PROJECT : BRAIN NETWORK ANALYSIS

FINAL REPORT

Semester : 4th
Name : Anjali Agarwal (B20AI051)
Department : School of AIDE
Under the guidance of : Prof. Dipanjan Roy and Dip Sankar Banerjee

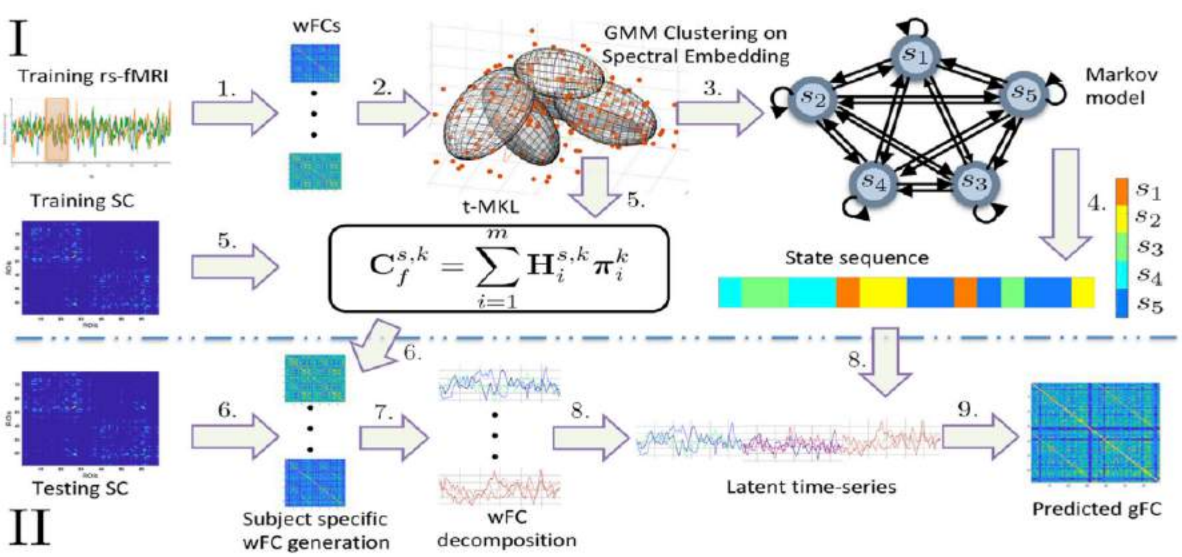
OVERVIEW :

Complex brain networks: graph theoretical analysis of structural and functional systems.

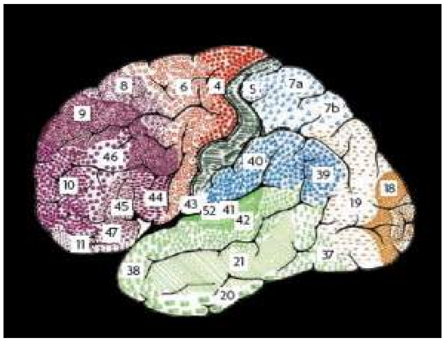
Brain networks can be defined based on structural connectivity or functional interdependence. The structural network organization of the brain is based on the anatomical linkage of its neurons

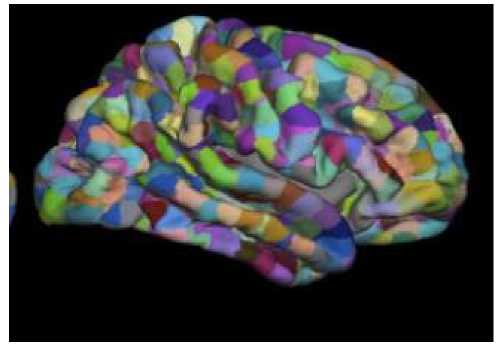
A graph is a mathematical entity consisting of a collection of nodes (vertices) and a collection of edges that connect pairs of nodes. Graphs are used to model pairwise relations between the nodes. A graph can be either directed, in which case the edges point from one node to another, or undirected, in which case the edges have no directionality.

A graph can also be weighted, in which case a numeric value is associated with every edge in the graph, or unweighted, in which case the edges are not distinguished by numeric value. A subgraph of a graph G is a graph for which the set of nodes is a subset of the set of nodes for G and the set of edges is a subset of the set of edges for G.



The structural MR image showing Nodes based on geometry





https://www.sciencedirect.com/science/article/pii/S1364661310000896#:~:text=Brain%20networks%20can%20be%20defined,axons%2C%20dendrites%20and%20gap%20junctions.

PROBLEM STATEMENT :

How do the brain networks and calculated network properties differ in different age groups?

For the given 3 tasks i.e watching movies, resting and sensory motor tasks for both young and old age groups.

AIM :

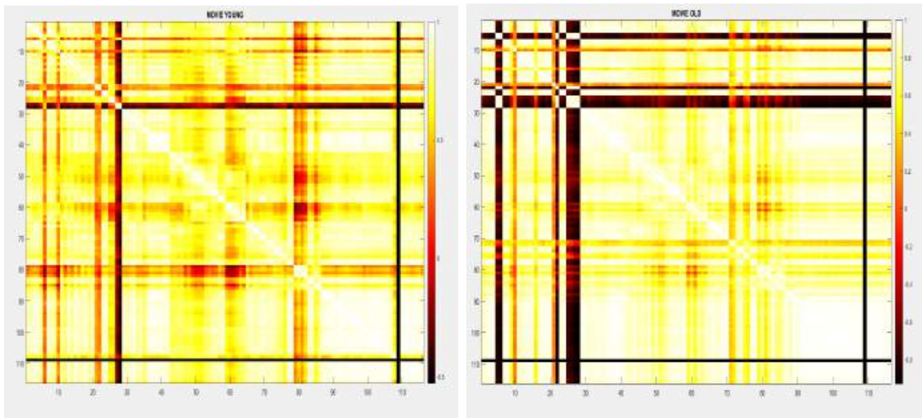
- To determine how functional network nodes defined by different recording modalities are related, and how they relate to structural network nodes.
- Compare and analyze the neural networks of different age groups using network properties.
- To predict whether the person is old/young based on time series data of 116 nodes.
- Compare the network properties like modularity(fMRI), clustering coeff, node degree and centrality of the 3 tasks namely movie, rest and sensory motor tasks for old and young people.

PROCEDURE :

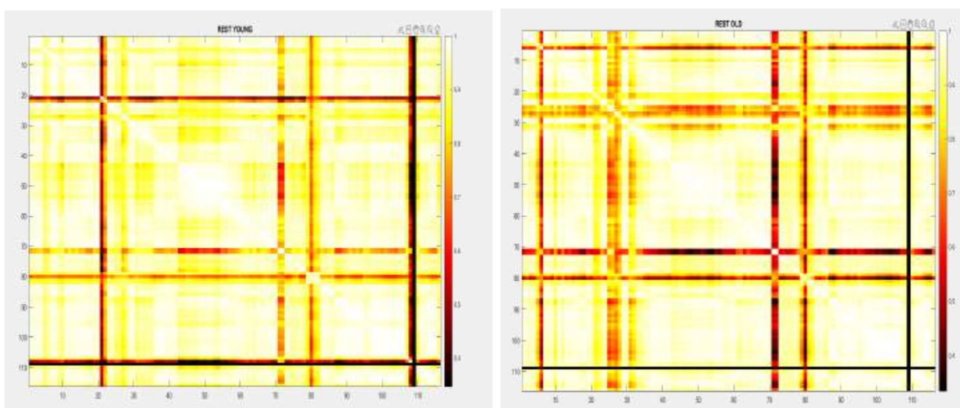
STEP 1

- Calculated the average of time series data over all the nodes for one person.
- There was no preprocessing required as the dataset was clean already.
- Calculated the adjacency matrix
- Converted the Matrix to mat format
- Plotted the adjacency matrix graph in matlab for all the 6 datasets
- Plotted the heatmap showing independence of nodes.

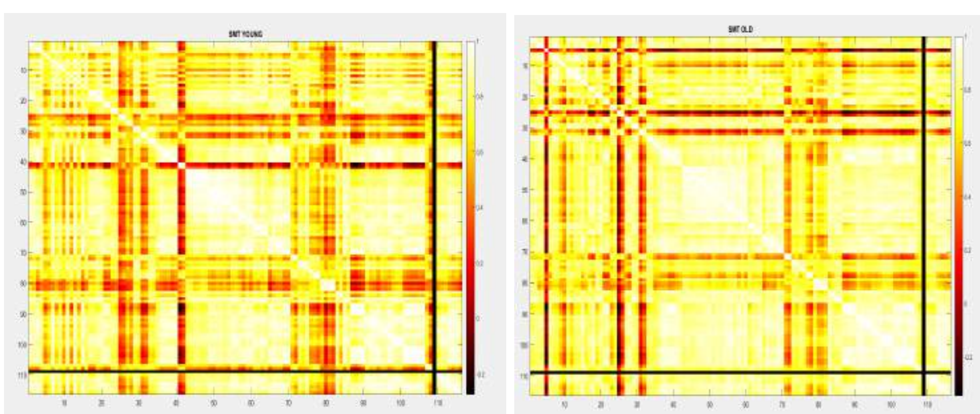
MOVIE YOUNG , MOVIE OLD



REST YOUNG, REST OLD



SMT YOUNG, SMT OLD



	0	1	2	3	4	5	6	7	8	9	...	106	107	108	109	110	111	112	
count	116.000000	116.000000	116.000000	116.000000	116.000000	116.000000	116.000000	116.000000	116.000000	116.000000	...	116.000000	116.000000	116.0	116.000000	116.000000	116.000000	116.000000	116.000
mean	0.756395	0.785804	0.788827	0.802492	0.733628	0.292091	0.778679	0.759033	0.722357	0.387479	...	0.761630	0.576374	0.0	0.804153	0.815692	0.815617	0.817948	0.813
std	0.222819	0.211677	0.234082	0.224975	0.185781	0.161420	0.234193	0.222446	0.210354	0.186100	...	0.201020	0.149624	0.0	0.241699	0.241943	0.238239	0.231615	0.239
min	-0.412369	-0.359334	-0.473954	-0.414402	-0.127322	-0.195127	-0.440774	-0.362741	-0.266224	-0.213718	...	-0.357099	-0.155213	0.0	-0.506994	-0.489357	-0.474715	-0.455891	-0.472
25%	0.753212	0.759002	0.735470	0.772533	0.697178	0.235592	0.756139	0.753434	0.692884	0.308647	...	0.759834	0.543238	0.0	0.767647	0.762301	0.781829	0.794878	0.775
50%	0.828418	0.860962	0.876468	0.886766	0.799171	0.275552	0.861540	0.821048	0.790398	0.374968	...	0.823955	0.602699	0.0	0.892592	0.910478	0.906801	0.903705	0.903
75%	0.874419	0.896158	0.926497	0.929988	0.836448	0.369806	0.907488	0.874183	0.839277	0.497908	...	0.863828	0.659452	0.0	0.951142	0.959737	0.957220	0.942407	0.959
max	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	...	1.000000	1.000000	0.0	1.000000	1.000000	1.000000	1.000000	1.000

8 rows × 116 columns

STEP 2 : ANALYSIS

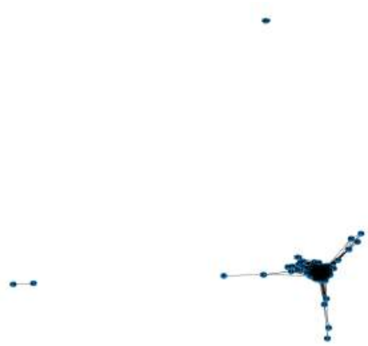
- This graph shows the interdependence between nodes

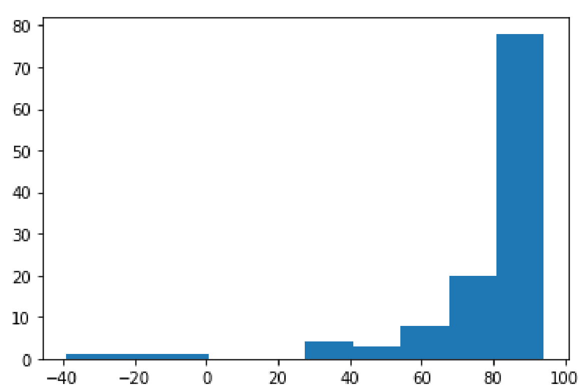
- The base here is the calculated covariance coeff. By which we calculated the covariance matrix

- The nodes that were actively connected while young people watching movies are not the same as old people.

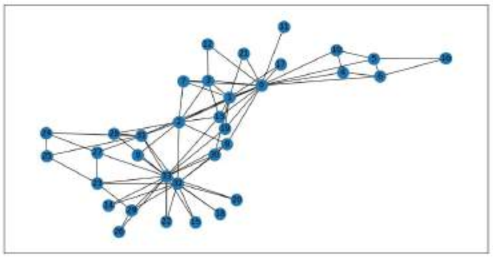
- Similarly we can analyze for every node, for all the 6 cases given.

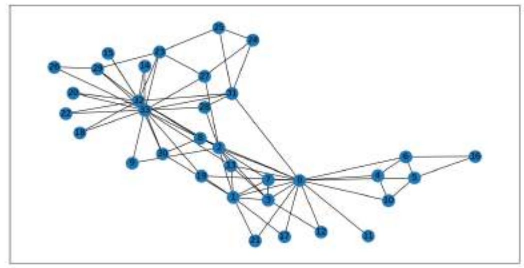
- Add the 116 nodes and then added the weighted edge , value of which was equal to the correlation value between 2 nodes.
- To select the highly dependent and connected nodes, we deleted the edges whose weight was lower than the average weight.
- Visualizing the edge weight using histogram

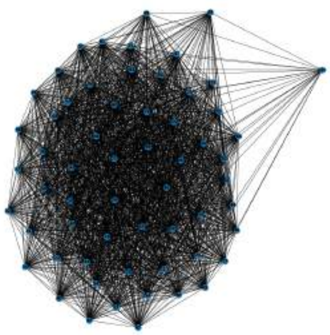


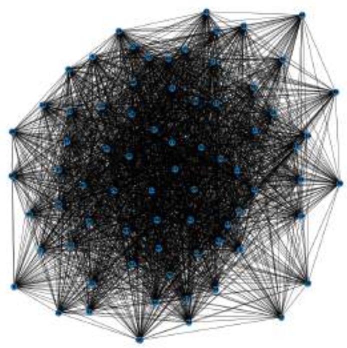


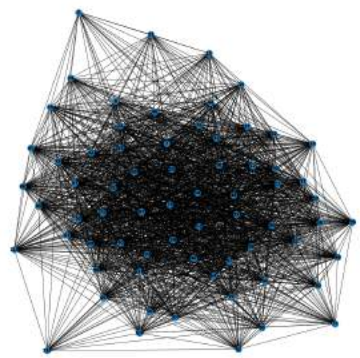
Analyzing centrality (taking 25 at a time, then one person)











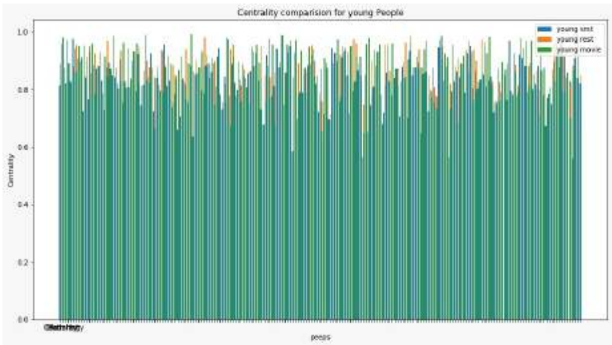
MOVIE OLD, REST OLD, SMT OLD

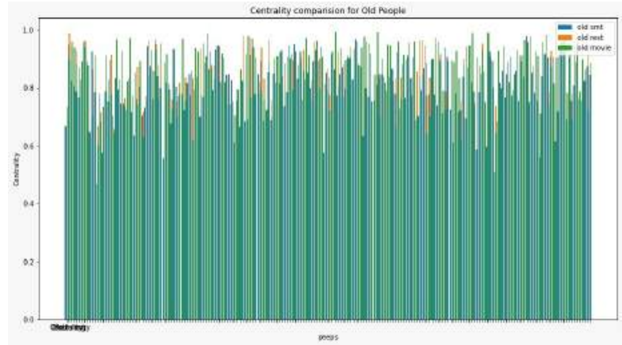
STEP 3: Network properties to be found

- Modularity：Modularity is a system property which measures the degree to which densely connected compartments within a system can be decoupled into separate communities or clusters.
- Clustering Coefficient：a clustering coefficient is a measure of the degree to which nodes in a graph tend to cluster together.
- Centrality：In graph theory and network analysis, indicators of centrality assign numbers or rankings to nodes within a graph corresponding to their network position.

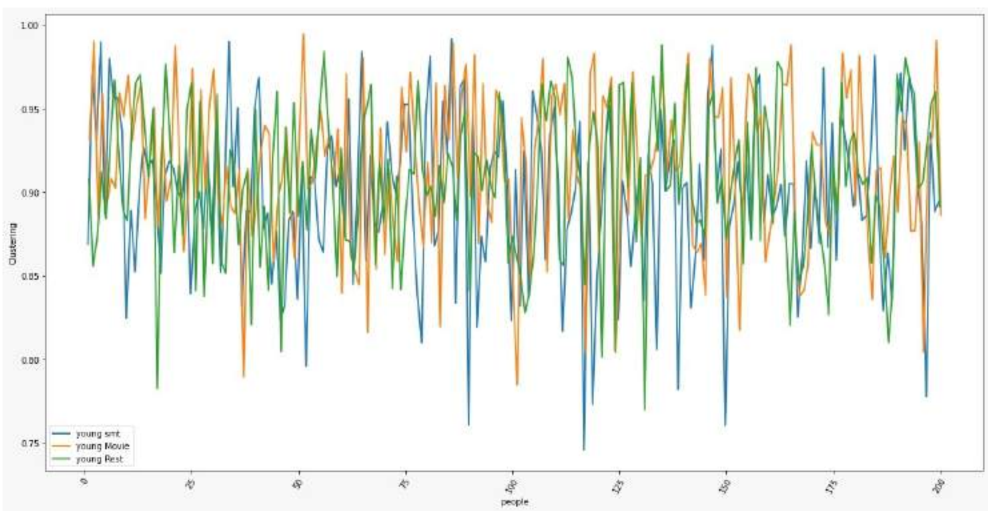
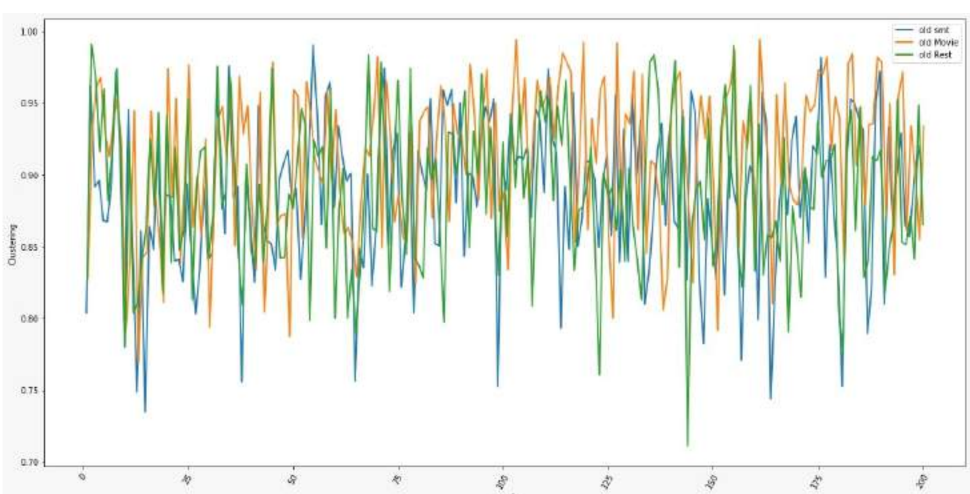
COMPARING CENTRALITY USING HISTOGRAM

YOUNG, OLD





COMPARING CLUSTERING COEFFICIENT USING LINE GRAPH

	**YOUNG**
	**OLD**

STEP 4 : STATISTICS

Noticing the overlap between the data of young and old people, while doing all the 3 tasks

- We subtracted the max value from the min value
- Then divided the whole set, by its max *m 100
- We got the % overlap

Movie young and Movie old : 61%

STEP 5: TRAINING THE MODEL

We calculated all the 3 network properties for :

1. Average
2. For each person

We created our final dataset by these network properties to finally train the data.

We used several models such as :

Fitted the data in those models using onevsrest classifier and reported the accuracy of each

Observed that Xgboost performed the best when we classified all six datasets whereas logistic regression performed better when the classification was in young and old only .

Here is the accuracy for each model in both the cases:

```
model : LogisticRegression(multi_class='ovr', solver='liblinear') accuracy score : 0.47411444141689374
model : KNeighborsClassifier() accuracy score : 0.5313351498637602
model : SVC(gamma='auto') accuracy score : 0.47411444141689374
model : RandomForestClassifier() accuracy score : 0.6049046321525886
model : DecisionTreeClassifier() accuracy score : 0.5231607629427792
model : AdaBoostClassifier() accuracy score : 0.5885558583106267
model : BaggingClassifier() accuracy score : 0.6294277929155313
model : GradientBoostingClassifier() accuracy score : 0.555858310626703
model : XGBClassifier() accuracy score : 0.670299727520436
model : LGBMClassifier() accuracy score : 0.6130790190735694
```

```
model : LogisticRegression(multi_class='ovr', solver='liblinear') accuracy score : 0.875
model : KNeighborsClassifier() accuracy score : 0.325
model : SVC(gamma='auto') accuracy score : 0.725
model : RandomForestClassifier() accuracy score : 0.22499999999999998
model : DecisionTreeClassifier() accuracy score : 0.07500000000000001
model : AdaBoostClassifier() accuracy score : 0.375
model : BaggingClassifier() accuracy score : 0.2
model : GradientBoostingClassifier() accuracy score : 0.2
model : XGBClassifier() accuracy score : 0.27499999999999997
model : LGBMClassifier() accuracy score : 0.175
```

We can try using a deep learning model to increase the accuracy, as the data provided is a very high dimensional data.

CONTRIBUTION :

- After downloading the time series data, I converted the data into readable format and calculated the adjacency matrix (by taking average of particular columns i.e for each person and finding the covariance)
- I further analyzed the matlab heatmaps
- I tried applying PCA on the dataset, but it wasn't beneficial as there was no node reduction required. We were supposed to monitor all the given 116 nodes.
- I plotted the histogram for both old and young data and we can check the overlap at each node.
- Plotting the line graph shows the similarity in clustering coefficients while doing the 3 different tasks.
- I used the networkx library and wrote all the functions to calculate the centrality, clustering coefficient and modularity.

ACKNOWLEDGMENTS :

https://www.sciencedirect.com/science/article/pii/S1364661310000896#:~:text=Brain%20networks%20can%20be%20defined.axons%2C%20dendrites%20and%20gap%20junctions.

https://www.youtube.com/watch?v=sxmMUA34CTw

https://networkx.org/

https://www.geeksforgeeks.org/python-visualize-graphs-generated-in-networkx-using-matplotlib/