

# EN.580.694: Statistical Connectomics

## Final Project Report

Kristin Gunnarsdottir · May 1, 2015

### Gender classification using networks based analysis of the human connectome

Table 1: The three lowest p-values of each feature, obtained using bootstrapping

Graph feature	Node (region) number	p-value
Degree centrality	6	0.0094
	20	0.0115
	44	0.0368
Weighted Degree	25	0.0378
	39	0.0212
	60	0.0474
Eigenvector centrality	23	0.0473
	25	0.0329
	36	0.0400
Cosine Similarity	68-70	0.0053
	49-69	0.0027
	66-68	0.0061

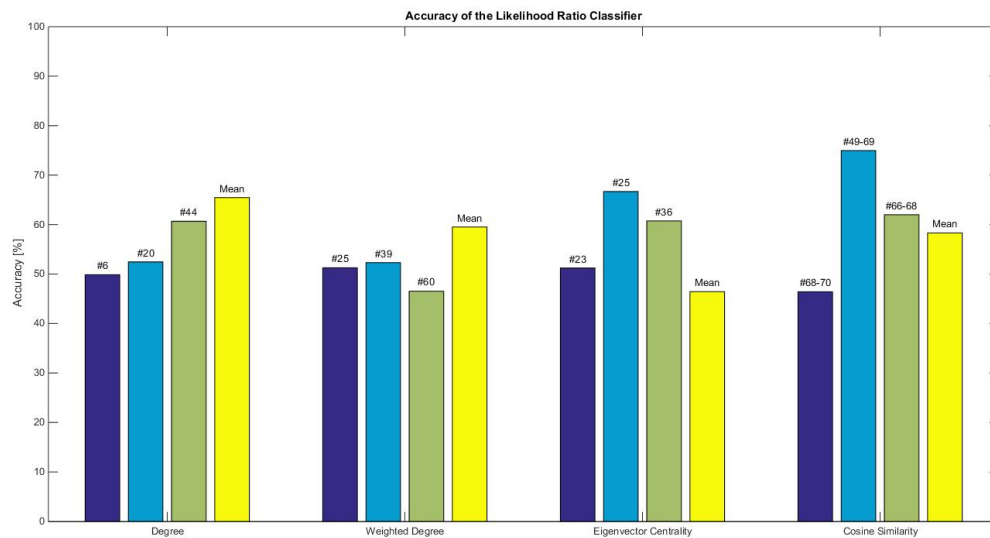


Figure 1: The Classification Accuracy of the Likelihood Ratio Classifier using the features that showed the most significant difference between males and females as an input

**Opportunity** Literature suggests that while there are many similarities in brain structure and function in men and women, there are important differences that discriminate male brains from female brains. Understanding the sex differences in the human brain can provide important information and better knowledge of the pathology or frequency of disease in men and women [1]. Using the human connectome to analyze the brain as a network provides a great opportunity to gain additional knowledge about the organization and interactions between brain structures and can demonstrate if gender difference shows altered connectivity patterns between brain regions.

**Challenge** Studies show sex differences in human brains, but currently the reasons for the connectivity difference are not as well understood. Additionally, because of a high variability between individuals in brain anatomy and activity, a certain degree of difference can be expected between subjects. The sample size is rather small, which could potentially make it harder to build a reliable classifier.

**Action** In this project I attempted to discriminate between male and female brains based on their brain region connectivity patterns using 42 small graphs (70x70 adjacency matrices) of 21 individuals [2]. In order to categorize each graph I implemented a Likelihood Ratio Classifier in Matlab R2014b. I attempted to capture the characteristics of each graph by using a few different network based measures as an input to the classifier. Specifically, the Degree centrality, the Weighted degree and Eigenvector centrality of each node were calculated, as well as the Cosine similarity between each pair of nodes in the network [3; 4]. For each measure I used bootstrapping (10,000 samples drawn with replacement) to identify the nodes showing the most significant difference between the groups (Table 1). The classifier was ran 500 times for each feature and the average accuracy then calculated (Fig. 1).

**Resolution** Figure 1 shows the classification accuracy using the selected features, as well as the average of all three nodes for each feature, as an input. A classification accuracy of up to 75.45% was obtained, indicating that there is an existing difference across males and females in network based features of the human connectome. However, the classifier was not as reliable for some of the features, with an accuracy less than 60%. The highest accuracy was obtained using the cosine similarity of nodes 49-69. This was also the value showing the most significant difference between the two groups, although a lower p-value did not necessarily yield a better accuracy.

**Future Work** Improving, or replacing, the classifier will likely increase the classification accuracy. For this project I chose to only look at the three nodes showing the most significant difference between the groups, although more nodes were significantly different. Since the results showed that the lowest p-values did not necessarily give the highest classification accuracy, the other nodes should be observed as well. Additionally, other graph measures should be tested, as well as using more than a single value as an input to the classifier each time. Moreover, research has shown that using a signal-subgraph instead of unique nodes can improve the classification accuracy significantly [5]. Finally, the dataset used was very small so using a larger dataset will provide more significant and conclusive results.

## References

- [1] K. Cosgrove, C. Mazure, and J. Staley. Evolving Knowledge Of Sex Differences In Brain Structure, Function, And Chemistry. 62:847–855, 2007.
- [2] Open Connectome Project. URL [http://openconnectome.org/data/public/MR/MIGRAINE\\_v1\\_0/](http://openconnectome.org/data/public/MR/MIGRAINE_v1_0/).
- [3] M.E.J. Newman. *Networks An Introduction*. Oxford University Press Inc., New York, United States, 2010. ISBN 978-0-19-920665-0.
- [4] Tore Opsahl, Filip Agneessens, and John Skvoretz. Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, 32(3):245–251, July 2010. ISSN 0378-8733. doi: 10.1016/j.socnet.2010.03.006. URL <http://www.sciencedirect.com/science/article/pii/S0378873310000183>.
- [5] J.T. Vogelstein, W.R. Gray, R.J. Vogelstein, and C.E. Priebe. Graph Classification using Signal-Subgraphs: Applications in Statistical Connectomics. 2013.