Summary

Studyforrest_dFC

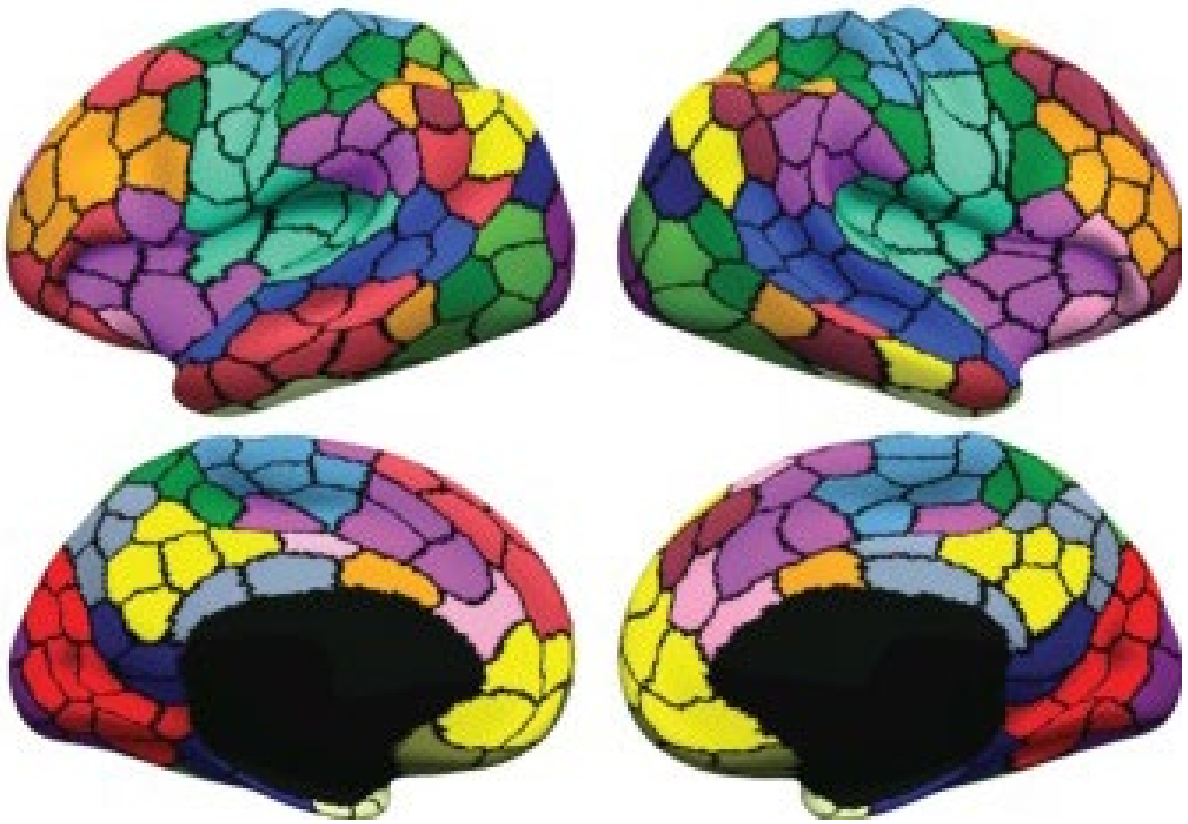
Data description

- Data were collected from the studyforrest project (studyforrest.org)
- 15 Participants were scanned using MRI while watching the entirety of the movie Forrest Gump
- fMRI activity were collected across the whole movie in 8 runs, in intervals (i.e. TRs) of 2 seconds

Question and Aim

- Does the pattern of connectivity between different brain regions, or the simple pattern of neural activity of brain regions, be better representative of movie scenes?
  - Connectivity measures how *related* patterns of activities are, not the magnitude of activity per se.
- To do so, I compared the classification accuracy of movie scenes between dynamic functional connectivity (dFC) and the pattern of BOLD activity across all Regions-Of-Interest (ROIs).

ROI delineation



The human brain divided (black lines) into 419 regions-of-interest (ROI). Figure adapted from Schaefer2018.

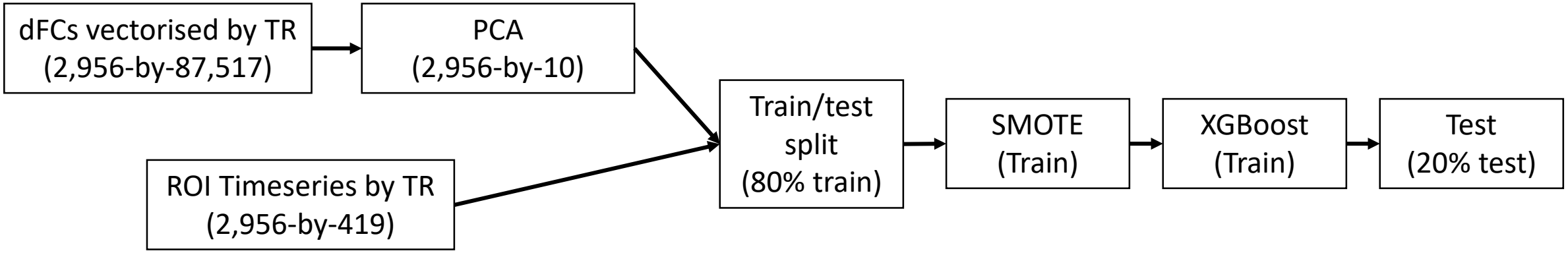
The neural activity within each ROI changes across time, and can be measured using fMRI.

Functional Connectivity (FC) correlates the activity of each ROI across time in a pairwise fashion with remaining ROIs, resulting in a 87,571 (lower-half of a 419-by-419 correlation matrix) unique correlations.

Preprocessing

- Both structural and functional data were projected into surface space using recon-all (Freesurfer 5.3.0), and functional data were preprocessed utilising the CBIG pipeline.
- dFC was calculated by taking a window of activity divided equally around the time point (TR) of interest, and then correlating the activity in this window amongst the 419 ROIs as delineated in Schaefer2018.
- ROI activity was simply the amplitude of the BOLD signal for each of the 419 ROIs at the corresponding timepoint (i.e. TR) of interest.

Overview of comparison flow

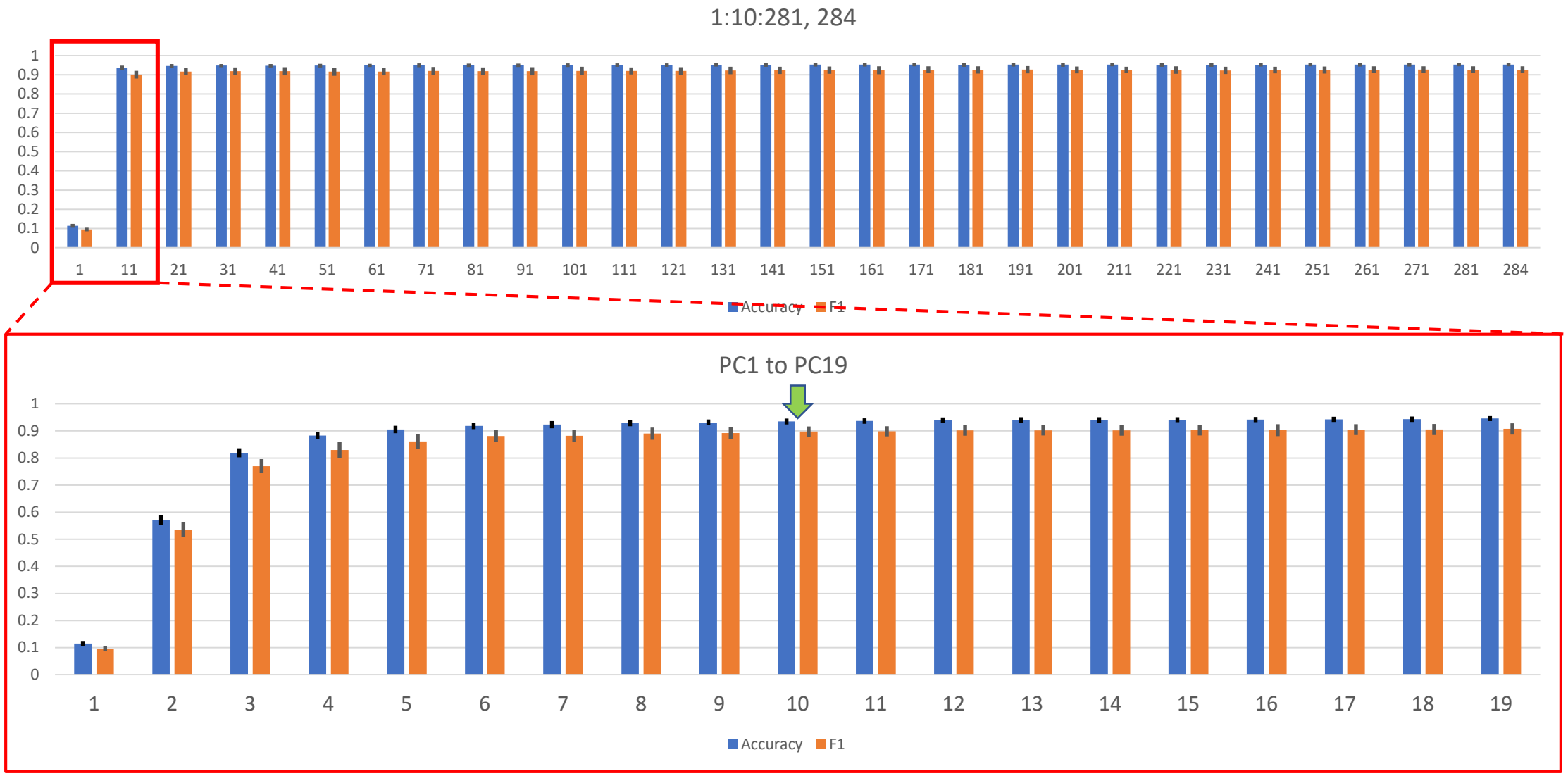


Data were transformed/normalised and averaged across subjects.

dFC data preperation

1. Generated dFCs using a 55 TR (110s) window, and averaged across subjects (total 3,167 TRs across 8 runs)
2. Vectorised dFCs per TR into rows (i.e. created matrix of 3,167 rows by 87,571 columns)
3. Labelled each dFC window with the scene according to the timestamp.
4. Dropped rows according to the criteria below, resulting in a 2,956-by-87,571 matrix (87,572 with labels):
    1. Scenes that were not consistent across the TR
    2. Scenes occurred less than 10 TRs in total across all runs
5. Performed PCA on the 2,956-by-87,571 matrix (excluding labels), selected the first 10 PCs
    1. 10 PCs were selected based on repeated testing of varying number of PCs, to find the least number of PCs where accuracy and F1 scores stabilised.
6. **Final Feature matrix: 2,956 vectorised dFC windows by 10 PCs**
7. **Target vector: 2,956 scene labels**

Selection of 10 PCs



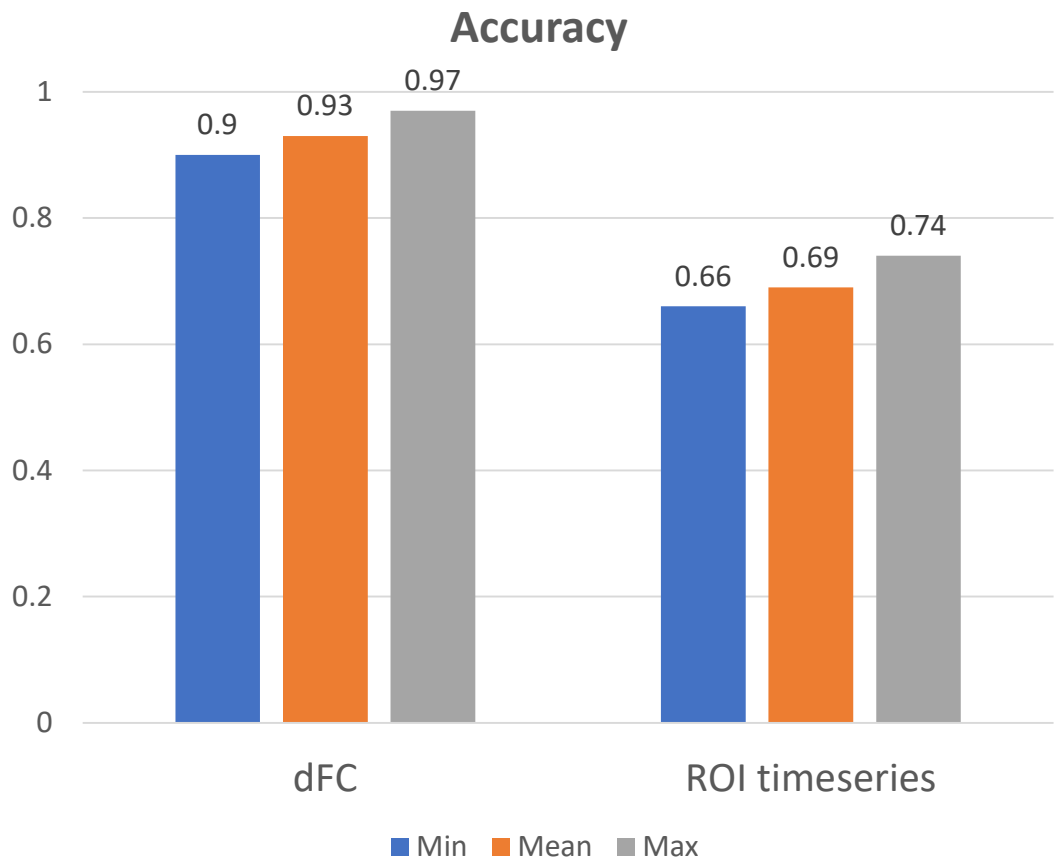
ROI activity data preparation

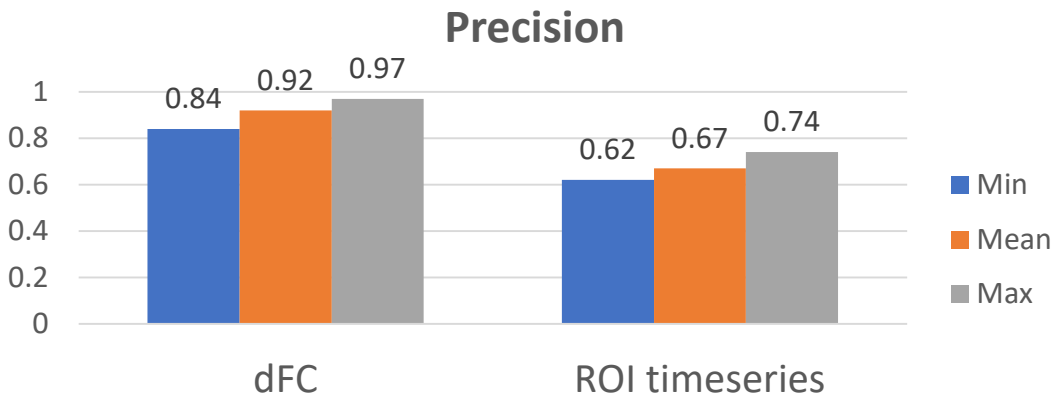
1. Generated averaged (standardised) timeseries of each ROI by TR across subjects (total 3,167 TRs across 8 runs)
2. Labelled data similarly to the dFC matrix, but DID NOT perform PCA.
3. **Final Feature matrix: 2,956 dFC windows by 419 ROIs**
4. **Target vector: 2,956 scene labels**

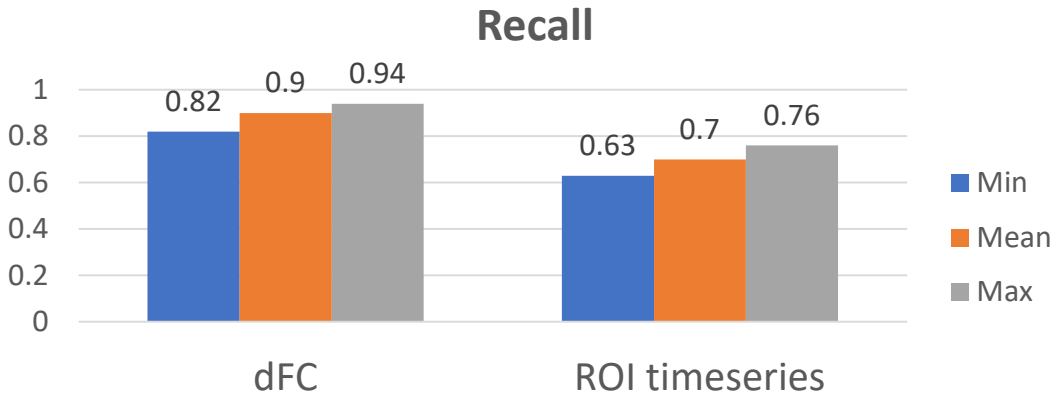
Common workflow

1. Features were randomly split into 80% train and 20% test matrices, stratified by target vector.
2. Synthetic data for minority classes in the Training dataset were generated using SMOTE (k neighbours = 5)
3. Model was trained using XGBoost, and tested on the remaining 20%.
4. Classification accuracy, precision and recall scores were measured.
5. This train/test workflow was repeated 50 times.

Results


Accuracy
1
0.8
0.6
0.4
0.2
0
0.9
0.93
0.97
0.66
0.69
0.74
dFC
ROI timeseries
Min
Mean
Max

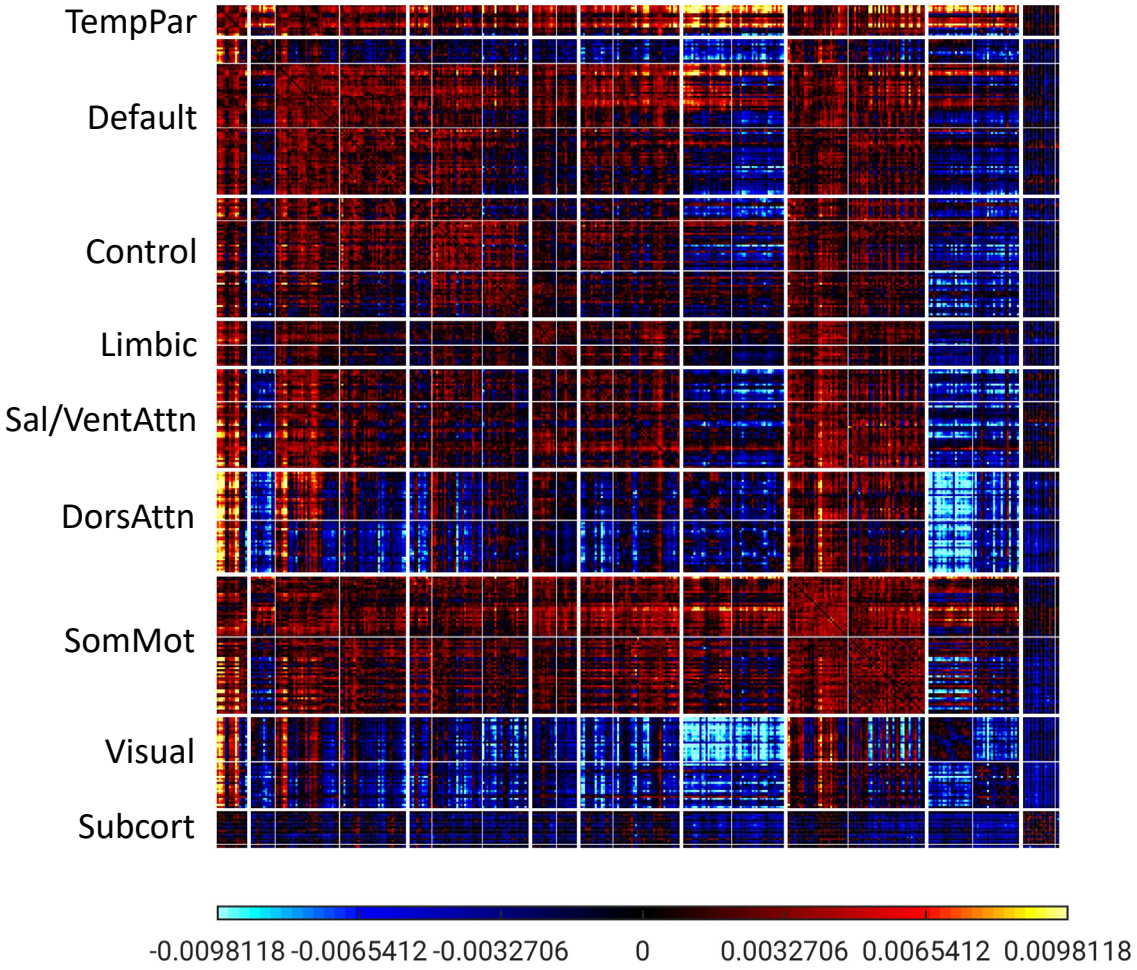

Precision
1
0.8
0.6
0.4
0.2
0
0.84
0.92
0.97
0.62
0.67
0.74
dFC
ROI timeseries
Min
Mean
Max


Recall
1
0.8
0.6
0.4
0.2
0
0.82
0.9
0.94
0.63
0.7
0.76
dFC
ROI timeseries
Min
Mean
Max

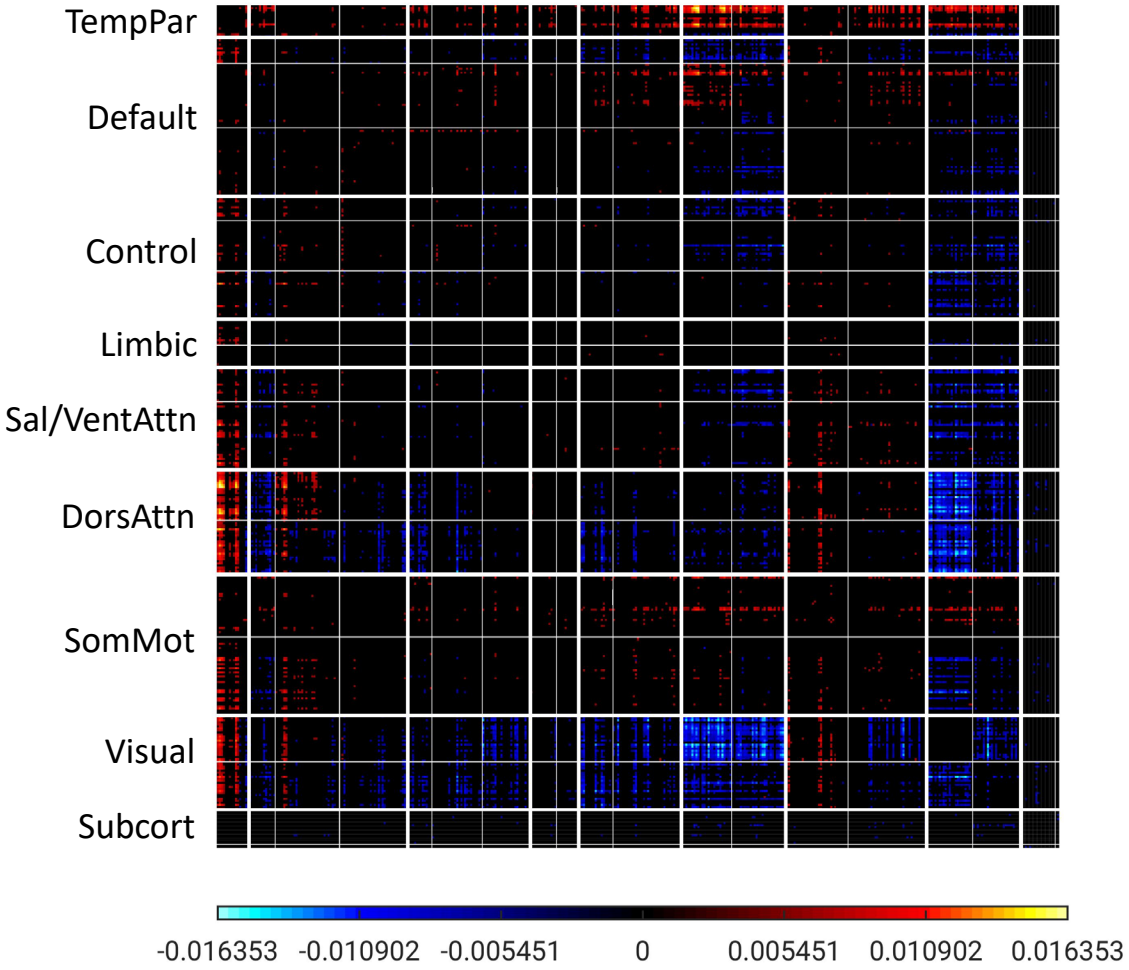
PC plots

- For each PC, the weights of each ROI-to-ROI connection are plotted as a 419-by-419 heatmap, with each ROI further grouped into 9 networks.
- The subsequent plots are thresholded to show the top 10%, in terms of absolute weight, of the importance of each connection between ROIs towards each PC.
- Figures were plotted in MATLAB 2018a according to the Schaefer2018 parcellation, **but importantly represent PC weights** and not the original intended purpose of correlation strengths.

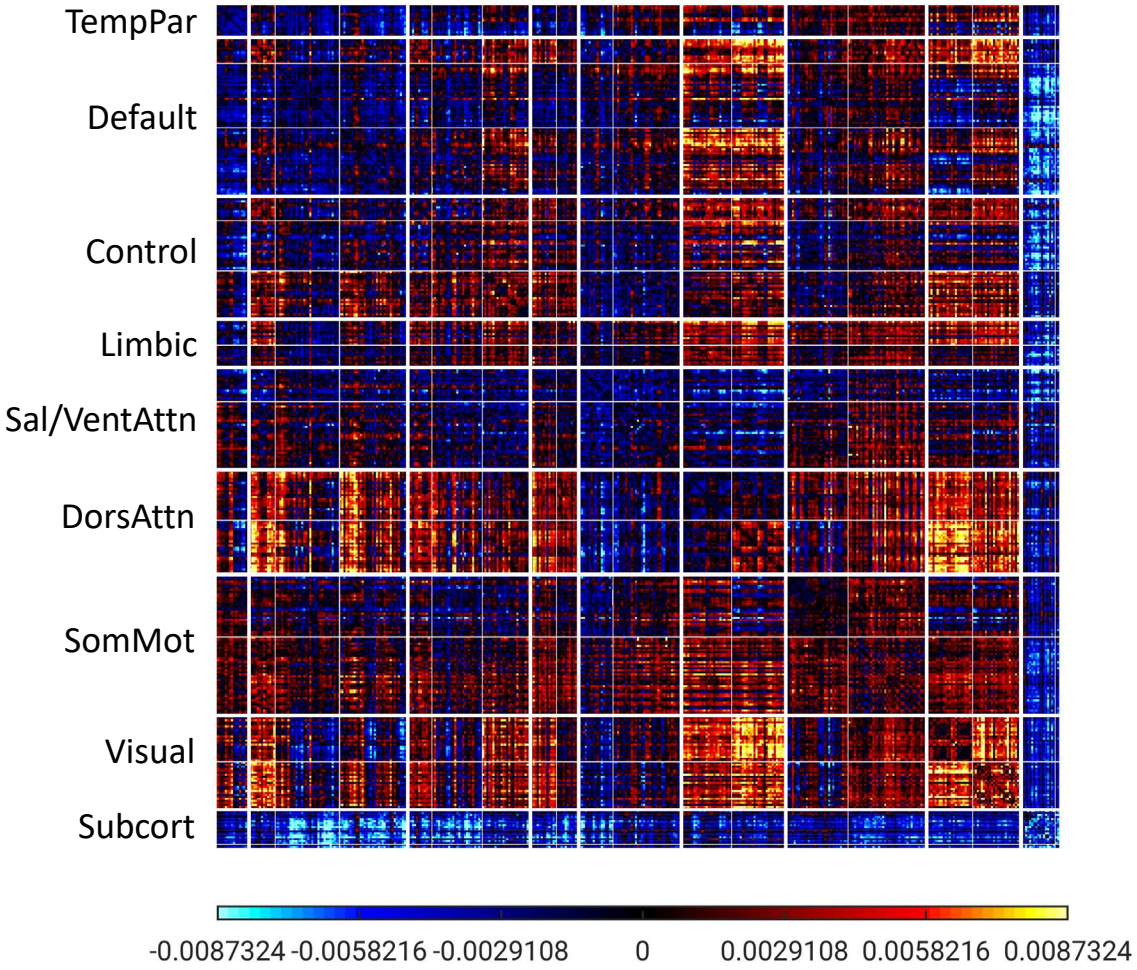
PC1 - unthresholded



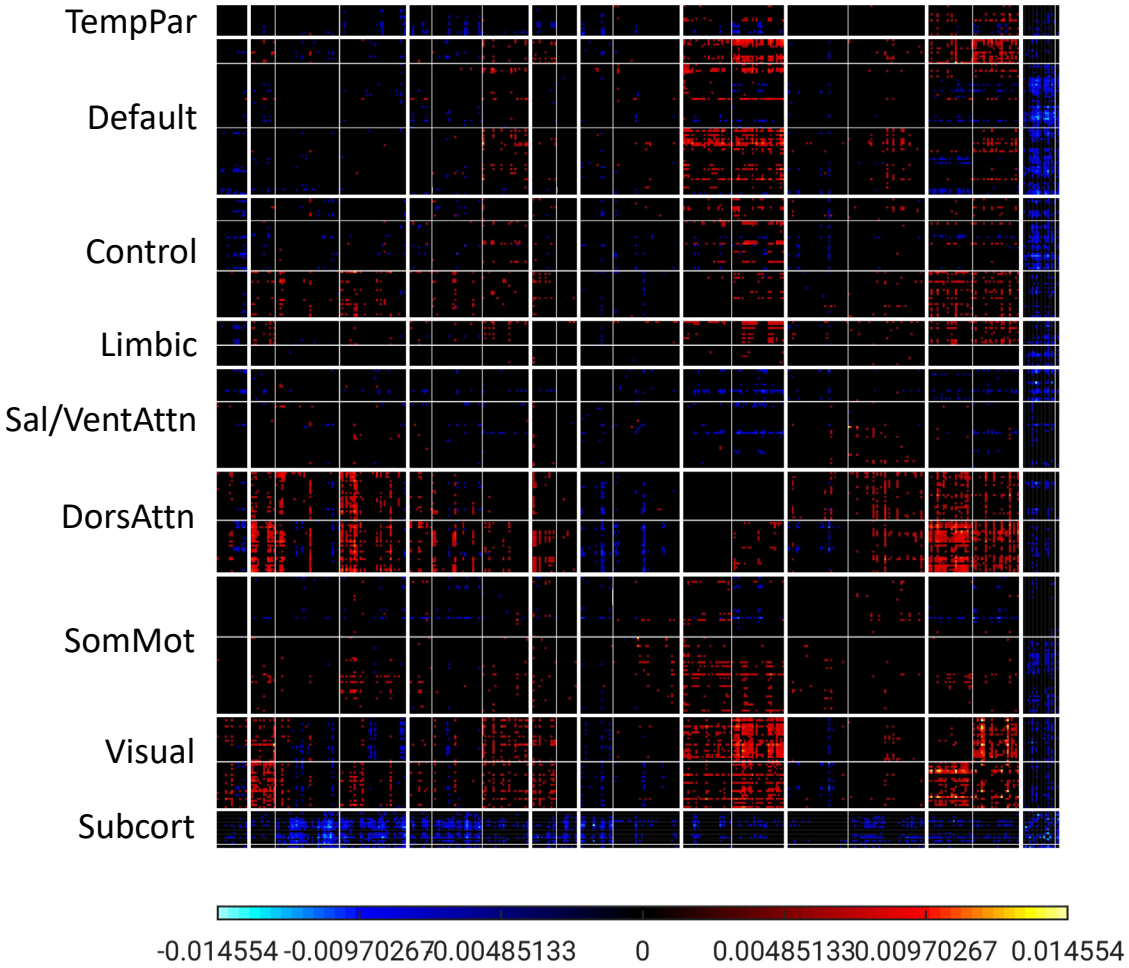
PC1 – thresholded (10%)



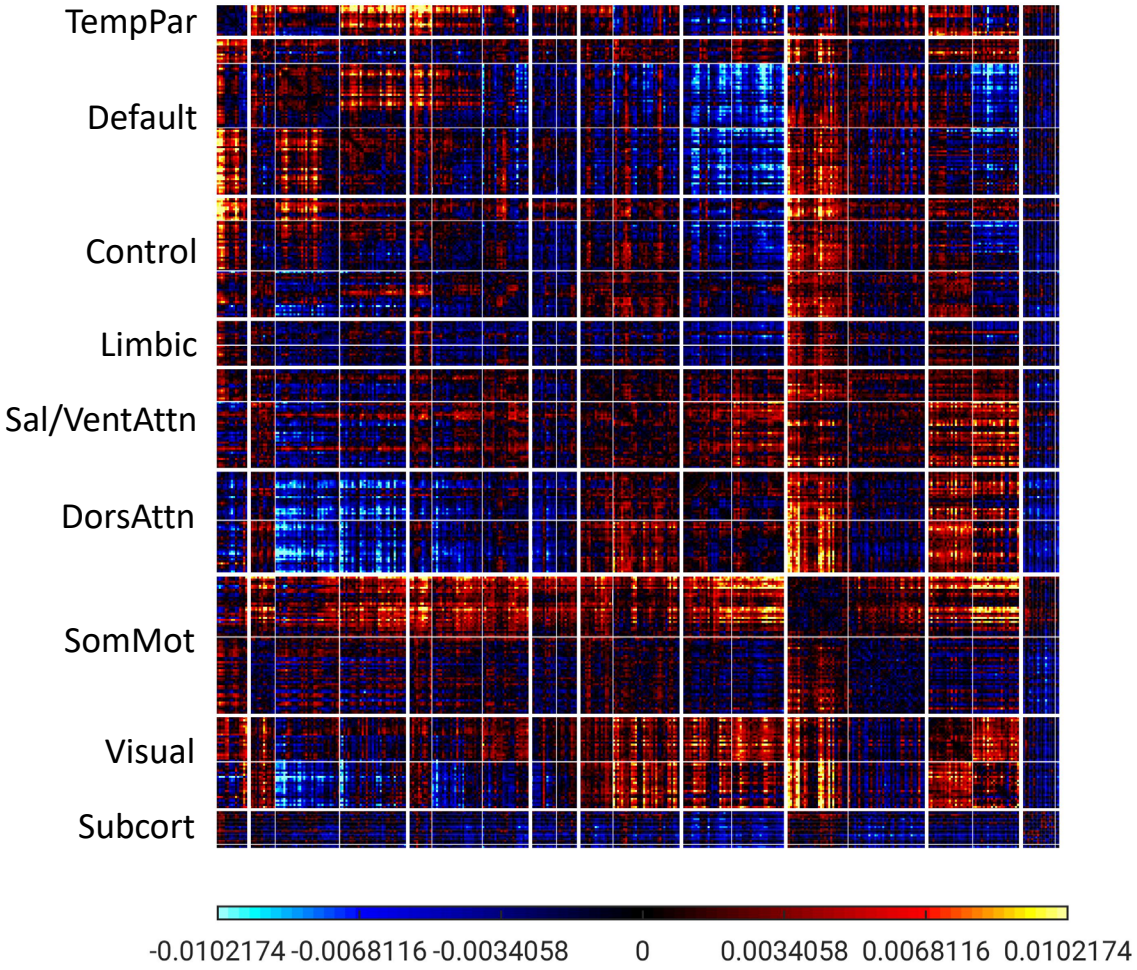
PC2 - unthresholded



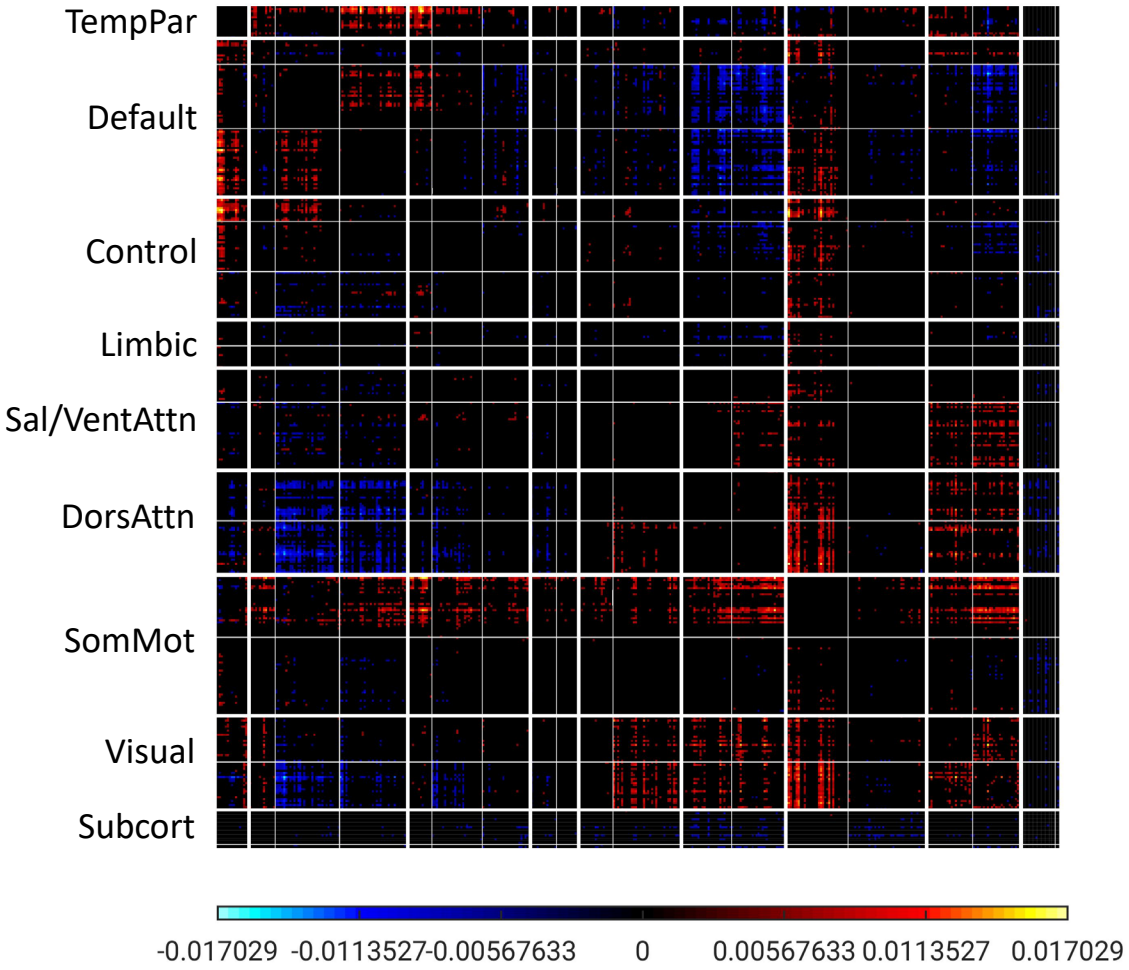
PC2 – thresholded (top 10%)



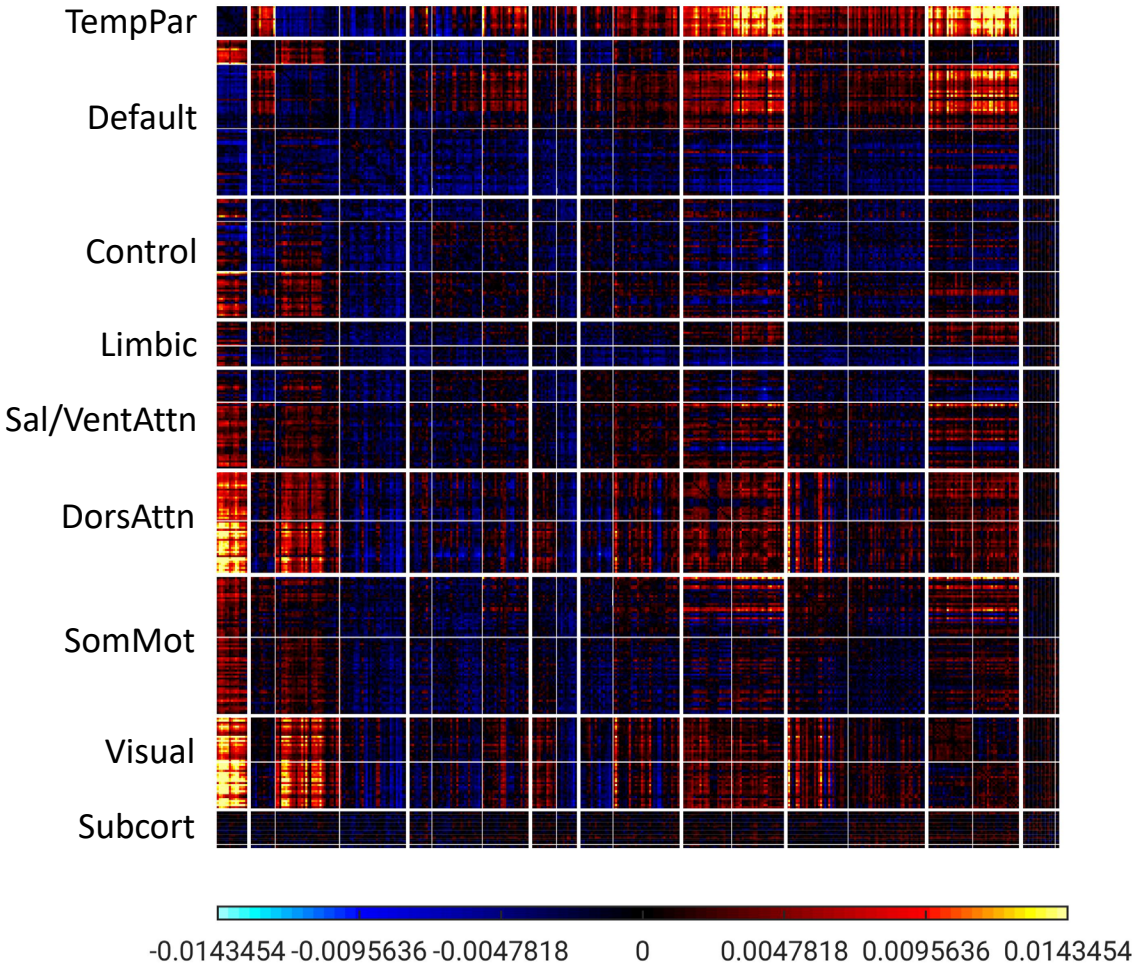
PC3 - unthresholded



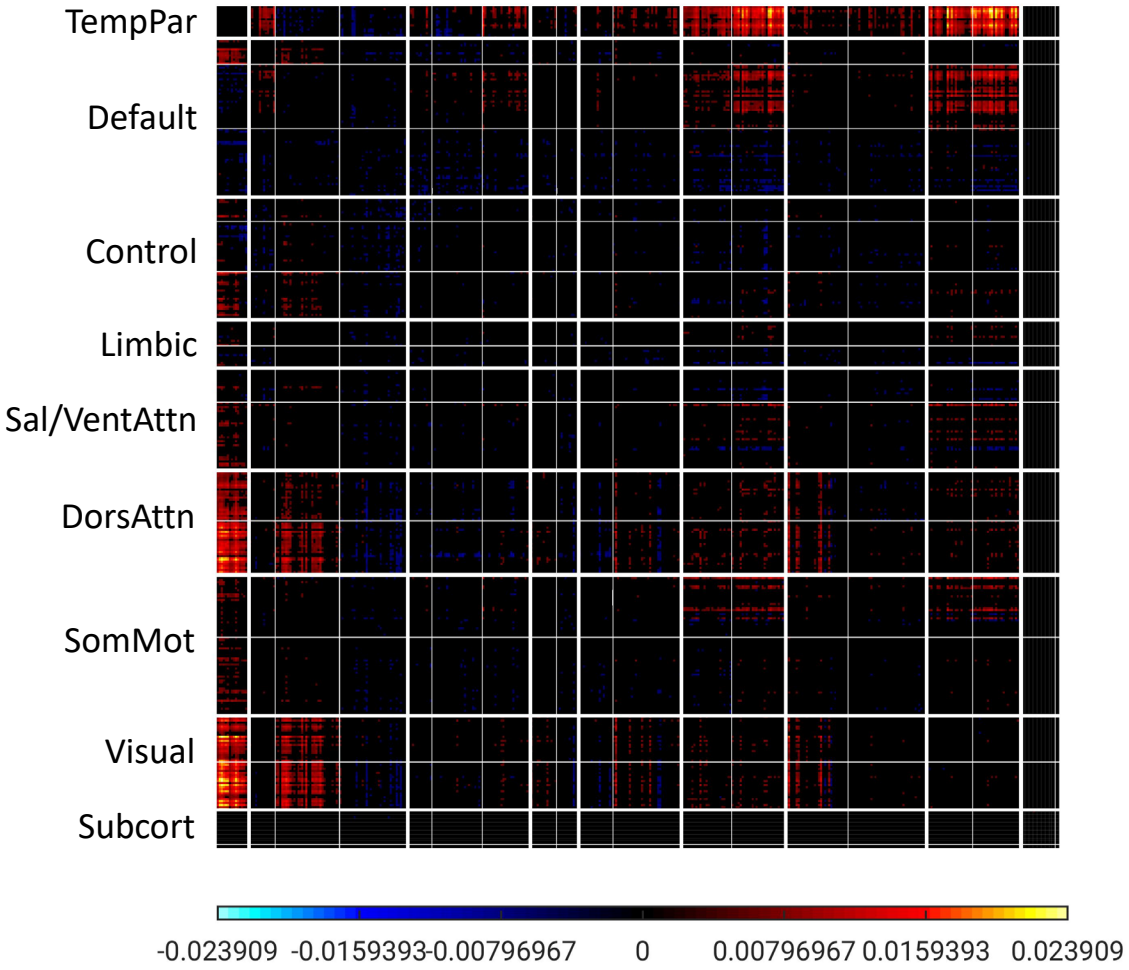
PC3 – thresholded (top 10%)



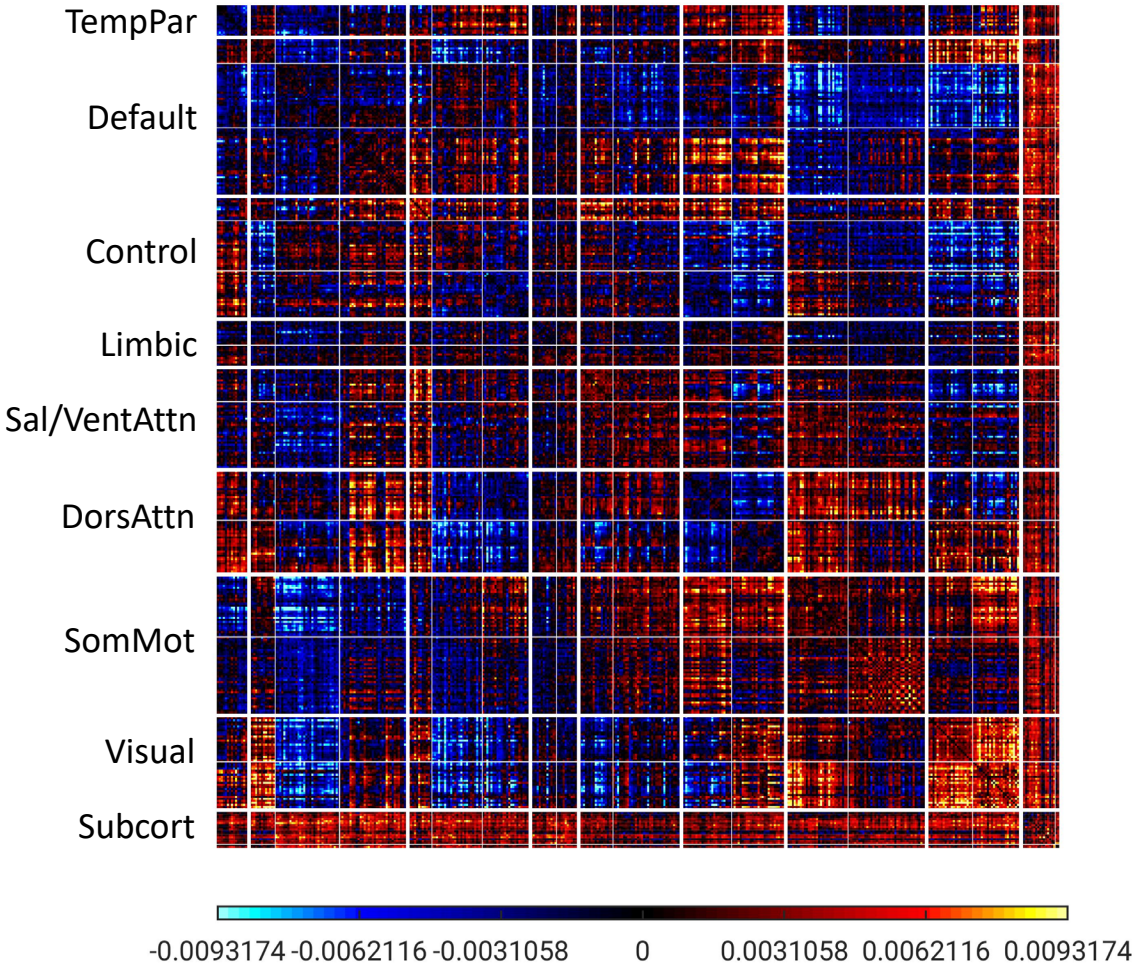
PC4 – unthresholded



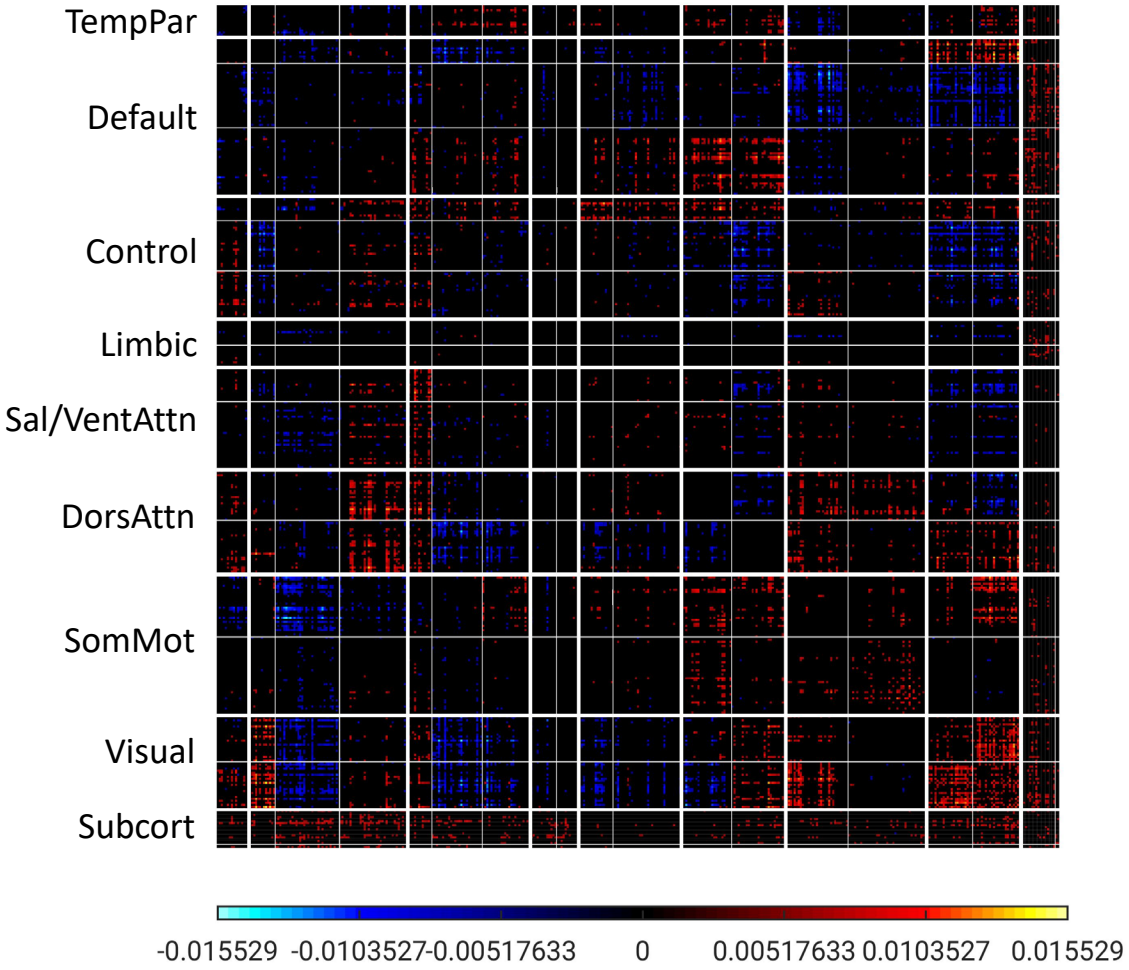
PC4 – thresholded (top 10%)



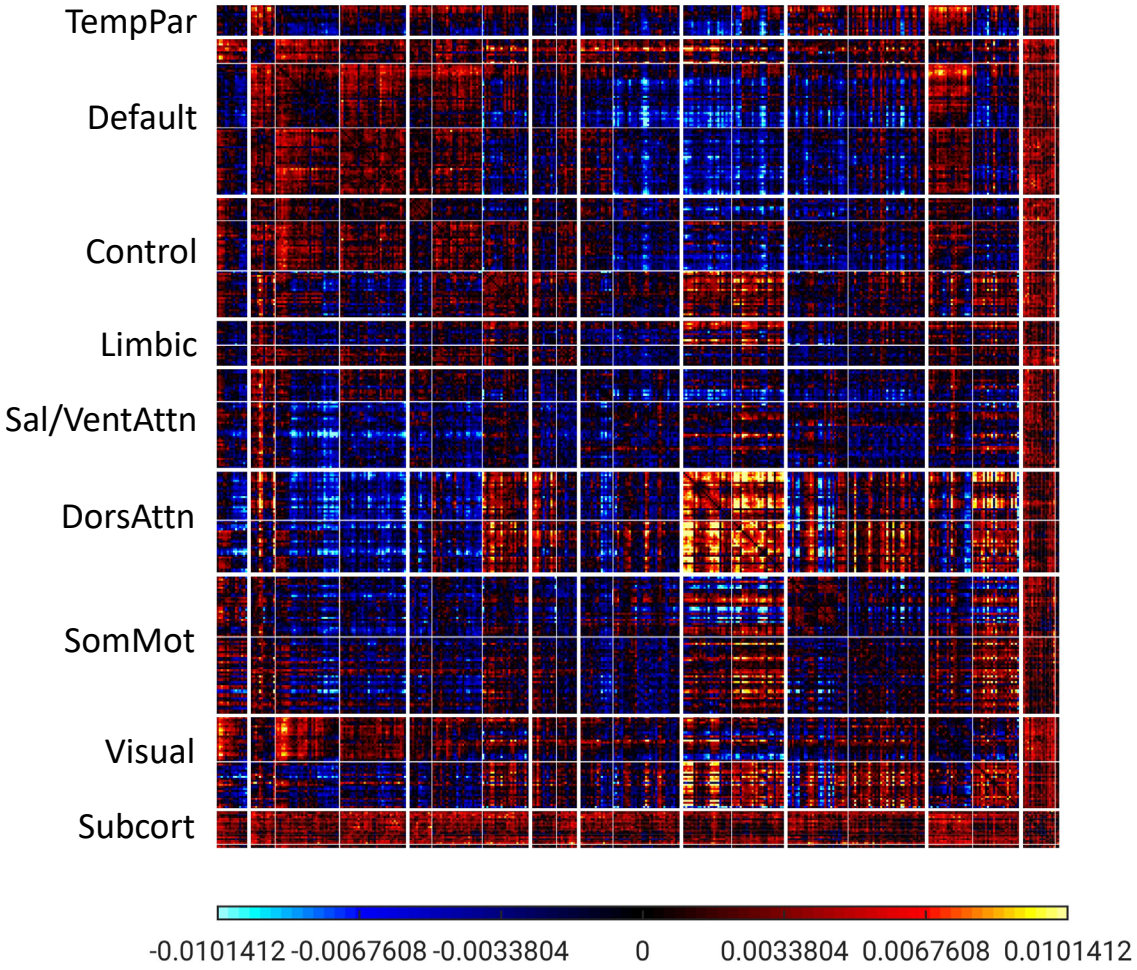
PC5 - unthresholded



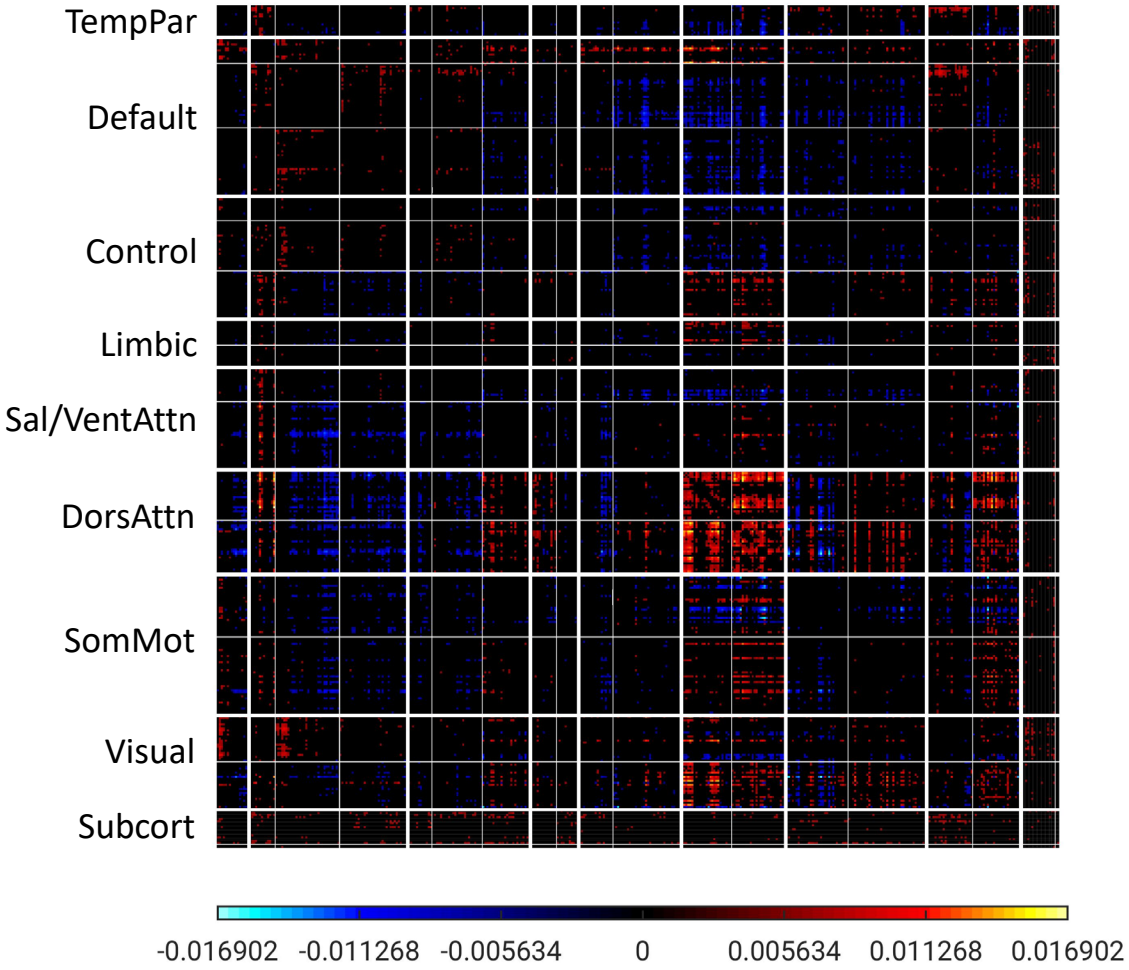
PC5 – thresholded (top 10%)



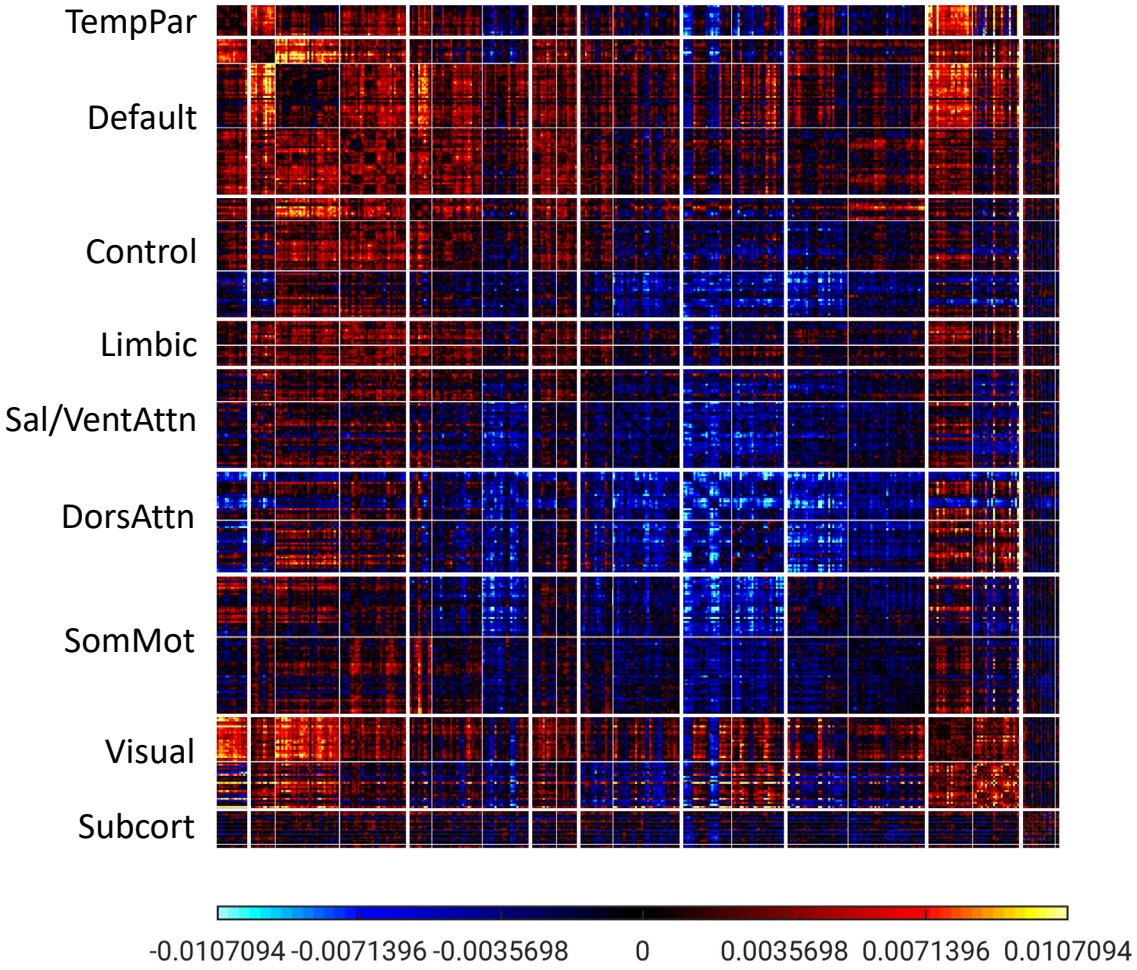
PC6 - unthresholded



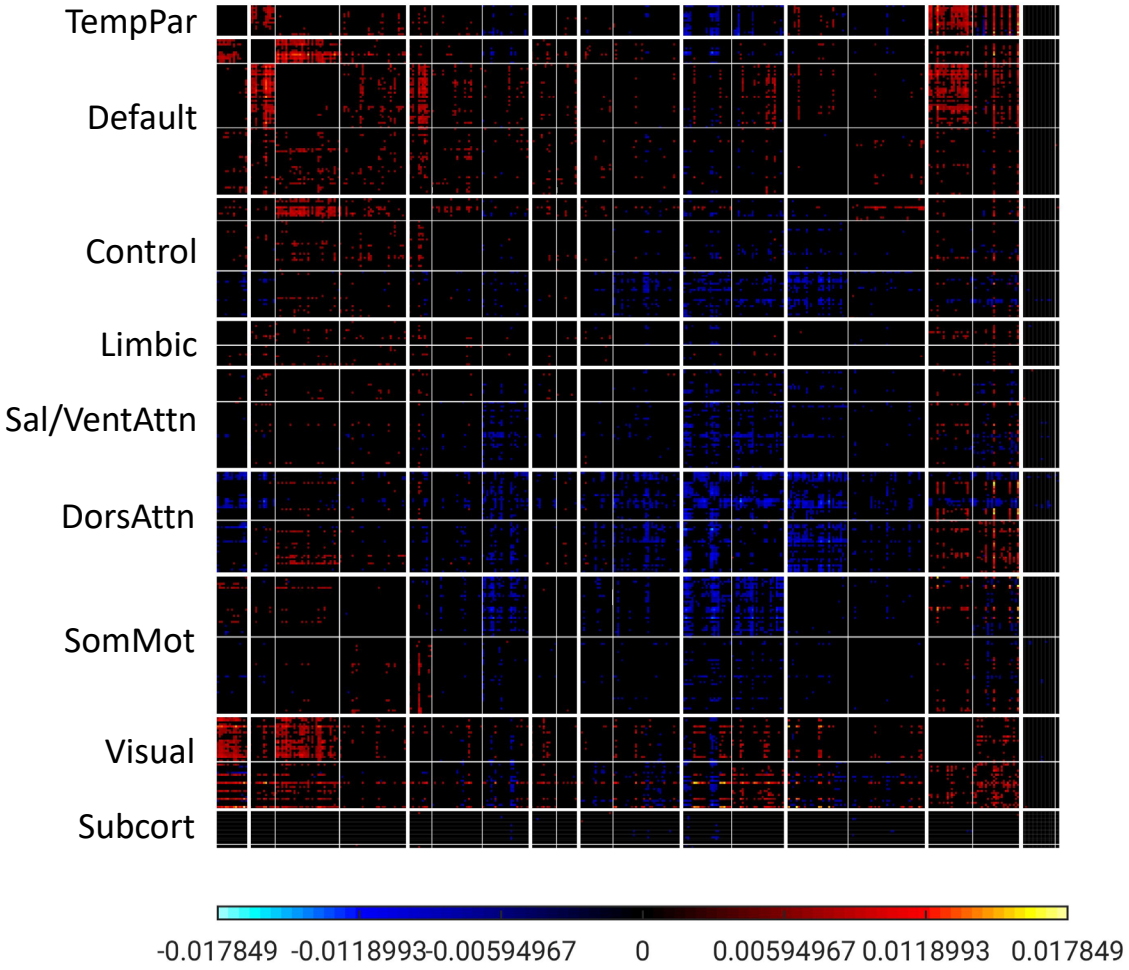
PC6 – thresholded (top 10%)



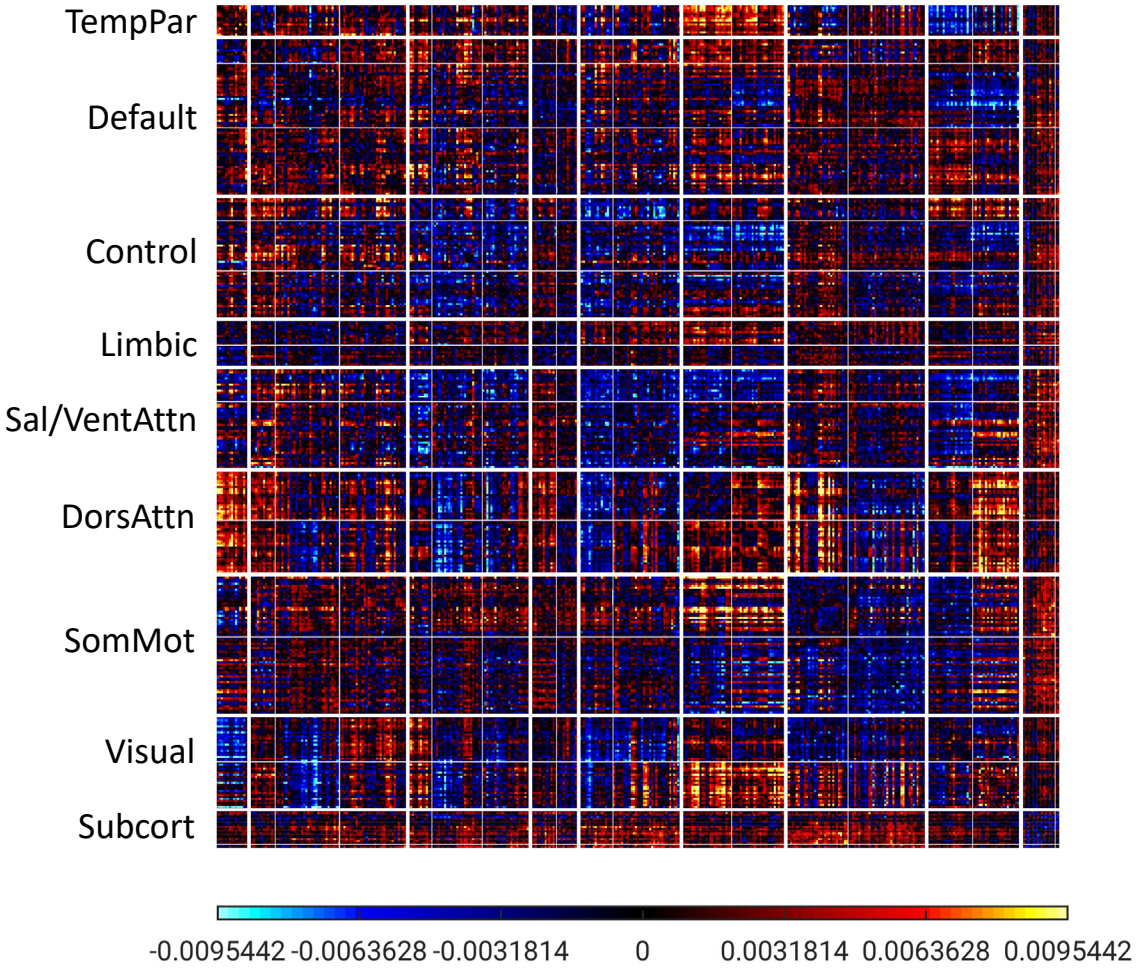
PC7 - unthresholded



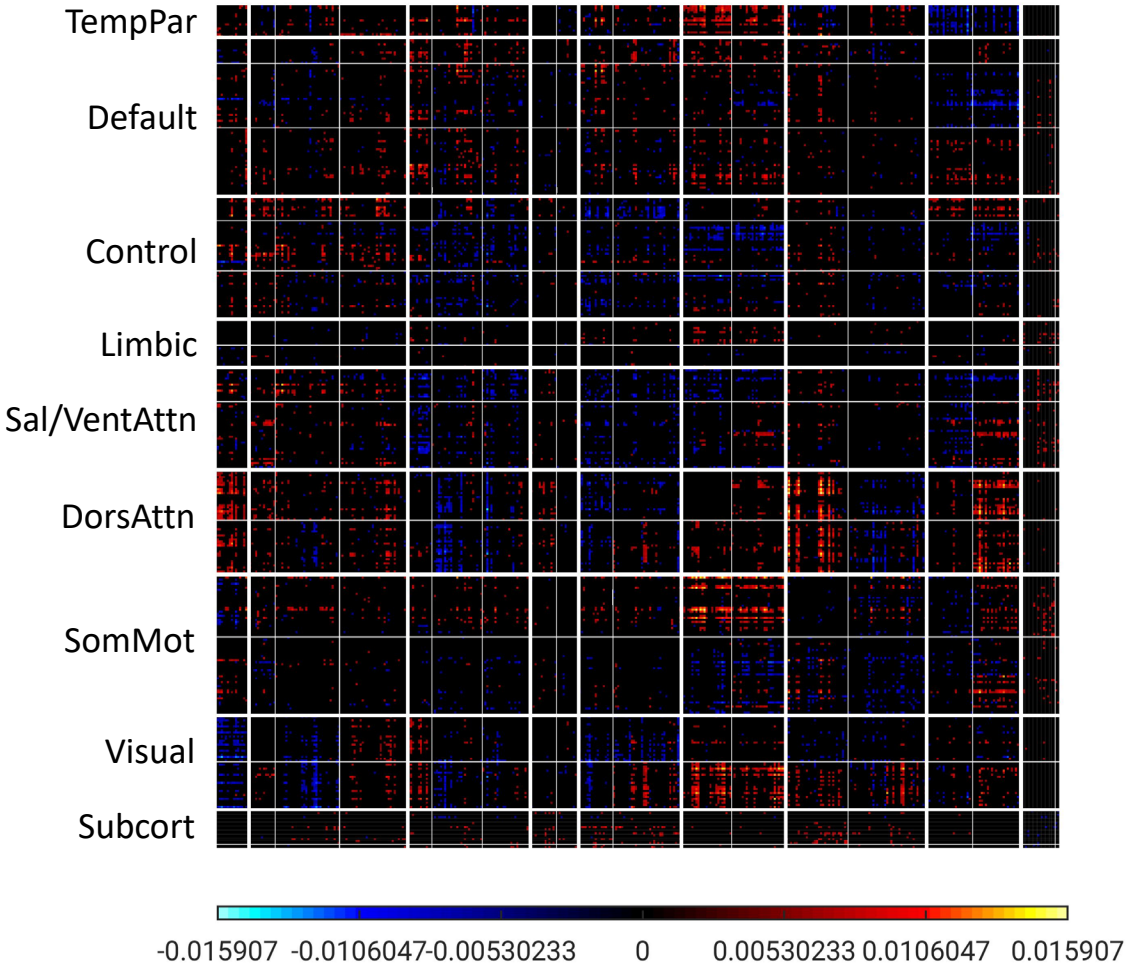
PC7 – thresholded (top 10%)



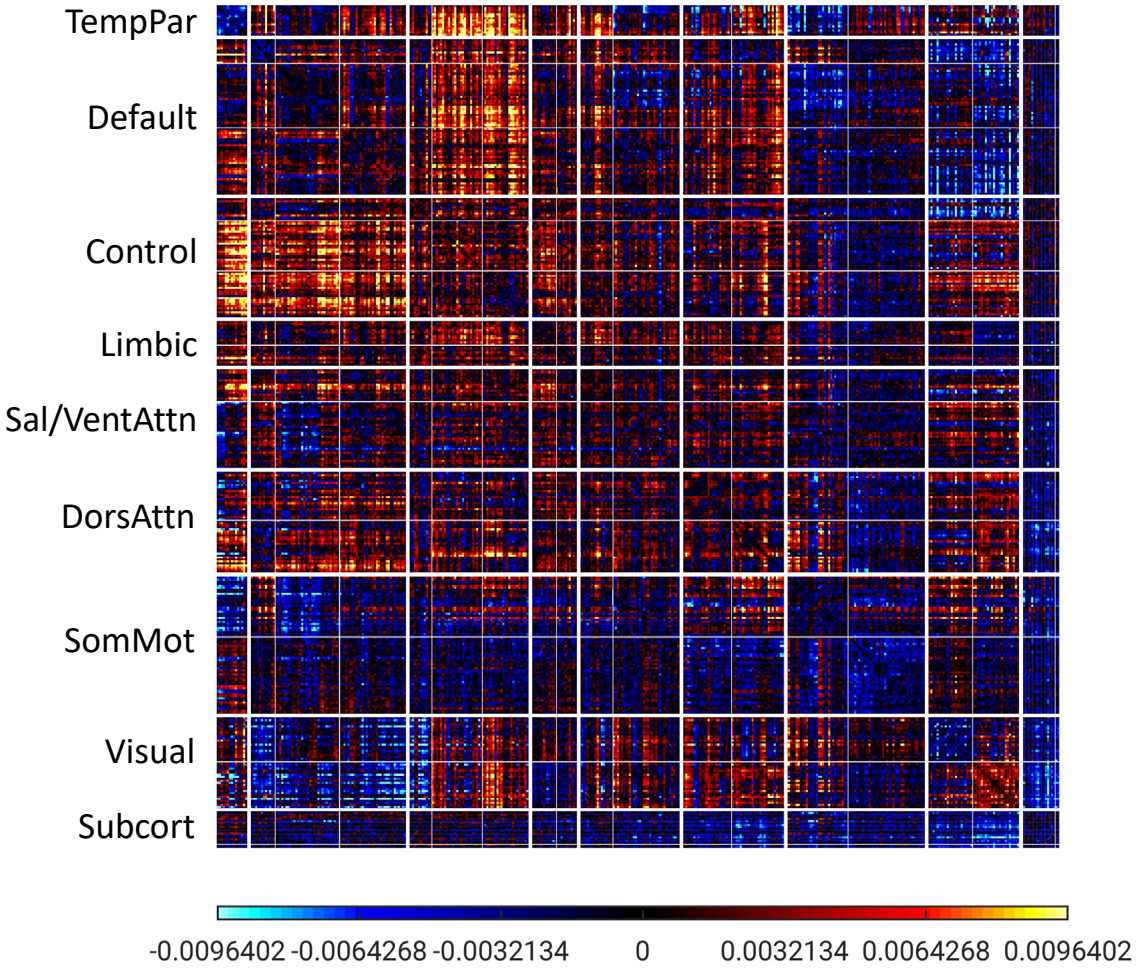
PC8 - unthresholded



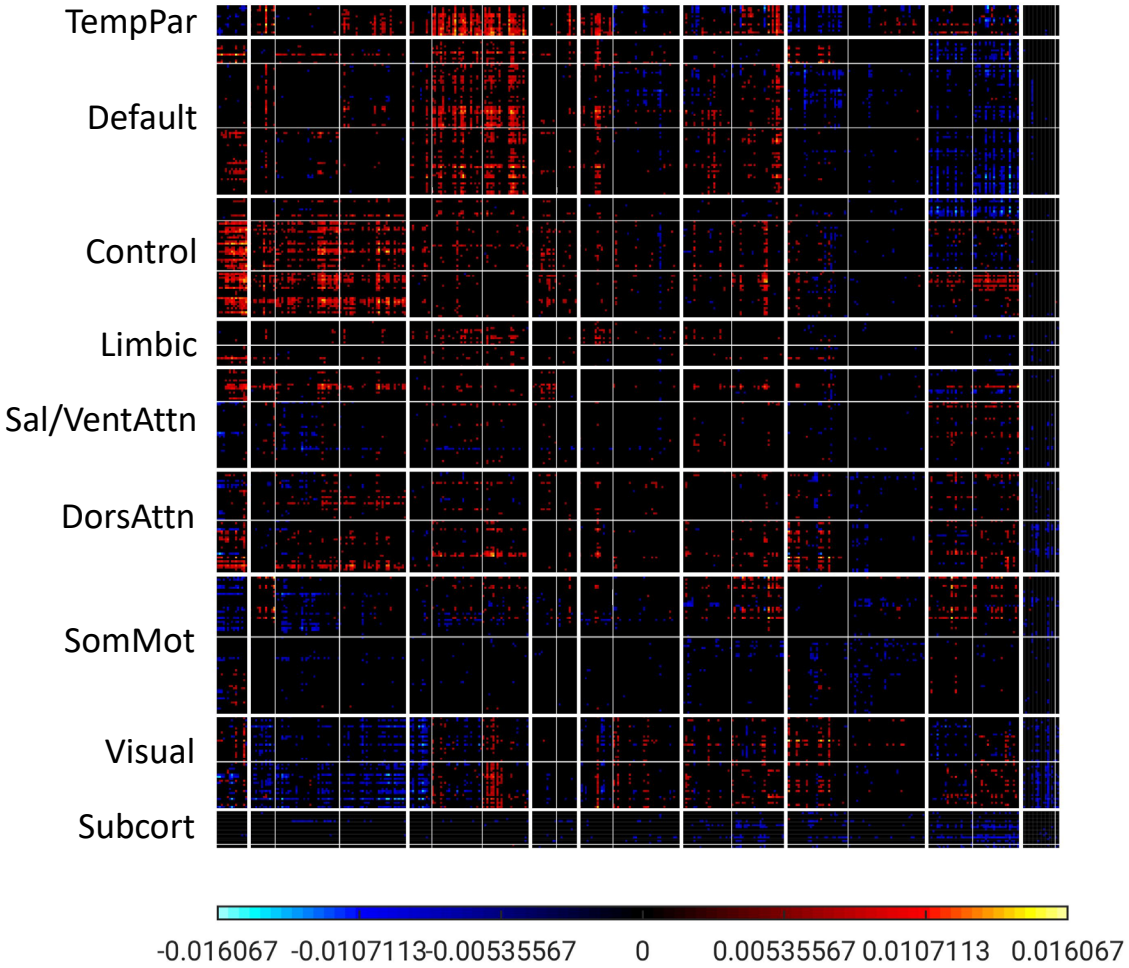
PC8 – thresholded (top 10%)



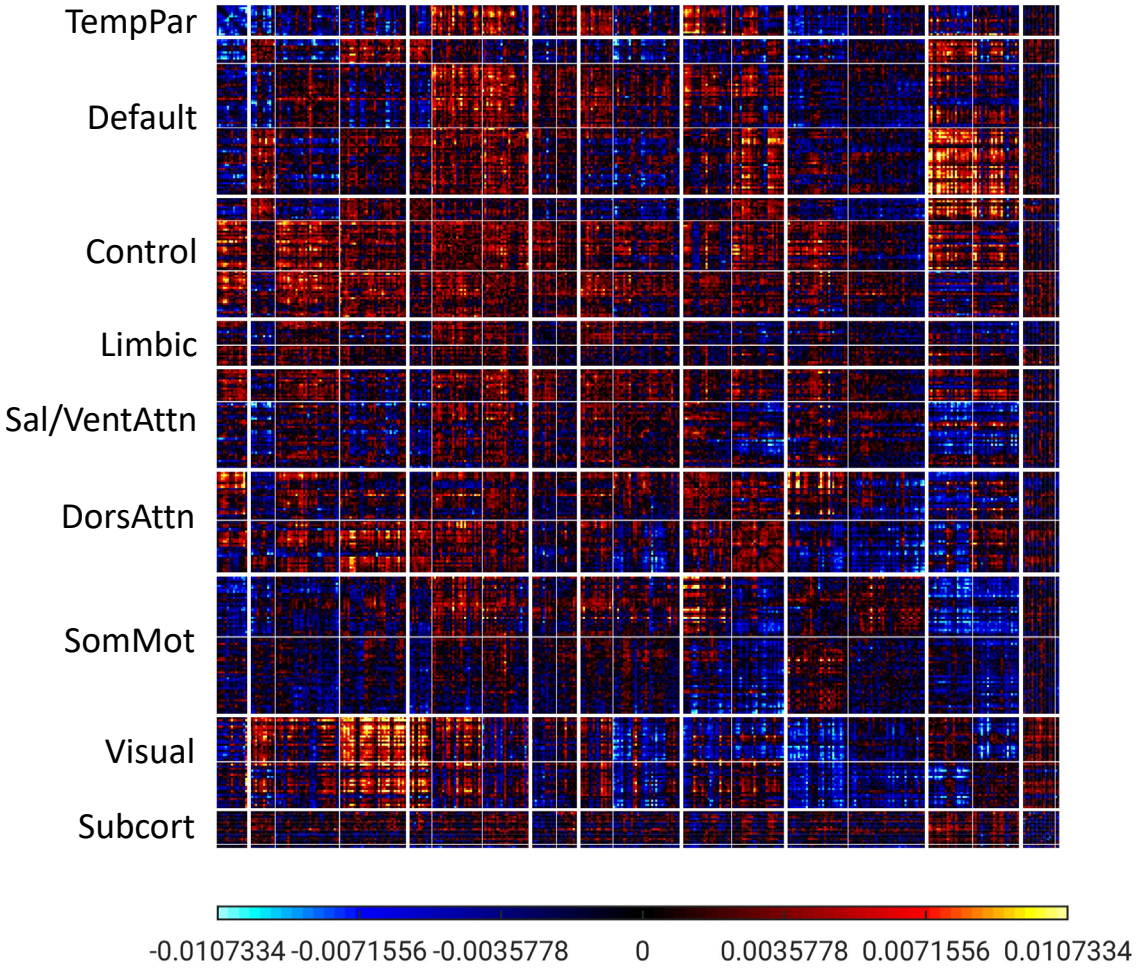
PC9 - unthresholded



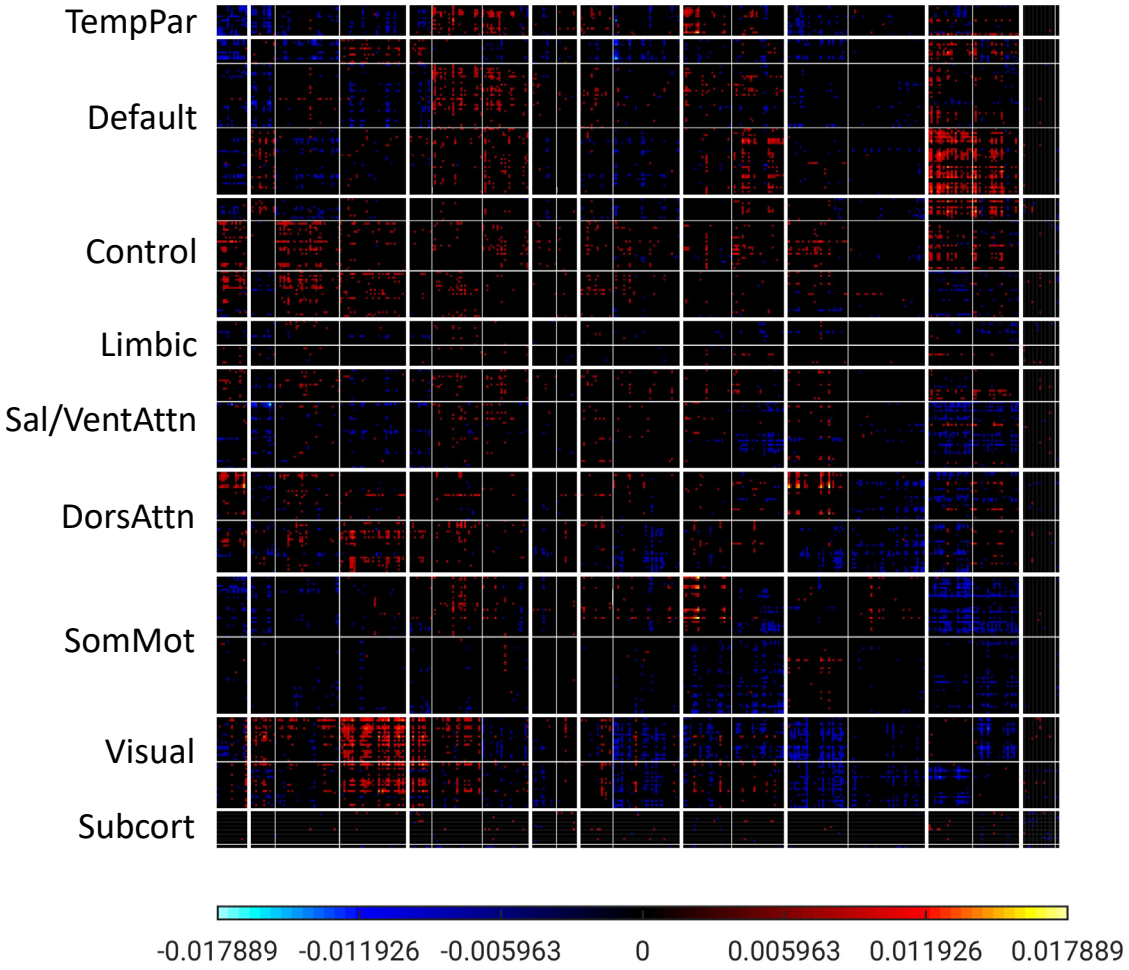
PC9 – thresholded (top 10%)



PC10 - unthresholded



PC10 – thresholded (top 10%)



Summary

- It would seem that connectivity between brain regions tend to be better in capturing information of the movie scene being viewed, rather than just the pattern of activity.
- The PC plots show that the connections *between* the Visual network and other networks, not just *within* the Visual network, tend to be more important in representing movie scenes.