

Graph Explorer: An Investigative Overview by Connectomically Challenged Children

BRIAN SOONG, TONY WU, DAVID YANG

1 Improvements and Observations for scienceinthe.cloud Demo

The run went relatively well. Although it took some time for the program to generate the data, we were able to successfully generate a connectome graph with 70 nodes, 788 edges, with an average degree of 22.5143 degrees. This data is reproducible and stays consistent when run multiple times. However, a lot of the data was hard to read, with some of it being incomprehensible. More explanation in what was actually being done would be helpful in understanding the processes going on. For example, there is a graphic in the beginning showing the ndmg pipeline but the graphic does little to explain each of the steps within the pipeline.

Suggestions for improvement:

1. More explanations of the process as well as explaining significance of data.
2. Better labels for the graph, currently unclear.
3. Take less time, if possible.

Things that went well:

1. The graphs at the very end were helpful and could be used for additional analysis.
2. The program compiled without error, gave indications that it worked properly.
3. The connectome graph is a good visual for all of the text above it.

2 Covariate and Connections

1. **nx.degree_centrality:** As male and female brains show different connectivities in different regions of the brain, degree centrality analysis would allow for differentiation between gender in brain graphs.
2. **nx.estrada_index:** The Estrada index, detailed by Ernesto Estrada in a 2007 paper, describes a measure of how well nodes are connected on a basis of strength and distance. Neural-related gender differences exhibit themselves as such.
3. **nx.current_flow_closeness_centrality:** An algorithm developed by Ulrik Brandes, this feature builds upon closeness centrality by supplementing the resistance between nodes in addition distance. Similar to degree centrality, this algorithm highlights the nuances of gender formulated neuropathology.
4. **skimage.feature.daisy:** The DAISY function, a rapid rough algorithm for density and depth, outputs numerous features on the order of 10^3 in order to reflect the minute differences between males and female graphs.
5. **np.sum:** Given already as a feature from the beginning, the sum of all the elements in the graph was the last feature used. Minute differences in the graph and its nodes would also be reflected.

Final Accuracy Achieved: 0.669

3 What We Learned

From this project, we learned multiple lessons. To begin with, we realized from mass experimentation that there were a lot of other classifying algorithms that we could use for our project other than Random Forest. For example, the scikit-learn package also contains other classifier algorithms, such as Naive Bayes, K-Nearest neighbors, etc.

Furthermore, we also realized that daisy is a really useful function, but the sad part is that out of the bag of features that the daisy function claims to assign to each pixel, they do not give any documentation on what any of the bag of features are. These types of functions definitely make us hesitate, as at some point, they become a last resort type of function if all else fails.

Lastly, it was very surprising that as humans, we could not tell the difference between the female and male graphs visually from the log graph, but when looking into the values in the matrices, the male and female averages were very noticeably different. This brings us to the conclusion that some differences and patterns cannot be easily detected by the human eye and that computations are very useful in helping us detect what we cannot see.