IDENTIFICATION OF FUNCTIONAL CONNECTIONS WITHIN ALZHEIMER'S PROTEIN-TO-PROTEIN INTERACTOME

1.0 BACKGROUND

Alzheimer's is a neurodegenerative disease pathologically characterized by the presence of amyloid plaques in the grey matter. It accounts for about 60 to 70% of dementia cases. Even though studies have identified misfolded Amyloid Beta (Aβ) protein and its precursor, Amyloid precursor protein APP(A4) to play a significant role in the plaque formation, the physiologic interactions of these proteins in the brain remains poorly understood. This knowledge gap hinders the development of effective drugs to halt the progression of the disease. In this regard, addressing this gap is crucial for identifying potential targets for effective drug development, advancing our understanding of the functional roles of proteins involved in Alzheimer's disease, reducing the socio-economic burden of the disease, improving care delivery for Alzheimer patients thereby reducing disease progression, and improving overall patient's quality of life.

2.0 OBJECTIVE

This study aims use network analysis to identify:

1) Important protein interactions that can be further explored using computational and bioinformatic approaches for the purpose of drug development and understanding of the disease.

3.0 METHODOLOGY

3.1 Data Preprocessing

The Alzheimer dataset is an interactome of proteins linked to the disease process. A compilation of protein interactions determined using various laboratory methods. The data set was downloaded from the EMBL-EBI INACT database. Dataset was filtered by; **interaction host** (Saccharomyces cerevisiae [Baker's yeast] and human neuron), **interactor species** (Homo sapiens), **interactor type** (protein to protein), and **detection method** (two hybrid array, two hybrid pooling, validated two hybrid, anti-tag coip). This brought the shape of the dataset to (55583, 7) despite which the data export button was disabled due to its large size. Attempted web-scrapping, but it was too slow, and I had to directly copy and save in an excel spread sheet from which I converted it into a pandas data-frame for use in my code. After deleting duplicates, the dataset was reshaped to (47359, 7).

In setting up the network, I set the nodes to proteins in the column named 'Molecule A' as the source node while the column named 'Molecule B' the target node. The edges were the confidence value of the interaction between the source and target nodes. The confidence value is derived from the Mi score which is computed based on the interaction detection method, interaction type and the number of publications reporting the interactions.

The python packages used for network analysis were Networkx, while Plotly and Matplotlib were used for visualization. The graph layout used in the visualizations were the spring layout, kamada_kawai_layout, and random layout.

Proteins function in a signaling cascade by activating or inhibiting other proteins hence, network properties in the protein-to-protein interactions that were of particular interests were the **betweenness centrality** which will help identify proteins that have the greatest influence in a signaling process based on the shortest path in the networks graph, **degree centrality** to identify the most connected node

i.e the central or important nodes to the network which is derived from the number of connections to the node, **closeness centrality** which helps to identify nodes that can easily interact with other nodes, these can be seen as key regulatory proteins, and **eigenvector centrality** which identify the most influential proteins based on their connections to other important proteins

4.0 RESULTS

Total number of nodes was 5751 while the total number of edges was 19208. Using the network properties mentioned above, a couple of nodes (proteins) were identified that are likely to be important to the network and their interactions can be further characterized with computational or bioinformatic approaches.

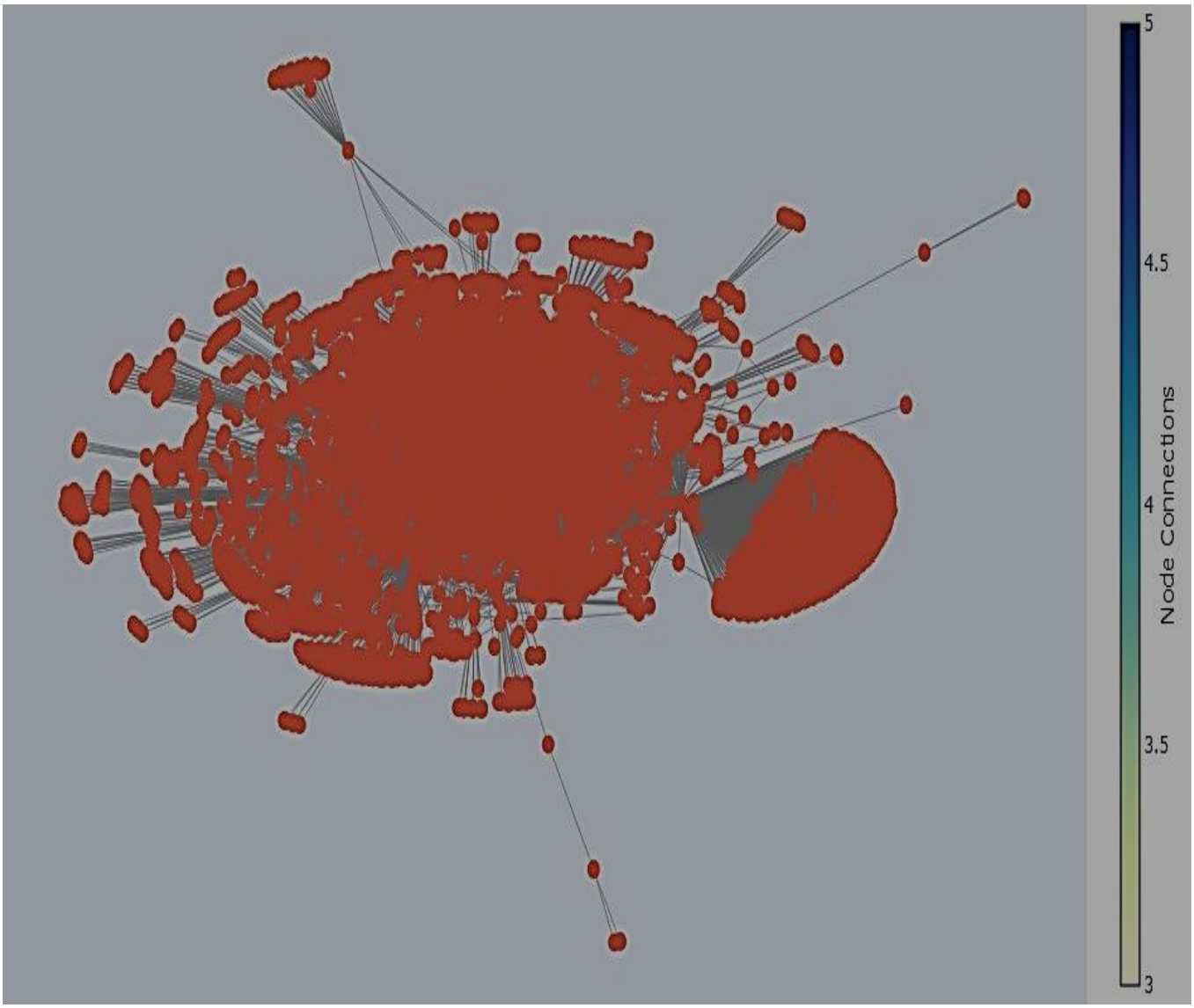

Fig 1.0 Spring layout graph network of Alzheimer's protein to protein interaction. Owing to the large size of the data set, labels would completely blur visualization of the network. To get a detailed view kindly refer to the graph in the Ipynb file.

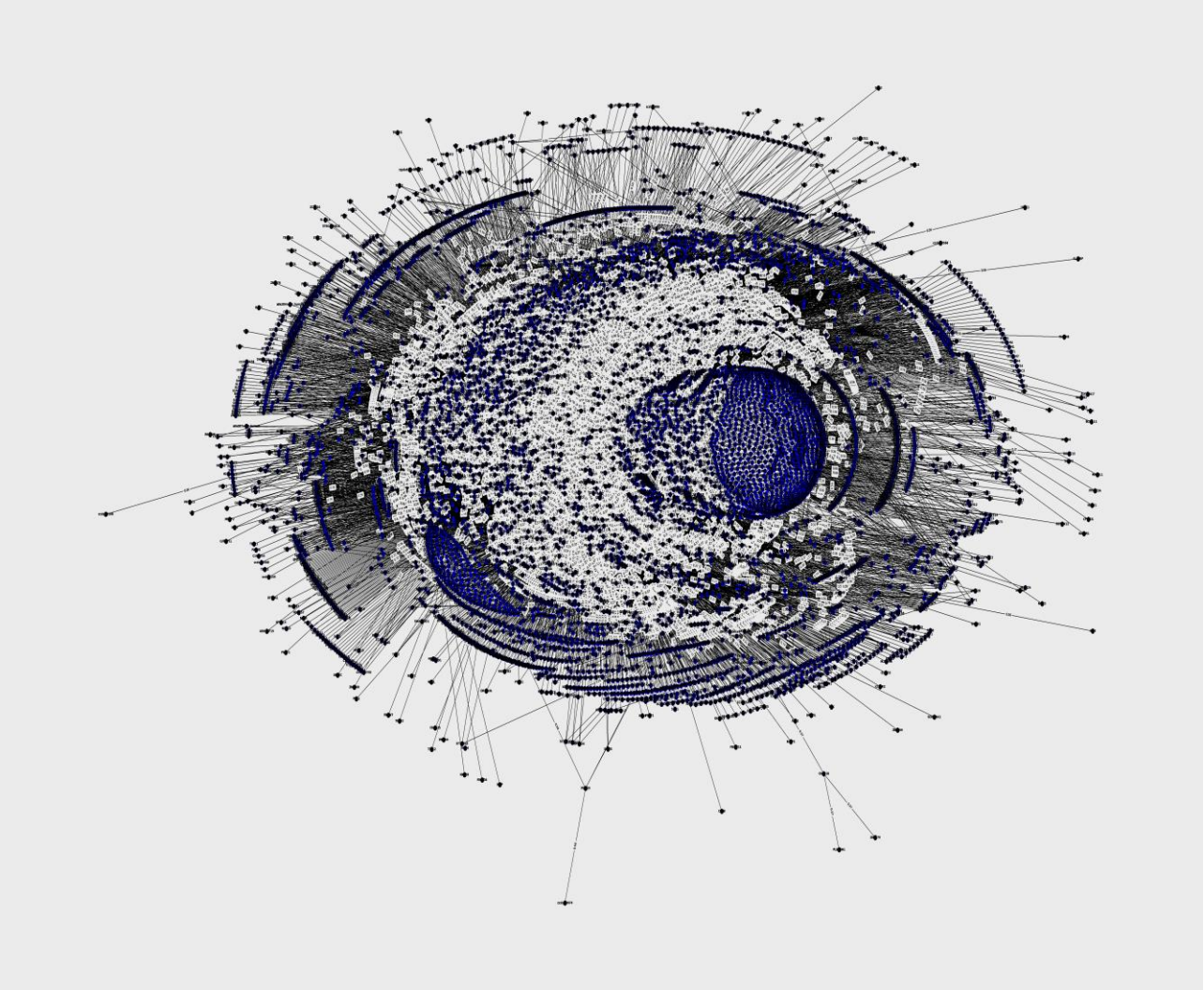

Fig 2.0 kamada_kawai_layout graph network of Alzheimer's protein to protein interaction. Owing to the large size of the data set, labels would completely blur visualization of the network. To get a detailed view kindly refer to the graph in the ipynb file.

MAPT (Microtubule Associated Protein Tau) which is involved in the stabilization and assembly of microtubules had the highest degree, betweenness, closeness and eigenvalue. This was followed by APP. This is quite as expected since many literatures have reported the key role they play in the development of Alzheimer's. However, of interest are the other less documented nodes (proteins) that exhibited similarly high centrality values. The closeness centrality, betweenness centrality, and eigenvector centrality identified the same set of proteins that exert the greatest influence on the network however, asides the MAPT protein, the degree centrality identified different sets of proteins. While the density distribution of the degree centrality had a single peak at 0.0, the betweenness, closeness, and eigenvector centrality had similar double peak distribution between 0.2 and 0.4.

4.1 DEGREE CENTRALITY

The degree centrality is defined by the number of edges that is connected to a node. It identifies the nodes that have the highest number of connections within the network. Computed by dividing the number of edges connected to the node by the total number of nodes in the network. A protein that is highly connected to other proteins can easily modify many aspects of the disease process thus exerting great influence on its progression. Proteins identified by degree centrality are listed in the horizontal bar chart in descending order of their degree (fig 5.0).

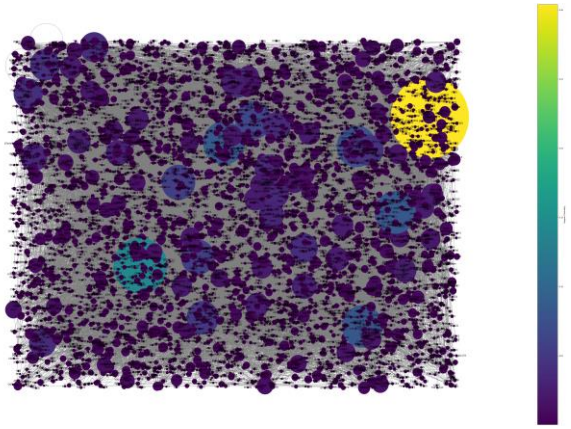

Fig 3.0 A Random layout network highlighting the degree centrality according to colour intensity and node size. The biggest and brightest (yellow) node being the MAPT protein.

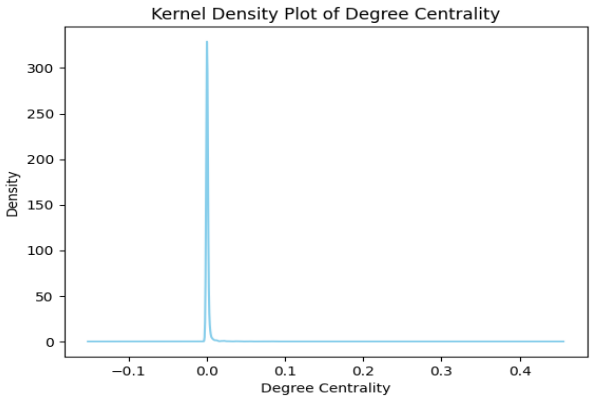

Fig 4.0 Density distribution of the degree centrality. Majority of the nodes had very few edges connected to them with few nodes having the greatest influence.

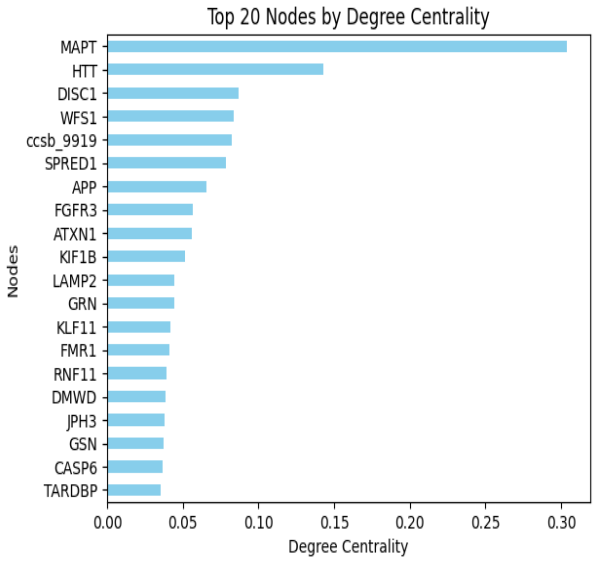

Fig 5.0 Horizontal bar chart of the top 20 nodes by degree centrality in descending order.

4.2 BETWEENESS CENTRALITY

The betweenness centrality is the number of shortest paths that passes through a node. It identifies nodes that directs the passage of information within a network. In protein-to-protein interactions, this may serve as signaling proteins that conveys information from one part of the cell to another.

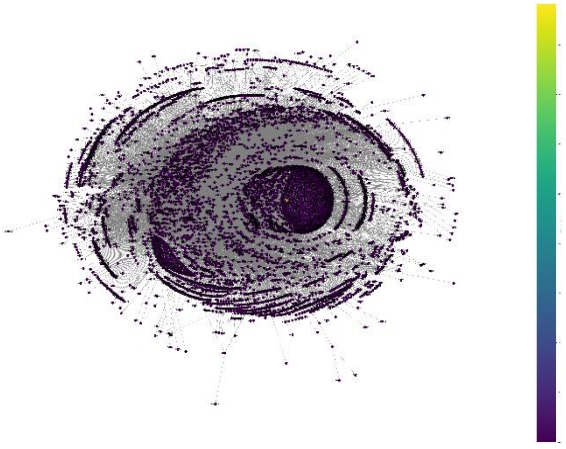

Fig 6.0 A kamada_kawai_layout network highlighting the betweenness centrality according to colour intensity. The brightest

(yellow) node being the MAPT protein at the centre of the network.

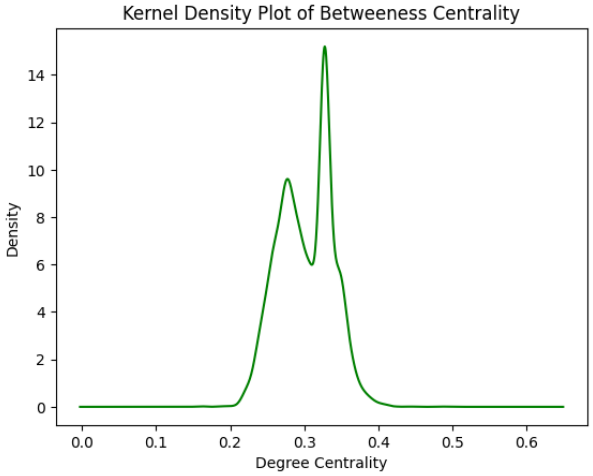

Fig 7.0 Density distribution of the betweenness centrality. Majority of the nodes are concentrated between 0.2 and 0.4. Unlike degree centrality, a good number of the proteins exhibited a high betweenness centrality hinting that many of them may be signalling proteins than regulatory proteins.

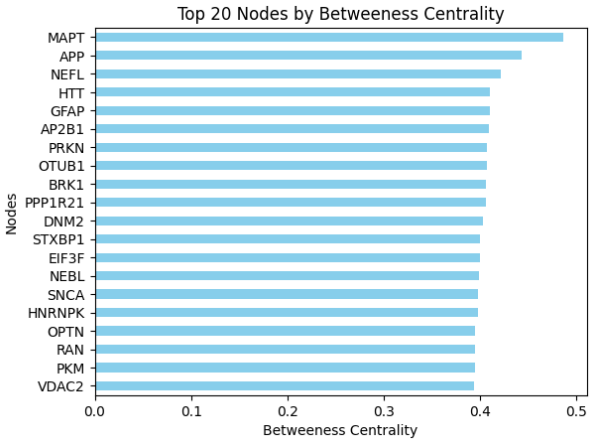

Fig 8.0 Horizontal bar chart of the top 20 nodes by betweenness centrality in descending order. As noted in the density plot most of the nodes fall within a range (0.2 to 0.4)

4.3 CLOSENESS CENTRALITY

The closeness centrality is a measure of how close a node is to other nodes. It is the reciprocal of the sum of the length of the shortest paths between a node and all other nodes within the network. It identifies proteins that can quickly interact with other proteins in the network for ease of signaling.

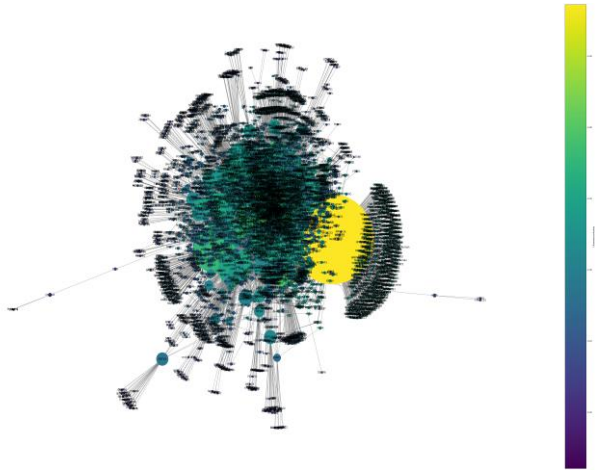

Fig 9.0 A spring_layout network highlighting the closeness centrality according to colour intensity. The brightest (yellow) node being the MAPT protein at the centre of the network.

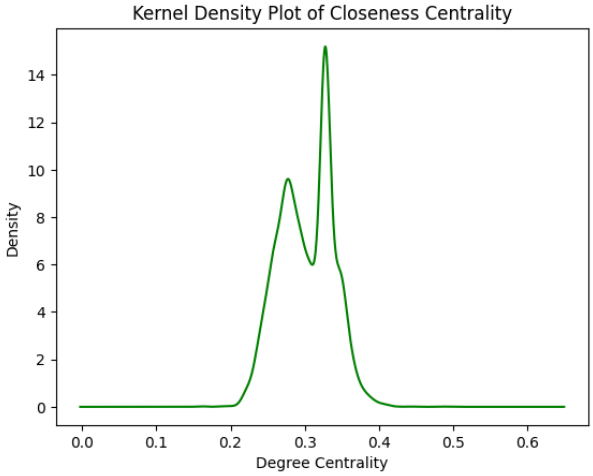

Fig 10.0 Density distribution of the closeness centrality. Similarly, majority of the nodes are concentrated between 0.2 and 0.4.

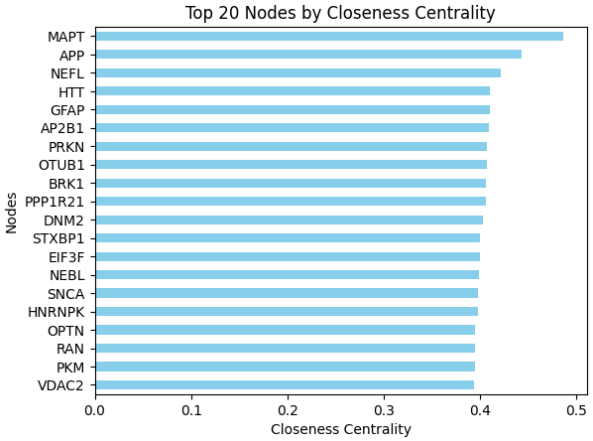

Fig 11.0 Horizontal bar chart of the top 20 nodes by closeness centrality in descending order.

4.4 EIGENVECTOR CENTRALITY

The eigenvector centrality measures the influence a protein has in the network based on its important connections. These proteins may be signaling proteins or even effector proteins that once activated or inhibited alters the disease process.

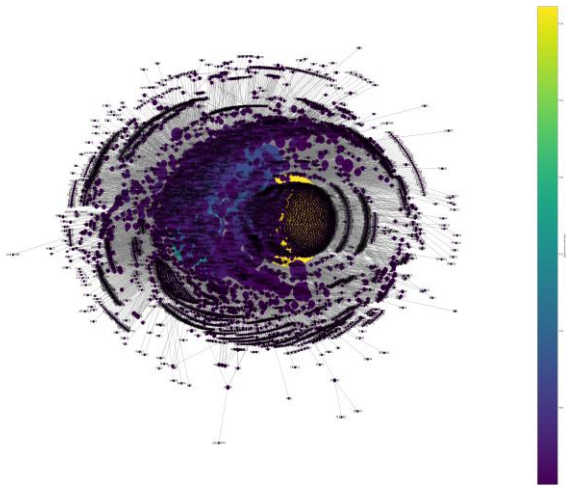

Fig 12.0 A kamada_kawai_layout network highlighting the eigenvector centrality according to colour intensity. The brightest (yellow) node MAPT.

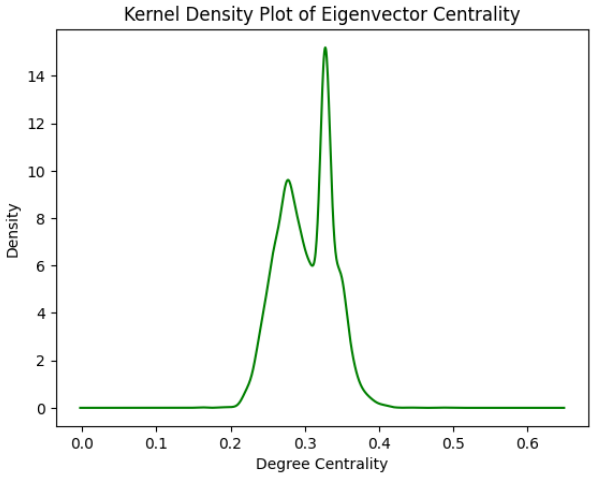

Fig 12.0 Density distribution of the eigenvector centrality. Similarly, majority of the nodes are concentrated between 0.2 and 0.4.

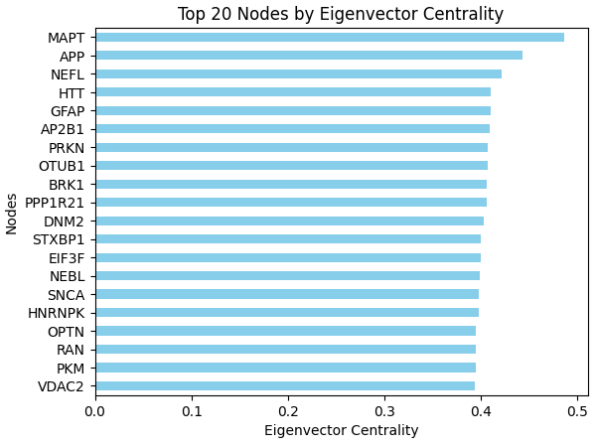

Fig 13.0 Horizontal bar chart of the top 20 nodes by eigenvector centrality in descending order. As noted in the density plot most of the nodes fall within a range (0.2 to 0.4)

5.0 DISCUSSION

From the properties of the Alzheimer protein-to-protein interaction network, about 40% of the proteins in the Alzheimer interactome exert significant influence on the network. Asides MAPT, APP which are the main proteins involved in the disease process, some other proteins identified in the interactome that worth further investigations include (**'NEFL', 'HTT', 'GFAP', 'AP2B1', 'PRKN', 'OTUB1', 'BRK1', 'PPP1R21', 'DNM2', 'STXBP1', 'EIF3F', 'NEBL', 'SNCA', 'HNRNPK', 'OPTN', 'RAN', 'VDAC2', 'DISC1', 'WFS1', 'ccsb_9919', 'SPRED1', 'FGFR3', 'ATXN1', 'KIF1B', 'LAMP2', 'GRN', 'KLF11', 'FMR1', 'RNF11', 'DMWD', 'JPH3', 'GSN', 'CASP6', 'TARDBP').** As observed in the results, similar proteins had the same betweenness, closeness and eigenvector centrality which may emphasize their role in the development of the disease. Notwithstanding, literatures have documented the roles of some of these proteins as either directly or indirectly involved in neuronal functionality however how they may interact to orchestrate the pathogenesis of Alzheimer's remain elusive. With the identification of these interacting proteins, research can be narrowed to further characterize the interactions and their role in the development of the disease. Documented roles of some of the identified proteins are listed in the table 1.0 and highlighted in the fig 14.0.

6.0 CONCLUSION

Alzheimer's is a neurodegenerative disease accounting for 60-70% of dementia cases however, its pathogenesis is not well understood. Using the betweenness, degree, closeness, and eigenvalue centrality scores, top 20 scoring nodes (proteins ['NEFL', 'HTT', 'GFAP', 'AP2B1', 'PRKN', 'OTUB1', 'BRK1', 'PPP1R21', 'DNM2', 'STXBP1', 'EIF3F', 'NEBL', 'SNCA', 'HNRNPK', 'OPTN', 'RAN', 'VDAC2', 'DISC1', 'WFS1', 'ccsb_9919', 'SPRED1', 'FGFR3', 'ATXN1', 'KIF1B', 'LAMP2', 'GRN', 'KLF11', 'FMR1', 'RNF11', 'DMWD', 'JPH3', 'GSN', 'CASP6', 'TARDBP']) in the Alzheimer's interactome were identified for further characterization using computational/ bioinformatic approaches.

7.0 LIMITATIONS

1. Large data size, difficulty downloading or scrapping data. Couldn't retrieve annotations for further analysis.
2. Low processing power, it takes about 2hrs to run the code on colab usingTPU.

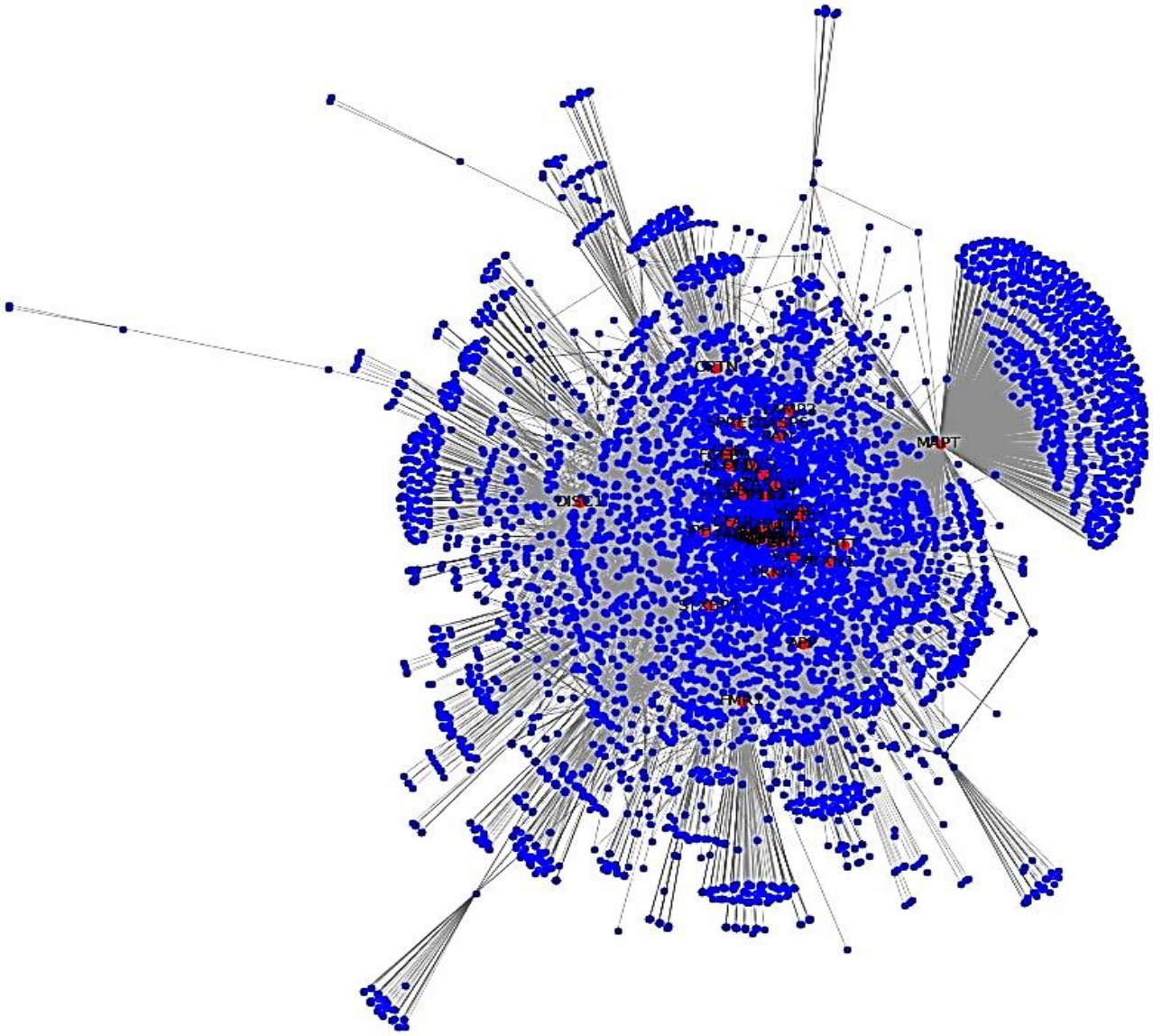

Fig 14.0 Graph of Alzheimer's Protein-to-Protein Interactome Highlighting Functional Connections Within the Network

Protein	Function
NEFL	Neurofilament light polypeptide; provide structural support to neurons
HTT	Huntingtin Protein; involved in axonal transport
GFAP	Glial fibrillary acidic protein; maintains astrocytes mechanical strength and shape
AP2B1	Adaptor related protein complex 2; links clathrin to receptors in coated vesicles
PRKN	Parkin; degradation of misfolded proteins by tagging them with ubiquitin.
OTUB1	Otubains; protein degradation
BRK1	Component of the Wave/Scar pathway involved in the branched nucleation of actin fibres
PPP1R21	Putative regulator of protein phosphatase 1; endosomal sorting process
DNM2	Dynamin 2; regulation of neuron morphology, axon growth
STXBP1	Syntaxin-binding protein 1; rregulates the release of neurotransmitters
EIF3F	Eukaryotic initiation factor 3; initiation of translation
NEBL	Nebulette; assembly of microfilaments
SNCA	Alpha-synuclein; regulates synaptic vesicle trafficking
HNRNPK	Heterogeneous nuclear ribonucleoprotein K; scaffold protein important relay of signals
OPTN	Optineurin; membrane trafficking
RAN	Small gtpase involved in nuclei transport in and out of the nucleus
VDAC2	Voltage-dependent anion-selective channel protein 2; mitochondrial membrane transport
DISC1	Disrupted in schizophrenia1; regulation of neural development
WFS1	Wolframin; intracellular regulation of calcium
ccsb_9919	Cytochrome c synthethase; antioxidants maintain membrane integrity
SPRED1	Sprouty related EVH1 domain containing 1; regulation of the Ras/MAPK signalling pathway.
FGFR3	Fibroblast growth factor receptor 3; mediates growth factor interactions in tissues
ATXN1	Ataxin-1; role poorly understood. Involvement in regulating gene expressions
KIF1B	Kinesin-like protein KIF1B; intracellular transport
LAMP2	Lysosome-associated membrane protein 2; maintains lysosomal membrane integrity
GRN	Granulin; involved in the development of the cerebral cortex
KLF11	Krueppel-like factor 11; involved in cellular differentiation
FMR1	Fragile X protein; involved in synapsis development
RNF11	Ringer Finger Protein; ubiquitination and proteosomal degradation
DMWD	Dystrophia myotonica WD repeat-containing protein; signal transducer
JPH3	Junctophilin-3; involved in the formation of junctional membrane complexes
GSN	Gelsolin; organization of actin filament
CASP6	Cysteine-aspartic acid protease; activation of caspases in apoptosis
TARDBP	Transactive response DNA binding protein; involved in alternative splicing

Table 1.0 Documented functional role of the identified proteins.

8.0 REFERENCES

1. Hasan M, Kumar N, Majeed A, Ahmad A, Mukhtar S. Protein-Protein Interaction Network Analysis Using NetworkX. Methods Mol Biol. 2023;2690:457-467. doi: 10.1007/978-1-0716-3327-4_35. PMID: 37450166.
2. Perreau VM, Orchard S, Adlard PA, et al. A domain level interaction network of amyloid precursor protein and Abeta of Alzheimer's disease. Proteomics. 2010 Jun;10(12):2377-2395. DOI: 10.1002/pmic.200900773. PMID: 20391539.
3. Dementia Fact sheet". World Health Organization. 15 March 2023. Retrieved 10 July 2023
4. Thal DR; Walter J; Saido TC; Fändrich M (2015). "Neuropathology and biochemistry of Aβ and its aggregates in Alzheimer's disease". Acta Neuropathologica. 129 (2): 167– 182. doi:10.1007/s00401-014-1375-y. PMID 25534025. S2CID 19701015
5. EMBL-EBI INtAct database, Alzheimer's

Link to Google Colab:

https://colab.research.google.com/drive/1ArA4Bsmo0C6zBYAh_dqeynzqbbkEjHrA?usp=sharing