Drosophila larva brain
Complex Networks project

Aldo Perri, Francesca Felici

February 1, 2024

Abstract

This project, undertaken for Complex Networks exam, is primarily focused on conducting an analysis of a complex network. Specifically, the selected network examines the neural connections within the larval brain of the fruit fly Drosophila melanogaster[1].
The dataset for Drosophila melanogaster has been sourced from *Netzschlauder*, the analysis extensively employs python with the packages *graph-tool* and *NetworkX*, and this report was typeset using LaTeX.

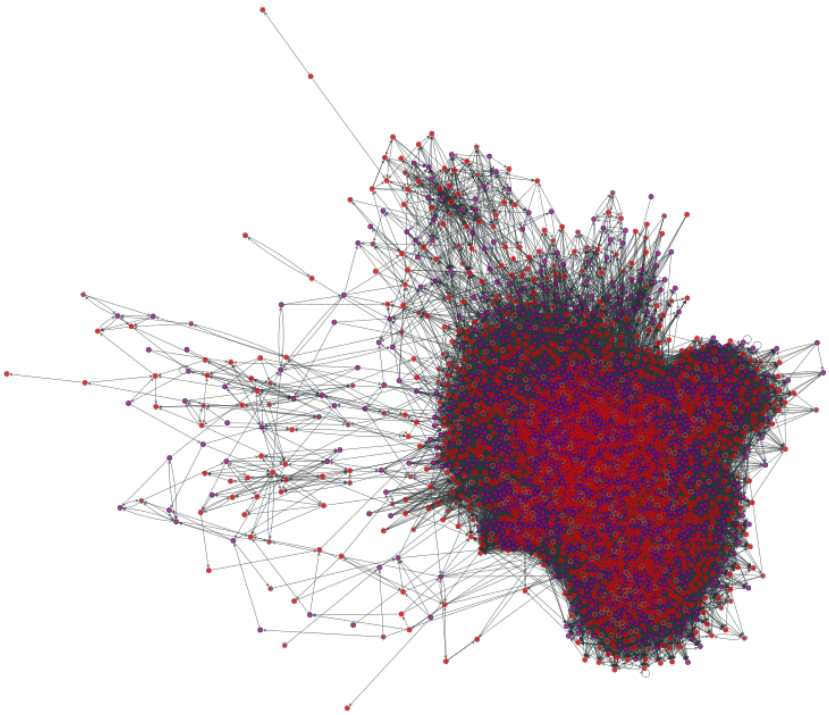


Figure 1: Graph illustrating connections, with red edges representing connections in the left hemisphere and blue edges for the right hemisphere.

1 Topological analysis

The introductory section consists of a concise review of all theoretical knowledge pertinent to the analysis of complex networks. The analysis is been split into distinct subsets to clarify the specific categories of graph features under examination.
The brain of the Drosophila larva can be modelled as a complex network: its features are captured by the *synaptic wiring diagram* also called *connectome*.

Neurons can be thought as nodes and synapses as the links/edges of the network. Each neuron is split into two compartments, axon and dendrite, resulting in a rich multiplexed network with four connection types. Therefore synaptic connections between brain neurons are categorized as axo-dendritic (a-d), axo-axonic (a-a), dendro-dendritic (d-d), or dendro-axonic (d-a) as shown in fig 2.
In our dataset we have a number of 2952 nodes connected by 116922 weighted links cathe-

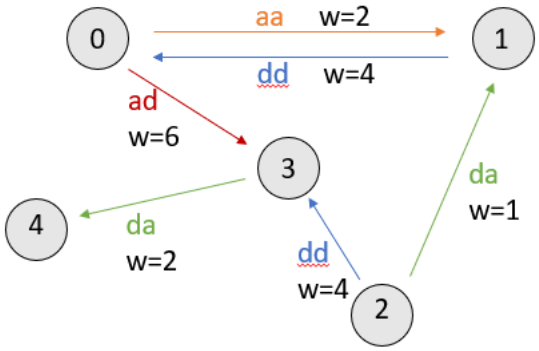


Figure 2: Example of multiple connection between nodes.

gorized by type connection, to which we refer as "etype" from now on: 'aa', 'ad', 'da', 'dd'. The effective number of a certain etype of synapses between two neurons is given by the weight of the edge of the corresponding etype: the total number of synapses is 352611.

1.1 Four connection types

From our datas, we constructed a directed multigraph containing all edge types and then we extracted four subgraphs corresponding to the four etype to perform analysis on each subgraph. In table 5 we can see for each subgraph the number of total nodes, the number

Etype	Nodes	Edges	Synapses	Density	Max Degree
A-D	2880	63545	234958	0.77 %	203
A-A	2887	40636	90800	0.49%	241
D-D	2204	9019	20415	0.19%	143
D-A	1907	3722	6438	0.10%	69

Table 1: Graph metrics for subgraphs comprising each connection type: number of nodes, number of edges, graph density and max degree.

of total edges, the density and the max degree. We can see that the majority of synapses are in 'ad' (66.6 %) and 'aa' (25.8 %) connection. However there is still a non negligible amount of synapses in the 'dd' (5.8 %) and 'da' (1.8 %) connections.
The density of a directed graph is the ratio between the actual directed edges in the graph over the number of all possible edges that can be present in the given graph: the 'ad' graph displays the highest density while the 'da' graph the lowest one.
The max degree is defined as the maximum number of connections (IN+OUT) from a single neuron in a given graph.

We examined the adjacency matrices of each subgraph to understand how nodes are interconnected in each case.

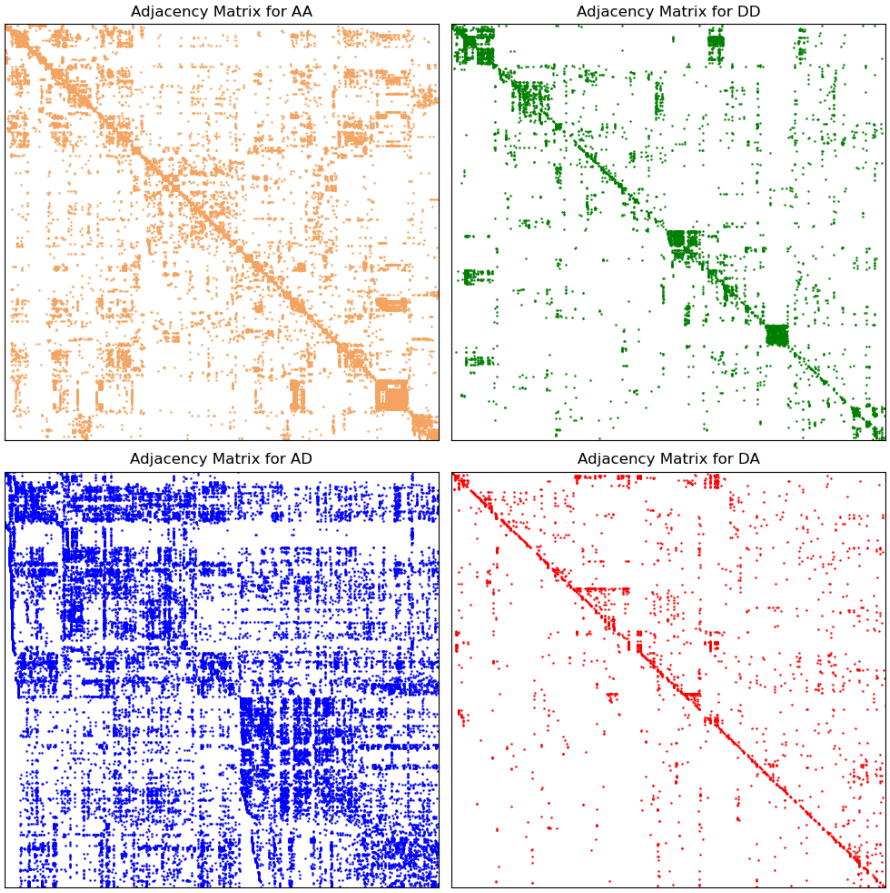


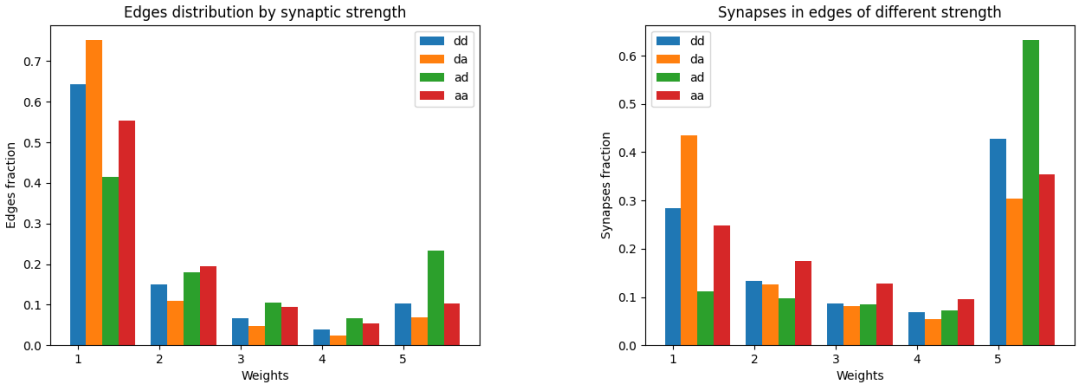
Figure 3: Adjacency matrices for each connection type.

We investigated the edge reciprocity of each sub-graph based on the etype. The reciprocity of a directed graph gives us a measure of the probability that two nodes are connected through mutual edges. Then we generalized this notion to graphs corresponding to different etype quantifying, for example, the probability that given an edge of etype 'aa' from node i to node j there is a mutual edge of etype 'ad' from j to i. The results are listed in tab 2.

We next investigated the distribution of the edge strengths for each etype. The result are shown in fig. 4a. We can see that most edges are weak for all etype. We also investigated the distribution of the number of synapses for edge strength for each etype. In this case we see, as shown in fig. 4b, that the majority of synapses are contained in

/	A-D	A-A	D-D	D-A
A-D	0.033	0.033	0.010	0.038
A-A	0.053	0.465	0.023	0.009
D-D	0.072	0.103	0.154	0.014
D-A	0.641	0.093	0.035	0.012

Table 2: Edge reciprocity between different edge types, i.e., fraction of forward edges that were coincident with different backward edge types.



(a) Edges distribution.

(b) Synapses distribution.

Figure 4: Distributions for edge strengths (weights). Note that the weight '5' actually refers to weight ≥ 5

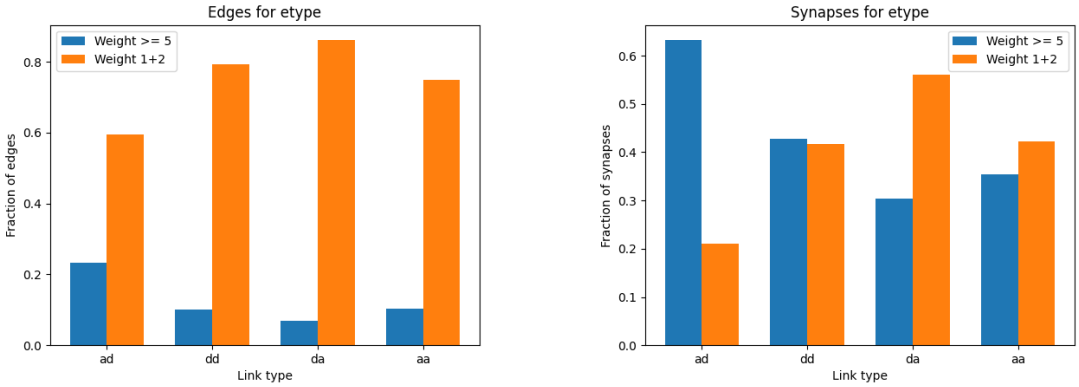
strong edges (weight ≥ 5). In fig. 5a and fig.5b we also plotted the fraction of edges and synpases in strong and weak edges. Regarding the edges fraction, all etype have the same qualitatively trend, while we can see some differences in the fraction of synapses. In fact we have that: for 'dd' and 'aa' we nearly have the same fraction in weak and strong edges; for 'ad' the majority of synapses are in the strong edges while for 'da' we have the opposite trend.

We next study if neurons are connected by different types of links: the majority of neurons are linked by only one etypes but there are several nodes which are linked by different combinations of edge types. In particular we compared the actual number of links per combination of etype to the expected number of a null model (E-R network). The comparison is shown in fig. 6.

1.2 Two Hemispheres

The nodes in the dataset are also labeled with the property 'Hemisphere' which can be 'left' or 'right' reporting in which brain hemisphere a given neuron is. We therefore constructed from the multigraph four subgraphs corresponding to the neural connections LEFT→LEFT, LEFT→RIGHT, RIGHT→LEFT, RIGHT→RIGHT. In figure 7 we plotted the adjacency matrix for the connections L→L, L→R, R→L, R→R with the nodes suitably sorted according to their position (Left or Right).

In table 3 we reported the main parameters of the four networks: number of nodes,



(a) Edges fractions.

(b) Synapses fractions.

Figure 5: Fraction of edges or synapses in strong (≥ 5 synaptic strength) or weak edges (1-2 synaptic strength), per etype

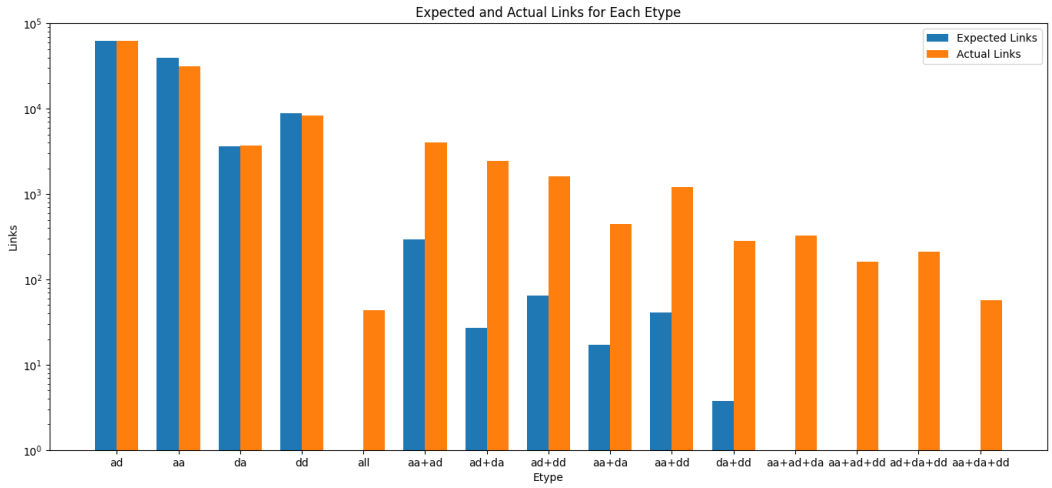


Figure 6: Number of edges displaying multiple connection types.

number of edges and number of synaptic sites. Comparing these values to the ones of the complete network we deduce that there are 346 nodes, i.e. neurons, which are neither in the left neither in the right hemisphere. Moreover we can see that there are 28354 edges between the two hemispheres corresponding to 48081 synaptic sites. Therefore we can conclude there are 21394 connections involving the 346 not localized neurons corresponding to 50304 synapses.

	L→L	**R→R**	**L→L+R→R**	**L→R**	**R→L**	**Overall**
Nodes	1299	1307	2606	2214	2212	2952
Edges	32598	34576	67174	14365	13989	116922
Synapses	104189	113327	217516	44006	40785	352611

Table 3: Number of nodes, edges and synaptic sites in each brain hemisphere compared with the overall numbers of the complete network and the connection between the two hemispheres.

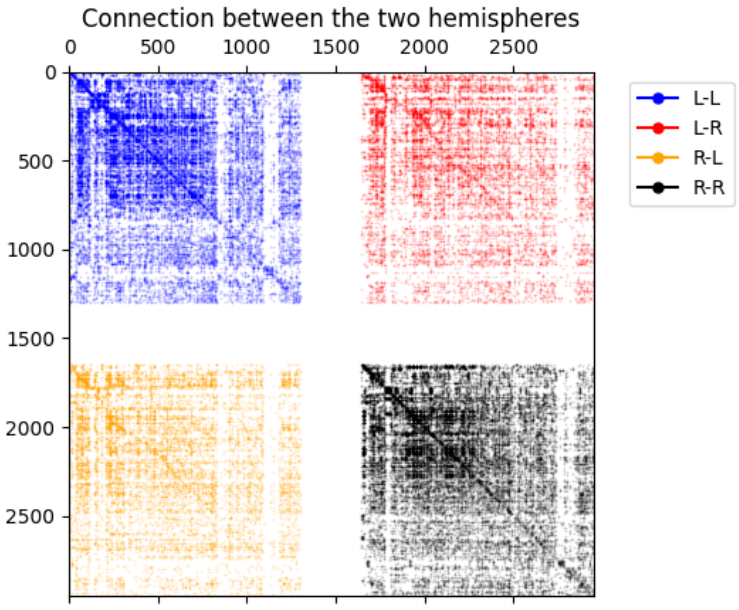


Figure 7: Adjacency matrices for the connections within and between the hemispheres. Note that the nodes are sorted according to 'Left' and 'Right' attributes. The white stripes represent the nodes which are not labeled with the attribute 'Hemisphere'.

1.3 Ipsilateral, bilateral and contralateral neurons

The presence of two hemispheres is a fundamental property of the brain. However how information is shared by the hemisphere is not yet well understood [1]. We can study the interhemispheric interaction starting from classifying neurons in three distinct cathegories:

- Ipsilateral neurons: neurons which are linked to neurons in the same hemisphere
- Bilateral neurons: neurons linked to both the hemispheres
- Contralater neurons: neurons localized in a hemisphere but linked only to neurons of the other side

In table 4 we summarize the number of ipsilateral, bilateral and contralateral neurons according to the connection type. We see that for Dendriditc-Dendriditc and Dendritic-

Etype	Ipsilateral	Bilateral	Contralateral
A-A	397	2114	32
A-D	320	2193	21
D-A	895	547	240
D-D	1698	210	9

Table 4: Number of ipsilateral, bilateral, contralateral neurons according to the connection type.

Axonic connection the most neurons are ispilateral. For Axo-Axonic and Axo-Dendritic connection most neurons are bilateral. There is a small fraction of contralateral neurons

for most etypes: 1.2 % (A-A), 0.8 %(A-D), 0.4 % (A-D). However we have a non negligible fraction of contralateral neurons in the D-A network with the 14.2%.

1.4 Spectral decomposition

We performed the spectral decomposition of the adjacency matrix, as described in 2.4, obtaining the matrix U containing the left singular vectors. The principal components extracted from U were visualized in the following plot. The left subplot shows the values of the first principal component, while the right subplot illustrates the second principal component.

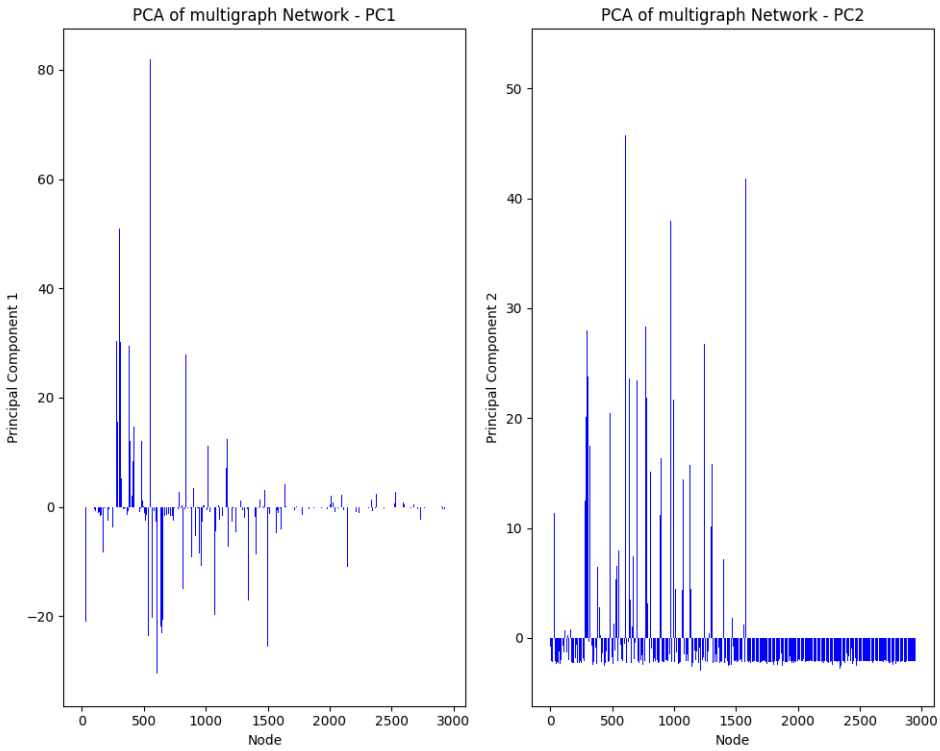


Figure 8: The spectral decomposition plot illustrates the principal components derived from the graph's adjacency matrix.

1.5 PCA analysis

We applied the PCA method to the adjacency matrix of the graph. The procedure we followed is:

- Compute the SVD of adjacency matrix.
- Extract the matrix U, to obtain a 2D representation of the graph.

The following plot provides the spatial distribution of nodes in the reduced 2D space, (with each node's position determined by the values in the corresponding rows of the matrix U). Moreover we also included color-based hemisphere information for the nodes, allowing a visual representation of the graph based on both connectivity patterns and

hemispheres.

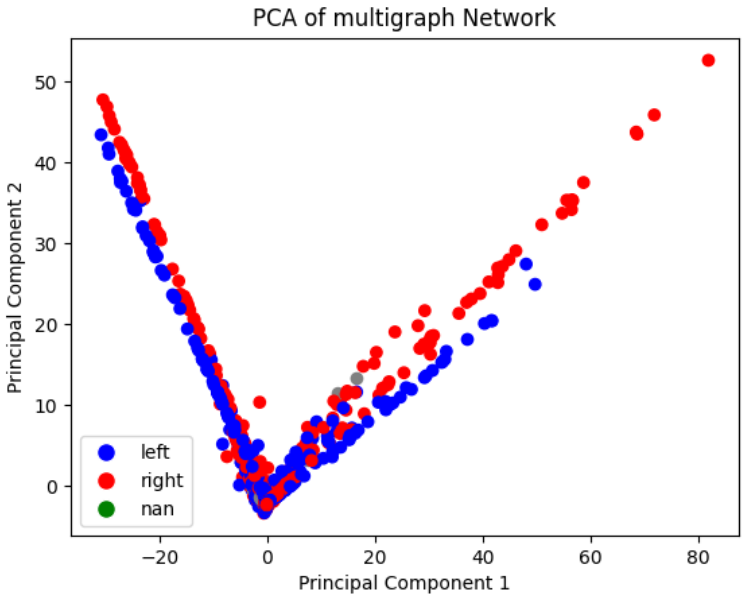


Figure 9: PCA plot with hemisphere Information

1.6 Multiplex

In order to better understand the connections structure we investigate separately the structure of the multiplex graph constructed only with the nodes involved in multiple type connections. In this graph we have 706 nodes in the left hemisphere and 716 nodes in the

Etype	Nodes	Edges	Synapses	Density	Max Degree
Multiple connection	1609	62498	189810	2.4 %	309

Table 5: Graph metrics for multiple connection graph: number of nodes, number of edges, graph density and max degree.

right one: therefore there are 187 neurons not localized in any hemisphere. Investigating the lateral nature of their connections we found:

- 3 contralateral neurons
- 1397 bilateral neurons
- 22 ipsilateral neurons

We applied a clustering algorithm in order to individuate communities in the multiplex. We individuated five communities with repectively with 751, 626, 228, 2, 2 neurons each

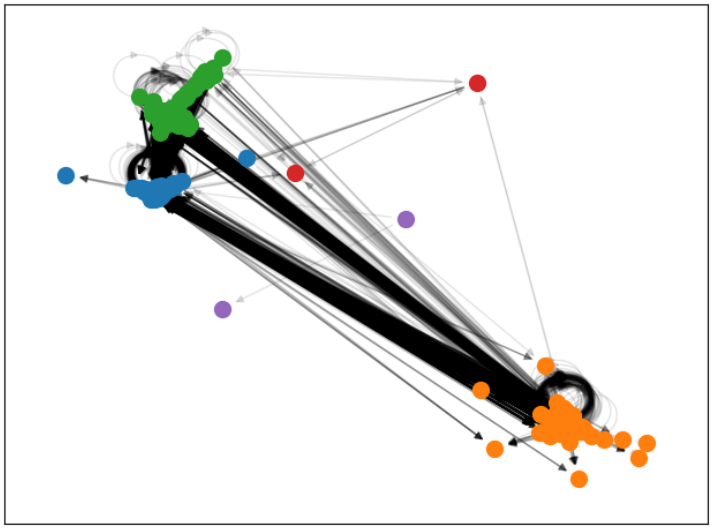


Figure 10: Five communities of the multiplex

as shown in fig.10. We see tha the most populated communities are higly inter-connected between each other.

We done a PCA analysis first on the adjacency matrix (fig.11) and then on various nodes measure (fig.12).

We can see that PCA analysis of the adjacency matrix in this case is able to distinguish between left and right hemispeheres.

With regard to the analysis on nodes measure we analyzed:

- Degree centrality
- Closeness centrality
- Betweenness centrality
- Eigenvector centrality
- Pagerank

The first two principal components are able to explain the 96 % of the variance.

2 Methods

2.1 Graph metric

In the case of a directed graph the density d for a graph of N nodes and m directed edges is given by [2]:

$$d = \frac{m}{N(N-1)} \tag{1}$$

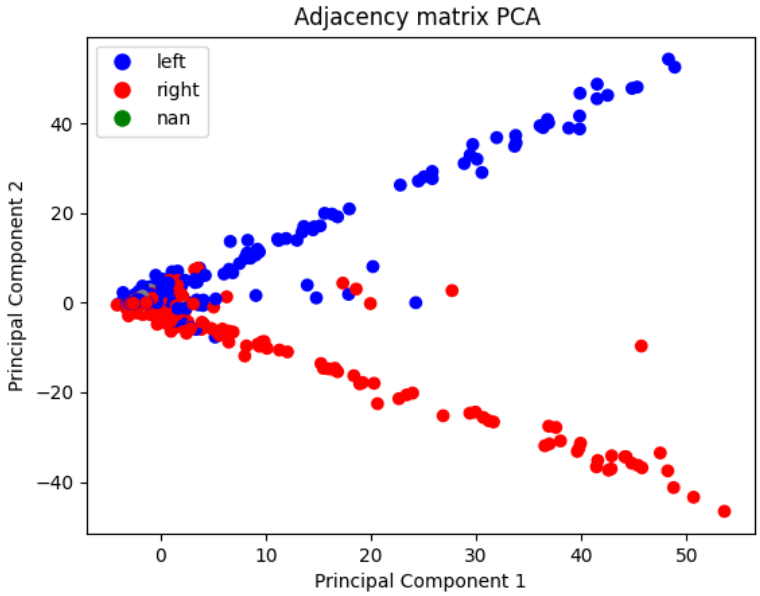


Figure 11: Projection of the PCA of adjacency matrix.

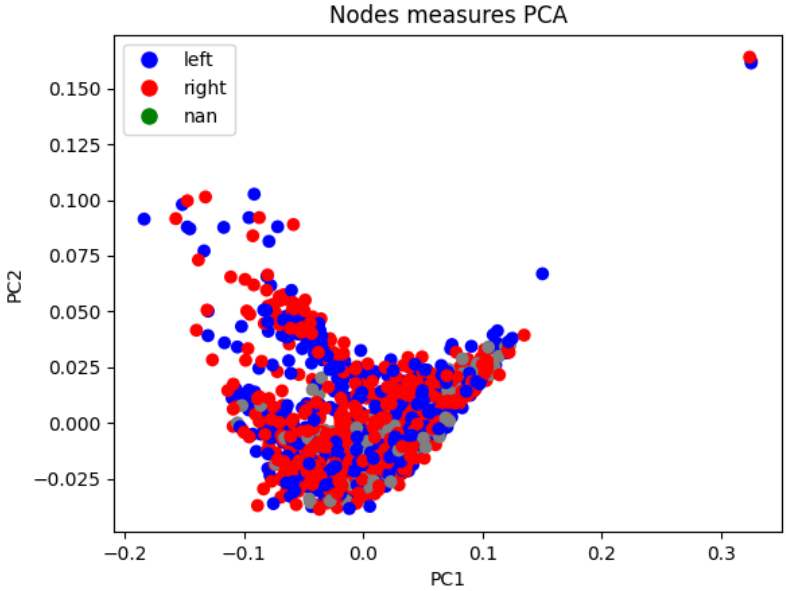


Figure 12: Projection along the first and second principal components of PCA analysis of multiple node measures.

2.2 Reciprocity

For each graph we extracted the loopless unweighted adjacency matrix A_{ij}^{etype} and we quantified the reciprocity r as:

$$r(etype) = \frac{1}{\sum_{ij} A_{ij}^{etype}} \sum_{ij} A_{ij}^{etype} A_{ji}^{etype} \tag{2}$$

which is the number of mutual edges divided by the total number of edges of a given graph. The result are in the diagonal of table 2. We then generalize the notion to graphs corresponding to different etypes. In this case the reciprocity between a source graph and a target graph is given by:

$$r(source, target) = \frac{1}{\sum_{ij} A_{ij}^{source}} \sum_{ij} A_{ij}^{source} A_{ji}^{target} \tag{3}$$

2.3 Null model

We compared the actual counted edges for each etype with the expected values one would have if all our network would be a random ER (Erdos-Renyi) network or null model. This particular model assumes that each of the four etype graphs are independent from each other: deviation from the random model would highlight interesting features characterizing our network.
In order to construct a null model we started from the computation of the *connection probability* p_a for each subgraph $A^{(a)}$ efined as:

$$p_a = \frac{1}{N^2} \sum_{ij}^{N} A_{ij}^{(a)} \tag{4}$$

where N is the number of nodes and $A^{(a)}$ is the unweighted directed adjacency matricx for the graph of etype $a(a = aa, ad, da, dd)$. In the assumption of independent connection we have that, for example, the expected number of connection of etypes 'aa' is given by $n_{aa} = N^2 p_{aa}(1-p_{ad})(1-p_{da})(1-p_{dd})$. Let be K a four vector that codifies the combination of different etypes: its entries are 0 (absence of a certain etype) or 1 (presence of a certain etype). The number of expected links corresponding to the combination K is:

$$n(K) = N^2 \prod_a p_a^{K_a} (1 - p_a)^{1-K_a} \tag{5}$$

The results are plotted in figure 6.

2.4 Spectral Decomposition and Principal Components

Given the adjacency matrix A of a graph, the Singular Value Decomposition (SVD) expresses A as a product of three matrices:

1. The left singular vectors matrix U:

$$A = U \Sigma V^T$$

where U contains the principal components as its columns.

2. The diagonal matrix of singular values Σ:

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_k \end{bmatrix}$$

where σ_i are the singular values in descending order, representing the amount of variance captured by each principal component.

3. The transpose of the right singular vectors matrix V^T:

$$V^T = \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_k^T \end{bmatrix}$$

2.5 PCA analysis

Principal Component Analysis (PCA) is a dimensionality reduction technique used to transform high-dimensional data into a lower-dimensional space, keeping the most significant information about the graph. PCA identifies the principal components, and then they're used to construct a new, reduced, coordinate system for the data.

2.6 Clustering

In order to detct the five commuinites of the multiplex we used the *greedy-modularity-communities* algorithm of NetworkX: this algorithm uses Clauset-Newman-Moore greedy modularity maximization [2] to find the community partition with the largest modularity.

3 Results

3.1 Deviation from ER network

Figure 6 shows that the connectome of the Drosophila larva brain deviates from the null model we constructed on the assumption of independent etype connections. A statistical analysis of the validity of the null model suggests a $\chi^2 = 10596$ with a p-value of nearly zero. In particular this is more evident in the multiple type connection: we have a significant number of edges sharing more than just one etype.
However the majority of connections displays only one type of connection among 'aa', 'ad', 'da' and 'dd' (90.7 %).

3.2 Edge symmetry

We investigated edge symmetry across the two brain hemisphere, i.e. we asked if a given connection in the left hemisphere is reproduced in the right one. We could do that since

in the dataset to each node in one hemisphere is associated a homologous node in the another one. Therefore we looked if, given an edge of a certain type in the left hemisphere between two nodes, there were in the right hemisphere an edge of the same type between the neurons homologous to the previous ones.
The fraction of edges which are conserved between the two hemispheres is given by 41.6 %: strong edges have the highest probability to be reproduced in both hemispheres while weak edges are mostly asymmetrical for all the four etypes as shown in figure 13. In particular the overall fraction of symmetrical edges with weight ≥ 10 is 86 % while for a weight ≥ 15 is 90 %.

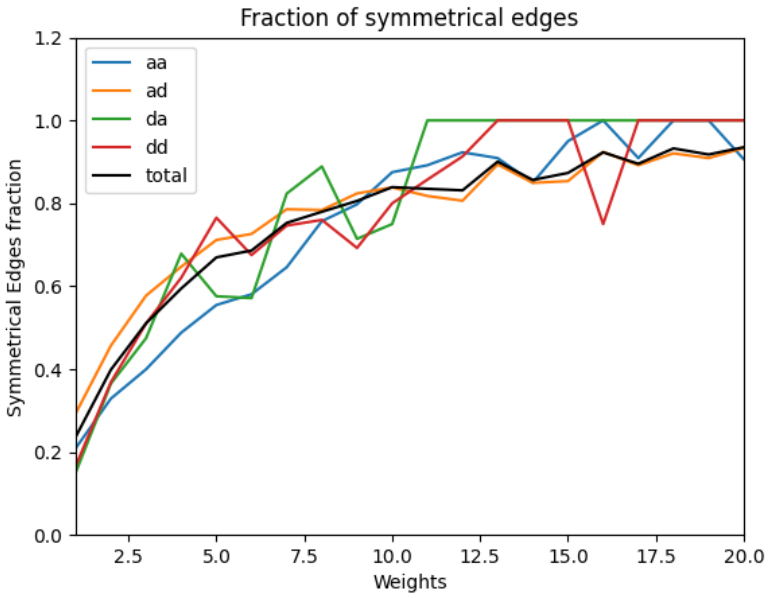


Figure 13: Stronger edges were more likely to have a homologous edge in the opposite hemisphere than weak edges.

3.3 Communication between hemispheres

Due to the presence of a high number of bilateral and contralateral neurons according to each etypes we can study the structure of interhemispheric connection to have a glimpse on how the two hemispheres shares information.
We investigated the connection probability between ipsilateral, bilateral and contralateral nodes for each of the etypes. In all networks we nearly have the same features: ipsilateral neurons seems to equally connect approximately with ipsi/bi/contra-lateral neurons. Bilateral neurons tends to connect more with other bilateral neurons while contralateral nodes prefers contralateral ones.
We also investigated if a contralateral neuron in one hemisphere is linked directly to its own homologue in the other hemisphere. We found that:

- A-A network: 20 % of probability that a contralateral neuron is linked to its homologue
- A-D network: 26 % of probability that a contralateral neuron is linked to its homologue

- D-A network: 48 % of probability that a contralateral neuron is linked to its homologue
- D-D network: 67 % of probability that a contralateral neuron is linked to its homologue

3.4 Multiplex

PCA analysis of the adjacency matrix correctly individuated the two hemispheres. However we can see that there's a set of 11 nodes approximately around $x = 0$ (circled in green box in fig.14) providing a particular behaviour. In order to investigate such a set,

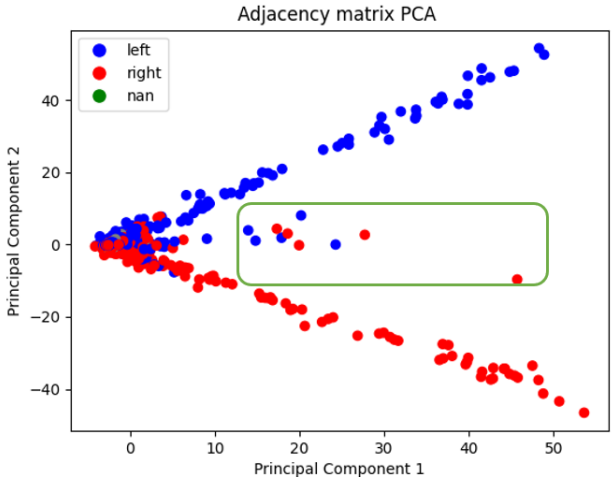


Figure 14: Caption

we constructed the subgraph made only by those nodes we found that it is indeed a connected graph, as shown in figure 15.

The interesting thing is that among these 11 nodes we have 4 couples of homologous neu-

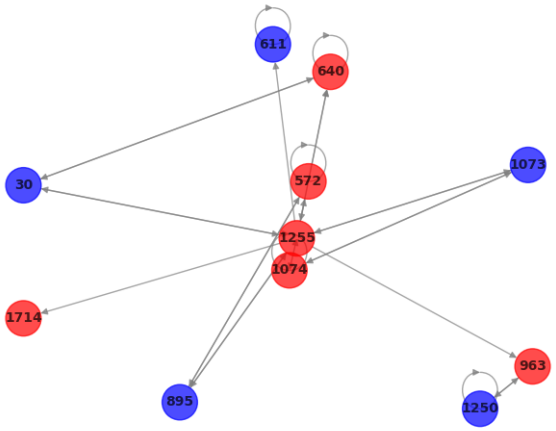


Figure 15: Connected graph of the nodes with particular behaviour coming from PCA analysis of adjcency matrix. Nodes ID are shown.

rons, namely (node ID in round brackets): (30, 640), (963,1250), (1073, 1074) and (572, 895). Moreover the most connected node (ID = 1255) of this subgraph has no homologue

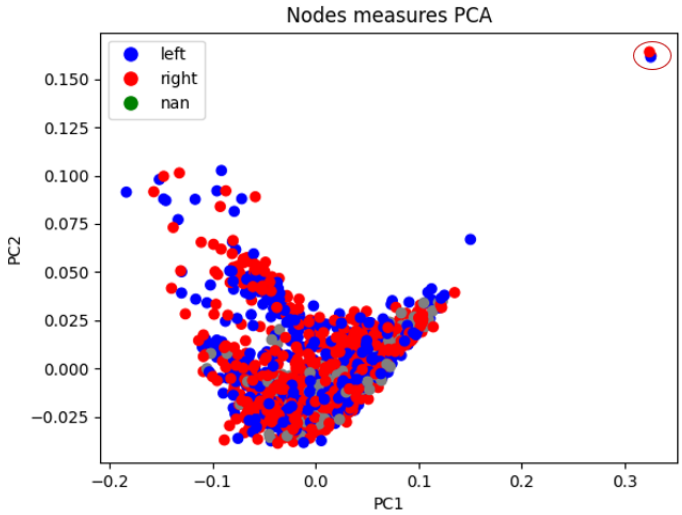


Figure 16: Couple of "strange-behaved" neurons.

in the left hemisphere, thus in a sense it has a unique role, maybe relevantly involved in the communication between hemispheres.
On the other hand PCA analysis on multiple node measures individuated a couple of left-right nodes which has singular behaviour as can be seen in fig.16 where the "strange" couple is circled in red. Investigating on such a couple of neurons we found that it is a homologous pair and moreover they're contralateral neurons. Additionaly they are exactly the same nodes composing the purple community in figure 10.

References

[1] Michael Winding, Benjamin D Pedigo, Christopher L Barnes, Heather G Patsolic, Youngser Park, Tom Kazimiers, Akira Fushiki, Ingrid V Andrade, Avinash Khandelwal, Javier Valdes-Aleman, et al. The connectome of an insect brain. *Science*, 379(6636):eadd9330, 2023.

[2] M. E. J. Newman. *Networks: an introduction*. Oxford University Press, Oxford; New York, 2010.