Novel Functional Interactomes of Proteins Across Brain Scales

Shreyank Kadadi

Program: Computational and Systems Biology

UID: 805573278

Degree Expected Term: Spring 2023

Faculty Mentor: Dr. Sharmila Venugopal, Department of Integrative Biology and Physiology

ABSTRACT

The complexity of neural systems necessitates advanced tools for understanding interactions within brain networks. Traditional biophysical models face limitations in scalability and specificity, while large-scale molecular network models require experimental validation. This study introduces the Functional Interactomes Score (FIS) Calculator, a novel tool for analyzing neural protein-protein interactions by emphasizing causal relationships. Through amalgamation and processing of data from over 150 studies, the FIS Calculator generates a directed network, where nodes represent proteins and edges signify causal relationships quantified by the FIS. The constructed network is then topologically analyzed, focusing on degree distribution and node centrality; through analysis, The FIS Calculator reveals critical nodes, termed hubs, and central intermediaries within neural protein-protein interaction networks. This approach provides a rigorous framework for investigating interactions in neural systems, offers insights into brain development, aging, and neurological disorders, and may inform the development of targeted therapies.

INTRODUCTION

Mathematical and computational models of neural systems are crucial tools in unraveling the complex biological mechanisms that underlie brain development, aging and chronic neurological diseases. Models help elucidate causal links between the constituent biological components while providing valuable insights and predictions of the biological and behavioral repercussions of these interactions at multiple scales of nervous system organization. Note that causality in this context signifies the capacity of one interactor to induce a transformation in another (Chen et al., 2017; Chen et al., 2015; Chen et al., 2011). At molecular and cellular scales,

causal interaction can manifest in an augmentation or reduction in abundance, alteration in signal transduction, reconfiguration of cellular structure and function, among other possible outcomes, to ultimately affect the whole brain and/or organism (Figure 1).

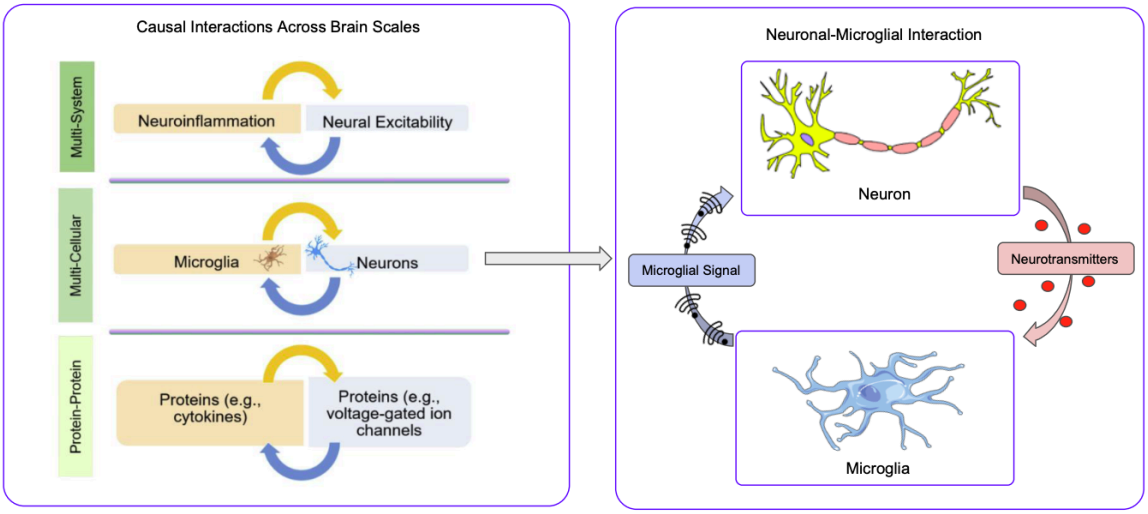

Figure 1: Overview of Neuronal Neural Causal Interactions

Models embodying the biophysical properties (e.g., Hodgkin-Huxley model (Hodgkin & Huxley, 1952)), renowned for their role in explicating the physiological basis of neuronal action potential generation and propagation (e.g., Rall's Cable Theory (Rall, 1977)) are mathematically defined using differential equations have been widely used in the field of neuroscience. However, these models are limited in scalability, nor can they explicitly decode specific neural pathways. This is because they model specific physical components in the brain, meaning that they cannot freely add external interactors.

In contrast, large-scale molecular and cellular network-based models administer a different approach, in which nodes symbolize interacting components and edges denote associations between the interactors (e.g., von Mering et al., 2005; Haenig et al., 2020). The

molecular network graph models have particularly catalyzed in-depth analysis of genome and proteome-wide interactors in many chronic human diseases (Breijyeh and Karaman, 2020; Mhyre et al., 2012). The key advantages of these network models reside in their scalability and amenability to topological analysis. However, they often use predictive measures to assess the likelihood of interaction. For instance, the widely-used STRING database computes predictive confidence scores for categories such as gene fusion, co-occurrence between species, among others, to generate a composite score to generate protein interaction network graphs (Szklyarczyk et al., 2011). Due to the predictive nature of these associations, subsequent experimental validation is essential to verify the causal relationships between interactors.

Hence, the demand for a methodology that can effectively model neurological interactions, which both highlights causal relations and offers scalability, becomes evident. This is precisely the approach taken by the Functional Interactomes Score (FIS) Calculator. To substantiate the methodology adopted by the FIS Calculator, we have selected a case study with a specific focus on causal protein-protein interactions.

METHODS

Python, a high-level, general purpose programming language was used on the Anaconda platform, to develop the modules for the FIS Calculator and network generation. Table 1 provides a list of tools used in relevant modules of this study. All the code, data and final outputs are available in a GitHub repository.

	Technological Tool/Library	**Web Link**	**Descriptors Used**

1	Python Software	https://www.python.org/	To develop the software for the calculator
2	NumPy	https://numpy.org	To easily perform mathematical and statistical functions on tabular data
3	pandas	https://pandas.pydata.org	To store protein data in a data frame for in-depth analysis
4	NetworkX	https://networkx.org	To generate the FIS network
5	Matplotlib	https://matplotlib.org	To visualize FIS network
6	Seaborn	https://seaborn.pydata.org	To visualize topological analyses performed on the FIS network

Table 1: Technological Tools and Libraries Used

Manually Assembling Experimental Evidence

We assembled an extensive compilation of causal associations between proteins that has been consolidated from more than 150 published studies. These studies were sources from the NCBI (National Center for Biotechnology Information) database, which contains the PubMed repository of scientific literature. We specifically filtered and used low-throughput mechanistic studies that focused on neural functional associations, which are physical changes caused by the interaction between neural proteins. In this specific study, we study two main functional associations: change in neural spiking and change in synaptic remodeling as a result of protein-protein interactions (Figure 2).

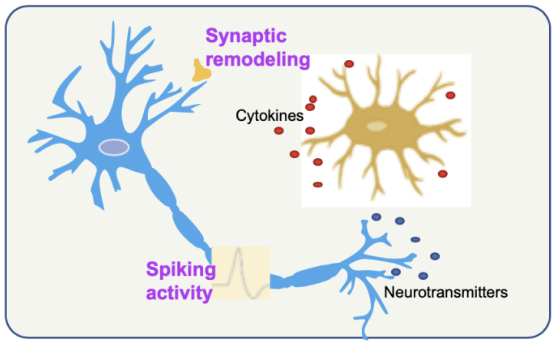

Figure 2: Visualization of Neural Functional Associations

Subsequently, the curated studies were methodically sorted into organized tables that we call Structured Scientific Meta Summaries (SSMS). The information within the SSMS were sorted based on relevant information on protein interactors, types of functional effects, and effect size. These tables were then further simplified to be converted into a CSV (comma-separated values) file with each entry containing the PubMed ID of the experiment of interest, the names of the proteins involved, the overall effect direction of the experiment, and the changes in the functional associations.

Calculating the Functional Interactomes Score

We initiate our analysis of interactions by importing this CSV file into the FIS calculator. The calculator converts the input file into a structured data frame by leveraging the capabilities of the pandas library, which is renowned for its efficient data manipulation tools.

Once the data is structured, we employ a combination of techniques from the pandas library for data processing. Alongside pandas, we utilize the statistical functions in the Numerical Python (NumPy) library for numerical computations.

This robust combination of techniques allow us to calculate our novel Functional Interactomes Score (FIS), a novel metric we devised to quantify the strength of interactions between neural proteins. Specific details surrounding the FIS and its underlying components will be comprehensively discussed in the results section.

Generating the Functional Interactomes Score Network

Having computed the FIS for each protein-protein interaction pair, we utilize the NetworkX library (Hagberg et al., 2008) to construct a directed protein-protein interaction network. In this network, the edges are assigned weights corresponding to the magnitude of the FIS, indicating the strength of the interactions.

For effective visualization of the constructed FIS network, we employ the Matplotlib and Seaborn libraries. These libraries are instrumental in generating clear and informative visual representations, aiding in the interpretation of the network's attributes.

RESULTS

Contextual Background of the Functional Interactomes Score

The Functional Interactomes Across Brain Scales (FIABS) approach focuses on illuminating the functional aspects of neural protein-protein interactions in order to uncover the underlying relation between numerous proteins in the brain. To quantitate the strength of these interactions, the FIABS approach applies the following formula for the FIS:

$$FIS_{A\rightarrow B} = D_M \times W_{A\rightarrow B}$$

As seen in the FIS formula, the strength of an association between a pair of proteins A and depends on two main factors: the median directionality of all experiments (D_M) and the

functional association weight ($W_{A\rightarrow B}$) between protein A and protein B. The directionality of a single experiment is simply the change in the quantity of protein B due to its interaction with protein A (Figure 2). The median directionality can then simply be found by taking the median of all directionalities pertaining to a certain interaction between protein A and protein B. The second component of the FIS is the functional association weight; this is provided by an external functional association lookup table which assigns weights based on the presence and importance of various functional associations (Figure 2). A zero in the functional association lookup table indicates that the particular functional association does not change as a result of the interaction between protein A and protein B whereas a one represents a change in the functional association due to the interaction between protein A and protein B. Based on the change in the various functional associations, a functional association weight is then deduced: for instance, a change in synaptic function and spiking would indicate a functional association weight of 2. By setting the weight in this manner, we can quantify the strength of the interaction as a function of individual associations as well as create a priority ranking between the different associations (in the example table, spiking has a higher priority than synaptic function). The two components of the FIS are then multiplied to output the final score for a given interaction.

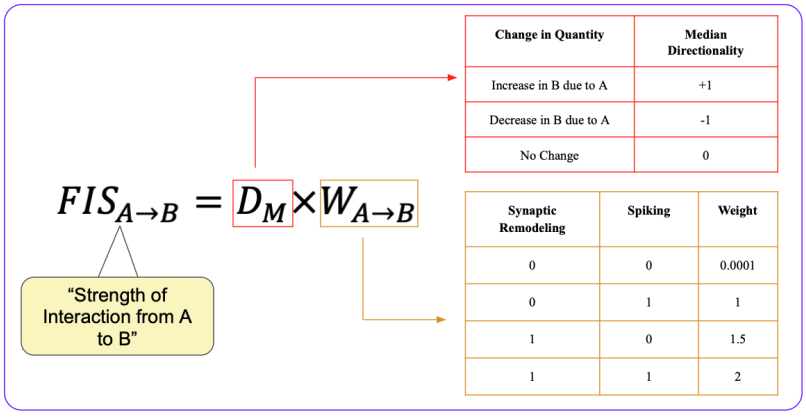

Figure 3: Breakdown of Functional Interactomes Score (FIS) Formula

Implementation of the Functional Interactomes Score Calculator

As the quantity of data increases, calculating and storing the FIS becomes a tedious task. As such, we automated the FIS calculator in order to speed up the scoring process and increase the efficiency of calculation (Figure 4).

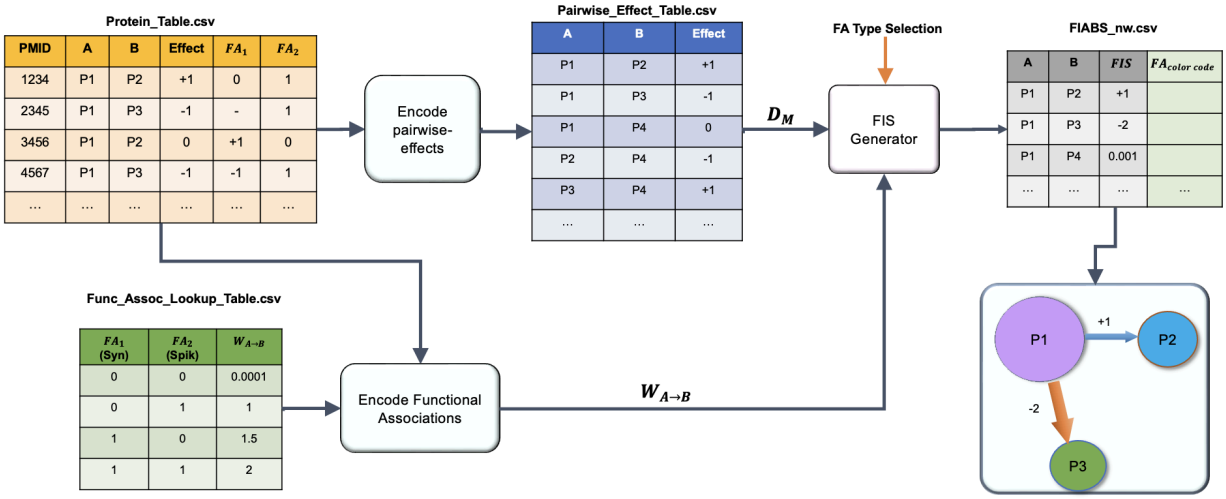

Figure 4: FIS Calculator Flowchart.

There are two tabular inputs to the FIS calculator: the protein table and the functional association lookup table. The input protein table (indicated by *Protein_Table.csv*) contains information about all the experiments, with fields such as PubMed ID of the experiment, the effect direction observed by the experiment, and the change in the various functional associations, if there were any. The functional association lookup table (indicated by *Func_Assoc_Lookup_Table.csv*) is the aforementioned table which provides the functional association weight (Figure 2).

With the input protein experiment table parsed and stored, the FIS calculator splits into two subroutines: first, the calculator generates the *Pairwise_Effect_Table.csv*, which gives us the median directionality D_M for every pairwise interaction. Second, the calculator references the

functional association lookup table to determine the functional association weight $W_{A\rightarrow B}$ for a pairwise interaction with respect to each of the functional associations.

Upon the completion of each of these subroutines, the calculator takes the product of the median directionality and the functional association weight to generate the FIS for each pair of interactions.

Functional Interactomes Score Network Visualization

We leveraged the NetworkX, Matplotlib, and Seaborn libraries to generate and visualize the FIS network:

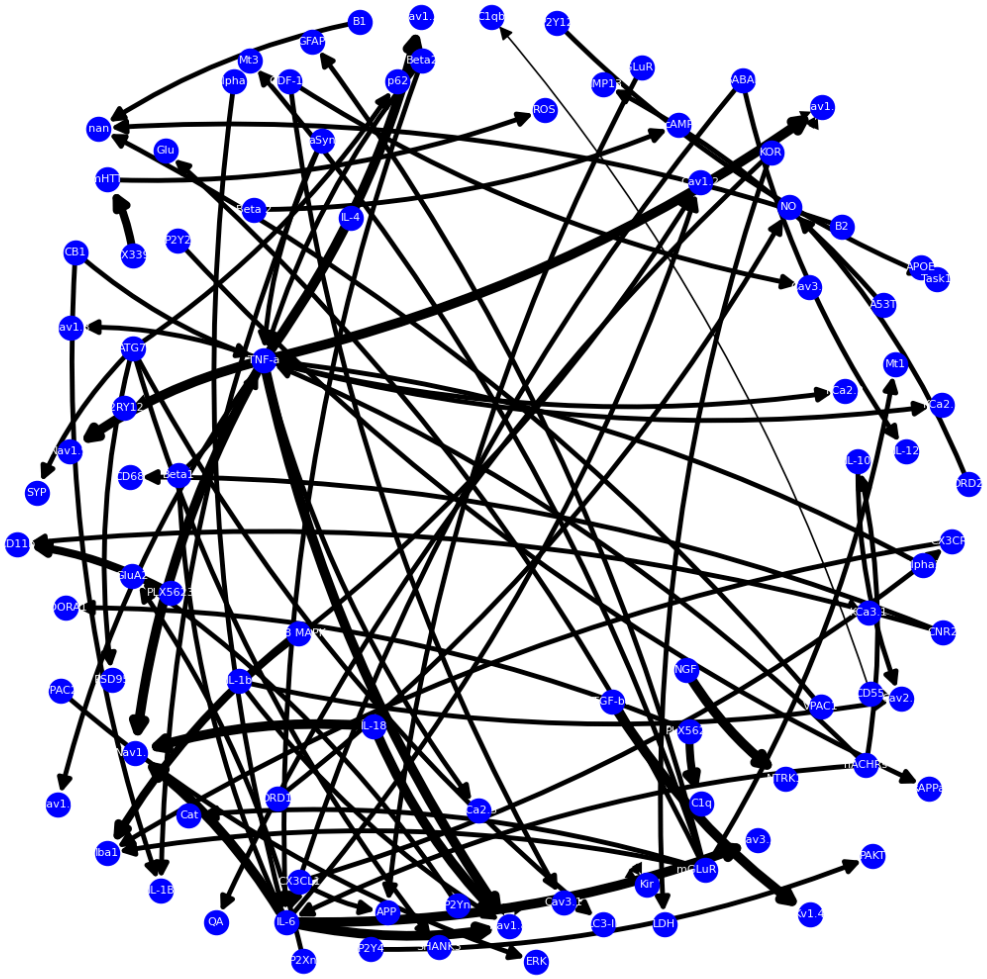

Figure 4: FIS Network

DISCUSSION

Upon generating the FIS network, we subsequently conduct a topological analysis of the FIS network using custom Python scripts. This analysis focused on key network properties such as degree distribution and centrality. which provides insight into the connectivity of the proteins in the network, and which help identify proteins that play a central role in the network.

Degree Distribution

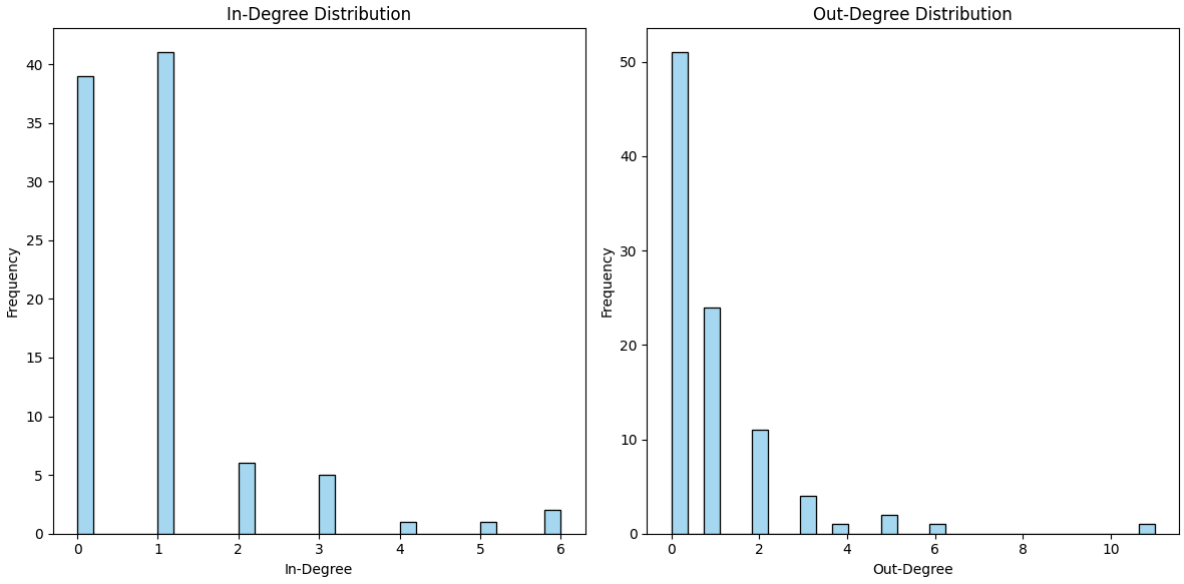

Figure 5: FIS Network Degree Distribution

To systematically study the degree distribution within the FIS network, we examine both the in-degree and out-degree distributions (Figure 5) in accordance with standard analytical approaches for directed graphs. The in-degree represents the number of edges directed towards a node, whereas the out-degree signifies the number of edges originating from a node.

Our analysis revealed that the degree distributions, for both in-degree and out-degree, follow a right skewed distribution. This suggests that the majority of proteins within the network have relatively low connectivity, with a small subset of proteins exhibiting high connectivity.

It is imperative to note that analyzing degree distributions is crucial for understanding the network's overall connectivity. Moreover, it facilitates the identification of protein hubs (Vallabhajosyula et al., 2009), which are characterized by high in-degree or out-degree values. These hubs are likely to play crucial roles within the network and may be involved in key biological processes (He & Zhang, 2006).

As such, the degree distribution analysis of the FIS network provides insights into its structural properties and enables the identification of potential protein hubs, which may serve as focal points for further biological investigations.

Node Centrality

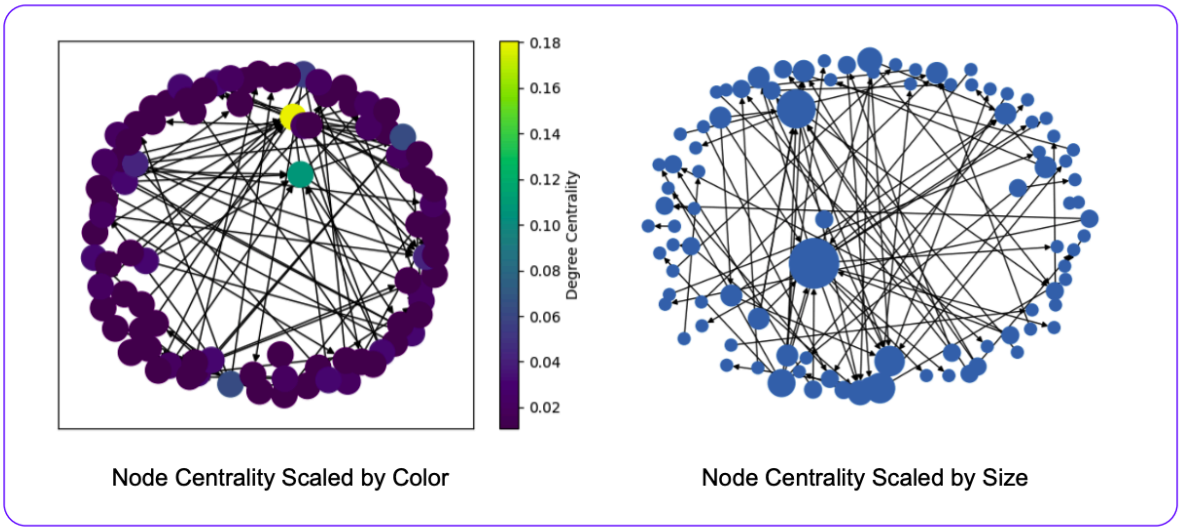

Figure 6: Node Centrality of FIS Network

To evaluate the significance of nodes in the FIS network, we focused our analysis on node centrality metrics, specifically betweenness centrality. Betweenness centrality is a measure defined as the sum of the fraction of all-pairs shortest paths that pass through a given node. Mathematically, the betweenness centrality (BC) of a node v is given by:

$$BC(v) = \Sigma \frac{\sigma(s,t;v)}{\sigma(s,t)}$$

Here, $\sigma(s,t)$is the total number of shortest paths from node s to node t, and $\sigma(s,t;v)$is the number of those paths that pass through node v. A high betweenness centrality indicates that a node acts as a bridge for many shortest paths, suggesting its role as a crucial intermediary in the network.

The FIS calculator has been configured to generate two intuitive visualizations of betweenness centrality. The first scales the color intensity of nodes in relation to their centrality values, while the second adjusts the size of the nodes proportional to their betweenness centrality (Figure 6).

Biologically, nodes with elevated betweenness centrality are likely to be crucial in the regulation and mediation of interactions within the network (Ashtiani et al., 2018). Identifying these nodes is essential, as they can provide insights into the structure and function of the biological pathways under investigation.

Using the FIS Network for Scientific Hypothesis Generation

The topological analysis undertaken in this study forms an essential foundation for subsequent, more advanced investigations. Through a detailed examination of the FIS network's structure, we acquire an enhanced understanding of its characteristics and underlying mechanisms. It is imperative to note that the FIS network uniquely models causal interactions predicated on cellular-level functional alterations, which represent a novel and robust approach to network analysis of neural systems.

The causal emphasis of the FIS network lends credibility to the results of our analysis, as it accounts for actual physiological changes that are directly consequent to the interactions of

proteins within the network. Furthermore, the modular architecture of the FIS network offers scalability and enables the seamless integration of additional proteins of interest; this makes the approach adaptable to expanding datasets and research objectives.

As such, the FIS network can be used for the formulation and rigorous evaluation of scientific hypotheses, particularly in relation to the complex interplay among biological systems such as neuroinflammation and neural excitability. This methodological approach is critical for guiding future research efforts, and holds potential implications for the development of targeted therapeutic strategies.

Future Directions of the FIS Calculator

One future enhancement to the FIS calculator involves the integration of a chromatic encoding strategy to the FIS network's edges for elucidating alterations in individual functional associations. In this strategy, red, blue, and black arrows would respectively denote increases, decreases, and no change in functional associations between proteins A and B (Table 3) . This streamlined visual representation will enable an expedient analysis of specific functional association impacts, thus fostering more efficient network interpretation for researchers.

Change in Functional Association	**Color Code**
Increase	Red
Decrease	Blue
No Change/Not Observed	Black

Table 2: Edge Coloring Based on Directionality of Functional Association

Secondly, our objective is to evolve the FIS calculator by constructing a web application with the FIS calculator integrated as the backend engine to expand accessibility. By providing an online platform, we aim to facilitate streamlined interaction analyses for scientists working in neural protein-protein research. The overarching vision behind this integration is to foster an ecosystem wherein researchers can effortlessly leverage the FIS calculator to drive breakthroughs in understanding of causal neural interactions, ultimately accelerating advancements in neuroscience.

REFERENCES

Ashtiani, Minoo, et al. “A Systematic Survey of Centrality Measures for Protein-Protein Interaction Networks.” *BMC Systems Biology*, vol. 12, no. 1, 2018, https://doi.org/10.1186/s12918-018-0598-2.

Breijyeh, Zeinab, and Rafik Karaman. “Comprehensive Review on Alzheimer’s Disease: Causes and Treatment.” *Molecules*, vol. 25, no. 24, 2020, p. 5789, https://doi.org/10.3390/molecules25245789.

Chen, Weiqiang, Jiangtao Sheng, Jingfang Guo, Fenfei Gao, et al. “Tumor Necrosis Factor-α Enhances Voltage-Gated Na+ Currents in Primary Culture of Mouse Cortical Neurons.” *Journal of Neuroinflammation*, vol. 12, no. 1, 2015, https://doi.org/10.1186/s12974-015-0349-x.

Chen, Weiqiang, Jiangtao Sheng, Jingfang Guo, Guoyi Peng, et al. “Cytokine Cascades Induced by Mechanical Trauma Injury Alter Voltage-Gated Sodium Channel Activity in Intact Cortical Neurons.” *Journal of Neuroinflammation*, vol. 14, no. 1, 2017, https://doi.org/10.1186/s12974-017-0847-0.

Chen, Xi, et al. “TNF-α Enhances the Currents of Voltage Gated Sodium Channels in Uninjured Dorsal Root Ganglion Neurons Following Motor Nerve Injury.” *Experimental Neurology*, vol. 227, no. 2, 2011, pp. 279–286, https://doi.org/10.1016/j.expneurol.2010.11.017.

Haenig, Christian, et al. “Interactome Mapping Provides a Network of Neurodegenerative Disease Proteins and Uncovers Widespread Protein Aggregation in Affected Brains.” *Cell Reports*, vol. 32, no. 7, 2020, p. 108050, https://doi.org/10.1016/j.celrep.2020.108050.

He, Xionglei, and Jianzhi Zhang. “Why Do Hubs Tend to Be Essential in Protein Networks?” *PLoS Genetics*, vol. 2, no. 6, 2006, https://doi.org/10.1371/journal.pgen.0020088.

HODGKIN, A, and A HUXLEY. “A Quantitative Description of Membrane Current and Its Application to Conduction and Excitation in Nerve.” *Bulletin of Mathematical Biology*, vol. 52, no. 1–2, 1990, pp. 25–71, https://doi.org/10.1016/s0092-8240(05)80004-7.

McKinney, Wes. “Data Structures for Statistical Computing in Python.” *Proceedings of the Python in Science Conference*, 2010, https://doi.org/10.25080/majora-92bf1922-00a.

Mhyre, Timothy R., et al. “Parkinson’s Disease.” *Protein Aggregation and Fibrillogenesis in Cerebral and Systemic Amyloid Disease*, 2012, pp. 389–455, https://doi.org/10.1007/978-94-007-5416-4_16.

“Networkx.” *Encyclopedia of Social Network Analysis and Mining*, 2018, pp. 1608–1608, https://doi.org/10.1007/978-1-4939-7131-2_100771.

Rall, Wilfrid. “Core Conductor Theory and Cable Properties of Neurons.” *Comprehensive Physiology*, 1977, pp. 39–97, https://doi.org/10.1002/cphy.cp010103.

Shannon, Paul, et al. “Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks.” *Genome Research*, vol. 13, no. 11, 2003, pp. 2498–2504, https://doi.org/10.1101/gr.1239303.

Szklarczyk, D., et al. “The String Database in 2011: Functional Interaction Networks of Proteins, Globally Integrated and Scored.” *Nucleic Acids Research*, vol. 39, no. Database, 2010, https://doi.org/10.1093/nar/gkq973.

Vallabhajosyula, Ravishankar R., et al. “Identifying Hubs in Protein Interaction Networks.” *PLoS ONE*, vol. 4, no. 4, 2009, https://doi.org/10.1371/journal.pone.0005344.

Von Mering, C. “String: Known and Predicted Protein-Protein Associations, Integrated and Transferred across Organisms.” *Nucleic Acids Research*, vol. 33, no. Database issue, 2004, https://doi.org/10.1093/nar/gki005.