

# EN.580.694: Statistical Connectomics

## Final Project

# Clustering Algorithms Engineering of the Mouse Neocortex Connectome

Figure 1: The unclustered data from the mouse connectome, identified by [1]. The anterograde and the retrograde tracer data was averaged together to create a combined matrix. Clustering algorithms were performed on these matrix.

**Challenge** Identifying subnetworks of the connectome is not a straight-forward task. There are countless algorithms for clustering which each have their pros and cons. It's difficult to compare which nodes form which networks and how they all fit together. This project will examine a few common algorithms and compare them to see what the paper [1] did and what might work better.

**Action** The raw data will need to be copied from Zingg, Hintiryan, et al. and processed into a binary, combined matrix. Two **kmeans** algorithms were ran, one with  $k=4$  and one with  $k=12$ . A major difficulty of **kmeans** is that it needs the number of clusters *a priori*. Four major subnetworks were identified by [1], each composed of many smaller modules, 12 in total. This is why  $k=4$  and  $k=12$  were chosen. These two clustering algorithms were then compared to in silhouette graphs and sorted bar graphs. Several hierarchical clustering algorithms were also ran. The hierarchical linkage methods used were: "simple", "complete", "average" and "weighted". There are then plotted in dendrograms and compared (Figure 2). Finally, a complete (non-binary) matrix was arranged and plotted (Figure 3).

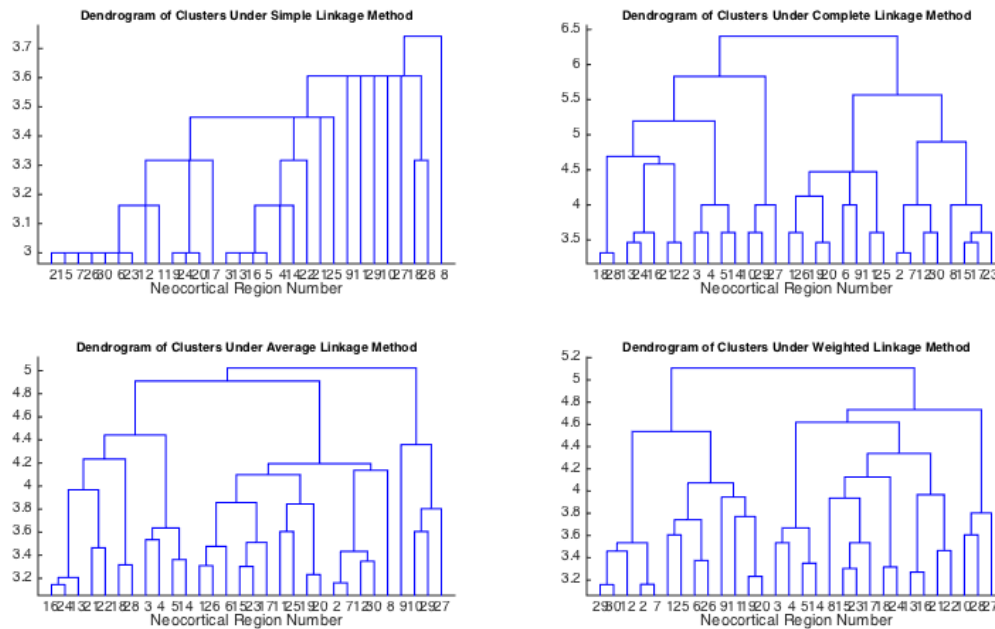


Figure 2: Four dendrograms for four different hierarchical clustering algorithms.

**Resolution** The paper identified 4 major subnetworks (Somatic Sensiomotor, Medial, Lateral, and Claustrum-Entorhinal), many composed of various smaller modules (12 in total). The results from the clustering algorithms show various ways to group the data. Which methods work the best were not analyzed empirically in this paper. However, different algorithms work for different purposes: a hierarchical clustering method makes a lot more sense when looking at how brain systems are related hierarchically, which a more simple **kmeans** is more useful for identifying larger modules.

**Future Work** The analysis of clustering algorithm used provides a tool to understand the assumptions made by the various algorithms which will allow more careful cognizance of what clustering of nodes means for the mouse neocortex. It's also a starting point to reexamine which clustering method is really the best when dealing with this data of this

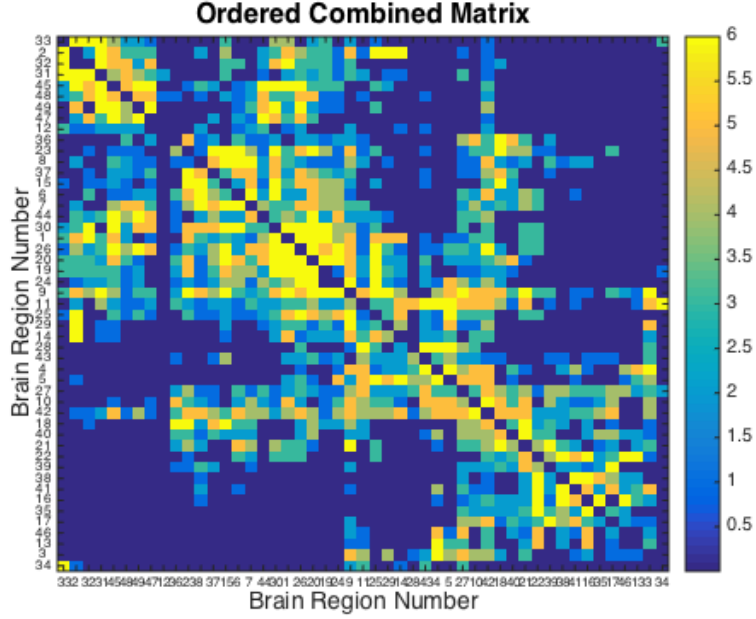


Figure 3: The final compiled matrix, the higher the number the more strong the connection. The brain region number maps to he labels included in the directory where this file lives.

time and how to implement it in future mesoscale connectomic research. The next step from this paper is to empirically analyze the clustering algorithms and see how they perform to the algorithms used in [1].

## References

1. Brian Zingg, Houri Hintiryan, Lin Gou, MonicaY. Song, Maxwell Bay, MichaelS. Bilenkowsky, NicholasN. Foster, Seita Yamashita, Ian Bowman, ArthurW. Toga, Hong-Wei Dong, Neural Networks of the Mouse Neocortex, Cell, Volume 156, Issue 5, 27 February 2014, Pages 1096-1111, ISSN 0092-8674, <http://dx.doi.org/10.1016/j.cell.2014.02.023>.
2. Sporns O, Tononi G, Ktter R (2005) The Human Connectome: A Structural Description of the Human Brain. PLoS Comput Biol 1(4): e42. doi:10.1371/journal.pcbi.0010042.