Determining Objective Measures of Human Intelligence using Brain Network Analysis

Abstract

Brain Network Analysis is growing in popularity in the field of medicine today. This methodology works on the premise that brain regions and their interactions can be modeled as brain network graphs. Graphs are a simple way of representing brain network organization across a wide range of scales of space and time. Using graph-based learning algorithms, studies are being performed to understand how differences in the brain network of humans affect the characteristics of humans such as gender, intellect or creativity. In this project, our task is to predict whether a person has high or low math capability, using his/her FSIQ (Full Scale Intelligence Quotient) score.

1. Introduction:

Brain networks show functional connectivity among different regions of the brain. Brain Network Analysis is growing in popularity in the field of medicine today. In the present day, we are introduced with many brain imaging techniques such as DWI, EEG, PET which give a detailed network of brain connections [Neuron-Neuron connections] and the structure and flow of the brain network. But it is still an enigma, how the differences in the brain network correspond to the difference in age, gender, Intelligence coefficient, Creativity etc. These brain networks are represented using graphs. So, we can use graph-based learning algorithms on the brain network how the brain connectomes are related to human characteristics such as gender, intellect or creativity.

A connectome is an exhaustive map of neural associations in the cerebrum. All the more comprehensively, a connectome would incorporate the mapping of every single neural association inside a nervous system.

Graph theory studies the properties and behavior of networks, which are frameworks comprising nodes linked by connections called edges. Numerous frameworks found in nature, running from social communications to metabolic systems and transportation frameworks, can be displayed inside this structure, indicating an arrangement of hidden likenesses among these varied frameworks. The human mind can likewise be depicted as a system called connectome, where cerebrum areas are nodes and links between them are the edges. The edges are gotten as probabilities of the relationship between the nodes.

Functional connectivity is the network between brain regions that share functionalities. All the more particularly, it tends to be characterized as the transient relationship between spatially remote neurophysiological events, expressed as a deflection from factual freedom over these events in dispersed neuronal gatherings and areas. This applies to both resting state and task state examines. Functional connectivity has been assessed utilizing the perfusion time series sampled with arterial spin labeled perfusion fMRI. Functional Connectivity MRI can be used to find mental health disorders, post-traumatic stress and evaluate the effect of treatment. The functional network has been recommended to be an output of the network behavior responsible for cognitive function.

In this project, we used the brain connectivity graphs from 114 human subjects to predict their math capability. These connectivity graphs were obtained by first dividing the brain into 70 regions and the connectivity between these regions were depicted using the edge weights of the graph. From the metadata of the human subjects we obtained their FSIQ scores and based on the premise of existing studies, we considered people having an FSIQ score of greater than or equal to 120 as highly math capable. Those having an FSIQ score lower than 120 were treated as having normal math capability. Thus, we transformed the problem into a binary classification problem for predicting high math capability vs. normal math capability. We then used the topological properties of the brain connectivity graphs as features to train supervised learning models to perform this binary classification.

This project report is summarized as follows:

1) We discuss the problem description of the project.
2) We will discuss about the methodology involved in implementing the project. In this step, we describe the functional connectivity between the regions of the brain and also how we generate features for the model using graph theoretical properties of the overall network.
3) We ran the data through different classifiers and visualized the results from different classifiers.
4) Then, we conclude which classifier is good for the given data and discussion about the future work.

2. Related Works:

Research work has been extensively carried out in the past to study the differences that result due variations in functionalities of the connectomes. Some of them include classifications and predictions based on sex, age, math capability, creativity and many other discriminative characteristics. Vivek Kulkarni et.al [1] examined whether a classification of a given connectome under one of the two different sexes is possible based on structural connectivity. Their results disclosed the statistical difference at the pars orbitalis of the connectome between sexes. Their models achieved 79% accuracy when used for discriminative features alone. Julio M et.al [2] used brain connectivity networks from HARDI tractography and analyzed differences in them due to sex and kinship. Their approach produced results of classification with 88.5% accuracy for kinship and 93% accuracy for sex.

In our project, we are analyzing structural differences of brain network data from a different data set and also look at how these discriminating brain network measures can be used to classify brain connectomes with respect to math capability of human subjects.

3. Problem Description:

The main goal of this project is to predict a person's math capability by analyzing their brain network. We have considered human subjects with an FSIQ score greater than 120 as highly math capable while the rest of the subjects are classified as having normal math capability. Supervised learning algorithms were implemented to perform this binary classification. Topological features of the brain network graph were extracted. Using this, we found the most pivotal nodes/edges of the graph with respect to our classification problem. The feature values at the pivotal nodes were used to classify a brain network corresponding to a human subject as having highly math capability or normal math capability.

Therefore, the project is composed of three main tasks:

1) Extraction of topological features from the brain network graph.
2) Determining the pivotal nodes/edges for our classification problem from the graph.
3) Classification of the human subjects as highly math capable or not using feature values of the pre-determined pivotal nodes/edges.

4. Dataset Description

In this project, we used the dataset published by Stony Brook University consisting of connectome data for 114 individuals. Each brain network graph was composed of a total of 70 nodes/voxels. The undirected edge weight is the number of fibers that pass through any pair of nodes. The dataset also contains the metadata for each human subject such as age, gender and FSIQ scores. Using the FSIQ scores, each sample or connectome in our dataset was assigned a class label (0 for normal math capability, 1 for high math capability), thus identifying the math capability with said connectome. Fig. 1. shows the split between the 2 classes in our dataset. Approximately 55% of the human subjects had high math capability and 45% of them had normal math capability.

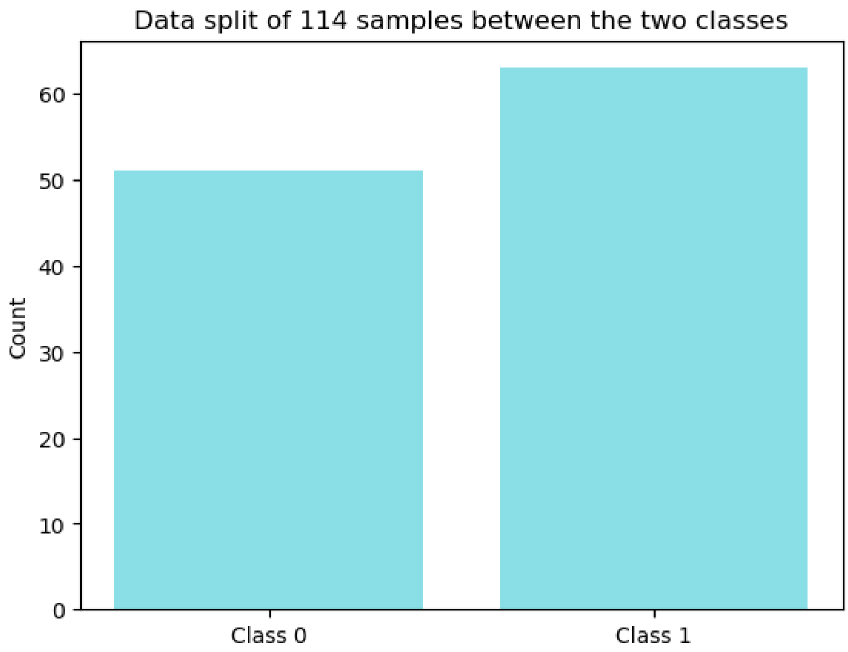

Fig. 1. Data split between the 2 classes

5. Methodology:

5.1 Feature Extraction:

Given that our problem is a graph learning problem, we extracted topological features of the brain networks graphs to be fed as the input to our classifiers. This was done using the NetworkX library in Python. We extracted local features such as clustering coefficient, local efficiency, participation coefficient and edge betweenness centrality pertaining to the individual components of the graphs (nodes and edges). We then examined the correlation between the features of each component and the output classes across our training examples and determined the components that best differentiated the 2 classes. The features pertaining to these pivotal regions were chosen as the input for our classifiers.

5.1.1 Clustering Coefficient

Clustering Coefficient is a measure of the degree to which nodes in a graph tend to cluster together. The local clustering coefficient of a node is given by the proportion of links between the vertices within its neighborhood divided by the number of links that could possibly exist between them.

In an undirected graph $G = (V, E)$, the local clustering coefficient of a vertex v_i is computed as follows:

$$C_i = \frac{2|\{e_{jk}: v_j, v_k \in N_i, e_{jk} \in E\}|}{k_i(k_i - 1)}$$

where, N_i is the neighborhood of vertex v_i and k_i is the number of neighbours of v_i.

We analyzed the mean clustering coefficient of each node and studied the distribution of these values across the 2 classes as shown in Fig. 2.

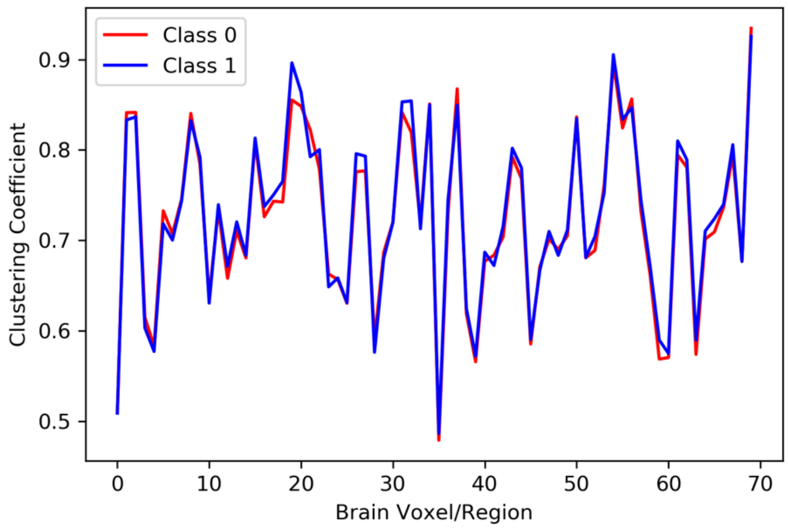

Fig. 2. Mean clustering coefficient of each node for the 2 classes

We noted that the mean clustering coefficient values of Nodes 19, 32, 21, 18 and 22 were significantly higher in highly math capable individuals (Class 1) than in individuals with normal math capability (Class 0). In order to rule out the effects of outliers we also studied the median values of clustering coefficient of each node. We got similar results with the median values as well. Since the clustering coefficient values of these nodes provided a good differentiation between the 2 classes these values were included in the feature vector to be fed to our classifiers.

5.1.2 Local Efficiency

The efficiency of a pair of nodes in a graph is the multiplicative inverse of the shortest path distance between the nodes. The local efficiency of a node in the graph is the average global efficiency of the subgraph induced by the neighbors of the node.

The local efficiency of a graph $G = (V, E)$ is defined as follows:

$$E(G) = \frac{1}{n}\sum_{i \in G} E(G_i)$$

where G_i is a local subgraph consisting only of a node immediate neighbors, but not the node {\displaystyle i} itself..

We analyzed the mean local efficiency of each node and studied the distribution of these values across the 2 classes as shown in Fig. 3.

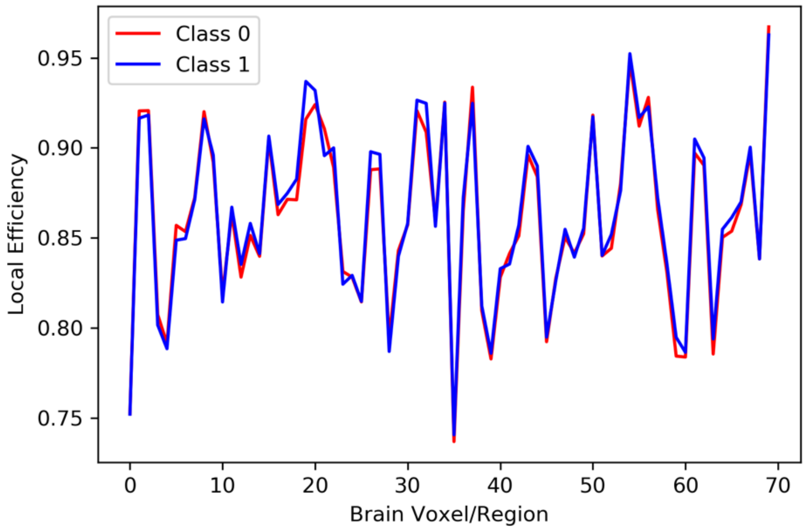

Fig. 3. Mean local efficiency of each node across the 2 classes

It was observed that the local efficiency values of Nodes 19, 32, 21, 18 and 22 were significantly higher in highly math capable individuals (Class 1) than in individuals with normal math capability (Class 0). In order to rule out the effects of outliers we also studied the median values of local efficiency of each node. Since the local efficiency values of these nodes are distinguishable across the 2 classes this was chosen as one of the input features for our classifiers.

5.1.3 Participation Coefficient

As explained in Vivek Kulkarni et al. [1], the participation coefficient is a measure based on modularity. It represents the diversity of inter-modular connections of a given node. Intuitively the participation coefficient of a node is close to 1 if its links are uniformly distributed across all modules and 0 if all its links are within its own module. A node with a high participation coefficient thus represents a connector hub in the brain.

We analyzed the mean participation coefficient of each voxel and studied the relationship of these values with the output variable as shown in Fig. 4.

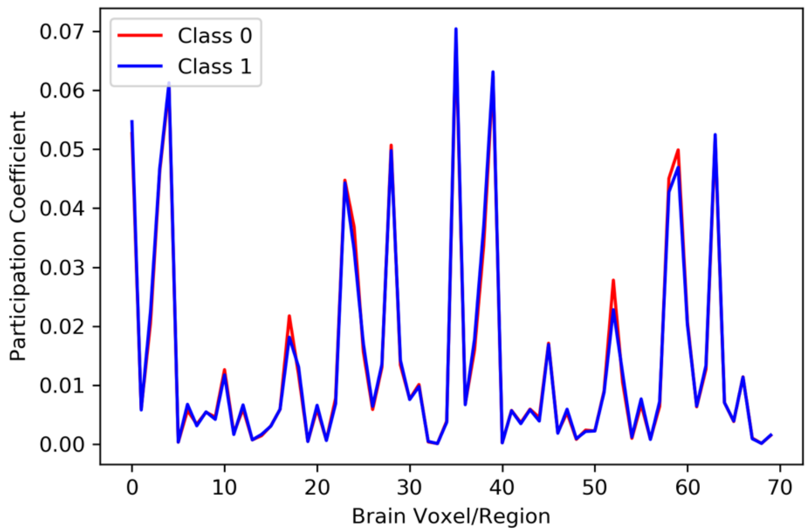

Fig. 4. Mean participation coefficient of each voxel across the 2 classes

We found that the mean participation coefficient value of Node 52 was discriminative across the 2 classes. This value was significantly higher in normally math capable individuals (Class 1) than highly math capability individuals (Class 0). Hence this feature value was included in our feature vector.

5.1.4 Edge Betweenness Centrality

The edge betweenness centrality of an edge is defined as the number of the shortest paths that go through an edge in a graph or network. An edge with a high edge betweenness centrality score represents a bridge-like connector between two parts of a network, and the removal of which may affect the communication between many pairs of nodes through the shortest paths between them.

We analyzed the mean edge betweenness centrality values of each edge and studied their relationship with the 2 classes as shown in Fig. 5.

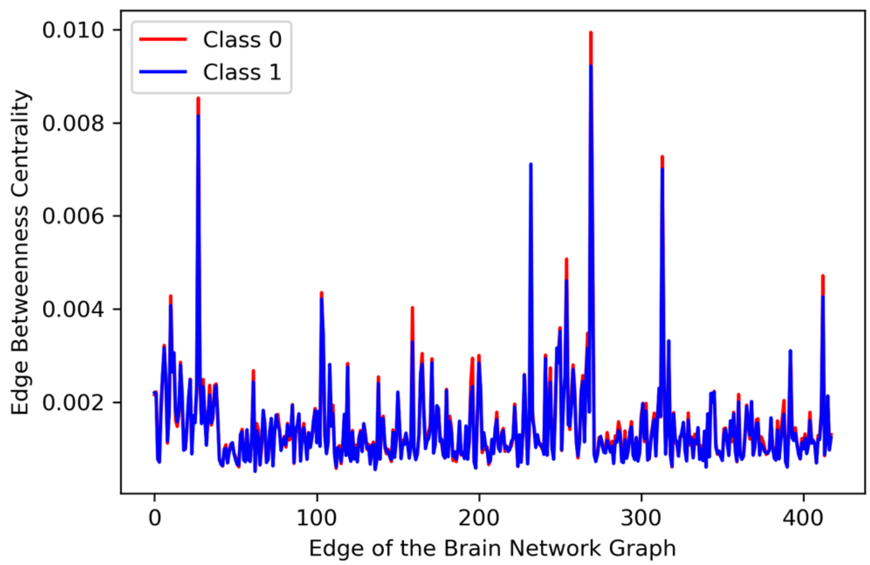

Fig. 5. Mean edge betweenness centrality for the graph edges across the 2 classes

It can be seen that the edge betweenness centrality values of Edges 232, 159, 269, 270 are significantly higher in individuals with normal math capability (Class 0) than in individuals with high math capability (Class 1). Our analysis of the edge betweenness centrality indicates that the above-mentioned edges are discriminative across the 2 classes. Hence these values were chosen as part of the input feature vector to train the classifiers.

5.2 Classifiers

5.2.1 SVM

Support Vector Machine (SVM) is a supervised machine learning algorithm. Consider an n-dimensional hyperspace (n is number of features), where every sample in our training dataset is represented by a single point, with the position of the point in the space corresponding to feature values of the sample. Some are positive examples, while others are negative. The goal of an SVM classifier is to find a separating hyperplane in this n-dimensional hyperspace, that differentiates the points corresponding to two classes (+ or -) very well. The hyperplane is characterized by certain points from the training dataset, which are called support vectors.Once this hyperplane is learned, classification is reduced to the simple task of finding the position of a test sample with respect to the hyperplane.

The inputs to SVM training function are:

- m training examples, appended with bias variable (x_i, i=1,..,m)
- m corresponding labels to each example (y_i, i $= 1, .. ,$ m)

- λ, the regularization parameter
- ε, the tolerance of the classifier.

SVM was implemented using Pegasos algorithm as mentioned in [4], where the unconstrained optimization problem is solved in its primal form using sub-gradient descent, while still leaving room for implementation of kernels.
The training objective function is as follows:

$$\min_{\alpha} \frac{\lambda}{2} \sum_{i,j=1}^{m} \alpha[i]\alpha[j]K(x_i, x_j) + \frac{1}{m} \sum_{i=1}^{m} \max\left\{0,1 - y_i \sum_{j=1}^{m} \alpha[j]K(x_i, x_j)\right\}$$

Where α, corresponds to Lagrangian multipliers and $K(.,.)$ represents the kernel function. Although this solution is presented in terms of α, sub-gradient descent is done with respect to weight vectors as explained in [4]. This algorithm is guaranteed to converge in $O\left(\frac{1}{\lambda\varepsilon}\right)$ iterations.
$\lambda = 1$ *and* $\varepsilon = 0.001$ were chosen during implementation, ensuring a valid solution in 1000 iterations.
The classification is done using: $y_{test} = sign(\sum_i y_i(\alpha[i])^T K(x_i, x_{test}))$
The use of bias variables eliminates the need for an intercept variable 'b'.
We implemented the radial basis function kernel, given by

$$K(x_i, x_j) = e^{-\gamma\left|\left|x_i - x_j\right|\right|_2^2}$$

Where γ defines the influence of each training example. $\gamma = \frac{1}{no\_of\_features} = \frac{1}{15}$ was chosen during implementation.

5.2.2 Logistic Regression (SGD)

Logistic regression is a supervised classification algorithm. At its core lies the logistic function (also known as the sigmoid function), which to takes any real value and maps it to a value between 0 and 1. The logistic function $f(x)$ is defined as follows:

$$f(x) = \frac{1}{1+e^{-x}}\,.$$

Logistic regression predicts the output by linearly combining weights and the input feature vector and feeding that value to the logistic function as input. So for a feature vector x and weight vector w, the class probability is predicted as:

$$p(y_i = 1) = \frac{1}{1+e^{-wx_i}}.$$

Where y_i is the predicted label of the i^{th} example and x_i is the feature vector of that example. Stochastic gradient ascent was used to train the weights. The update formula for the weights is as follows:

$$w_i = w_i + \eta \sum_{l} X_i^l (Y^l - P(Y^l = 1|X^l, w)$$

A learning rate (η) of 0.0001 was used and the following objective function was used to compute the cost.

$$l(w) = \sum_{l} (Y^l - 1)(w_0 + \sum_{i=1}^{n} w_i X_i^l) - \log(1 + e^{-(w_0 + \sum_{i=1}^{n} w_i X_i^l)})$$

If the difference in the costs of two successive iterations equal to or less than zero, then we assume convergence and the weights from that iteration are used as the trained weights.

5.2.3 kNN

K-Nearest Neighbour is an instance-based learning algorithm where all computations are done with respect to a single example at a time locally and they are kept aside until classification is done. k-NN can be used both for regression and classification. When K = 1 each example defines its own space, creates a Voronoi partition of the space. This algorithm is non-parametric as model is defined based on input data.

The inputs to the KNN algorithm:

- m training examples.
- m corresponding labels for each example.

Euclidean distance (L2 distance) metric was used as the base distance to find the nearest neighbour. If $x=(x_1, x_2, \ldots \ldots \ldots x_n)$ and $y= (y_1, y_2, \ldots \ldots \ldots y_n)$ are two points in Euclidean space, the L2 distance between them is given by :

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 \ldots \ldots \ldots \ldots . + (x_n - y_n)^2}$$
$$= \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

Specific to the problem in this project $x= (x_1, x_2, \ldots \ldots \ldots x_n)$ is a test example and $y= (y_1, y_2, \ldots \ldots \ldots y_n)$ is a training example. Both the training and test example have n features and the distance between each feature for each example is calculated.

Data points with minimal distance were found and the label with corresponding to majority number of neighbours was assigned as the label for the test data point. The model was tested for k values ranging from 1 to 10 and the best results was produced by k = 5.

6. Results:

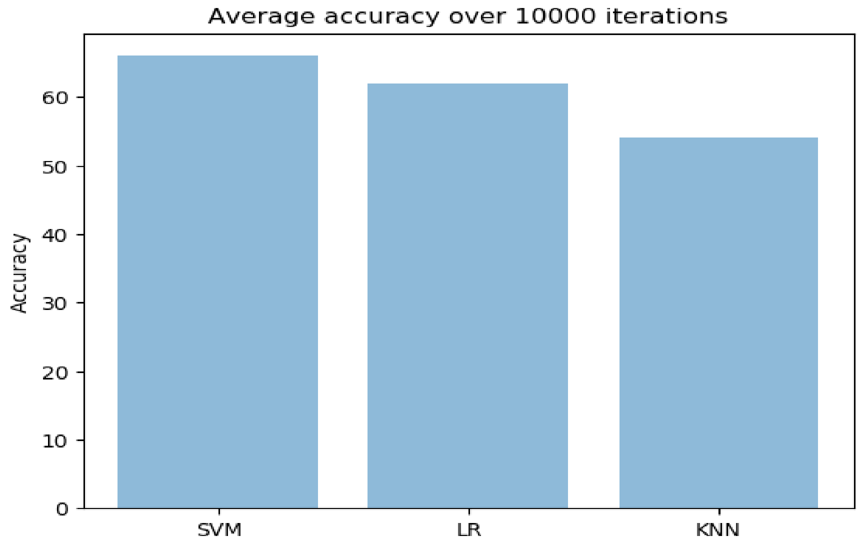

Fig. 6. Prediction Accuracy of Different Classifiers

7. Conclusions and Future Work:

In this project, we are trying to predict if a human subject has high or low math capability from the brain network graphs of the human subjects. To support our predictions, we use brain connectivity graphs from 114 human subjects containing connections between different regions of the brain, which has metadata of human characteristics like age, gender etc. We label the subjects with math capability more than 120 as 1 and others 0 for the classification purpose. Then we were able to make a prediction about the math capability of a human subject.

We tried different models like SVM, Logistic Regression, KNN to predict the math capability of the human subjects. We found that SVM provides the highest prediction accuracy among the other classifiers. But as the dataset is not big enough, the results are not satisfying enough to apply it in real world.

It is important to use larger dataset for training and testing the classifier as the dataset we have used is small and may not have enough discriminating power to classify the features. It would also be useful to use advanced learning algorithms like ensemble learning and deep learning methods for classification. We could extend our analysis to study the differences in creativity, neuroticism, agreeability etc.

8. References:

[1] Vivek Kulkarni, Jagat Sastry, "Sex differences in the Human Connectome", International Conference on Brain and Health Informatics BHI 2013: Brain and Health Informatics pp 82-91

[2] Julio M.Duarte-Carvajalino, Neda Jahanshad, Christophe Lenglet, Katie L.McMahon, Greig I.de Zubicaray, Nicholas G.Martin, Margaret J.Wright, Paul M.Thompson, Guillermo Sapiro, "Hierarchical topological network analysis of anatomical human brain connectivity and differences related to sex and kinship" NeuroImage Volume 59, Issue 4, 15 February 2012

[3] Shai Shalev-Shwartz, Yoram Singer, Nathan Srebro, Andrew Cotter, “Pegasos: Primal Estimated sub-GrAdient SOlver for SVM”, Mathematical Programming, March 2011, Volume 127, Issue 1, pp 3–30.

[4] Marcus Kaiser, “A tutorial in Connectome Analysis: Topological and Spatial Features of Brain Networks”, Available at “https://arxiv.org/ftp/arxiv/papers/1105/1105.4705.pdf”

[5] Xi-Nian Zuo, Ross Ehmke, Maarten Mennes, Davide Imperati, F. Xavier Castellanos, Olaf Sporns, Michael P. Milham,“Network Centrality in the Human Functional Connectome”,Cerebral Cortex, Volume 22, Issue 8, 1 August 2012, Pages 1862–1875, https://doi.org/10.1093/cercor/bhr269.

[6] Achard S, Salvador R, Whitcher B, Suckling J, Bullmore E, “A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs”, J Neuroscience , 2006, vol. 26 (pg. 63-72)

[7] S. Achard, E.T. Bullmore, “Efficiency and cost of economical brain functional networks”, PLoS Computational. Biology., 3 (e17) (2007)

[8] Eugene M. Izhikevich,“Simple Model of Spiking Neurons”, IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 14, NO. 6, NOVEMBER 2003

[9] Joshua T. Vogelstein, William R. Gray, R. Jacob Vogelstein, Carey E. Priebe, “Graph Classification using Signal-Subgraphs: Applications in Statistical Connectomics” available at “https://arxiv.org/abs/1108.1427”