web/packages/edfReader/

[6] https://physionet.org/cgi-bin/atm/ATM

1. **Spectral analysis**
    1.1. (**mandatory**) Select relevant channel and estimate Power Spectral Density (PSD) using one of the methods introduced during the course. Justify your choice of channel selection and parameters used for PSD estimation.
    1.2. (class 'B') Compare two different methods of PSD estimation.
    1.3. (class 'C') Compare different choices of estimation parameters.
    1.4. (class 'D') Make a multivariate comparison of PSDs, considering simultaneously two independent variables (channels resting state).
    1.5. (class 'E') Assign a statistical significance value to the differences in PSDs between the two rest conditions.
2. **Connectivity graph**
    2.1. (**mandatory**) Estimate functional brain connectivity among 64 channels using one of the MVAR estimators: Partial Directed Coherence (PDC), Direct Transfer Function (DTF). Select one relevant frequency value. Apply a threshold so that the resulting binary connectivity matrices have network density equal to 20%. Create a graphical representation of the binary adjacency matrix.
    2.2. (class 'A') Perform task 2.1 using both estimators (PDC and DTF).
    2.3. (class 'A') Perform task 2.1 using thresholds yielding the following density values: 1%, 5%, 10%, 20%, 30%, 50%.
    2.4. (class 'D') Considering the subset of 19 channels suggested in Figure 1 and Table 2, estimate the connectivity using PDC or DTF and apply a statistical validation method (asymptotic statistics[7], resampling procedure[8],...) to filter out values that are not significantly different from 0 ($PDC(i,j) \neq 0 \; with \; p < 5\%$).
    2.5. (class 'C') Make a topographical representation of of the networks (see example in Figure 2). Cartesian coordinates of planar representation of EEG channels are available in Table 3 (see also the file channel_locations.txt).
    (the choice of this task is advised in the case of 19-channel networks and/or density ≤ 5%).
    2.6. (class 'B') Perform task 2.1 considering a second frequency value belonging to a different EEG rhythm with respect to the first choice.
3. **Graph theory indices**
    3.1. (**mandatory**) Compute binary global (clustering coefficient, path length) and local (degree, in/out-degree) graph indices. List the highest 10 channels for local indices.
    3.2. (class 'D') Search in the literature a definition of small-worldness index (i.e. an index describing the small-world organization of a network) and compute it.
    3.3. (class 'B') Compare the global indices extracted from PDC and DTF connectivity estimations.
    3.4. (class 'C') Study the behaviour of global graph indices in function of network density (see point 2.3 for density values).
    (the choice of this task is advised in the case of selection of task 2.3).
    3.5. (class 'B') Make a topographical representation of local indices.
    3.6. (class 'B') Compare the networks obtained with the analysis 2.6 in terms of graph indices.
    (the choice of this task is advised only in the case of selection of task 2.6).
    3.7. (class 'C') Perform point 3.1 considering the weighted version of the graph indices definitions.
4. **Motif analysis**
    4.1. (**mandatory**) Perform motifs analysis to investigate the presence of 3-node configurations in the networks: determine their frequency and statistical significance (motifs, anti-motifs).
    4.2. (class 'C') For the motif with pattern $A \rightarrow B \leftarrow C$, create a topographical representation of the networks considering only the connections involved in this configuration.

---

[7] as in http://www.lcs.poli.usp.br/~baccala/pdc/
[8] as in http://connectivipy.readthedocs.io/en/latest/tutorial.html

4.3. (class ‘C’) Choose a channel selected in parieto-occipital scalp region and determine the motifs which involve it.

4.4. (class ‘E’) Perform the same analysis described in task 4.1 considering 4-node motifs.

5. **Community detection**

5.1. (**mandatory**) Determine number and composition (i.e. list of nodes) of the communities obtained applying one of the algorithms introduced during the course.

5.2. (class ‘B’) Make a graphical representation of the community structure in both rest conditions.

5.3. (class ‘C’) Compare the community structure obtained by means of two different methods (modularity-based vs information theory-based approaches).

Format of the report

No template is provided. Use the freedom in choosing your format wisely (i.e. sloppy formats will be penalized).

List the authors’ names and the group’s number on the first page.

The expected length of the report’s text is 4 pages. Body text font sizes smaller than 11pt should be avoided.

Any number of figures and tables (with their respective legends) can be added in excess of this page limit. Please arrange them in a compact layout to save pages.
A table containing a list of the tasks chosen must be included in the report (see Table 0 for an example).

Do acknowledge any source you used [9], such as software code, third party figures, cited text, contribution by non-authors, etc. Failure to comply with this requirement will be considered plagiarism, and may result in (severe) penalties.

Evaluation criteria

The mark will be based on: clarity of writing, accuracy of the methods’ description, completeness and appropriate presentation of results (including quality of figures and tables), appropriate discussion of the outcome of comparisons, overall structure and format of the report.

Use as guideline the material of the ‘Scientific and technical writing’ seminar held during the course by Dr. Petti.

Selecting tasks in number and difficulty class above the required minimum has a positive impact on the mark.

Submission procedure

Each group will submit two files:

i. a PDF file containing the report;
ii. a compressed archive (zip, rar, 7z) containing the software code used to perform the analyses. The latter archive may contain intermediate results (unedited figures, raw output files, etc), if appropriate. Do not include EEG files in the archive.

The files must be named according to the following scheme:

`bioinf_proj_neuro_group-<nn>.<ext>`,

where <nn> is the two-digit group number and `<ext>` is either ‘`pdf`’ or {‘`zip`’|‘`rar`’|‘`7z`’}[10]

Files will be handed in using a Piazza private note addressed to all instructors and to the co-authors of the project. See Figure 3 for hints on how to fill in the note.

[9] With the exception of course material provided by the instructors

[10] For instance: `bioinf_proj_neuro_group-04.pdf`

The report must be handed in by **Friday 12 January 2018**. Late submissions (up to one week) will be subjected to mark penalties, increasing for every day of delay.

Figures and Tables

Table 0. Example of list of tasks chosen for the project

Task	Class
1.1	mandatory
1.3	C
2.1	mandatory
2.2	A
2.4	D
3.1	mandatory
3.4	C
4.1	mandatory
4.3	C
5.1	mandatory
5.2	B

Table 1. Association between project groups and experimental subjects

Group	1	2	3	4	5	6	7	8
Subject	S002	S003	S004	S040	S050	S059	S064	S038

Group	9	10	11	12	13	14	15	16
Subject	S070	S071	S072	S076	S079	S086	S089	S095

Table 2. List of labels of the reduced set of 19 channels

#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Label	Fp1	Fp2	F7	F3	Fz	F4	F8	T7	C3	Cz	C4	T8	P7	P3	Pz	P4	P8	O1	O2

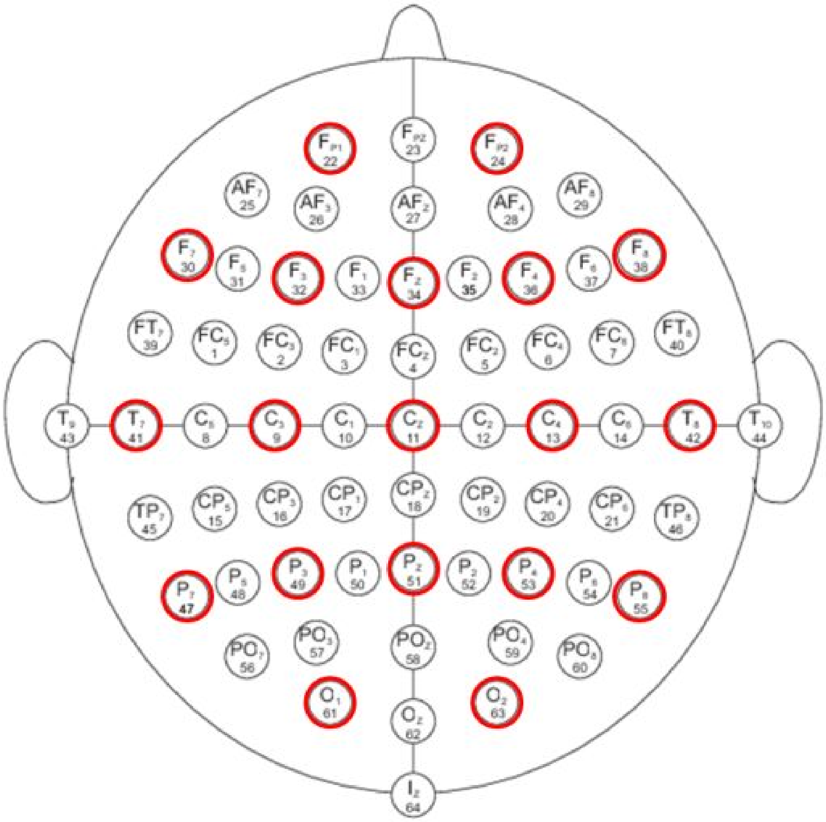

Figure 1. Topographical representation of electrode position and labels, according to the international 10-20 system. Red circles point out the reduced set of 19 channels.

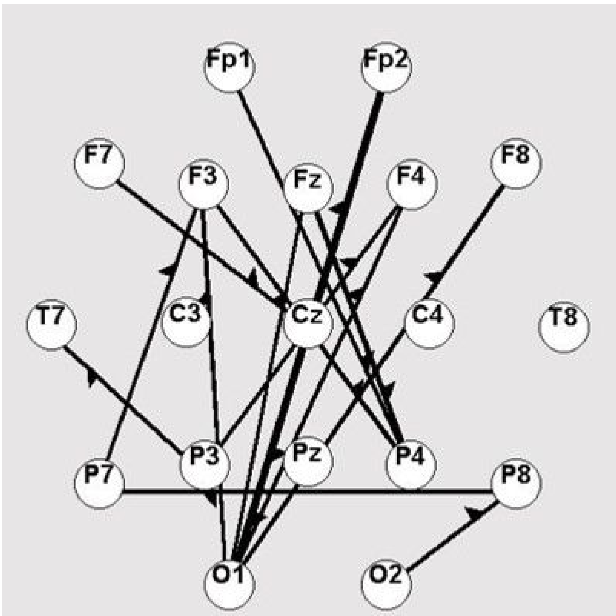

Figure 2. Mockup of 19-channel binary directed connectivity network with density d=.05 (top view, nose up).

Table 3. Cartesian coordinates of planar representation of EEG channels (top view, nose up)[11]

#	label	x	y
1	Fc5.	-0.28	0.11
2	Fc3.	-0.19	0.11
3	Fc1.	-0.09	0.10
4	Fcz.	0.00	0.10
5	Fc2.	0.09	0.10
6	Fc4.	0.19	0.11
7	Fc6.	0.28	0.11
8	C5..	-0.30	0.00
9	C3..	-0.20	0.00
10	C1..	-0.10	0.00
11	Cz..	0.00	0.00
12	C2..	0.10	0.00
13	C4..	0.20	0.00
14	C6..	0.30	0.00
15	Cp5.	-0.28	-0.11
16	Cp3.	-0.19	-0.11
17	Cp1.	-0.09	-0.10
18	Cpz.	0.00	-0.10
19	Cp2.	0.09	-0.10
20	Cp4.	0.19	-0.11
21	Cp6.	0.28	-0.11
22	Fp1.	-0.13	0.39
23	Fpz.	0.00	0.41
24	Fp2.	0.13	0.39
25	Af7.	-0.24	0.33
26	Af3.	-0.13	0.32
27	Afz.	0.00	0.30
28	Af4.	0.13	0.32
29	Af8.	0.24	0.33
30	F7..	-0.33	0.24
31	F5..	-0.26	0.25
32	F3..	-0.16	0.23

#	label	x	y
33	F1..	-0.28	0.11
34	Fz..	-0.19	0.11
35	F2..	-0.09	0.10
36	F4..	0.00	0.10
37	F6..	0.09	0.10
38	F8..	0.19	0.11
39	Ft7.	0.28	0.11
40	Ft8.	-0.30	0.00
41	T7..	-0.20	0.00
42	T8..	-0.10	0.00
43	T9..	0.00	0.00
44	T10.	0.10	0.00
45	Tp7.	0.20	0.00
46	Tp8.	0.30	0.00
47	P7..	-0.28	-0.11
48	P5..	-0.19	-0.11
49	P3..	-0.09	-0.10
50	P1..	0.00	-0.10
51	Pz..	0.09	-0.10
52	P2..	0.19	-0.11
53	P4..	0.28	-0.11
54	P6..	-0.13	0.39
55	P8..	0.00	0.41
56	Po7.	0.13	0.39
57	Po3.	-0.24	0.33
58	Poz.	-0.13	0.32
59	Po4.	0.00	0.30
60	Po8.	0.13	0.32
61	O1..	0.24	0.33
62	Oz..	-0.33	0.24
63	O2..	-0.26	0.25
64	Iz..	-0.16	0.23

[11] See also text file channel_locations.txt

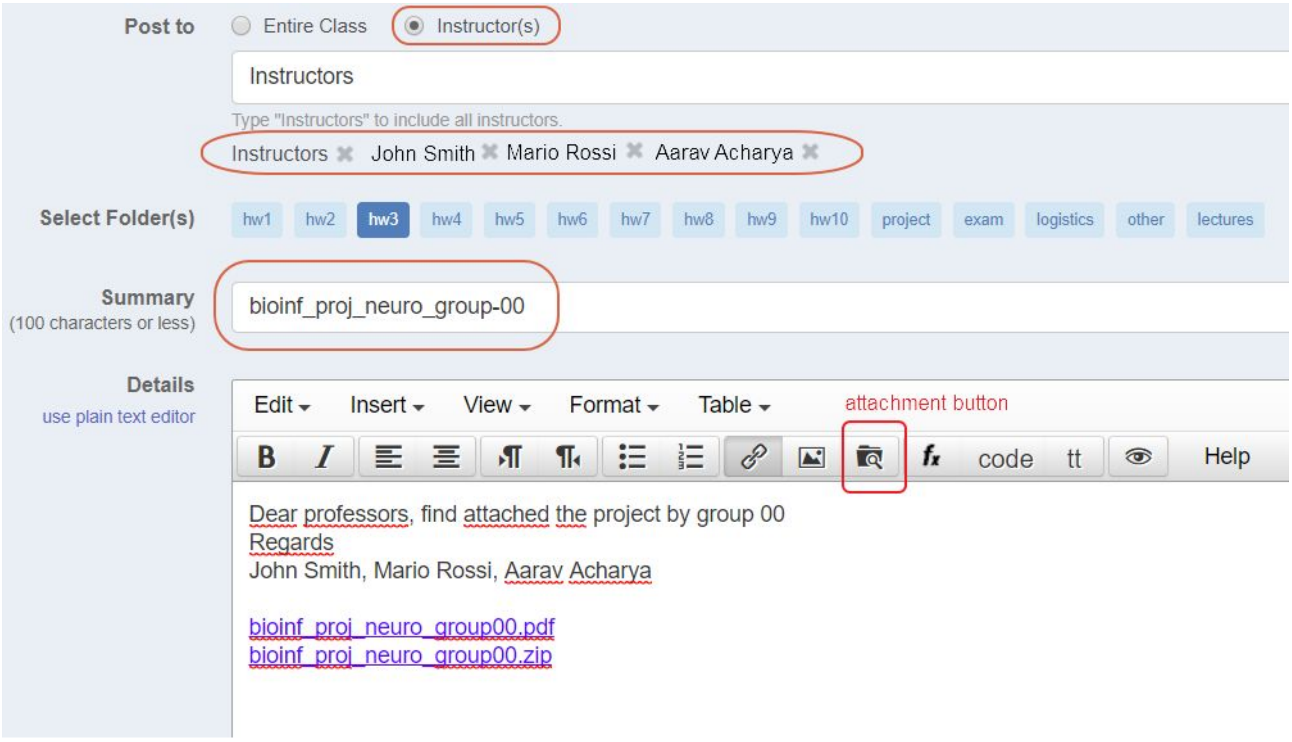

Figure 3. Screenshot of the submission message using a Piazza private note. Note that: recipients are (i) all instructors and (ii) all coauthors; the subject of the note has the same format as the filenames described in section 'Submission procedure'; files (pdf and archive) are attached using the attachment button.