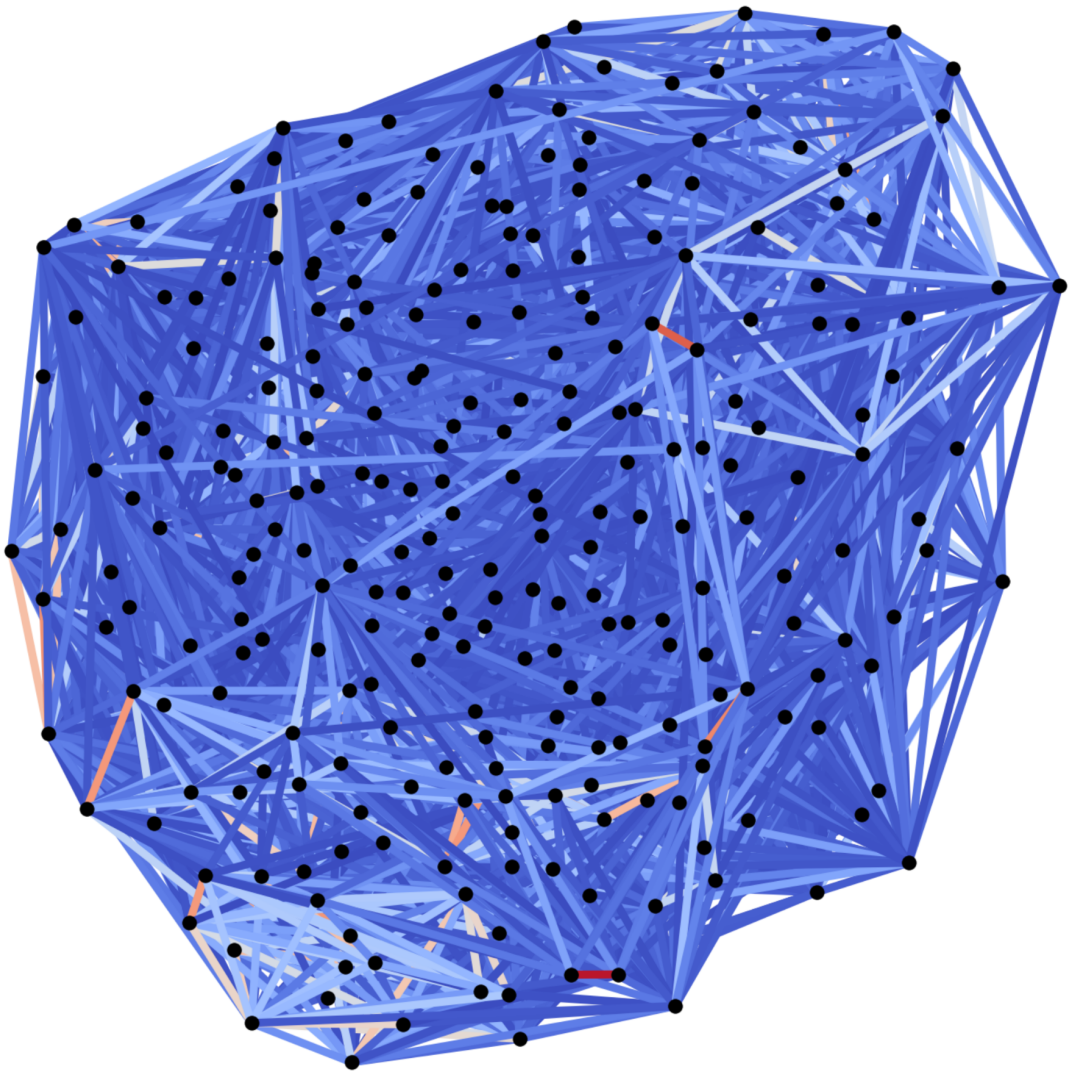
# rs-fMRI Graph Theoretical Exploration of Traumatic Brain Injury Subjects

**CSB199 Report**

**Daniel Frees**
**UID: 605104056**
**12/07/2021**



Img: Brain Graph for Subject 1 (Spring Layout)

## Abstract

This project was designed and undertaken with the goal of determining potentially significant differences in graph theory measures of injured brains as compared with healthy brains.

To accomplish this goal, the first step was extracting weighted adjacency matrices from resting-state functional magnetic resonance imaging (MRI) images. BrainSport's CBF dataset provided 104 participants with complete imaging sessions, and CONN Toolbox was used to preprocess these images, run connectivity analysis, and output connectivity matrices with edge weights corresponding to Fisher's Z values between ROIs. In addition to connectivity matrices, relevant covariates (including the target: traumatic brain injury (TBI) versus health control (HC)) were extracted. Data was cleaned, imputed, thresholded, and wrangled through MATLAB and Python, primarily using Python's pandas package. Once the data was cleaned, Python's NetworkX package and custom scripts were used to calculate graph theory measures for each subject in the dataset, with a focus on global metrics and amygdala region of interest (ROI) metrics. Finally, averaged measures were compared between mild TBI (mTBI) and HC groups, and statistics were calculated to determine significant graph theoretical differences between groups.

Several findings bordering on significance came out of this exploratory analysis. Chief among these was the finding that transitivity is different in TBI brains compared to healthy control brains (2-sided $p<0.075$). Transitivity was also found to be moderately negatively correlated with TBI, suggesting that brain cliques might break down with injury.

Future steps include performing a bootstrapped analysis to increase the statistical power of this study, given the relatively small sample size available, and designing a machine learning algorithm to classify brain injury.


## Introduction

Resting-state functional MRI (rsfMRI) is currently under research for its potential uses for diagnosing TBI and predicting symptom outcomes following insult. Images are taken at a resting-state, meaning that subjects perform no specific tasks while undergoing imaging, and are usually told to think about nothing in particular. rsfMRI measures blood-oxygen-level-dependent (BOLD) signal throughout the brain, and can be used to compute correlations between brain regions under the assumption that BOLD signals correlate over time between regions that are more highly correlated (brain regions with BOLD signals which anticorrelate are considered to be negatively correlated).

Despite the resting-state nature of rsfMRI, controlling covariates to easily differentiate between brain injury patients and healthy control subjects remains a challenge. Differences in brain shape, movement, thought processes, personality, gender, handedness, lesion status, injury

severity, chronicity (time since injury), and other variables can all affect rsfMRI results. As such, rsfMRI studies have found inconsistent trends when investigating specific brain networks/ regions. For this reason, graph theoretical measures which summarize larger scale patterns in the brain might be a valuable tool for investigating rsfMRI results.

Studies aiming to diagnose autism using machine learning models have found greater success when incorporating topological graph theory metrics from patient resting-state functional MRI images[1]. The goal of this study is to extend a similar approach to the classification of traumatic brain injury. This report marks the first milestone of a project to develop a machine learning (ML) algorithm which will diagnose brain injury based on topological brain graph properties. In this report, an exploratory data analysis is undertaken to identify graph theoretical properties which might serve as useful features to input into an ML classification algorithm, as well as to determine significant differences between mTBI and HC brains using classical statistical methods.

## Methods

*Extraction:* Using MATLAB's CONN Toolbox[2], images from time point 1 of BrainSports' CBF dataset were preprocessed with standard regressions, smoothing, and motion correction. Regions of interest (ROIs) were selected based on the Brainnetome 264 Atlas[3]. Analyses were performed to calculate Fisher's z value correlations between all pairs of ROIs for each subject, and this data was exported in .mat files containing adjacency matrices and numerous other results. Covariates were also exported in an .xlsx file.

MATLAB was used to extract adjacency matrices from amongst the large variety of data contained in the .mat results output file. Next, all data was imported into a Python Jupyter Notebook and imported as pandas dataframes. Data was cleaned, imputed, thresholded, and wrangled primarily using Python's pandas package. Once the data was cleaned, Python's NetworkX package and custom scripts were used to calculate graph theory measures for each subject in the dataset, with a focus on global metrics and amygdala ROI metrics. The following analysis steps are unique to this project, but based largely on prior precedents outlined in the primary paper establishing BRAPH (Brain analysis using graph theory) software[4].

*Graph Theory Metric Calculation & Graph Theory Based Processing:*

1. ROI-ROI correlations zeroed for ROIs to themselves (zero diagonal of adj. mat.)
2. Negative correlations zeroed
3. All subject adjacency matrices passed through a dynamic correlation threshold designed to make all graphs have density ~ 0.25
4. ROIs of interest (nodes_of_interest) extracted from the entire list of ROIs. For this, all amygdala ROIs were extracted as ROIs of interest.

*For each subject:*

5. Covariate data matched to subject
6. NetworkX graph generated from subject adjacency matrix
7. Weighted and unweighted clustering coefficients calculated
8. Characteristic path length calculated
9. Density calculated to verify that all patients have density around 0.25 (minor fluctuations may occur due to limitations of float precision and ties around the dynamic threshold)
10. Transitivity calculated
11. Node degree and node strength calculated for all ROIs of interest
12. Closeness centrality calculated for all ROIs of interest
13. Shortest path calculated between left dorsal amygdala ROI and left MB ROI
14. Shortest path calculated between left dorsal amygdala ROI and right MB ROI

Upon completion of graph theory processing and metric calculations, histograms were visualized for all variables in the dataset, as well as separate histograms for the TBI data subset and HC data subset. Means and standard deviations were calculated for each metric within the TBI and HC subsets. Two-sample Welch's t-test using two-sided alternative hypotheses were calculated to compare features across the TBI and HC groups. Finally, correlations were visualized for the entire dataset.

## Results

*Distribution of Features (Complete Dataset):*

Some interesting distributions appeared when visualizing all features using histograms. First of all, it should be noted that the raw dataset is imbalanced, with approximately ~⅓ of the data corresponding to TBI patients and ~⅔ to HC subjects. This should not affect the current analysis, but will important for future directions with machine learning. Average clustering coefficient was approximately normally distributed for both weighted and unweighted clustering coefficients. Characteristic path length was strongly right skewed. Transitivity did not match any canonical distributions but was approximately right skewed. Node degree for both the left dorsal amygdala ROI and right dorsal amygdala ROI were bimodally distributed. Closeness centrality for the left lateral amygdala was bimodally distributed as well, as was closeness centrality for the right medial amygdala.

See distributions on the following page.

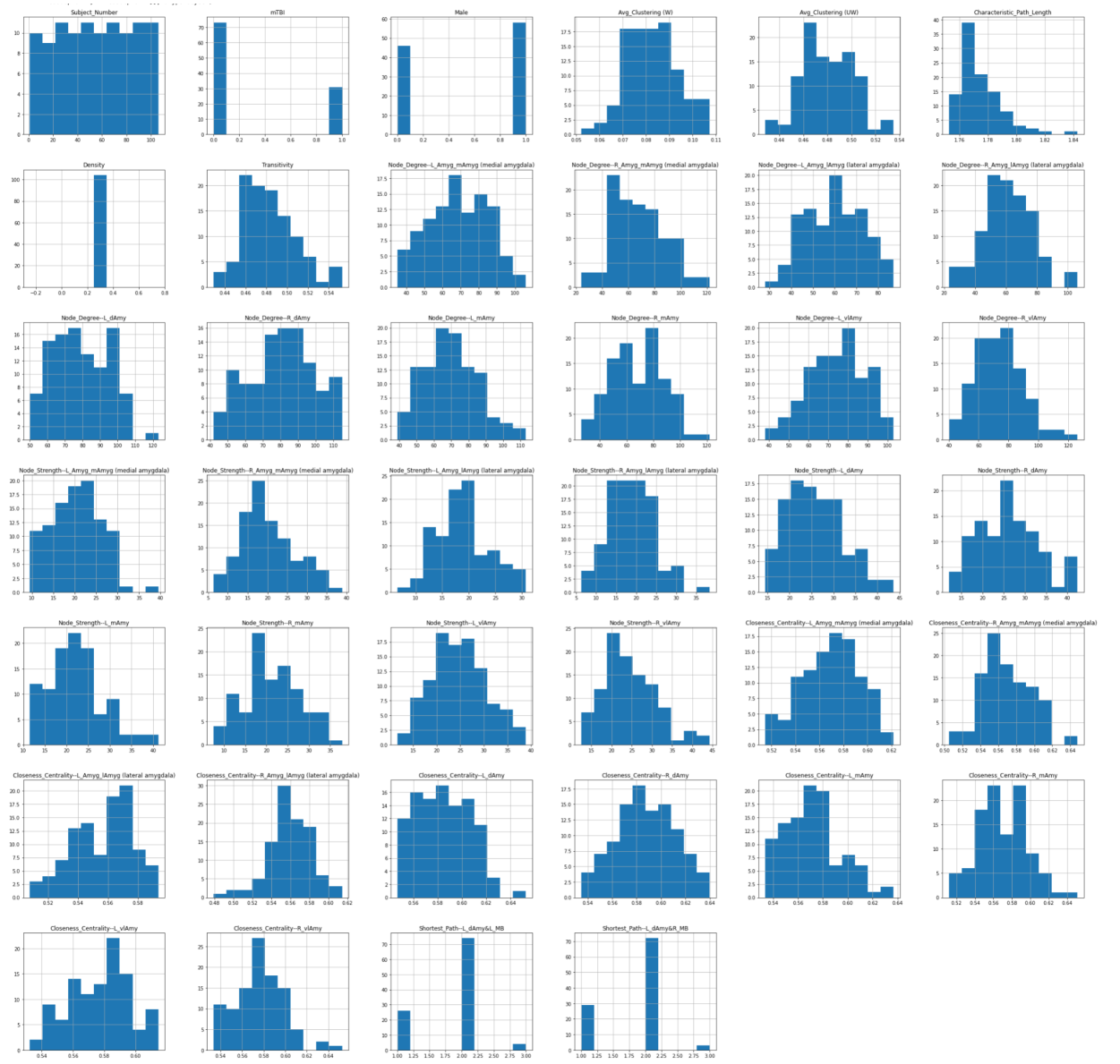



**Figure 1: Distributions of Graph Theory Features Across Full Dataset**

*Correlation Analysis (All Data):*

For the following correlation analysis, note that presence of mild TBI (mTBI) was indicated with a 1, and absence was indicated with a 0. Thus, positive correlations between any feature and *mTBI* indicate that said feature trends higher in the *mTBI* population.

mTBI is moderately negatively correlated with average unweighted clustering coefficient, characteristic path length, transitivity, both node degree and node strength for the right dorsal amygdala ROI, closeness centrality for the right dorsal amygdala ROI, and shortest path length between the left dorsal amygdala and left MB.

Furthermore, mTBI is moderately positive correlated with closeness centrality of the left lateral amygdala.

See the following correlation outputs. Note that all values reflect Pearson's r values, and all correlations visualized below are with *mTBI.*


**Figure 2: Visualization of Correlations (Pearson's r) of All Features with mTBI**

# Discussion

*Feature Distributions:*

The current analysis is not sufficient to explain the underlying causality of these various distributions, however the existence of bimodal distributions is promising. It may be the case that either TBI or some subset of TBI correlates to one of the modes in some of these distributions.

*Classical Statistical Analysis:*

*Subject_Number* is significantly different between groups only because subject numbers were ordered from mTBI participants to HC participants, and therefore indicates nothing. *mTBI* is significant because it is the way the groups were formed, and also indicates nothing. *Density* is significant only because of minor differences in node thresholding (tied nodes can ever so slightly affect the manually chosen 25% density conformity requirement imposed in the dynamic thresholding function). There is no difference between groups, but the standard deviation of density is so low that analysis yields a very low p-value.

*Transitivity* on the other hand, is an interesting group difference finding. While the p-value would only be semi-significant at alpha = 0.05, this is still a promising finding. Transitivity is a measure which can be interpreted in terms of cliquiness of a graph, whereby higher transitivity indicates that node relationships are more likely to form a clique. Specifically, transitivity measures the chance that if node x knows node y (is correlated), and y is correlated with z, then x is also correlated with z. Based on this data, transitivity seems to be higher for the healthy control group (mu=0.4857) than the traumatic brain injury group (mu=0.47704), indicating that perhaps traumatic brain injury breaks down cliques. This is also indicated by the negative correlation between *mTBI* and *Transitivity* in the correlation analysis. Transitivity should certainly be hypothesized to be an important feature for future machine learning modeling.

*Correlation Analysis:*

Correlation calculation can be misleading, so no conclusions will be drawn firmly from the presented correlation values. However, with future statistical modeling and feature transformation (especially transformation to higher dimensional feature spaces with models such as RBF kernel support vector machine (SVM)) might reveal correlated features to be important). For now, it can be said that average unweighted clustering coefficient, characteristic path length, transitivity, node degree and node strength for the right dorsal amygdala ROI, closeness centrality for the right dorsal amygdala ROI, shortest path length between the left dorsal amygdala and left MB, and closeness centrality of the left lateral amygdala are all hypothesized as potentially important features for future ML modeling.

# Conclusion

This paper identifies a potentially significant finding of reduced transitivity in subjects with mild traumatic brain injury, suggesting that cranial insult may lead region-based brain cliques to break down. Furthermore, numerous graph theoretical measures are hypothesized as potentially important features for future machine learning modeling, and a data pipeline is established to set up ML modeling using weighted adjacency matrices extracted from the BOLD signal correlations of resting-state functional MRI images.

*Limitations:*

The study size is relatively small (n=104), giving the current analysis low statistical power. As such, there may be features which truly differ between mild TBI and healthy control groups that were not identified as significant.

Furthermore, numerous graph theory metrics which might prove to be useful descriptors for injured versus healthy brains cannot be calculated using the current standard processing methodologies. For example, weighted shortest path length between nodes could theoretically be calculated, but with the data in its current state (edge weights = Fisher's z values), a shorter weighted path length between nodes would simply indicate that those nodes were connected via an uncorrelated path, which is not intuitive.

*Future Directions:*

To counteract the small study size, the current analyses could be repeated using bootstrapping aimed to increase statistical power.

For the goal of calculating further graph theory metrics and retaining the intuition these metrics have for canonical graph problems, additional data processing could be undertaken for the calculation of different measures. For example, Fisher's z values could be passed through an inverse transformation (after dynamic thresholding) so that more highly correlated ROIs are treated as 'closer' and less highly correlated ROIs are treated as 'farther'.

Lastly, the current project is set up as a precursor to a machine learning modeling approach to classify mild traumatic brain injury patients based on the topological properties of their brains. This is the most obvious next step to take, and my current plan for self-guided research (CSB199) in the Winter quarter.

# Works Cited


1. URL: (https://www.frontiersin.org/articles/10.3389/fnins.2018.01018/full)
2. URL: (https://www.liebertpub.com/doi/10.1089/brain.2012.0073)
3. URL: (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4961028/)
4. URL: (https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0178798)


**Code for this project can be found at:** https://github.com/danielfrees/TBI_BrainGraph_ML
Note: Most of the important work for this project occurs in T1_CBF_GraphExploration.ipynb, and I added HTML markups and a table of contents to guide navigation through the code and results in this file.