Project Drosophila

Marco Ballarin, Francesco Bianco
marco.ballarin.6@studenti.unipd.it
francesco.bianco.5@studenti.unipd.it

Abstract

In this work we will analyze the hemibrane of o Drosophila, commonly called "small fruit fly". This analysis will be fascinating since we will apply all the techniques proper of the network science field to a neural connectome, thus connecting two fields that can really benefit of each others. One of the most challenging section of this project will be handling the dataset, since it amounts of more than $2 \cdot 10^4$ neurons and $2 \cdot 10^6$ synapses. We will so develop a technique to reduce the network size, trying to induce the smallest possible variation to the important quantities. Finally, we will pose particular interest in the analysis of the communities, trying to understand if we are able to distinguish the different macro-areas of the brain.

Contents

1 Introduction

In this work we will analyze the structural connectome of a Drosophila, commonly known as "small fruit fly". We will use the hemibrane dataset provided by the FlyEM project[instert citation], which is composed by 20000 neurons connected by 200000 synapses. The size of the resulting network is overwhelming, and lots of the more advanced analysis are not feasible in this framework. For this reason, borrowing the terminology from statistical physics, we will apply a coarse graining to the network, reducing its size. We will then apply several different analysis. This report will be organized as follows:

1. An introduction on the data available, focusing on the macro-areas of the Drosophila brain. We strongly believe that, even if a deep physiological understanding is not necessary, it will really enhance the understanding of the measures we will apply, making this work not a simple application of the network science tools to a different topic, but an interesting case of study;
2. A definition of the coarse graining technique that we will use to reduce the size of the network;
3. An analysis of the simplest network characteristics, on both the grained and the original network, focusing on the evolution of such quantities;
4. A research for the different communities in the networks;
5. The robustness of the network, simulating brain damage. What happens if we damage the reduced network and then expand it to the original size? Is this equivalent to remove the same number of nodes from the original network?
6. A conclusion to wrap up our results.

The introduction is not finished yet. However, it is difficult to write it before the end of the work.

2 Drosophila brain regions

1. **OL**: Optical lobe. sits behind the arthropod eye (mostly compound eyes) and is responsible for the processing of the visual information. It is made up of three layers:
   (a) Lamina (ganglionaris): responsible for contrast enhancement through lateral inhibition
   (b) Medulla: processes movement and shows movement direction sensitivity. Possesses local motion detectors
   (c) Lobula: integrates information from large areas of the visual field to abstract visual information and object recognition
2. **MB**: Mushroom body. They are also known to play a role in olfactory learning and memory. In most insects, the mushroom bodies and the lateral horn are the two higher brain regions that receive olfactory information from the antennal lobe via projection neurons
3. **CX**: Central complex, navigation, sleep, learning, memory, nociception
4. **LX**: Lateral complex, I think that has the same function of the central complex but it is lateral.
5. **VLNP**: Ventrolateral neuropils
6. **LH**: Lateral horn. It is one of the two areas of the insect brain where projection neurons of the antennal lobe send their axons. The other area is the mushroom body. Several morphological classes of neurons in the lateral horn receive olfactory information through the projection neurons.n lateral horn, axons of pheromone-sensitive projection neurons are segregated from the axons of plant odor-sensitive projection neurons. In addition, the dendrites of lateral horn neurons are restricted to one of these two zones, suggesting that pheromones and plant odors are processed separately in the lateral horn.
7. **SNP**: Superior neuropils
8. **INP**: Inferior neuropils
9. **AL**: antennal lobe. The antennal lobe is the primary (first order) olfactory brain area in insects. The antennal lobe is a sphere-shaped deutocerebral neuropil in the brain that receives input from the olfactory sensory neurons in the antennae and mouthparts
10. **VMNP**: ventromedial neuropils
11. **PENP**: periesophageal neuropils
12. **GNG**: gnathal ganglia, taste and feeding;

There is no clear definition for the neuropils regions, and so we provide a definition for what the neuropils are. Neuropil (or "neuropile") is any area in the nervous system composed of mostly unmyelinated axons, dendrites and glial cell processes that forms a synaptically dense region containing a relatively low number of cell bodies. We can so think of them as an information bottleneck, the bridges that connects other brain regions. We present in Figure 1 a quick and not exact representation of the network between these high level regions.

3 Coarse graining

It is necessary to define rigorously a way to apply the coarse graining procedure, i.e. our way to combine different neurons in super-neurons. We stress that this procedure is not simple at all, and present different degrees of freedom that can (and will) be chosen with respect to the performances on the data. We start by defining these degrees of freedom:

- the **metric** d_{ij} that we use to define when two different neurons are similar;
- The way in which we aggregate the weights after we combine two nodes.

We will so present two different algorithms that approaches in different ways the second degree of freedom. We stress that the first algorithm is not really working, but it is however instructing analyzing why it is so.

3.1 Random aggregation

We so present the steps of the first algorithms. Calling $n_i^{(\alpha)}$ the i-th neuron at the α-th iteration in the coarse graining algorithm and $w_{ij}^{(\alpha)}$ the links between neurons $n_i^{(\alpha)}$, $n_j^{(\alpha)}$:

1. Pick a neuron $n_i^{(\alpha)}$ at random;
2. Compute the distance $d_i^{(\alpha)} = (d_{i1}^{(\alpha)}, d_{i2}^{(\alpha)}, \dots, d_{iN}^{(\alpha)})$;
3. Combine the two nearest neurons:
$$n_i^{(\alpha+1)} = \left\{ n_i^{(\alpha)}, \min_{d_i} \left[n_k^{(\alpha)} \right] \right\}$$
4. Build the new network connections:
$$w_{ij}^{(\alpha+1)} = \left\{ w_{ij}^{(\alpha)}, \bar{w}_{kj}^{(\alpha)} \right\}$$
where $\bar{w}_{kj}^{(\alpha)}$ are a random subset of the connections $w_{kj}^{(\alpha)}$ such that the density of the network is conserved.
5. Start again from point 1.

This constraint on the density was needed, since it avoids a proliferation of connections. It is, however, a bound that is not easy to fulfill. In particular, it can be implemented by randomly extract the correct number of connections. This does not work in all cases, since not all nodes have the same degree: there are a lot of nodes we very small degree, that have less connections than the ones needed. Another problem of this algorithm is that it is an iterative algorithm and it is so really slow: at each iteration we need to compute all the distances and we can eliminate only a neuron at each iteration.

3.2 Clustering aggregation

The second algorithm is simpler and more effective. We make use of a hierarchical agglomerative clustering technique, i.e. an algorithm that recursively merges the pair of clusters that minimally increases a given linkage distance. This is really similar to the idea developed for the community detection using the dendogram technique. We so perform the following steps:

- Select the final number of neurons N', i.e. the number of clusters for the agglomerative clustering;
- Compute the distance matrix d_{ij};
- Apply the clustering algorithm;

- Combine nodes in the same cluster in a supernode I, which is connected to another supernode J if at least a node in I was connected in a node in J. The weight of the connection is the sum of the weights fo the single nodes, i.e.:

$$W_{IJ} = \sum_{k \in I, l \in J} w_{kl}$$

This algorithm has several advantages over the previous one:

- We do not have to impose the density constraint to avoid the proliferation of the connections;
- We have to compute the distance matrix only once. This can not seem an advantage. In the random algorithm we had to compute the distance from a node $N \cdot (N - N')$, where N is the total number of nodes and N' the final number of nodes after the coarse graining, while in the clustering case we must compute the distance N^2 times. However, if we want to try different N' or run the algorithm multiple times in the first case we should start from scratch for each graining, while in the second case we only need to compute the distance matrix once;
- The clustering algorithm used is optimized for even larger systems, and it is so really fast.

We decided to use two different metrics for this procedure. We only have to remember that we are minimizing a distance in these algorithms, so strongly connected nodes will be nodes considered nearer.

1. The weights, with
$$d^(w)_{ij} = \frac{1}{w_{ij}}$$
This is a really naive distance metric, and we do not expect a really good performance for the algorithm. However, it has the advantage of being already computed;
2. The local page rank, with:
$$d^{pr}_{ij} = \frac{1}{\mathrm{pr}_{ij}}$$
This is a measure often used in network science, and takes into account both the outcoming and incoming links. We expect much better results with this metric.

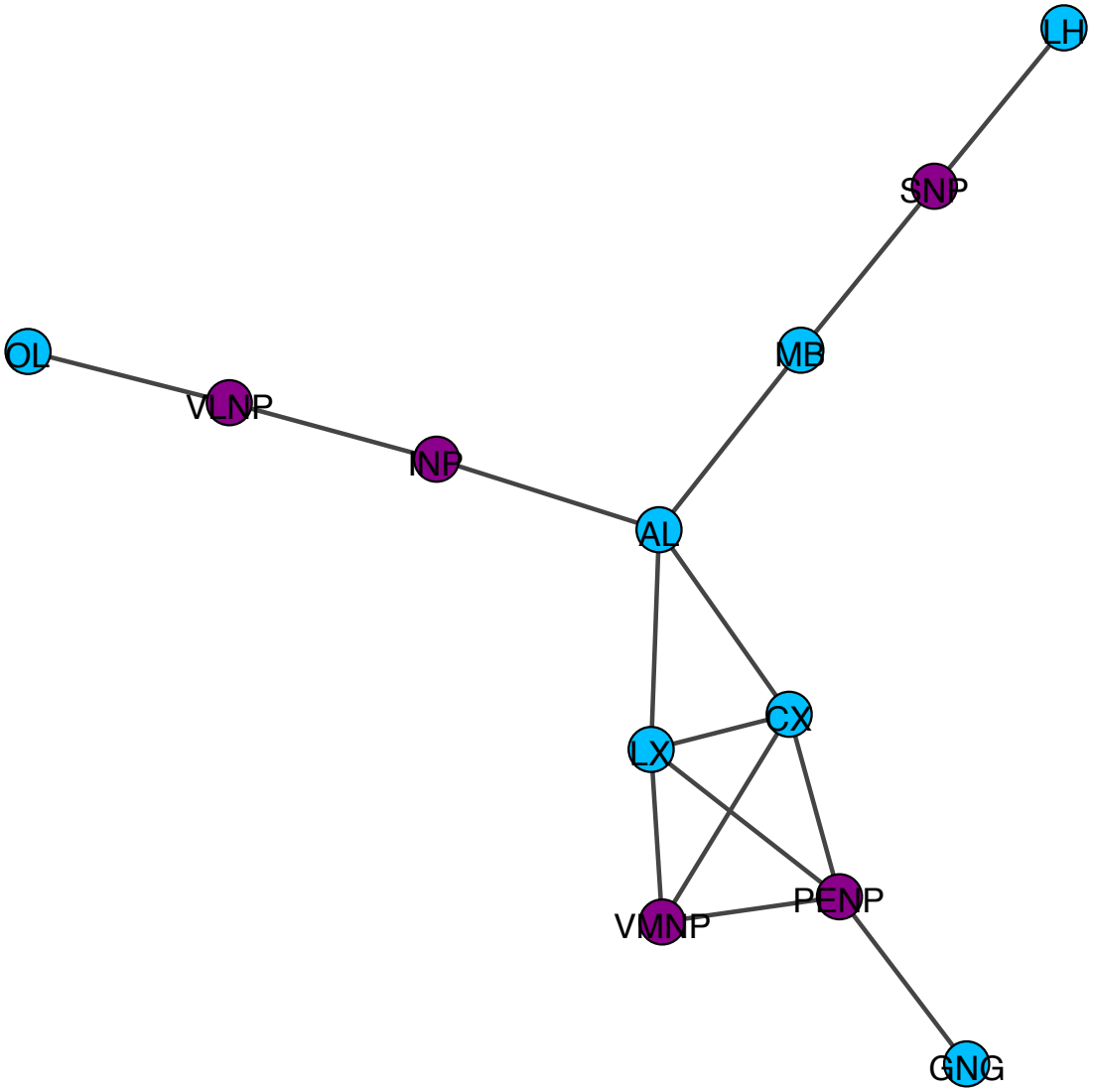

Figure 1: Approximate graph of the high level brain regions. We reported in magenta the neuropils. We can see even in this sketch that they act as bridges between brain regions. Using another layout (tree network) we can see that the central complex is the central one.