NBS-Predict Tutorial

Brief Outline

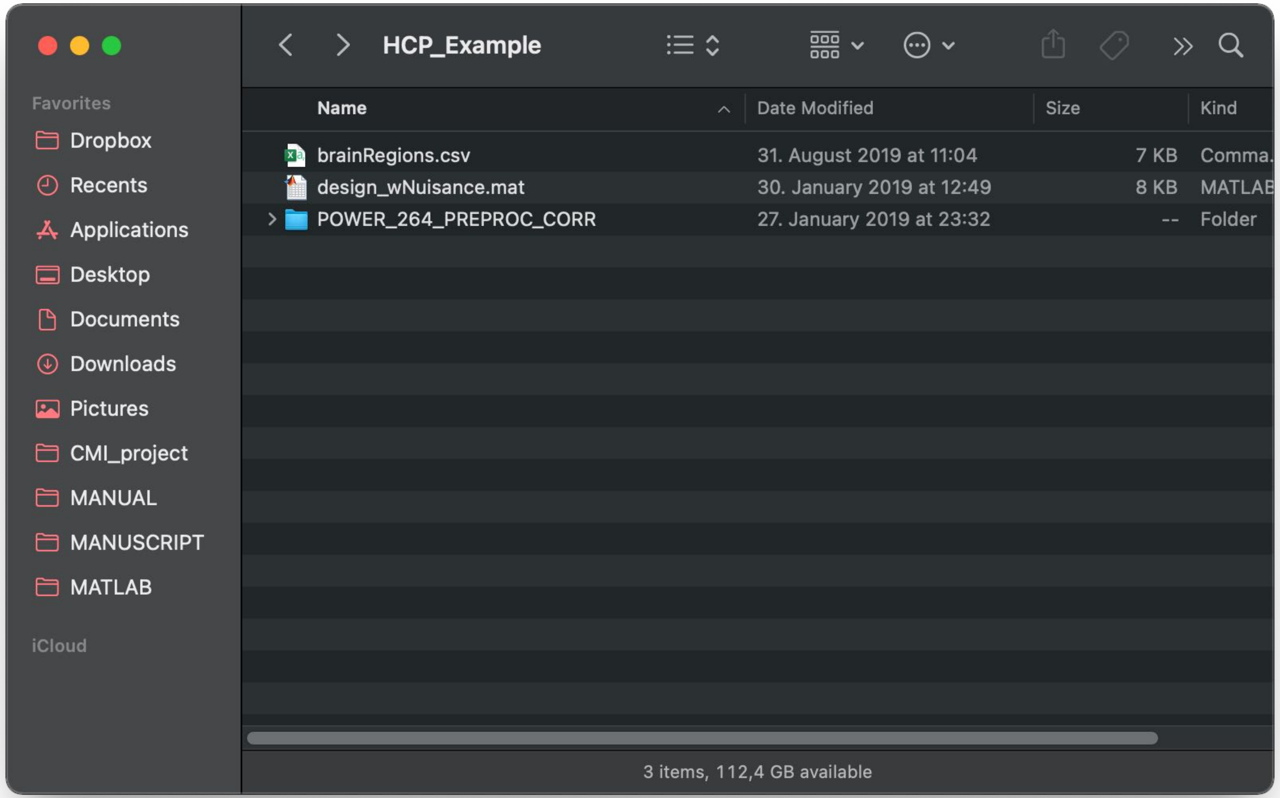
- **Aim**: To extract biomarkers of intelligence from resting-state fMRI (rs-fMRI) images.
- **Method:**
    - The sample consists of 897 individuals (μ_{age} = 28.76, σ_{age} = 3.69, 406 males) from the 1200-subject release of the Human Connectome Project (Glasser et al., 2016).
    - General intelligence scores are used as target variables (Heaton et al., 2014).
    - Functional connectivity matrices are used as features.
    - The rs-fMRI data was processed as in Kruschwitz et al., 2018, and nodes were delineated using a 264-region functional atlas (Power et al., 2011)
    - Correlations between time series from each region were computed to create 264x264 connectivity matrices.
    - NBS-Predict (10-repetitions 5-fold CV, p = 0.01) with hyperparameter optimization is applied.
    - Age and gender are regressed out, as shown in Snoek et al., 2019.

Glasser, M.F., Smith, S.M., Marcus, D.S., Andersson, J.L.R., Auerbach, E.J., Behrens, T.E.J., Coalson, T.S., Harms, M.P., Jenkinson, M., Moeller, S., Robinson, E.C., Sotiropoulos, S.N., Xu, J., Yacoub, E., Ugurbil, K., Van Essen, D.C., 2016. The Human Connectome Project's neuroimaging approach. Nature Neuroscience.
Heaton, R.K., Akshoomoff, N., Tulsky, D., Mungas, D., Weintraub, S., Dikmen, S., Beaumont, J., Casaletto, K.B., Conway, K., Slotkin, J., Gershon, R., 2014. Reliability and validity of composite scores from the NIH toolbox cognition battery in adults. Journal of the International Neuropsychological Society.
Kruschwitz, J.D., Waller, L., Daedelow, L.S., Walter, H., Veer, I.M., 2018. General, crystallized and fluid intelligence are not associated with functional global network efficiency: A replication study with the human connectome project 1200 data set. Neuroimage.
Power, J.D., Cohen, A.L., Nelson, S.M., Wig, G.S., Barnes, K.A., Church, J.A., Vogel, A.C., Laumann, T.O., Miezin, F.M., Schlaggar, B.L., Petersen, S.E., 2011. Functional Network Organization of the Human Brain. Neuron.

Input Directory

Input directory should include:

- file[1] containing brain regions.
- file[1] containing design matrix.
- Subdirectory containing correlation matrices[1]



[1] .csv or .mat

Brain Regions

Brain regions file[1] should include 4 columns:

- First 3 columns represent x, y, z MNI coordinates of regions.
- 4th column contains labels.

brainRegion... Open with Microsoft Excel

-25	-98	-12	Uncertain
27	-97	-13	Uncertain
24	32	-18	Uncertain
-56	-45	-24	Uncertain
8	41	-24	Uncertain
-21	-22	-20	Uncertain
17	-28	-17	Uncertain
-37	-29	-26	Uncertain
65	-24	-19	Uncertain
52	-34	-27	Uncertain
55	-31	-17	Uncertain
34	38	-12	Uncertain
-7	-52	61	Sensory/somatomotor Hand
-14	-18	40	Sensory/somatomotor Hand
0	-15	47	Sensory/somatomotor Hand
10	-2	45	Sensory/somatomotor Hand
-7	-21	65	Sensory/somatomotor Hand
-7	-33	72	Sensory/somatomotor Hand
13	-33	75	Sensory/somatomotor Hand
-54	-23	43	Sensory/somatomotor Hand
29	-17	71	Sensory/somatomotor Hand
10	-46	73	Sensory/somatomotor Hand
-23	-30	72	Sensory/somatomotor Hand
-40	-19	54	Sensory/somatomotor Hand
29	-39	59	Sensory/somatomotor Hand
50	-20	42	Sensory/somatomotor Hand
-38	-27	69	Sensory/somatomotor Hand
20	-29	60	Sensory/somatomotor Hand
44	-8	57	Sensory/somatomotor Hand
-29	-43	61	Sensory/somatomotor Hand
10	-17	74	Sensory/somatomotor Hand
22	-42	69	Sensory/somatomotor Hand
-45	-32	47	Sensory/somatomotor Hand
-21	-31	61	Sensory/somatomotor Hand
-13	-17	75	Sensory/somatomotor Hand
42	-20	55	Sensory/somatomotor Hand
-38	-15	69	Sensory/somatomotor Hand
-16	-46	73	Sensory/somatomotor Hand
2	-28	60	Sensory/somatomotor Hand
3	-17	58	Sensory/somatomotor Hand
38	-17	45	Sensory/somatomotor Hand

[1] .csv or .mat

Design Matrix File

Design matrix[1] should comprise at least two columns, which represents:

- One-hot encoded group labels (for classification) or
- Intercept + Target (for regression)

Confounds could be included as additional columns.

Here, we have (1) intercept and (2) general intelligence scores **(target)** as well as (3) age and (4) gender[2] as **confounds**.

1	2	3	4
1	113.7443	27	0
1	103.0871	27	1
1	113.7671	33	0
1	116.3485	27	0
1	83.3218	35	1
1	107.2565	22	0
1	98.9149	29	0
1	102.7344	29	0
1	99.3000	35	1
1	112.2568	24	0
1	101.3789	27	0
1	99.3443	26	1
1	94.6005	30	0
1	114.9372	23	0
1	108.3393	26	0
1	95.5643	30	1
1	95.6871	25	0
1	103.9179	30	0
1	103.6322	34	0
1	94.3286	25	1
1	85.0929	28	1
1	104.5243	32	1
1	90.2893	26	1
1	118.5184	31	1
1	95.7895	36	1
1	100.4214	29	1
1	106	31	0
1	121.1486	27	0
1	116.8977	34	1
1	101.5736	33	1
1	105.9314	33	1
1	109.8143	28	0
1	119.8156	29	1
1	114.2565	22	0
1	101.3628	26	1
1	100.7611	26	1
1	119.6529	25	1
1	119.2894	22	0
1	96.2090	35	1

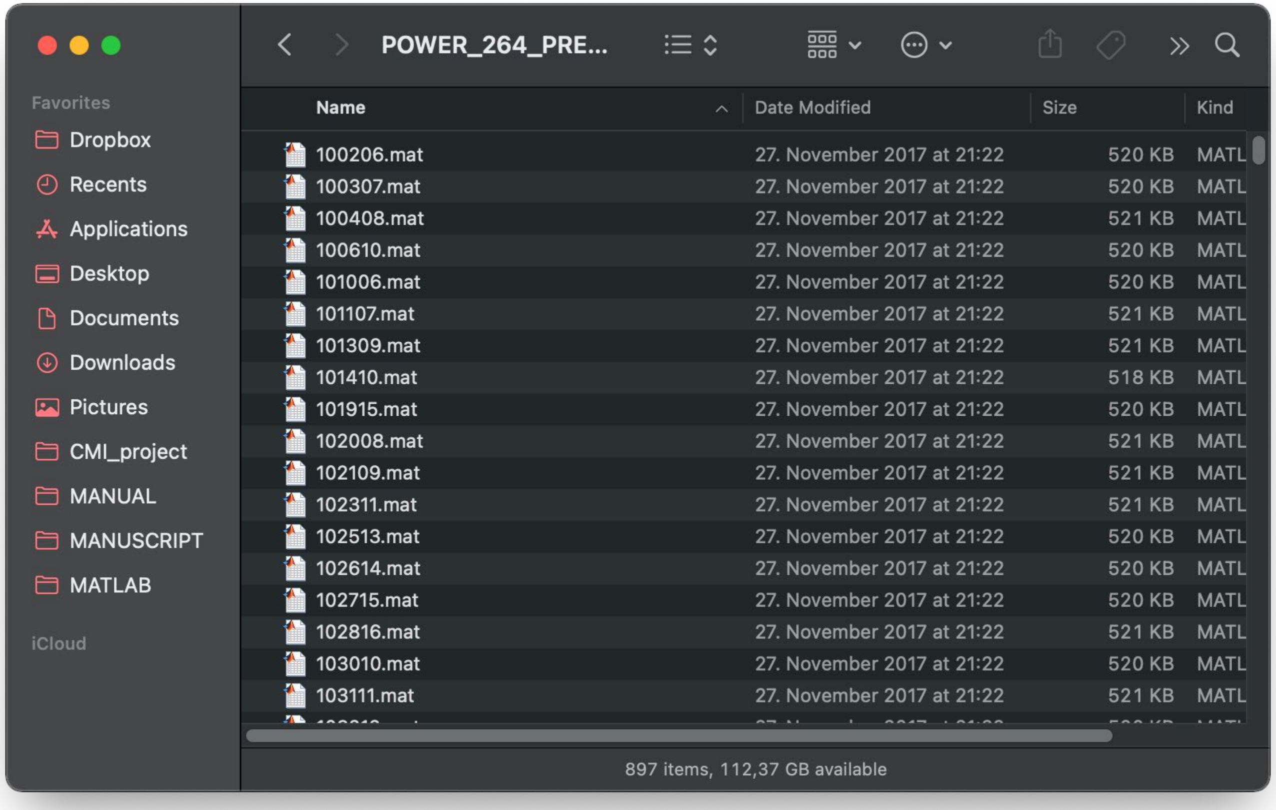
[1] .csv or .mat

[2] Categorical variables should be one-hot encoded.

Correlation Matrices

Subjects' correlation matrices[1] should locate under a distinct subdirectory.

Simply use the GraphVar Toolbox to generate connectivity matrices if you don't have them yet.



[1] .csv or .mat

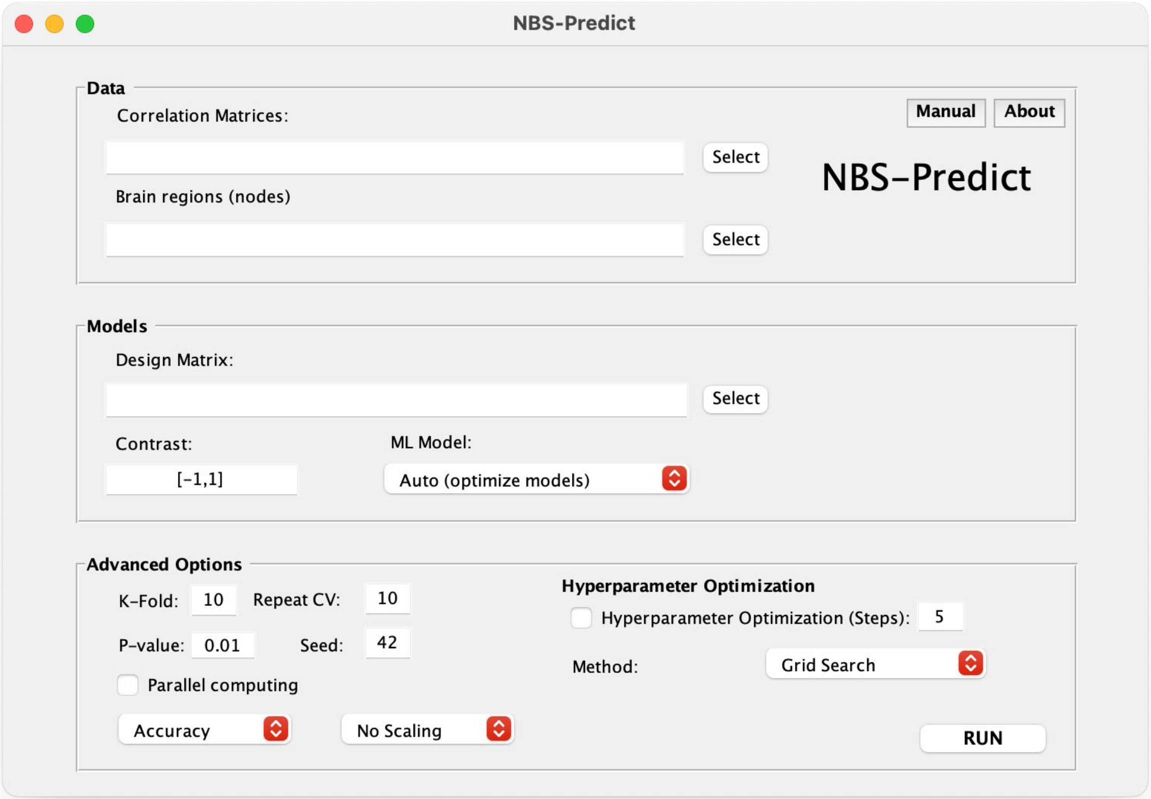
Kruschwitz, J. D., List, D., Waller, L., Rubinov, M., & Walter, H. (2015). GraphVar: A user- friendly toolbox for comprehensive graph analyses of functional brain connectivity. Journal of Neuroscience Methods, 245, 107–115.

Analysis Setup

Open MATLAB and type

"start_NBSPredict"[1] in command window.

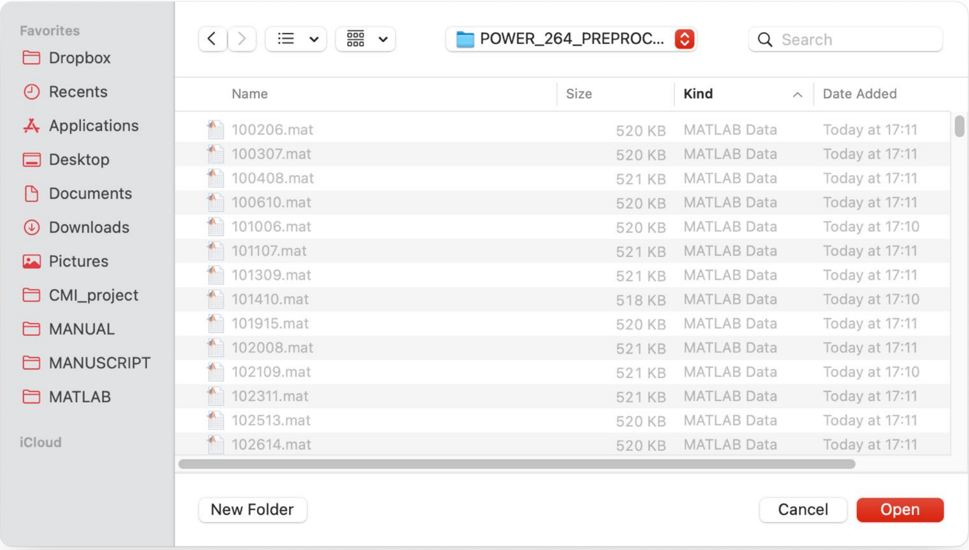
The GUI appears on the screen.

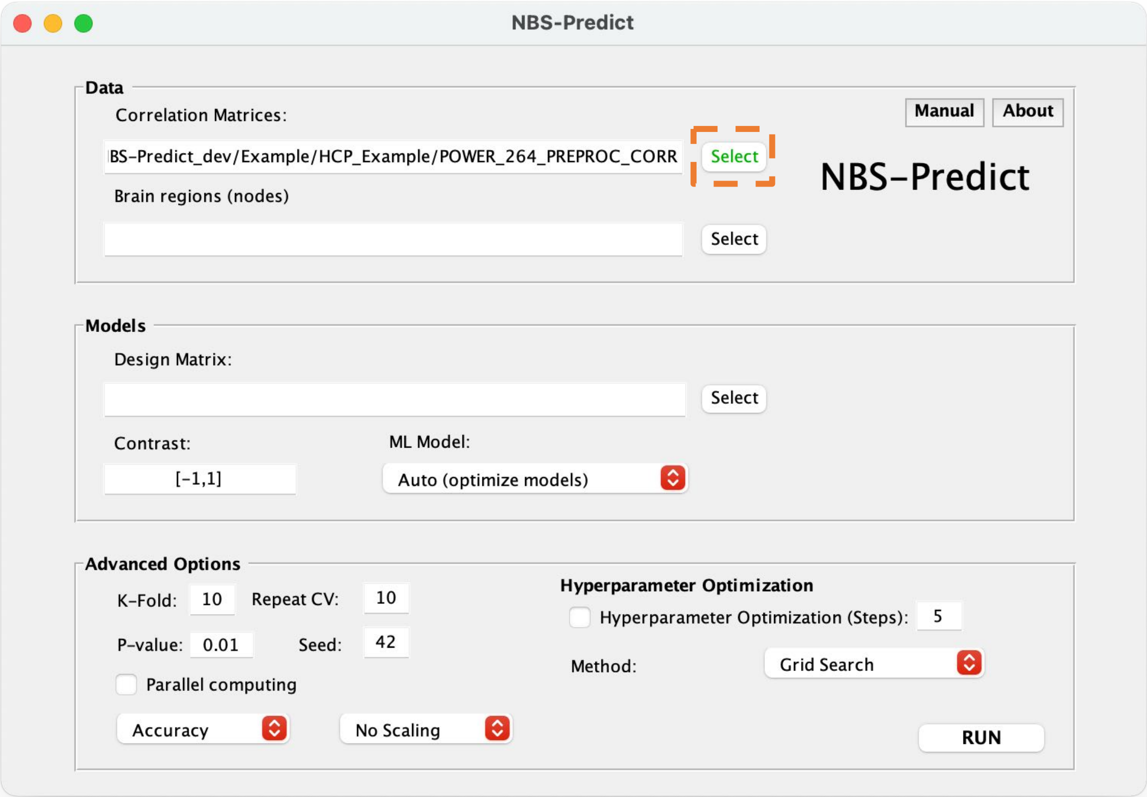


[1] Make sure you already added NBS-Predict to your MATLAB path.

Data Panel

Select directory containing correlation matrices[1].

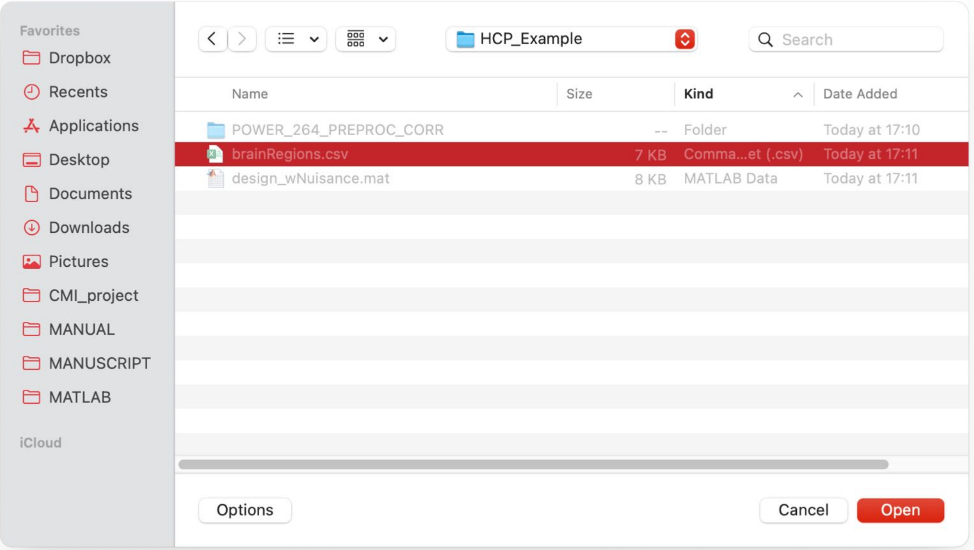


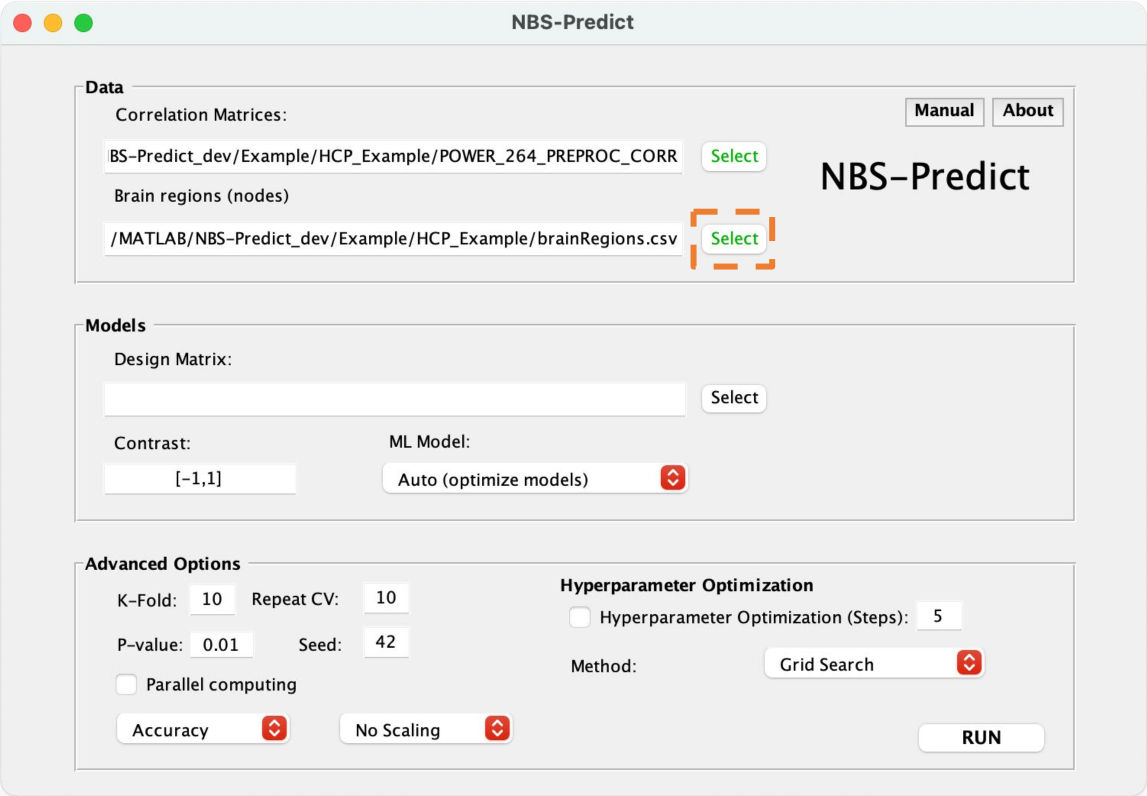


[1] .csv or .mat

Data Panel

Select a brain parcellation file[1].

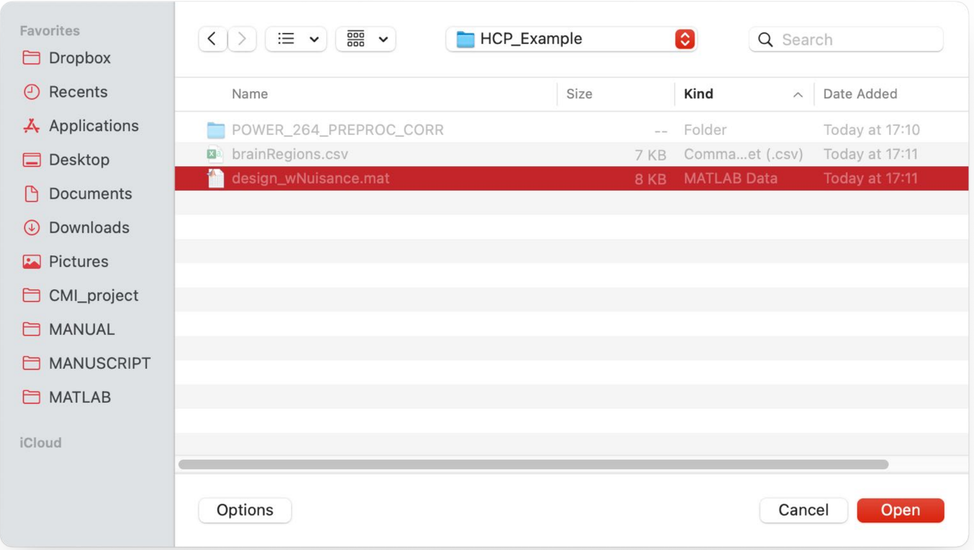


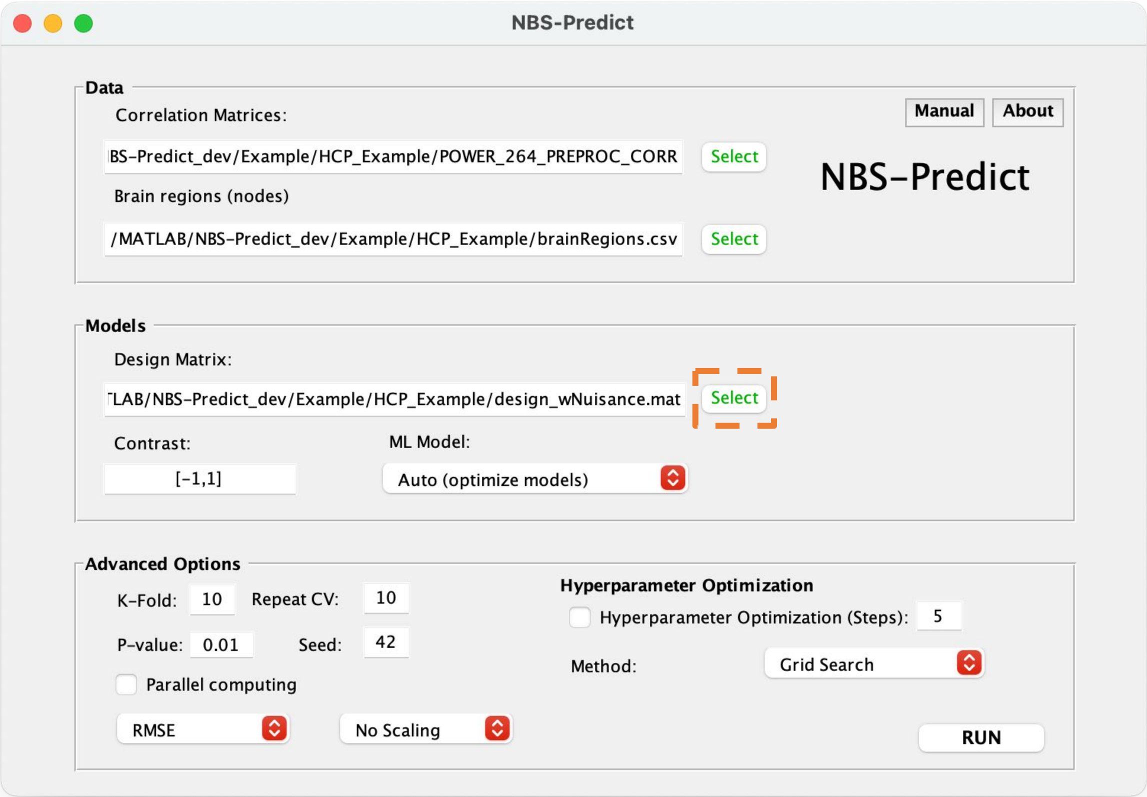


[1] .csv or .mat

Models Panel

Select a design matrix[1].



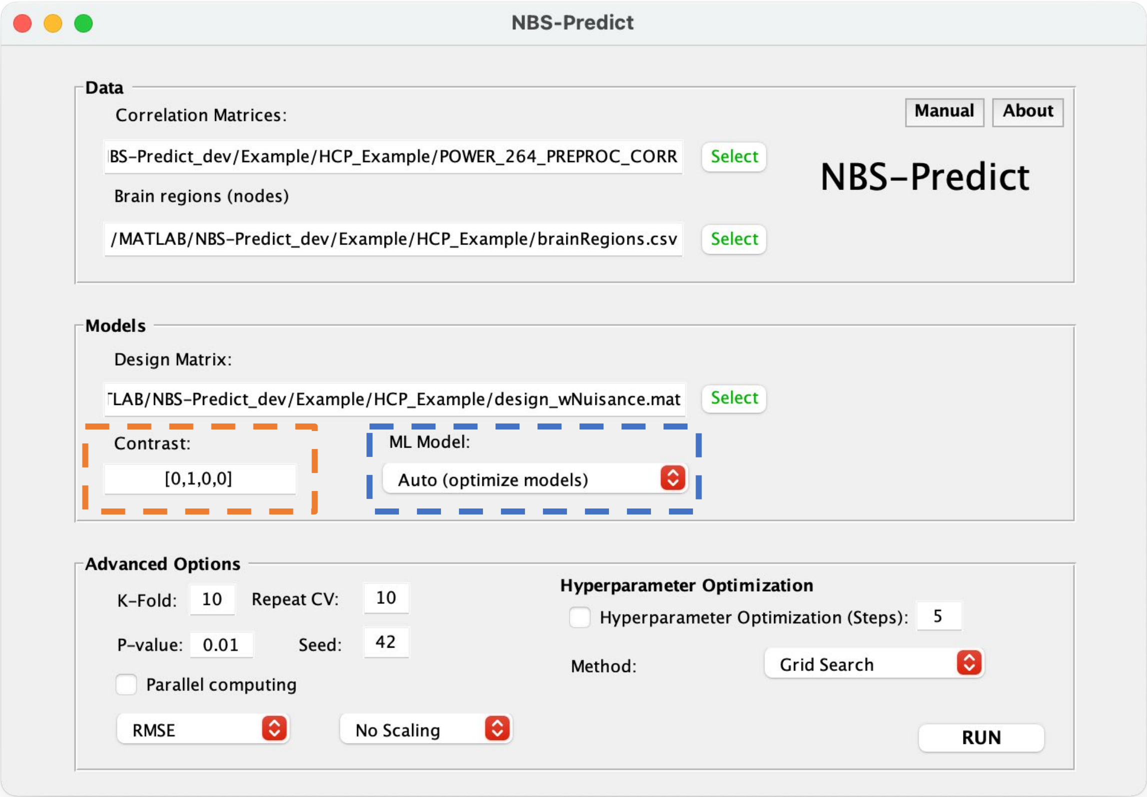


[1] .csv or .mat

Models Panel

1. We specified contrast as [0,1,0,0], meaning general intelligence scores will be used as the **target**, and age and gender as **nuisance variables**.

2. NBS-Predict allows us to use Decision Tree Regressor, SVM Regressor, and Linear Regression for our specific design (i.e. regression). Simply leave as **Auto (optimize models)**.

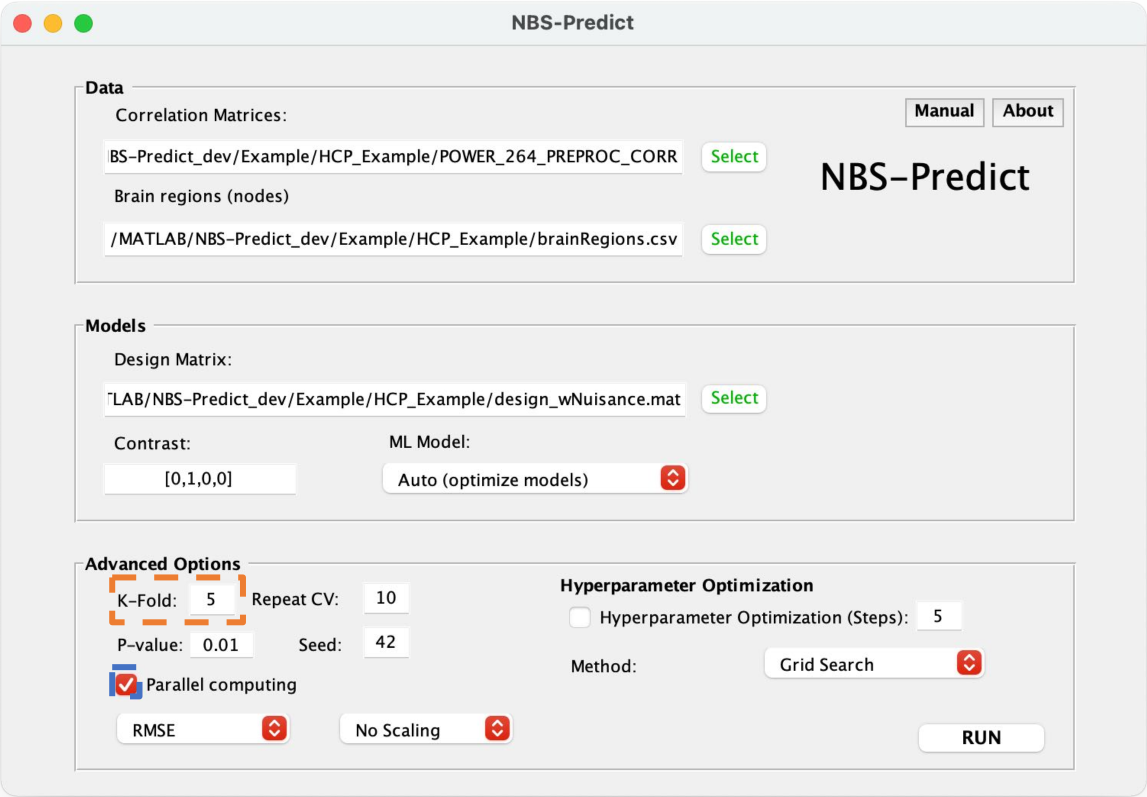


Advanced Options

1. We set K-fold to 5 (for faster computation) and number of CV repetition to 10.

2. Leave p-value (0.01) and seed (42) as default.

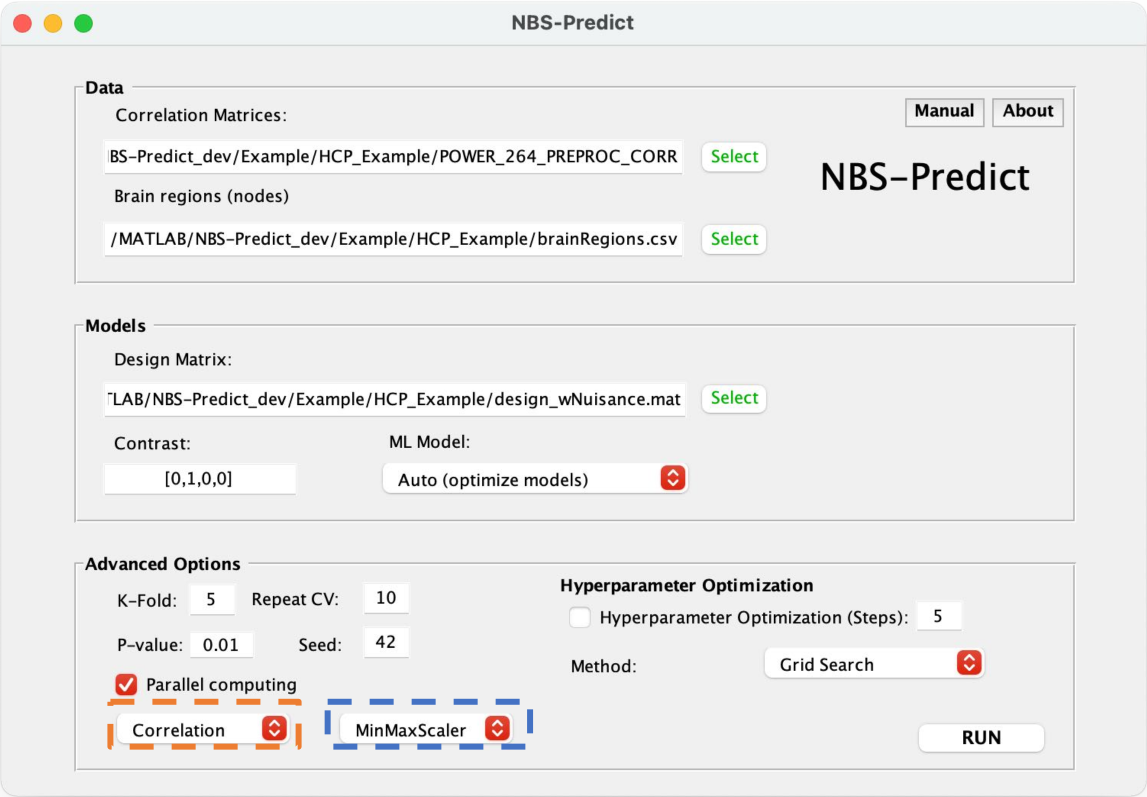
3. Activate parallel computing (if you have MATLAB Parallel Computing Toolbox installed on your computer).



Advanced Options

1. NBS-Predict offers wide selection of performance metrics suitable for our design. Simply select “Correlation”.

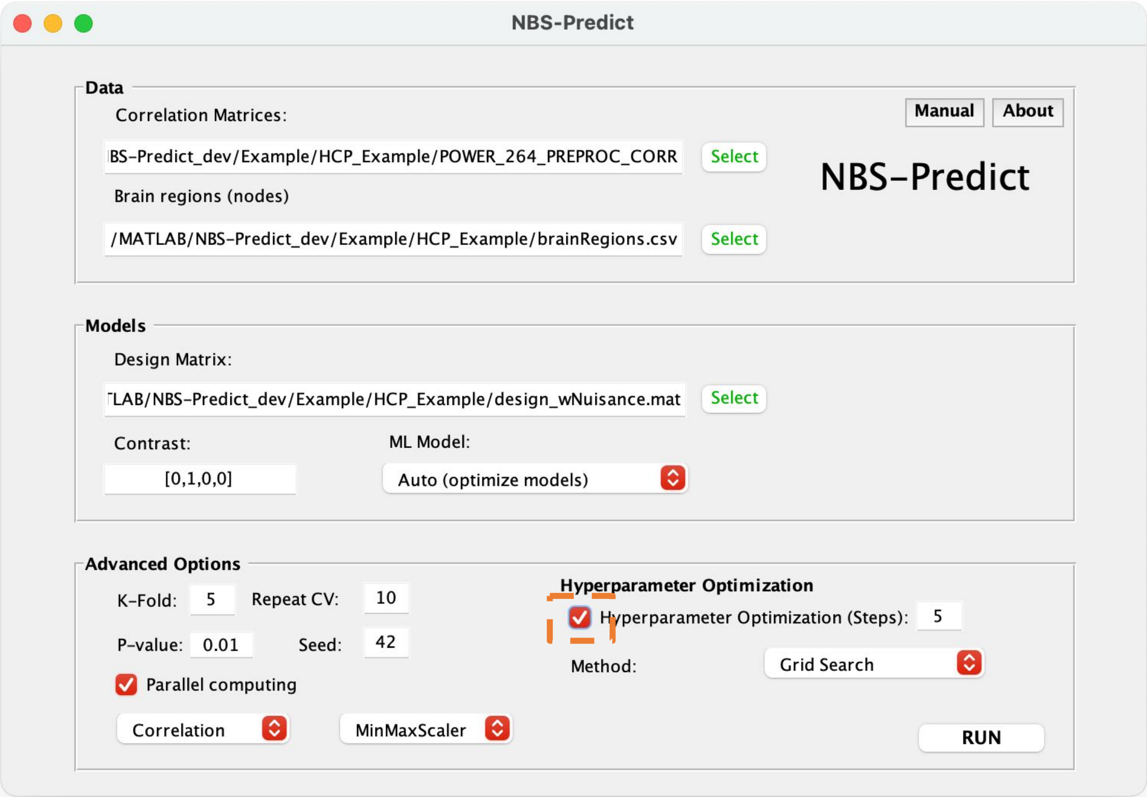
2. Regarding scaling, MinMaxScaler, MaxAbsScaler and StandardScaler are provided. Select “MinMaxScaler”.



Advanced Options

1. Activate Hyperparameter Optimization, and leave optimization steps as default.

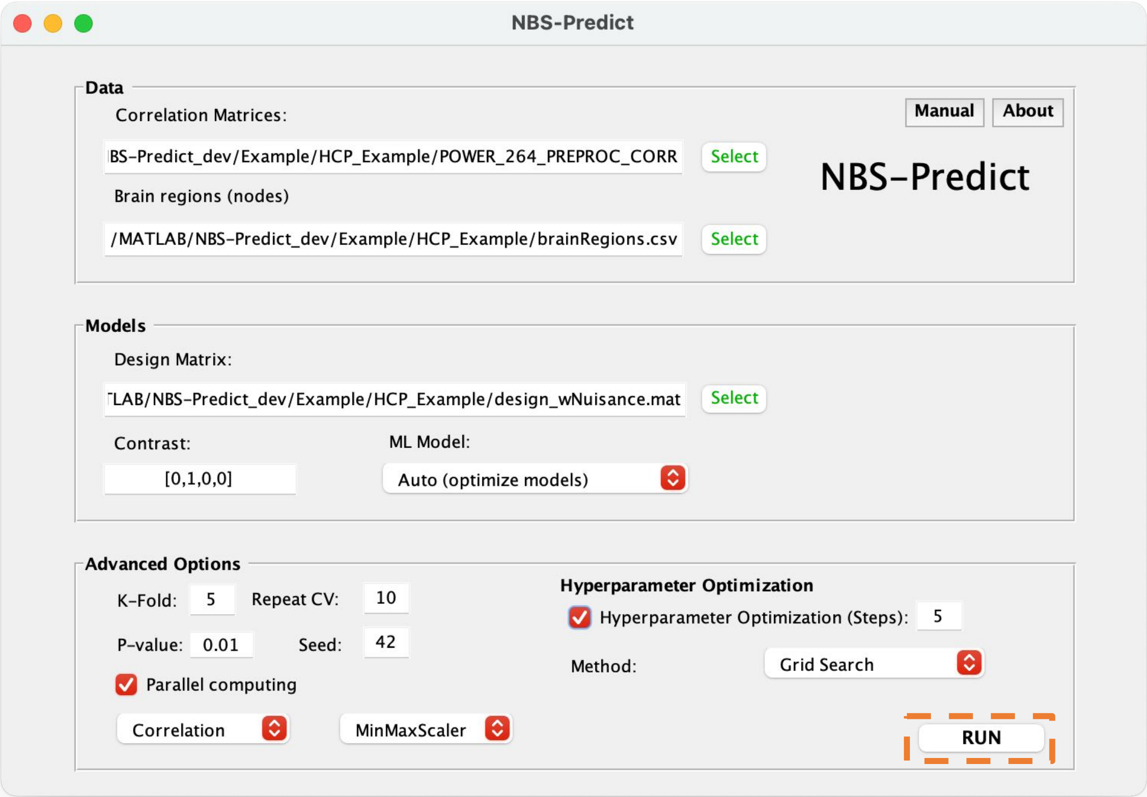
2. You can use Grid Search, Random Search and Bayesian Optimization as searching algorithm. Simply leave as Grid Search.



Run!

We are now ready to run NBS-Predict!

Hit the ”RUN” button!



Analysis

You can check the progress of the analysis

on MATLAB's Command Window.

```
ESTIMATOR: LinReg
Searching Algorithm: gridSearch
METRIC: correlation
Number of Folds: 5
Number of Repetitions: 10
-------------
|   Score   |
-------------
|   0.214   |
|   0.227   |
|   0.215   |
|   0.232   |
|   0.202   |
|   0.190   |
|   0.210   |
|   0.174   |
|   0.191   |
```

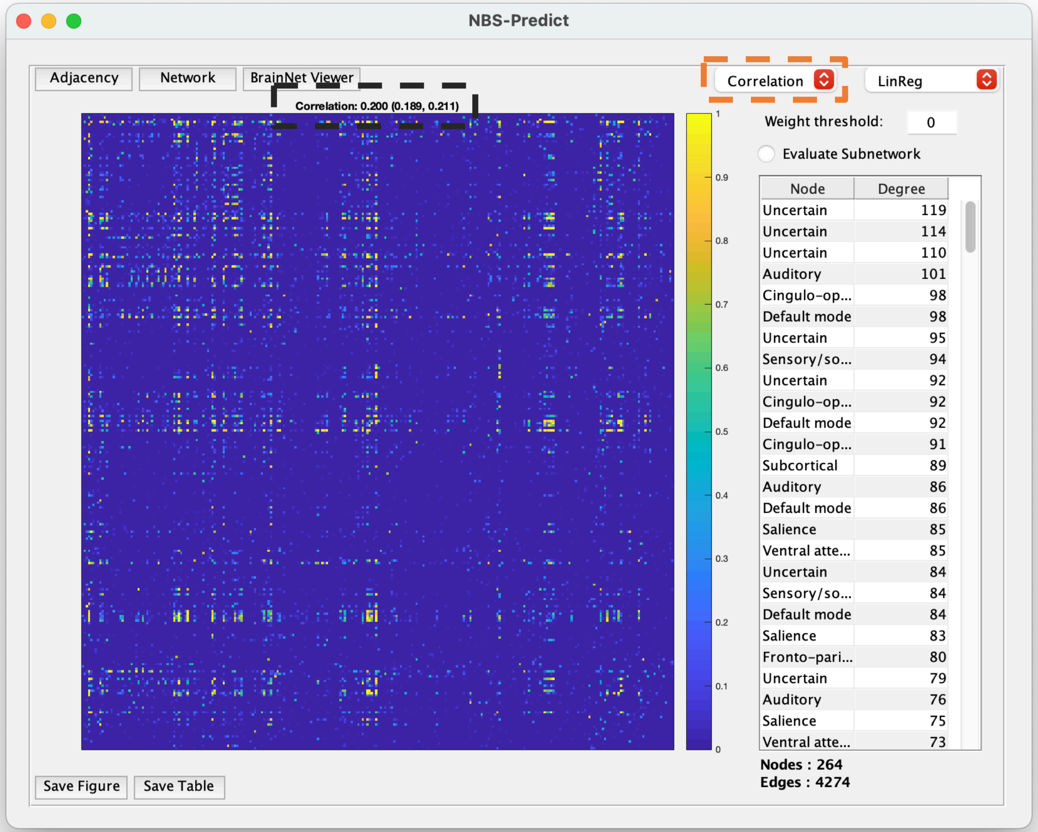
Results Viewer

After the analysis is finished, Results Viewer will automatically pop-up.

Notice that NBS-Predict predicted general intelligence with Pearson’s CC of 0.200.

You can also check other performance metrics (e.g. RMSE is 8.690).

Notice that Linear Regression is the best performing algorithm on this specific dataset.

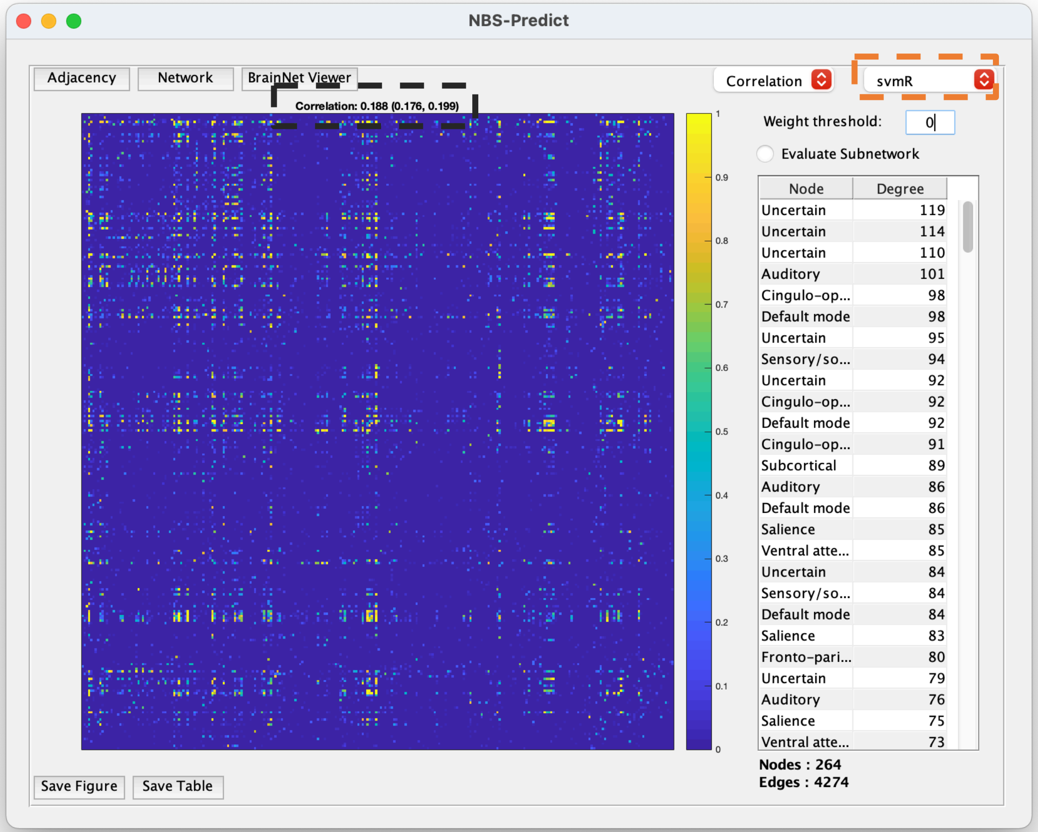


Results Viewer

You can also check results from the other algorithms (e.g. SVM Regressor)

Notably, the network would be almost the same[1], because the edges are selected using a connectome-based filter-based selection algorithm.

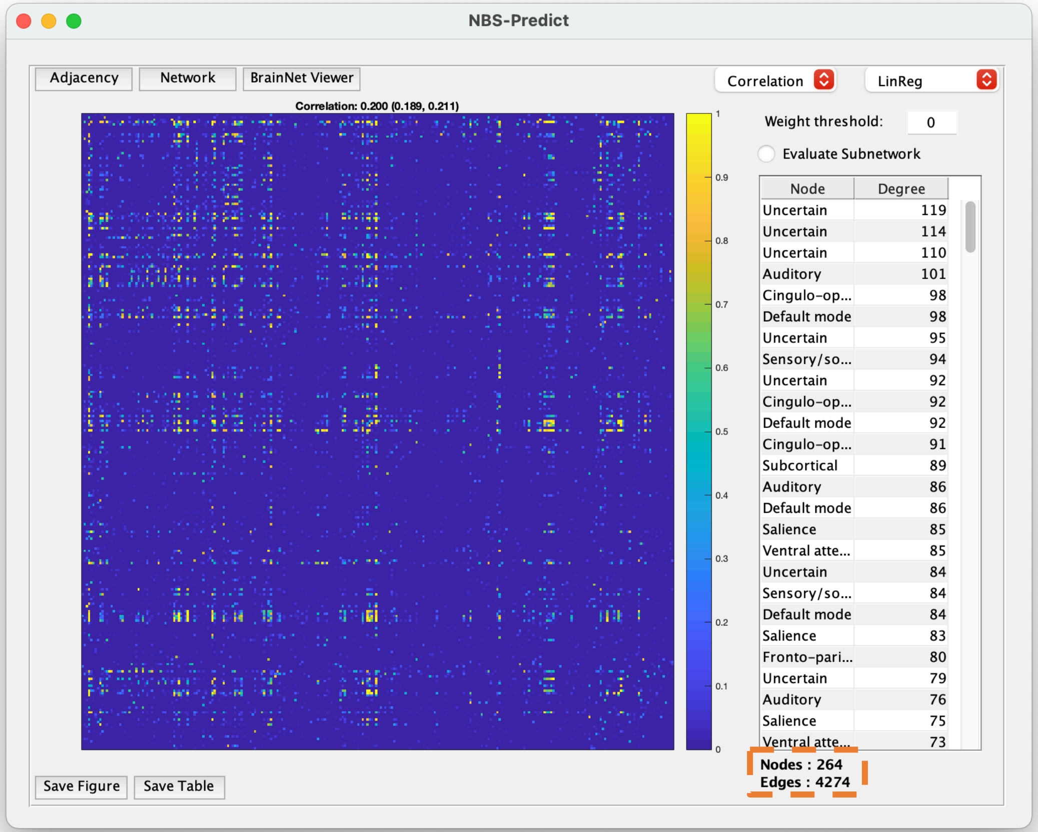
In this case, it looks exactly the same.



[1] Possibly minor variation due to randomized K-fold CV

Results Viewer

You will see a weighted adjacency matrix. In the weighted adjacency matrix, **4274 edges among 264 nodes** have weight values above 0. That is, almost 30k edges have not been selected in any folds during analysis.

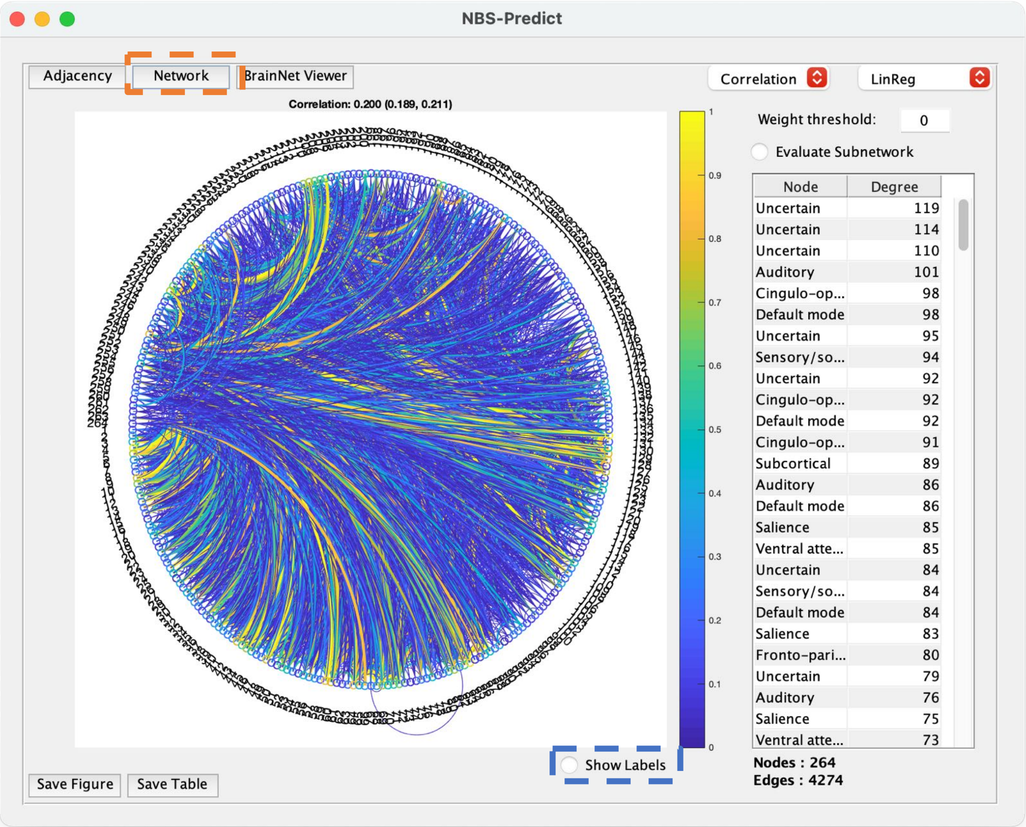


Weighted Network

Now, have a look at the network!

Simply click the “Network” button.

Click ”Show Labels” if you want to see the labels.

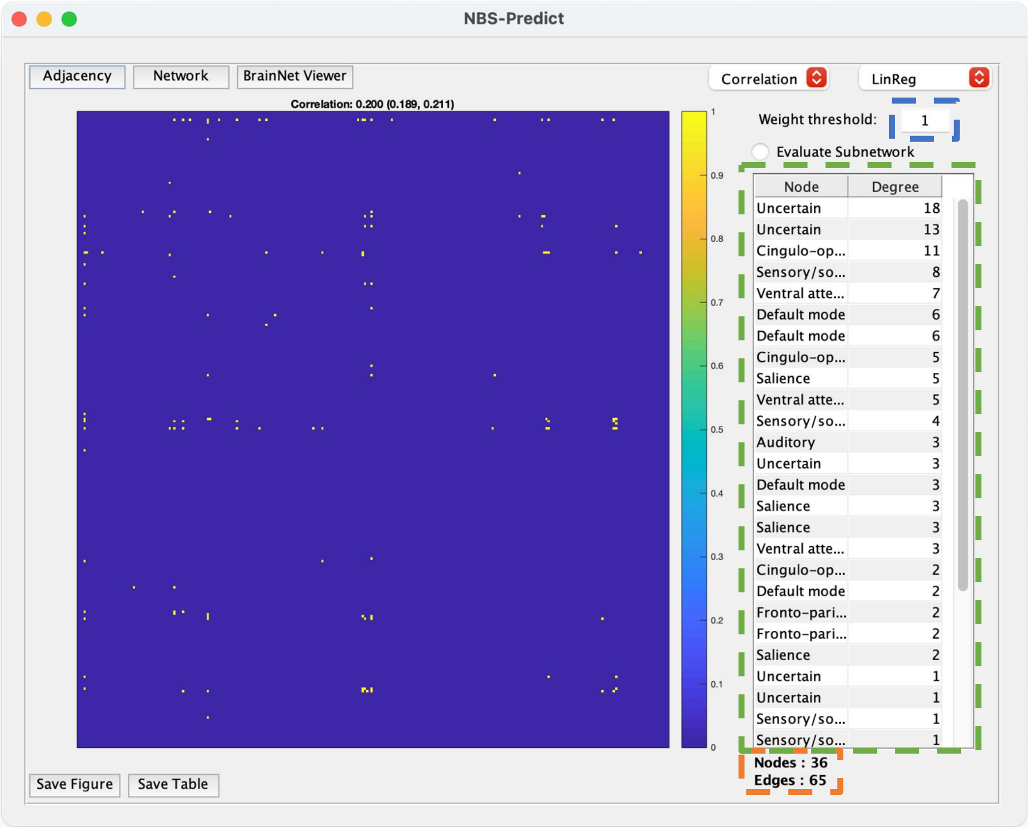


Subnetwork

Set the “Weight Threshold” to 1 to check the most relevant subnetwork!

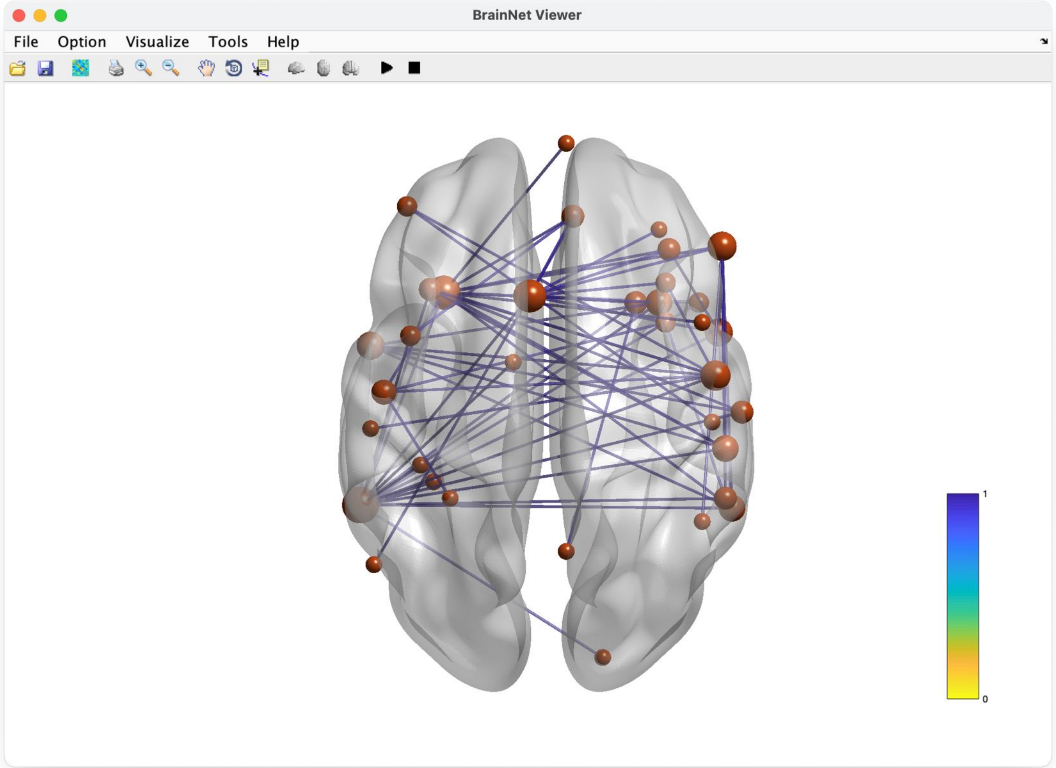
There are 65 connections between 36 regions associated with general intelligence.

We note that the subnetwork comprises a large scale of functional networks across the brain.



Subnetwork

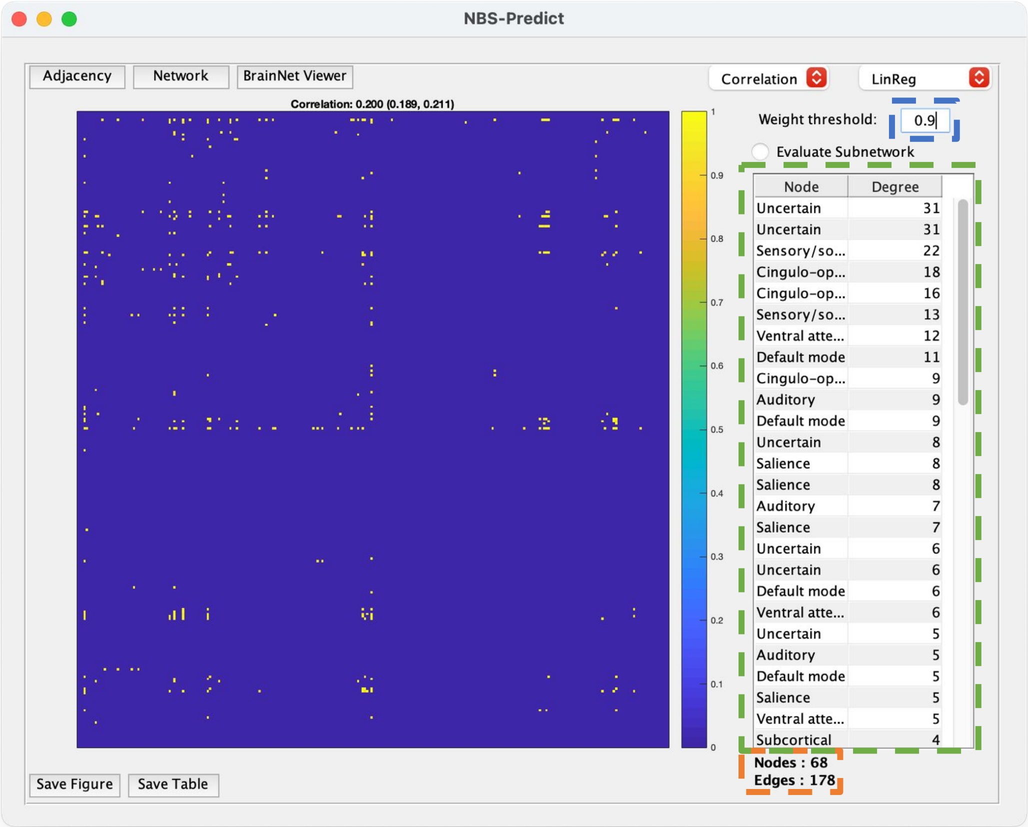
Network plots of the most relevant network.



Subnetwork

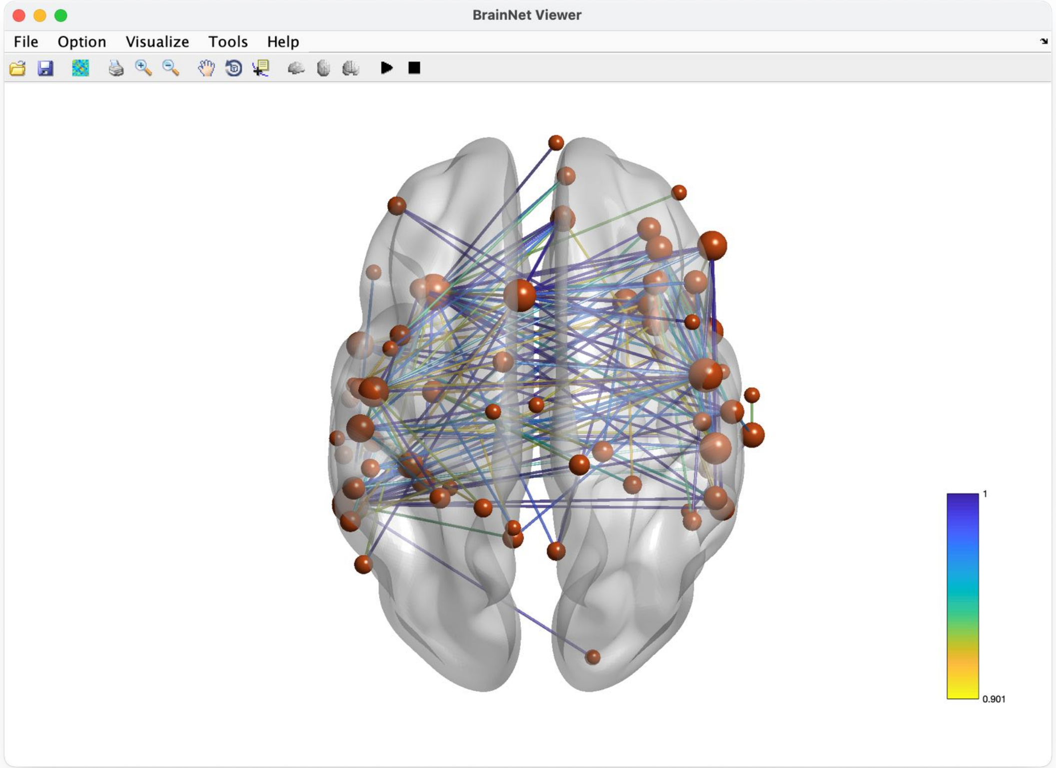
As the weight threshold of 1 might be too conservative, let’s set it to more a lenient value of 0.9.

Now, it gives us a subnetwork comprising 178 edges associated with general intelligence.



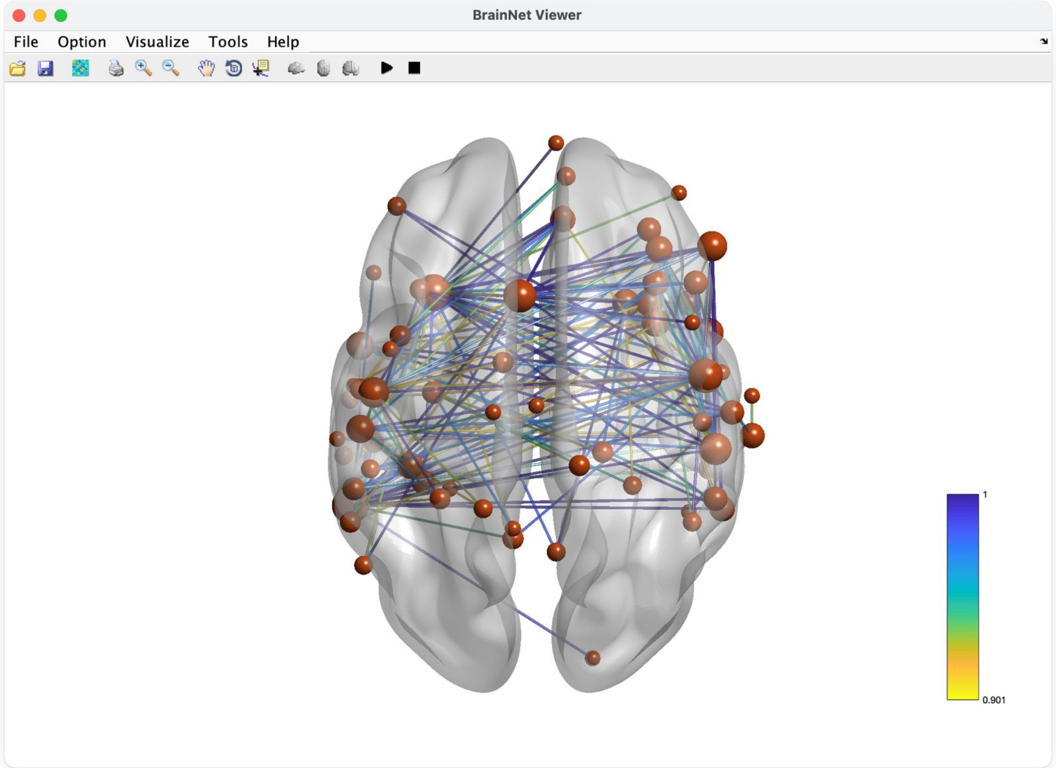
Subnetwork

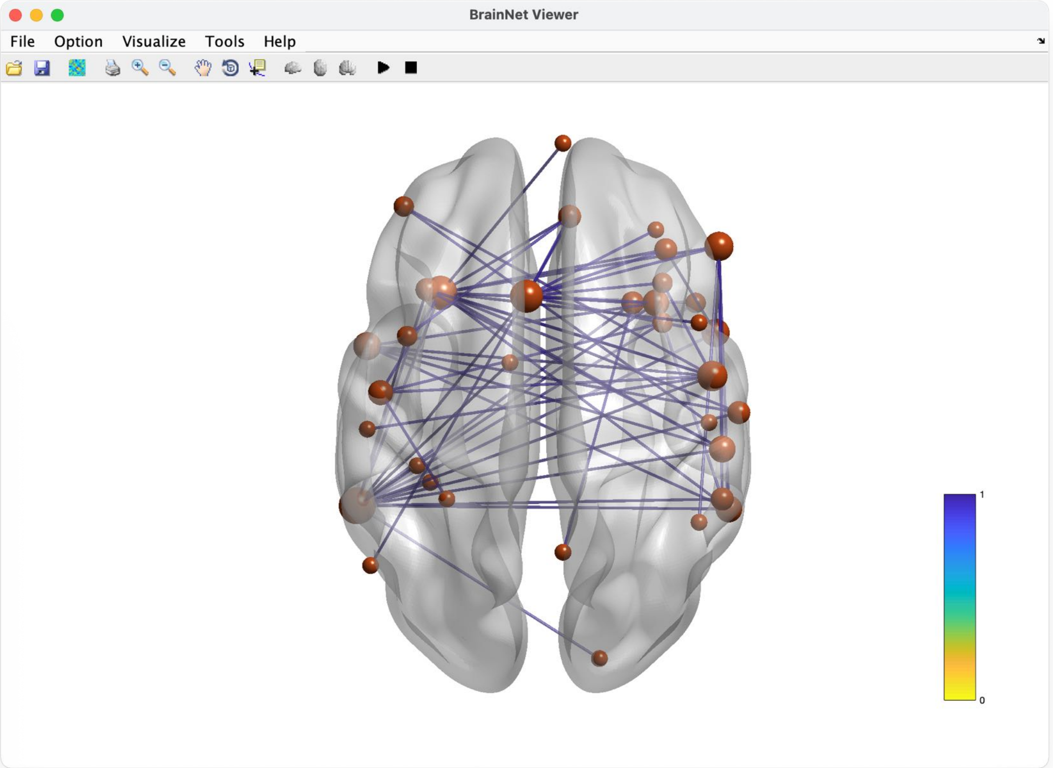
Again, the subnetwork comprises a large scale of connections.



Results

Obviously, mostly temporal, frontal and parietal regions are relevant.





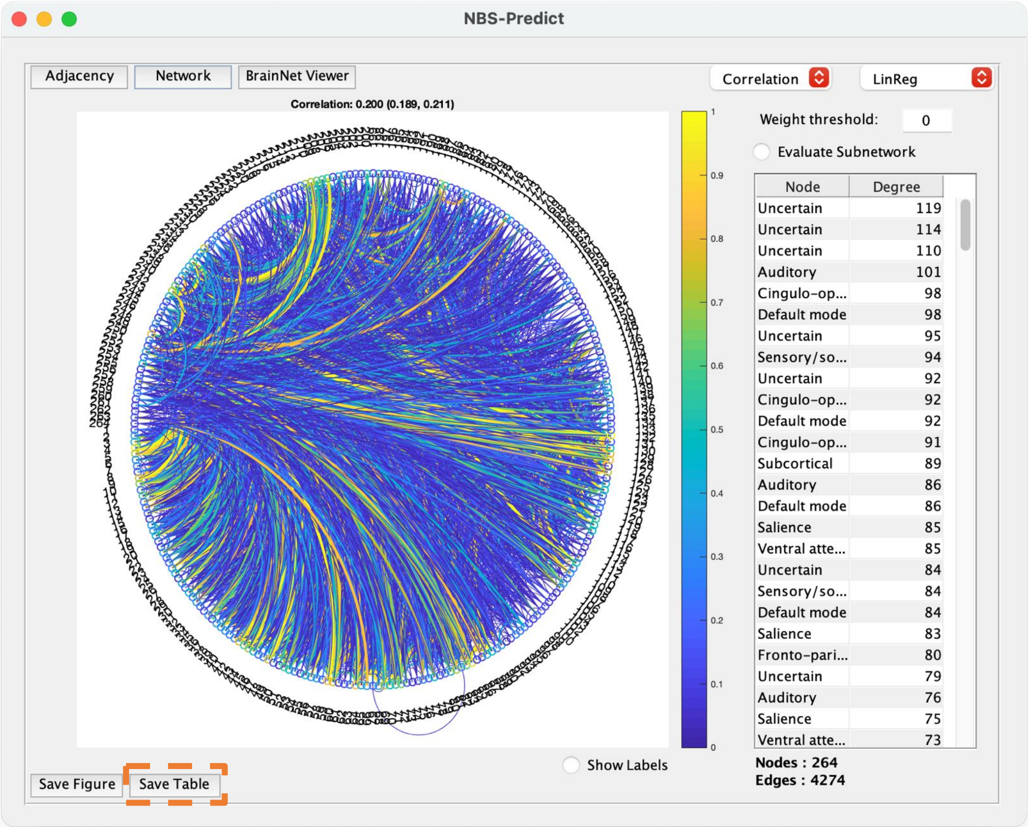
However, the posterior regions (esp. early visual regions) do not seem to be involved.

Output

You after the analysis and results visualization, it is time to save the results.

NBS-Predict allows you to save the figures[1] and nodal degree tables[2] in various formats.

For the nodal degree table, simply hit the “Save Table” button and save the file in the desired location.

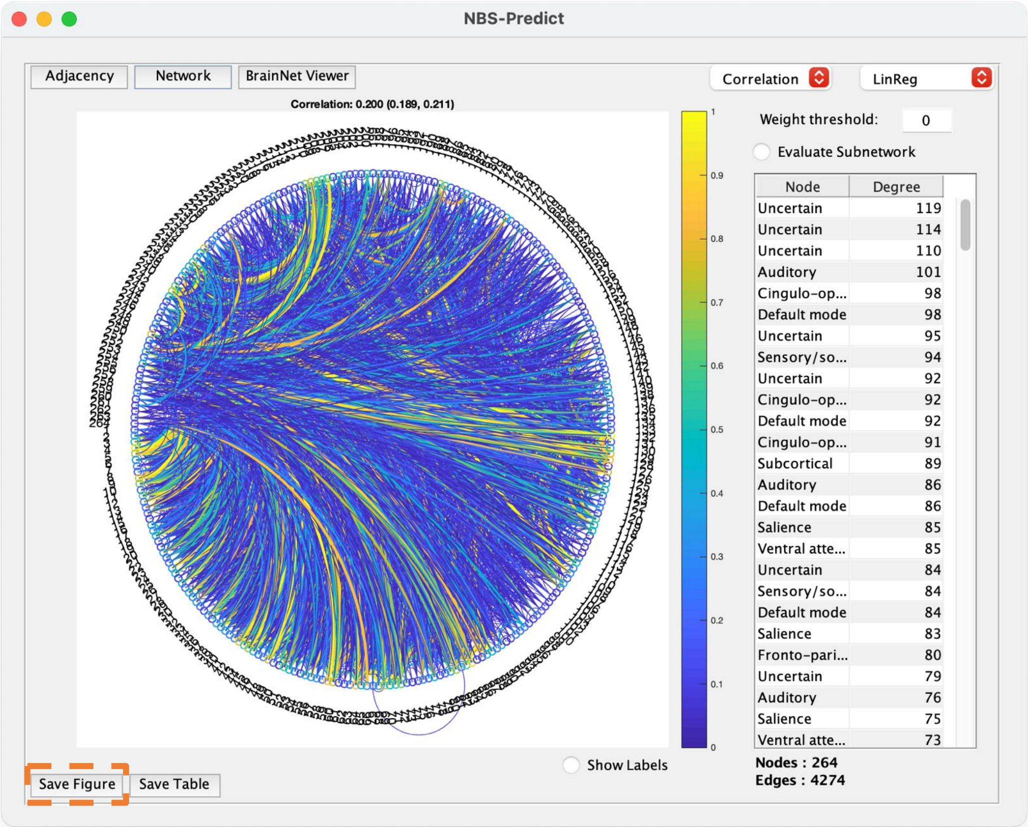


[1] Available formats are: .png, .fig, .jpeg, .tiff
[2] Available formats are: .txt, .mat, .xlsx, .csv

Output

When you click “Save Figure”, the figure (adjacency matrix or circular network graph) in the current axis will be saved.

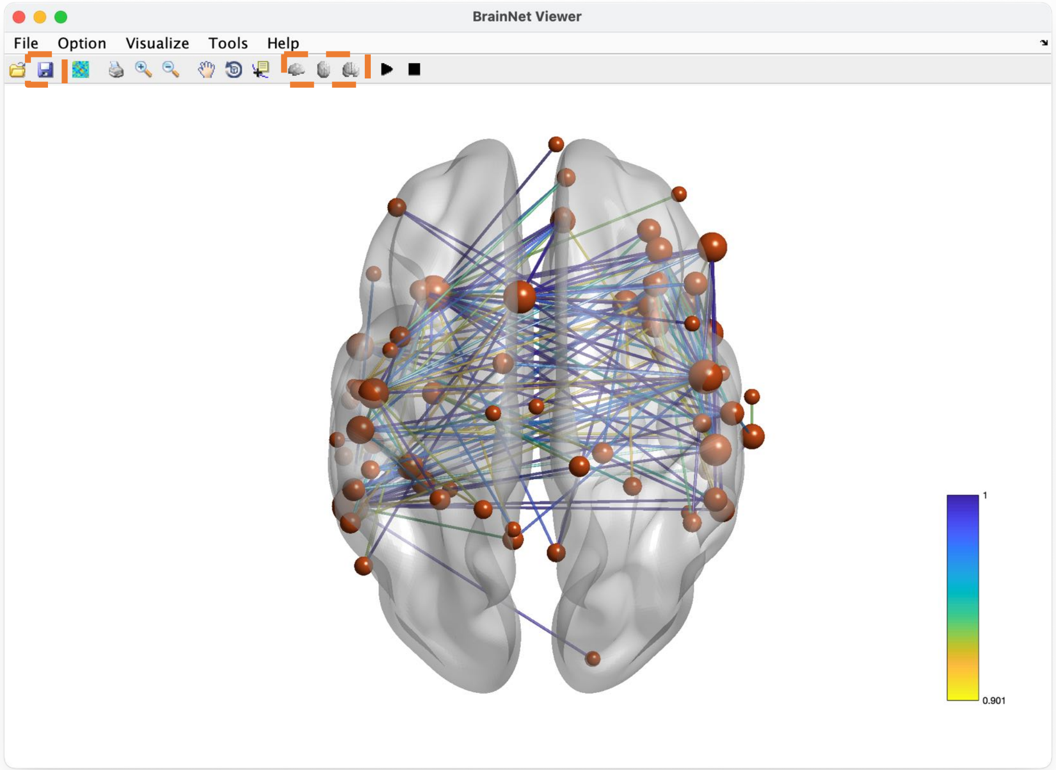
Note that, it might take a few minutes to save the circular network graph (esp. if you would like to save the fully weighted network) as it is a very dense plot.

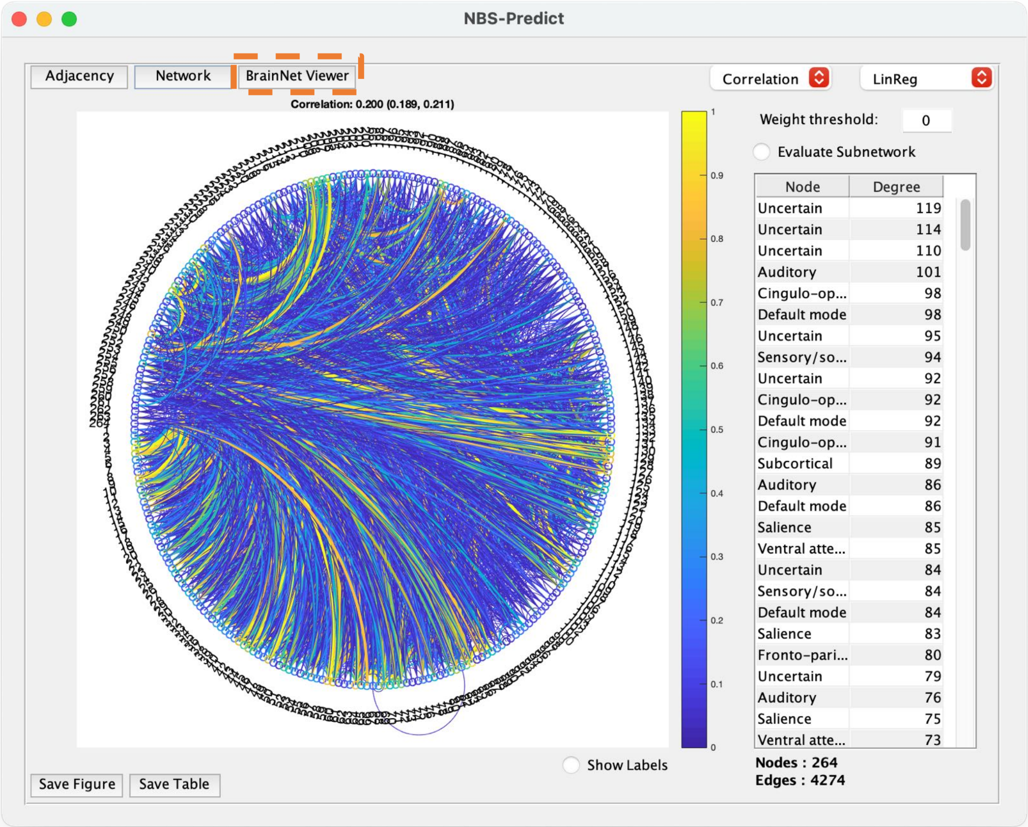


Output

The network on 3D Brain can be saved using the BrainNet Viewer GUI.

1. Click "BrainNet Viewer" to open its GUI.



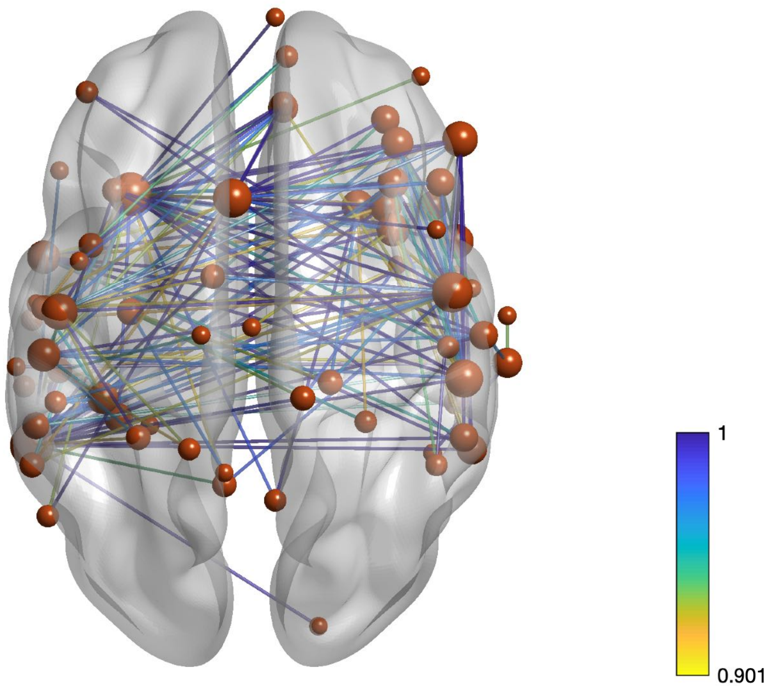


2. Select a desired anatomical plane.

3. Click the "Save" button

Summary Results

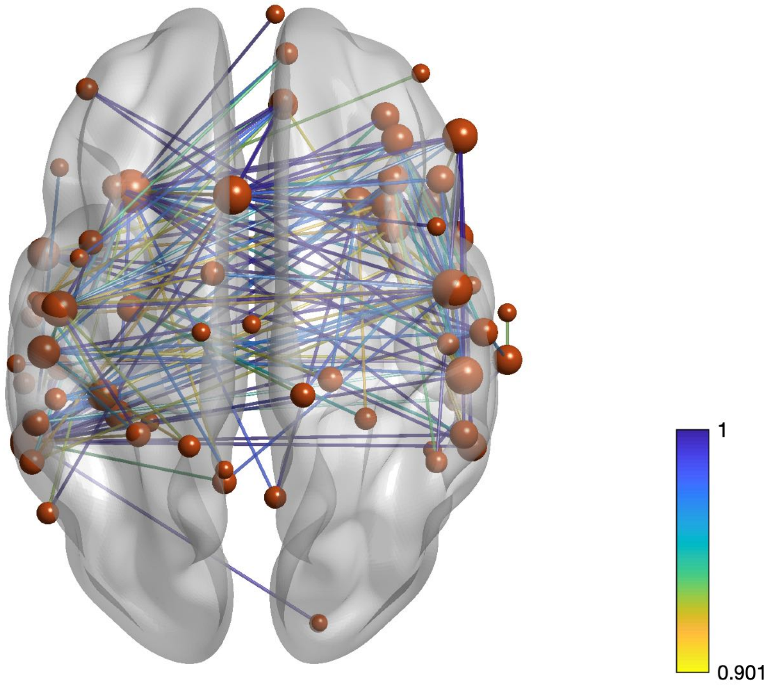
- General intelligence is predicted from subjects' connectivity matrices with a Pearson's CC of 0.200.
- The most relevant subnetwork comprises 65 edges among 36 nodes.
- Less conservative subnetwork (threshold = 0.9) comprises 178 connections between 68 regions.



Summary Results

- The regions are most from the default mode, cingulo-opercular task control, somatomotor, ventral attention, salience, and auditory networks.

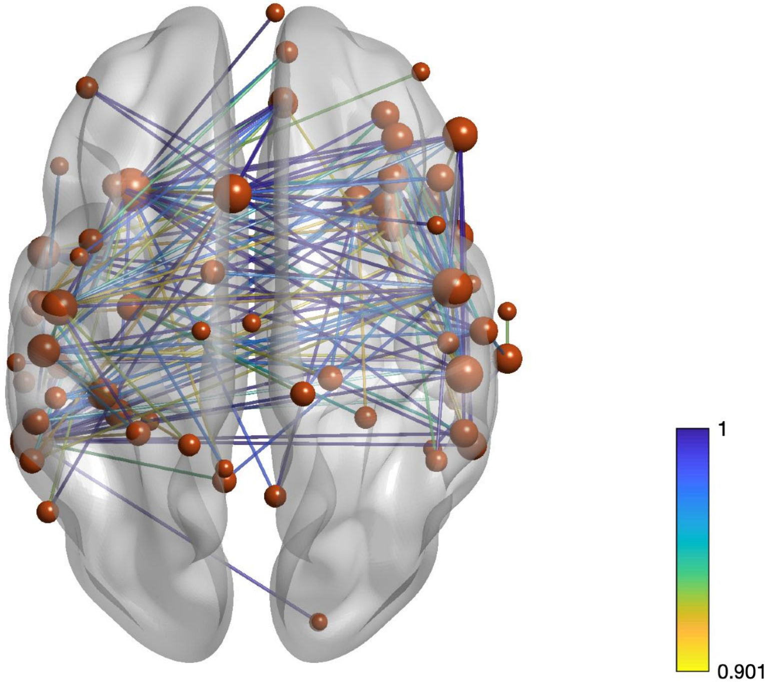
- There are also plenty of brain regions that are not associated with any specific functional networks.



Summary Results

- Our results show that the neural underpinnings of intelligence is characterized by a large scale of interconnections between brain regions.

- In parallel with the literature (Hearne et al., 2016; Song et al., 2008; van den Heuvel et al., 2009)



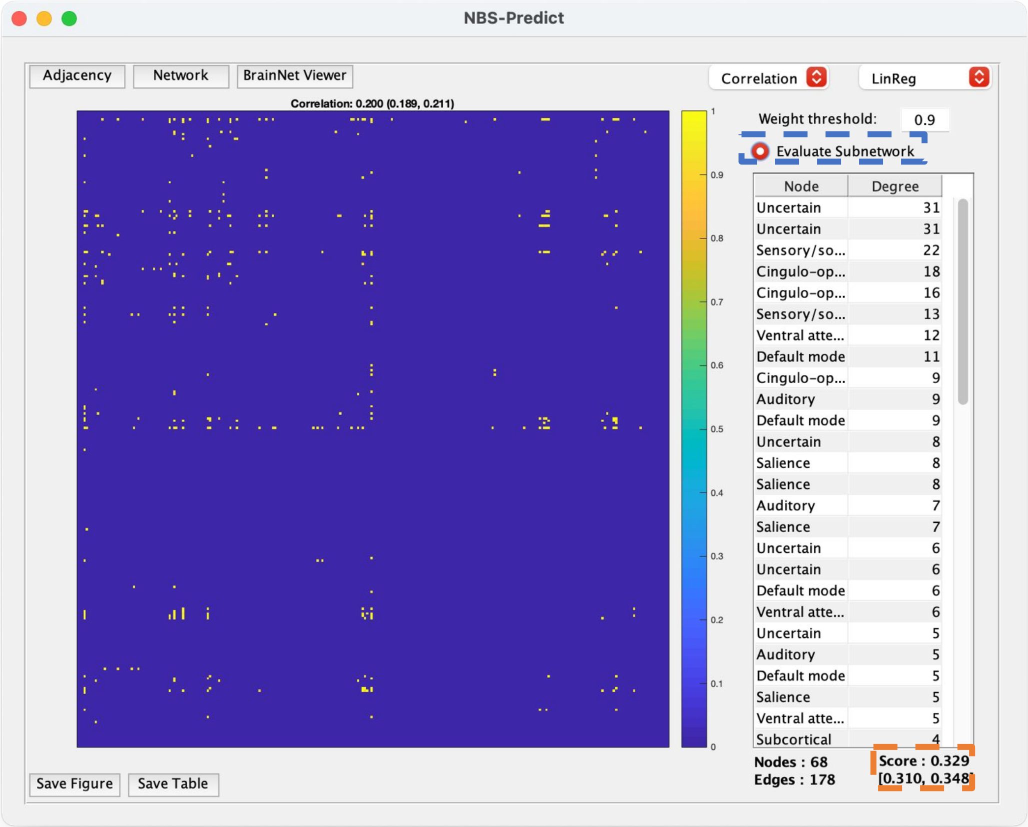
Hearne, L.J., Mattingley, J.B., Cocchi, L., 2016. Functional brain networks related to individual differences in human intelligence at rest. Scientific Reports.
Song, M., Zhou, Y., Li, J., Liu, Y., Tian, L., Yu, C., Jiang, T., 2008. Brain spontaneous functional connectivity and intelligence. Neuroimage 41, 1168–1176.
van den Heuvel, M.P., Stam, C.J., Kahn, R.S., Hulshoff Pol, H.E., 2009. Efficiency of Functional Brain Networks and Intellectual Performance. J Neurosci 29, 7619–7624

Subnetwork Evaluation (Optional!)

Let’s predict general intelligence using the suprathreshold subnetwork!

Click ”Evaluate Subnetwork”, and NBS-Predict will run Linear Regression in a 10-repeated 10-fold CV structure[1].

The suprathreshold subnetwork yields correlation score of 0.329, which is a significant improvement.



[1] NBS-Predict performs scaling and deconfounding, but not hyperparameter optimization.

Subnetwork Evaluation (Optional!)

Critical Warning: Do not use this performance score to select “optimal” weight threshold to visualize as it could lead to a serious overfitting problem!

Subnetwork evaluation only aims to give a very rough idea of the extent to which the suprathreshold subnetwork would perform, **not to find the optimum subnetwork**.