Systems/Circuits

Hierarchical Modular Structure of the *Drosophila* Connectome

Alexander B. Kunin,[1,2] Jiahao Guo,[3,4] Kevin E. Bassler,[3,4,5] Xaq Pitkow,[2,6,7,8,9] and Krešimir Josić[5,10]

[1]Department of Mathematics, Creighton University, Omaha, Nebraska 68178, [2]Department of Neuroscience, Baylor College of Medicine, Houston, Texas 77030, [3]Department of Physics, University of Houston, Houston, Texas 77204, [4]Texas Center for Superconductivity, University of Houston, Houston, Texas 77204, [5]Department of Mathematics, University of Houston, Houston, Texas 77204, [6]Department of Electrical and Computer Engineering, Rice University, Houston, Texas 77005, [7]Center for Neuroscience and Artificial Intelligence, Baylor College of Medicine, Houston, Texas, [8]Neuroscience Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, [9]Department of Machine Learning, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, and [10]Department of Biology and Biochemistry, University of Houston, Houston, Texas 77204

The structure of neural circuitry plays a crucial role in brain function. Previous studies of brain organization generally had to trade off between coarse descriptions at a large scale and fine descriptions on a small scale. Researchers have now reconstructed tens to hundreds of thousands of neurons at synaptic resolution, enabling investigations into the interplay between global, modular organization, and cell type-specific wiring. Analyzing data of this scale, however, presents unique challenges. To address this problem, we applied novel community detection methods to analyze the synapse-level reconstruction of an adult female *Drosophila melanogaster* brain containing >20,000 neurons and 10 million synapses. Using a machine-learning algorithm, we find the most densely connected communities of neurons by maximizing a generalized modularity density measure. We resolve the community structure at a range of scales, from large (on the order of thousands of neurons) to small (on the order of tens of neurons). We find that the network is organized hierarchically, and larger-scale communities are composed of smaller-scale structures. Our methods identify well-known features of the fly brain, including its sensory pathways. Moreover, focusing on specific brain regions, we are able to identify subnetworks with distinct connectivity types. For example, manual efforts have identified layered structures in the fan-shaped body. Our methods not only automatically recover this layered structure, but also resolve finer connectivity patterns to downstream and upstream areas. We also find a novel modular organization of the superior neuropil, with distinct clusters of upstream and downstream brain regions dividing the neuropil into several pathways. These methods show that the fine-scale, local network reconstruction made possible by modern experimental methods are sufficiently detailed to identify the organization of the brain across scales, and enable novel predictions about the structure and function of its parts.

Key words: clonal units; community detection; connectome; Drosophila; machine learning

---

Significance Statement

The Hemibrain is a partial connectome of an adult female *Drosophila melanogaster* brain containing >20,000 neurons and 10 million synapses. Analyzing the structure of a network of this size requires novel and efficient computational tools. We applied a new community detection method to automatically uncover the modular structure in the Hemibrain dataset by maximizing a generalized modularity measure. This allowed us to resolve the community structure of the fly hemibrain at a range of spatial scales revealing a hierarchical organization of the network, where larger-scale modules are composed of smaller-scale structures. The method also allowed us to identify subnetworks with distinct cell and connectivity structures, such as the layered structures in the fan-shaped body, and the modular organization of the superior neuropil. Thus, network analysis methods can be adopted to the connectomes being reconstructed using modern experimental methods to reveal the organization of the brain across scales. This supports the view that such connectomes will allow us to uncover the organizational structure of the brain, which can ultimately lead to a better understanding of its function.

Introduction

Understanding how brains function requires understanding how they are wired, and how this wiring underpins neural computation (Bullmore and Sporns, 2009). Advances in biology, imaging, and machine learning have led to a proliferation of vast, highly detailed connectomes of brain tissue from insects (Scheffer et al., 2020), mammals (Turner et al., 2022), and humans (Shapson-Coe et al., 2021). Dense reconstructions of neural tissue at synaptic resolution present new opportunities to interrogate the wiring principles of different brains. However, the enormous volume of the data presents fundamental challenges since it is difficult to identify relevant features of a large network and understand how these features interact.

The brains of sufficiently complex animals, including insects, are composed of interacting structural units that often have distinct functions. These "communities" are sometimes, but not always, anatomically distinct and have traditionally been identified using painstaking methods that involve the tracing of individual cells and their connections. However, as the number and size of connectomes grow, we will need automated methods to uncover a brain's functional units and their interaction. Networks of the size and connectedness of connectomes pose fundamental technical challenges to many existing network community detection methods, the speed and accuracy of which can scale poorly with network size and often cannot resolve structures below a certain limiting size (Fortunato and Hric, 2016).

Insect brains are an excellent target for methods that automatically identify the salient features of complex networks. They are small enough and stereotyped enough that complete, validated connectomes are within reach (Scheffer et al., 2020). Much is known about the organization of insect brains already, as anatomists have been able to characterize them at the level of single cells, circuits, and regions, and understand the computational roles and interactions between these components (Hanesch et al., 1989; Lai et al., 2008; Li et al., 2020). Understanding the structure of insect brains also provides insights into general organizational and computational principles of other brains (Haberkern and Jayaraman, 2016; Kim et al., 2017; Takemura et al., 2017a). Recently, automated methods have accelerated and amplified the abilities of scientists to identify neurons and their interactions at high resolution and in large quantities (Turner et al., 2020). The fly Hemibrain (Scheffer et al., 2020), a synapse-level reconstruction of approximately two-thirds of the volume of the brain of an adult female *Drosophila* fruit fly, is the largest (by number of neurons) connectome published to date.

Here we use an automated method to uncover cell communities that may constitute different signaling paths, or may be devoted to distinct computations in the Hemibrain network. We identify communities of neurons as the sets of neurons that are more densely connected than expected in a random network. The community structure is found by partitioning the neurons in a way that maximizes a modularity density measure (M. Chen et al., 2013, 2014; Botta and del Genio, 2016; T. Chen et al., 2018; Guo et al., 2023). To find the maximizing partition, we use a recently introduced machine-learning algorithmic scheme Reduced network Extremal Ensemble Learning (RenEEL) (Guo et al., 2019) that enables fast and accurate analysis of the Hemibrain network. By increasing a tunable parameter in our modularity density measure, we resolve the structure of the network on increasingly smaller scales.

We identify communities that range in size from thousands to only a few neurons and find a roughly hierarchical organization of the structure (K. Ito and Awasaki, 2008). At the coarsest scale, we automatically identify well-known brain regions and functional networks. At smaller scales, our analysis reveals how these brain regions are organized into subnetworks. It also identifies specific connectivity patterns among modules and brain regions. For example, we automatically recover the layered structure of the fan-shaped body (FB) (Hanesch et al., 1989) in an unsupervised way. By considering the cell type composition of the communities, we are able to identify potentially biologically relevant networks and cell type-specific wiring patterns. Thus, we provide a scalable, automated method to identify structure in connectomes composed of tens of thousands of neurons, or more.

Materials and Methods

Datasets analyzed. The larval mushroom body dataset is a dense reconstruction of all neurons in the mushroom body of a larval female fruit fly, *Drosophila melanogaster* (Eichler et al., 2017). The network we studied consisted of 365 neurons, selected for having at least one synapse in the synapse table published in (Eichler et al., 2017). From the synapse table, we constructed a weighted, directed graph with neurons as nodes and edge weights defined by total number of synapses between neurons. For input to the community-detection algorithm, we combined antiparallel edges by summing their edge weights to obtain an undirected graph.

The Hemibrain dataset is a dense reconstruction of approximately one-half of the central brain of an adult female fruit fly, *D. melanogaster* (Scheffer et al., 2020). The network we analyzed was based on version 1.1 of the dataset and consists of 21,733 nodes and 2,872,500 undirected, weighted edges. As for the larval dataset, the undirected edge weights in the network were the total number of synapses between each pair of neurons, in either direction.

Detecting communities with generalized modularity density maximization. We have identified communities of nodes with clusters C_1, C_2,... in a partition of the nodes $C = \{C_1, C_2,...\}$ that maximizes Q_g, the Generalized Modularity Density measure (Guo et al., 2023) as follows:

$$Q_g(\mathcal{C}; \chi) = \frac{1}{2m} \sum_{C \in \mathcal{C}} \left(2m_C - \frac{K_C^2}{2m} \right) \rho_C^\chi. \qquad (1)$$

Here m is the sum of the weights of all links in the network, m_C is the sum of weights of all links between nodes in community C, K_C is the weight-degree sum of nodes in C (the sum of the weights of the links connected to each node in C), and ρ_C is the relative density of connections in C, as follows:

$$\rho_C = \frac{2m_C}{n_C(n_C - 1)}, \qquad (2)$$

where n_C is the number of nodes in C. The exponent χ is a tunable control parameter. We consider $\chi \geq 0$. At $\chi = 0$, Q_g equals the classical modularity score (Newman, 2006). Modularity is the difference of fraction of the network's links that are within communities of the partition C minus what that fraction would be, on average, if the links were randomly shuffled. For $\chi > 0$, a power-law function of the relative density of each cluster C weights the cluster's contribution to Modularity in the Q_g measure.

For networks of the size and density we consider here, Modularity ($\chi = 0$) is typically maximized by partitions that consist of a relatively small number of large clusters. Indeed, it can be difficult to resolve clusters that are small by maximizing Modularity, even if those clusters are very well connected. This well-known resolution limit problem with maximizing Modularity (Fortunato and Barthélemy, 2007; Traag et al., 2011) is mitigated by maximizing Generalized Modularity Density at positive values of χ (Guo et al., 2023). As χ is increased, the clusters in partitions that maximize Generalized Modularity Density tend to subdivide into smaller more tightly linked clusters (Guo et al., 2019).

RenEEL algorithm for maximizing generalized modularity density. Finding the partition that maximizes Generalized Modularity Density can be a challenging and computationally expensive problem, especially for larger and denser networks. An exact solution for an arbitrary network has an NP-Complete computational complexity (Brandes et al., 2008). Thus, it is necessary to use an approximate method that is fast, with polynomial time complexity, but still gives accurate results. We used an algorithm that has been shown to give accurate results for networks as large and dense as we consider in this paper (Guo et al., 2019). This algorithm is based on RenEEL, which uses a machine learning paradigm for graph partitioning, Extremal Ensemble Learning (EEL). ELL evolves an ensemble of partitions toward consensus by replacing the" worst" partition with a new one. RenEEL efficiently generates the new partition by expending effort only where there is disagreement within the ensemble's existing partitions. The speed and accuracy of our algorithm enables analyses of networks with tens of thousands of nodes, such as ours, that previously were not possible.

The RenEEL scheme for community detection first uses a very fast base algorithm to create an ensemble of partitions that try to maximize a modularity measure. Then an iterative learning process is used to update the ensemble. A reduced network is formed by combining the groups of nodes that the ensemble partitions agree should be clustered together into single nodes. The base algorithm is then used to partition the reduced network. The ensemble is updated either by using the new partition to replace the ensemble partition with the lowest modularity or by reducing the size of the ensemble if the new partition matches one already in the ensemble or has a lower modularity than any in the ensemble. The learning process continues until only one partition remains in the ensemble; thus, consensus is reached on what partition maximizes modularity. In our implementation of the RenEEL scheme, we used a randomized greedy algorithm (Clauset et al., 2004; Newman, 2004) for the base algorithm and typically started with 100 partitions in the ensemble.

To ensure validity of our results, we tested its sensitivity to initial conditions and to incompleteness of the underlying data. We tested the robustness of the clustering results by repeating our analysis multiple times and found that consistently clustered pairs of neurons are orders of magnitude more common than inconsistently clustered pairs in the Hemibrain network (Fig. 1). We also repeated the analysis on a perturbed network from which we excluded synapses that were marked as low-confidence in the data, and found qualitatively similar results to the original network (Fig. 2).

Information measures. Clusters define a partition $C = \{C_1, C_2, ...\}$, and cell types define a partition $T = \{T_1, T_2, ...\}$ of the nodes in the network. By comparing these partitions, we are able to identify potentially biologically relevant networks.

Cluster heterogeneity quantifies the variety of cell types within a cluster, and is measured using Shannon entropy. We defined the heterogeneity of a Cluster C as follows:

$$\text{Het}(C) = -\sum_{T \in \mathcal{T}} \frac{n_{C \cap T}}{n_C} \log_2 \left(\frac{n_{C \cap T}}{n_C} \right) \tag{3}$$

where n_C is the number of cells in Cluster C and $n_{C \cap T}$ is the number of cells of type T in Cluster C, and the sum is taken over cell types T. If a cluster consists entirely of cells of one type, the cell type distribution has zero bits of entropy, while a cluster composed of equal numbers of cells of n different types has $\log_2(n)$ bits of entropy, the maximum for a cluster composed of n types.

Cluster completeness measures how well represented each cell type is in a given cluster. This is the fraction of a cell type present in a cluster averaged across the cell types within that cluster, defined by the following:

$$\text{Com}(C) = \sum_{T \in \mathcal{T}} \frac{n_{C \cap T}}{n_C} \cdot \frac{n_{C \cap T}}{n_T}. \tag{4}$$

Here n_C and $n_{C \cap T}$ are the same as in Equation 3, and n_T is the number of cells of type T in the whole network. A completeness score close
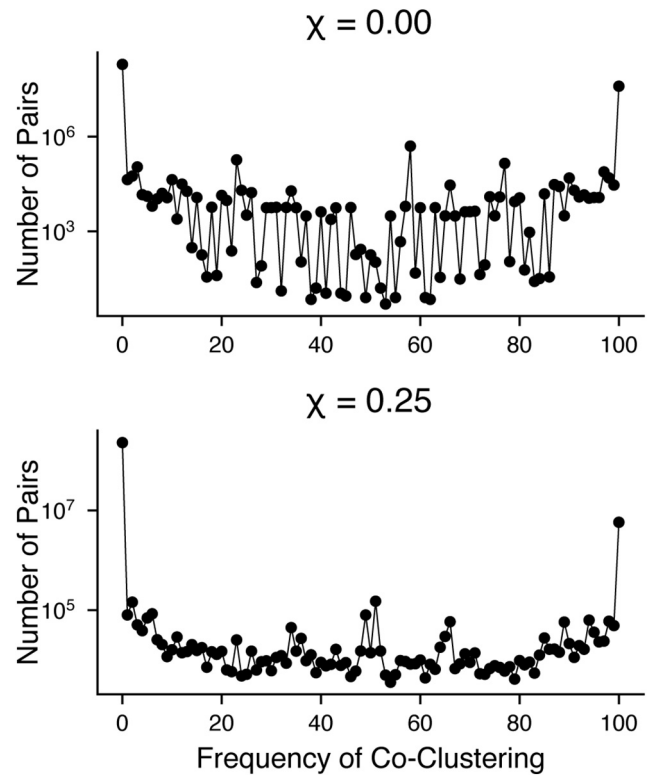

Figure 1. Robustness of clustering. RenEEL was run with a different random seed 100 times; shown here is the number of pairs of neurons (on the *y* axis) that appear in the same cluster a number of times given on the *x* axis. Pairs of neurons that are consistently clustered have a frequency of co-clustering of 0 or 100, meaning that they are never in the same cluster or are always in the same cluster. Consistently clustered pairs of neurons are orders of magnitude more common than inconsistently clustered pairs.

to 1 means that, among cell types represented in a cluster, nearly all of the cells of those types belong to that cluster.

We defined analogous measures for cell types. To identify cell types that may be partitioned into dense subnetworks, we introduce the Fraction of Type measure, defined as follows:

$$\text{FT}(T; \chi) = \sum_{C \in \mathcal{C}(\chi)} \left(\frac{n_{C \cap T}}{n_T} \right)^2. \tag{5}$$

Here $n_{C \cap T}$, and n_T are defined as in Equation 4. This is the fraction of cells of type T that belong to each cluster, averaged over clusters that contain cells of type T. If $\text{FT}(T; \chi)$ is close to 1, then cells of type T form a single dense network at resolution scale χ. Lower values of $\text{FT}(T; \chi)$ indicate that cells of type T are partitioned into multiple densely connected communities.

To identify subnetworks which may represent repeated wiring patterns, we introduce the Fraction of Cluster measure, defined as follows:

$$\text{FC}(T; \chi) = \sum_{C \in \mathcal{C}(\chi)} \frac{n_{C \cap T}}{n_T} \cdot \frac{n_{C \cap T}}{n_C}. \tag{6}$$

This the fraction of each cluster that is composed of cell type T, averaged over clusters which contain cells of type T. A score near 1 indicates that clusters containing cells of type T contain only cells of type T, regardless of how many such clusters there are.

Visualization of volumes. To visualize the relationship between cluster identity and brain regions, we considered the locations of synaptic sites on the neurons. Brain regions annotated in the Hemibrain partition the dataset into disjoint volumes. The presynaptic and postsynaptic sites of each neuron are given coordinates, which lie in one of those disjoint volumes. Thus, we can identify all synaptic sites lying within a given
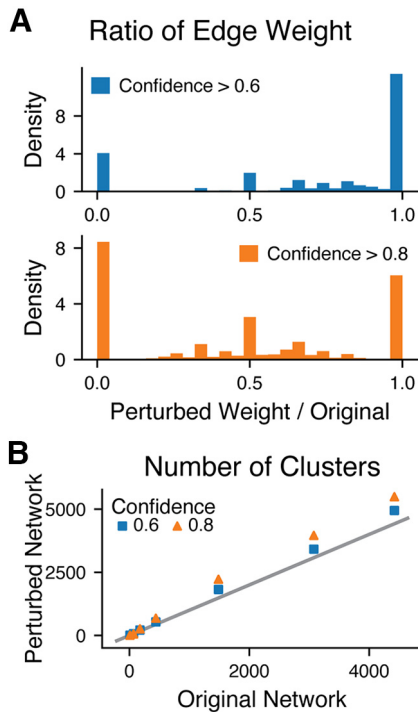
A

Ratio of Edge Weight

B

Number of Clusters

Figure 2. Effects of network perturbation. ***A***, Synapses in the Hemibrain each have a confidence score, indicating the confidence of the machine learning algorithm which automatically identified them. We perturbed the network by excluding synapses with a confidence score below a certain threshold. Each edge in the perturbed network had a weight, which was a fraction of its original weight; shown here is the distribution of these weight ratios. This perturbation resulted in overall weaker edges, with higher thresholds also severing more edges (counted in the bin at 0.0). ***B***, The number of communities found in the perturbed networks compared with the number in the original network. Gray line indicates equality. At higher resolution scales, as the perturbed graph became weakly connected, more clusters were found relative to the original network.

brain region, and associate each to a specific neuron and thus a specific cluster. This is used to compute the fraction of the volume belonging to each cluster as in Figure 5B. The regions are represented as triangulated 3D surfaces. To represent the fraction of the volume belonging to a cluster, we color that fraction of the triangles in the mesh as in Figures 5C and 7A.

Visualization of clusters. To better visualize and compare our clustering results, we used a simulated annealing method to arrange cells in the adjacency matrix plots. In Figure 3B, the clustering results of three different values of parameter χ are visualized to show the evolution of clustering as the parameter increases, and the spontaneous emergence of the hierarchical structure. Specifically, for visualization, we ordered the nodes to minimize a function H of the Euclidean distance d_{ij} between matrix elements (i, j) and the closest point on the diagonal (under periodic boundary conditions). This function takes the following form:

$$H = \sum_{i<j} s_{ij}\, d_{ij}^{\alpha}, \tag{7}$$

where s_{ij} is the link weight between node i and j. We analyzed undirected networks, so the sum is over the elements above the diagonal. The parameter α controls the nonlinear dependence of H on d_{ij}. Figure 3B is generated by setting $\alpha = 3$.

For the order we used, we started with the partition at $\chi = 0.5$, and used simulated annealing to swap pairs of clusters and pairs of nodes within clusters until the order of nodes and clusters minimized the cost function. Fixing this partition, we then repeated this process with the partition at $\chi = 0.25$, now only swapping clusters from the finer partition ($\chi = 0.5$) within the clusters defined by the coarser partition ($\chi = 0.25$). This process was then repeated for the partitions at $\chi = 0.0$ and

$\chi = 0.25$. The final order thus obtained was used for plotting all three plots in Figure 3B.

Software accessibility. A *C* implementation of RenEEL to maximize Q_g (Guo et al., 2023) is available at https://github.com/prameshsingh/generalized-modularity-density.

The Python code and data to generate these figures are available at https://github.com/josiclab/flybrain-clustering.

Many of the figures in this paper are simplified to be printer-friendly, and, necessarily, static. They are available at https://josiclab.github.io/flybrain-clustering/ rendered as interactive plots using javascript so zooming, panning, and mouse-over information is available.

Results

To infer the community organization of a connectome, we treat it as an undirected graph whose nodes are neurons and whose edge weights are defined as the total number of synapses between neurons. By initially treating the connectome as undirected, we obtain a liberal measure of communities, and we subsequently evaluate directed motifs on this undirected scaffold. We discover these communities based on the strength and density of connections between individual cells. Strong connections between neurons may be important for information flow in the network, and highly connected groups of neurons may represent distinct computational circuits. We therefore assume that groups of strongly interacting neurons are likely to form functional units and use a multiresolution community detection method to identify groups of densely connected cells. To validate our clustering approach, we compare our results with previously identified structures in the fly connectome, and apply our method to the connectome of the mushroom body of a *Drosophila* larva.

Many biological networks are hierarchical. Our method is designed to uncover such organization in an unsupervised way, without assuming *a priori* that a network is structured hierarchically. We identify communities in the network by partitioning the network into clusters in a way that maximizes a global measure, the Generalized Modularity Density $Q_g(\chi)$ (Guo et al., 2023). This measure increases with the density of connections within clusters, and depends on a tunable parameter $\chi \geq 0$, which governs the resolution scale of the communities identified. At $\chi = 0$, $Q_g(0)$ is equivalent to the classical Modularity, Q (Newman, 2006), and a relatively small number of large communities are identified (eight in the Hemibrain). As χ increases, maximizing $Q_g(\chi)$ identifies progressively smaller, more densely connected communities (Table 1; see Materials and Methods). Thus, the resolution scale of the community structure within the network varies with χ. The number of communities identified at any particular value of χ is not predetermined but is a result of the optimization. The communities we identify at larger χ are generally subsets of the communities we identify at smaller χ; thus, the community structure is generally hierarchical. To find the partition that maximizes $Q_g(\chi)$, we use RenEEL (Guo et al., 2019), a recently introduced machine-learning algorithmic scheme for graph partitioning that enables fast and accurate results for networks of the size and density of the Hemibrain (Guo et al., 2019). At this scale, without this speed and accuracy of RenEEL, this study would not have been possible.

Modular structure in the larval mushroom body is driven by cell type and anatomy

To validate our method, we applied it first to the connectome of the mushroom body of a larval fruit fly (Eichler et al., 2017). This network consists of 365 neurons and is composed of two
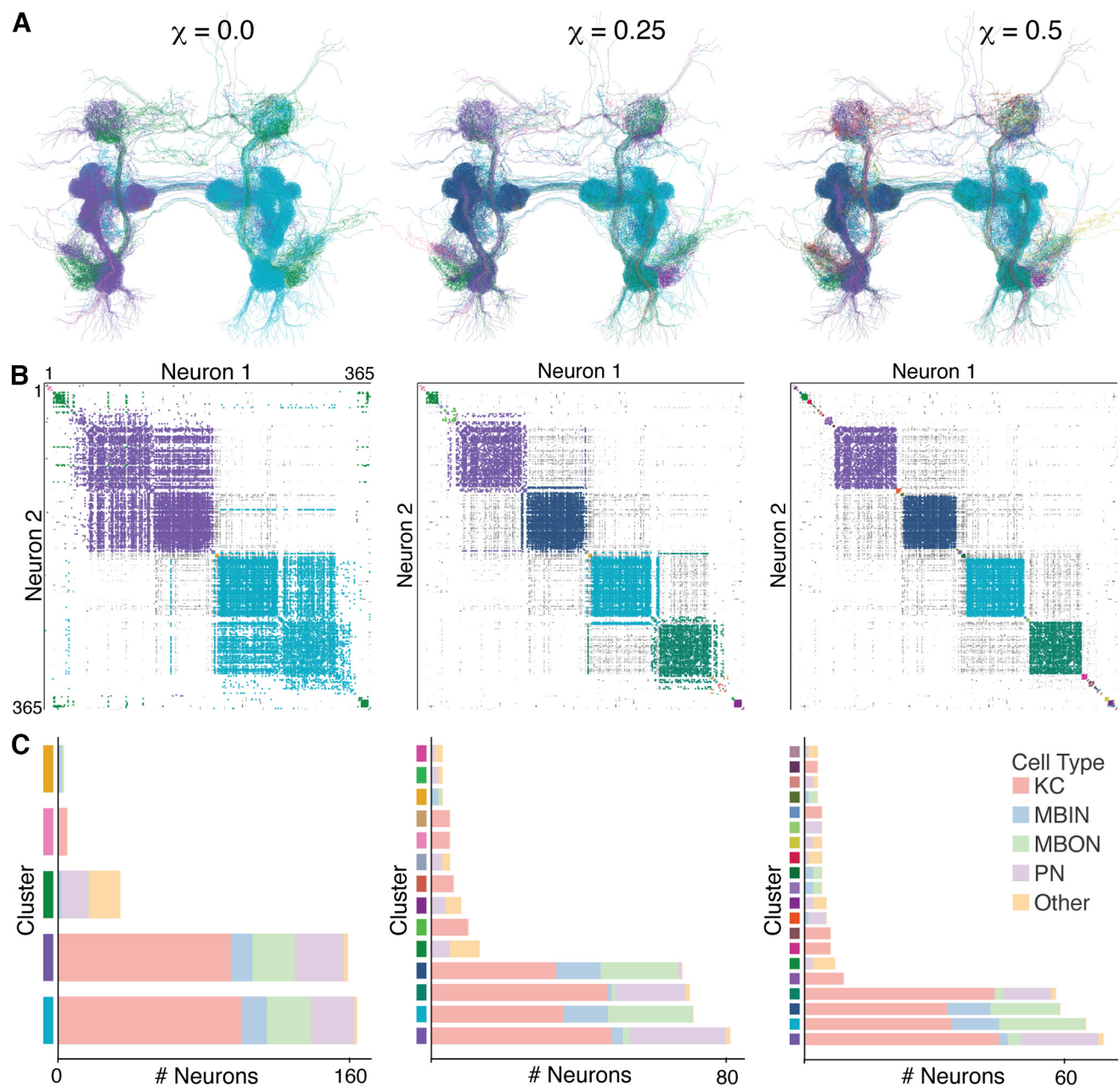
Figure 3. The larval mushroom body network exhibits hierarchical modularity which aligns with anatomy and cell type. **A**, Morphologic reconstructions of the neurons in the larval mushroom body network, colored by cluster membership, viewed from a posterior and slightly dorsal viewpoint. Cluster coloring is preserved down each column. From left to right, clustering was performed by maximizing generalized modularity density, $Q_g(\chi)$, with $\chi = 0.0, 0.25, 0.5$ (see Materials and Methods). The three largest communities are shown in isolation in Figure 4. **B**, Undirected adjacency matrix of 365 neurons in the larval mushroom body. Rows and columns correspond to neurons ordered using a simulated annealing method for cluster visualization. Pixel intensity corresponds to edge weight on a log scale. Colors highlight within-cluster edges. Neuron ordering is preserved across panels to show the hierarchical organization: with increasing χ, larger clusters break into smaller, more tightly linked clusters. **C**, Cell type composition of clusters. Clusters, identified by color bars on the *y* axis, are ordered by size; clusters consisting of 1 or 2 neurons are not shown. Each hemisphere is separated into two primary clusters: with one containing the input/output neurons and the other containing the sensory projection neurons.

symmetric hemispheres with lobes extending in the anterior and dorsal direction (Fig. 3A). At the coarsest level, $\chi = 0$, our method partitioned the mushroom body into the left and right hemispheres (Figs. 3A, 4, purple and blue clusters), as well as a bilateral dorsal-posterior bundle of 34 neurons with projections to both halves (green). In addition, it identified two small clusters: one in the anterior bridge connecting the two hemispheres and one in the left hemisphere. The larger cluster bridging the two halves consists of sensory and other projection neurons that often come in left/right pairs. The small cluster in the anterior bridge consists of a left-right pair of dopaminergic neurons and

one output neuron, with strong internal connections and strong connections to both hemispheres. The five-neuron cluster consists of very young Kenyon cells (KCs) in the left hemisphere with few synaptic connections to each other and to the remaining neurons.

Increasing the resolution by increasing χ reveals a nested hierarchy in the hemispheres that follows anatomy and cell type. Higher values of χ separate each of the two hemispheres into anterior and posterior halves (Fig. 3A,B). These clusters are also composed of distinct cell types. The network we analyzed is composed of 201 KCs, 48 output neurons, 30 input neurons, 66 projection neurons, and 20 cells of various other types (Eichler et al.,

Table 1. Generalized modularity in the Hemibrain network[a]

χ	No. of communities	Mean size	Mean weighted density (IQR)
0.0	8	2716.6	0.60 (0.30-0.64)
0.05	68	699.5	2.15 (0.63-1.60)
0.1	175	277.4	2.70 (0.72-2.99)
0.25	440	66.9	8.16 (1.00-5.00)
0.5	1485	16.1	23.52 (2.00-12.33)
0.75	3076	7.5	27.62 (3.48-20.00)
1.0	4421	5.1	30.35 (4.33-23.00)

[a]Statistics of communities in the Hemibrain network found by maximizing generalized modularity density $Q_g(\chi)$. Increasing χ results in smaller, more tightly-connected communities. Weighed density is the sum of edge weights divided by the number of possible edges within a community; mean and interquartile range (IQR) exclude communities consisting of a single node.

2017). Distinct populations of KCs have different connection probabilities to the other cell types (Eichler et al., 2017), and our method assigns the input and output neurons to one community and the projection neurons to another, with KCs split between those two communities based on strength of connectivity (Fig. 3C). As χ increases, a large community persists in each hemisphere composed of KCs, input neurons, and output neurons, reflecting a repeated tightly connected microcircuit (Eichler et al., 2017) (Fig. 3C). These results highlight the contrast between modularity maximization methods and spectral embedding methods of graph clustering. The latter groups neurons based on common connectivity properties; in the larval mushroom body, the clusters found this way are generally composed of a single cell type (Chung et al., 2021).

At the coarsest scale, our method thus identifies the clear, bilateral structure of the network. However, it also identifies two less obvious bilateral structures and isolates several very weakly connected KCs. At finer resolutions, the structure separates into anatomically meaningful subsets with distinct cell type compositions, grouping sensory projection neurons separately from modulatory input and output neurons. The inferred organization is approximately hierarchical; at successively finer resolution, the communities we find are typically nested (Fig. 3B). The agreement between the communities we detected and previously identified anatomic and cell type organization (Eichler et al., 2017) provides support for our method.

We next turn to the Hemibrain connectome, which is two orders of magnitude larger in size.

In the hemibrain, communities at the largest scale correspond to anatomically identified structures

At the coarsest resolution, $\chi = 0$, we identified eight communities in the Hemibrain. Some of these communities correspond to individual anatomic structures in the brain, while others span several brain regions (Fig. 5A).

The fly brain is composed of many interconnected neuropils, which have been further partitioned into ROIs over the last decades by expert anatomists (K. Ito et al., 2014). In Figure 5B, we compare the communities we identified with the previously identified brain regions in the Hemibrain. Box height is proportional to the fraction of the brain region contained in the cluster, while width is proportional to the number of neurons synapsing in that region. We find that clusters we identified make or receive synaptic connections in restricted subsets of previously identified regions, and many of those regions are primarily contained in one cluster.

Several anatomic structures in the brain correspond to an entire community identified at this coarsest scale (Fig. 5C). For example, the mushroom body, a structure that plays a role in olfaction and memory (Turner et al., 2008), forms a single cluster

(Cluster 5). The FB, which plays a role in spatial navigation and modulation of internal states (Hulse et al., 2021), likewise forms a single cluster (Cluster 2). Other structures jointly form a larger cluster, such as the right lateral horn, antennal lobe, and superior lateral protocerebrum, which form a cluster with connections to other lateral structures on the right side of the Hemibrain (Cluster 4). This is a primary sensory pathway by which olfactory information is transmitted from the antennae to the central brain (Jefferis and Hummel, 2006; Vosshall and Stocker, 2007; Pereanu et al., 2010). The superior lateral protocerebrum is part of the superior neuropils, which are split into two clusters (Clusters 3 and 4). The right superior intermediate and medial protocerebra cluster together with their left pairs in Cluster 3. This cluster makes bilateral connections to the left and right superior and inferior neuropils. Cluster 7 consists of the anterior visual pathway, which transmits polarized light information from the optic lobe to the ellipsoid body (EB) (Otsuna et al., 2014; Omoto et al., 2017), as well as the EB itself. The protocerebral bridge (PB), which connects the ellipsoid body and the FB, is split between Clusters 7 and 2, with the greater portion grouped with the EB in Cluster 7. The remaining ventral neuropils and optic lobe are mostly split between Clusters 1 and 6. The lobula plate, a structure in the optic lobe that is not fully contained in the Hemibrain volume, forms a cluster consisting of just 88 cells (Cluster 8).

Much of the network topology of the fly brain connectome is closely tied to its spatial topography (M. Ito et al., 2013). However, the community structure we find is not simply a reflection of spatial proximity. For example, the anterior and posterior visual pathways (parts of Clusters 7 and 6, respectively) span nearly the whole width of the Hemibrain volume. Likewise, Clusters 3, 4, and 5 (making up the superior neuropils and the mushroom body), while spatially colocalized, reflect the distinct roles of their constituent structures in the network: the superior intermediate and medial neuropils form a pathway connecting the FB to the mushroom body (Li et al., 2020; Hulse et al., 2021), while the superior lateral protocerebrum is part of a network connecting the antennal lobe to the mushroom body.

Thus, at the coarsest level of resolution, our algorithm automatically identified well-known anatomic subunits of the fly brain. However, it also placed some anatomic structures with distinct but related functions into the same cluster, suggesting that these structures are tightly linked. We next show that increasing the partitioning resolution can resolve the finer structure of the fly brain, and reveal the hierarchical organization of communities.

Generalized modularity density reveals the hierarchical organization of the fly brain

Clustering is a form of dimension reduction for network data that allows us to project a larger network onto a network of clusters and the connections between them. The reduced graph of the clusters we found at the coarsest level, $\chi = 0$, reveals high-level organization of the network. In Figure 6A, each node represents one of the eight clusters, with the size of a node indicating the number of neurons in the corresponding cluster. The thickness of the edges is proportional to the weighted edge density between clusters (black) and within clusters (gray). Weighted density is how the measure of connection strength is defined for the Generalized Modularity Density, $Q_g(\chi)$, and we see the within-cluster connections are an order of magnitude stronger than between-cluster connections. Indeed, if we used the same edge scale for both within and intercluster connections, the latter would be nearly invisible.

We reran our RenEEL community detection algorithm using several values of χ to find smaller, more densely connected
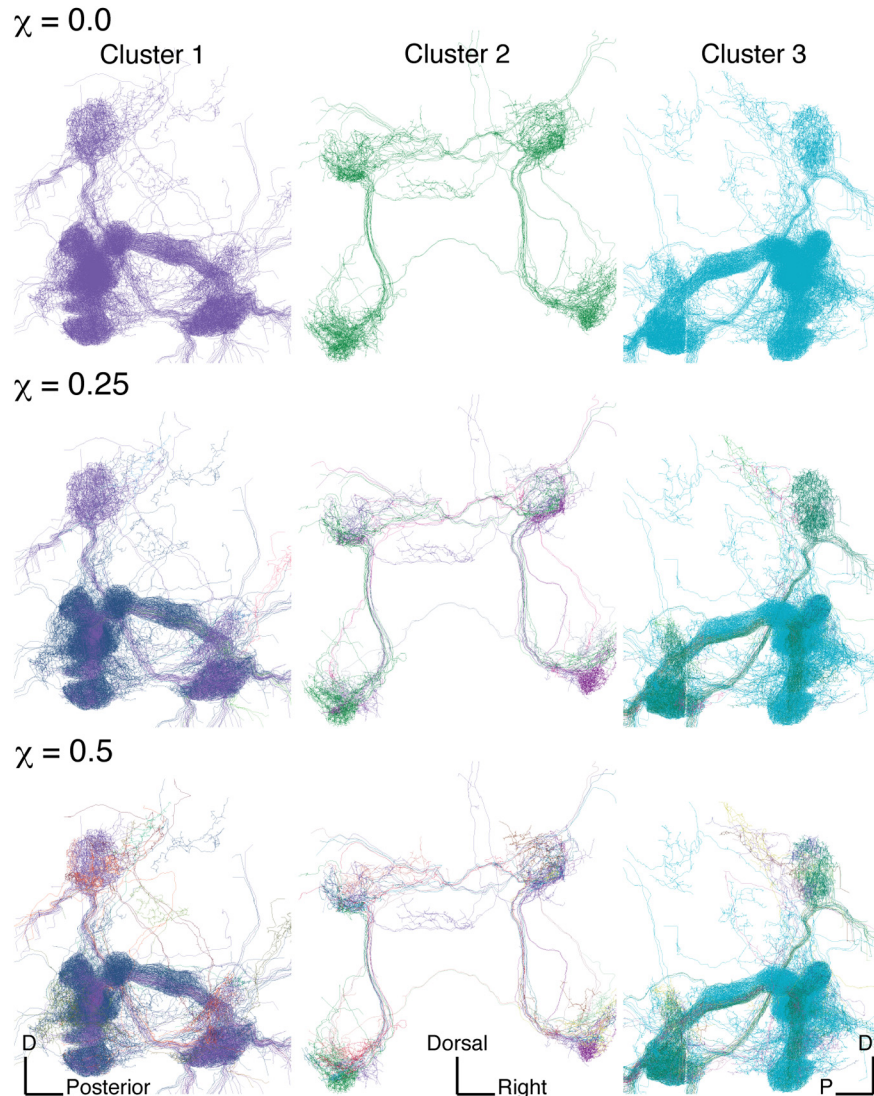
Figure 4. The three primary clusters of the larval mushroom body, shown in isolation. Left column, Cluster 1 in a lateral view from the left. Middle column, Cluster 2 from the same posterior view as in Figure 3. Right column, Cluster 3 in a lateral view from the right.

clusters. The clusters identified using the classical modularity score, $Q = Q_g(0)$, roughly broke into smaller clusters as we increased χ, revealing more specialized subclusters. While increasing χ generally resulted in a refinement of the partition (i.e., a community $C_i(\chi)$ at a larger value of χ is usually a proper subset of some community $C_j(\chi')$ at a smaller value, $\chi' < \chi$), some clusters appearing at higher values of χ are composed of subnetworks from two or more clusters at lower values of χ. For example, 3796 (87.3%) of the 4345 neurons in the largest cluster found using $\chi = 0.05$ (Fig. 6B, bottom right, large purple node) belong to Cluster 1 (Fig. 6A, purple node), while 490 neurons (11.3%) belong to Cluster 6 (red). However, the vast majority of clusters were broken into smaller subclusters: on average, clusters identified at resolutions $\chi > 0$ were at least 85% contained in one of the eight clusters found with $\chi = 0$ (Fig. 6F).

Thus, increasing the resolution of the modularity measure allows us to identify subclusters that may have important functions in the network. For instance, Figure 6A shows the strong within-community connections in the anterior visual pathway-EB-PB circuit (represented by the thick gray loop on the brown node) and its strong connection to the FB (the thick black edge

connecting brown and green nodes). This circuit splits into two stages in Figure 6B (brown nodes). The first stage consists of the inputs from the optic lobe and is very weakly connected to the FB, while the second, more densely connected cluster, contains the EB and PB and retains the strong connection to the FB. Increasing χ resolves these clusters further, revealing the modular organization of the visual inputs and of the FB. The visual inputs are organized into a densely connected network of clusters which feed into a single densely connected cluster, which in turn connects strongly to three of the four subclusters of the FB (Fig. 6C,D). The clusters in the FB correspond to distinct anatomic layers, which we discuss further in the next section. We discuss the inputs to the visual pathway further in Clustering Reveals Cell type-specific Wiring Patterns.

Our approach shows that the fly brain is composed of hierarchically organized communities. The function of many of these communities, and the reason for this structure, is not completely understood. However, the inferred structure suggests a functional role for this organization that could be tested in follow-up experiments. We next take a closer look at the organization of the FB revealed by our method.
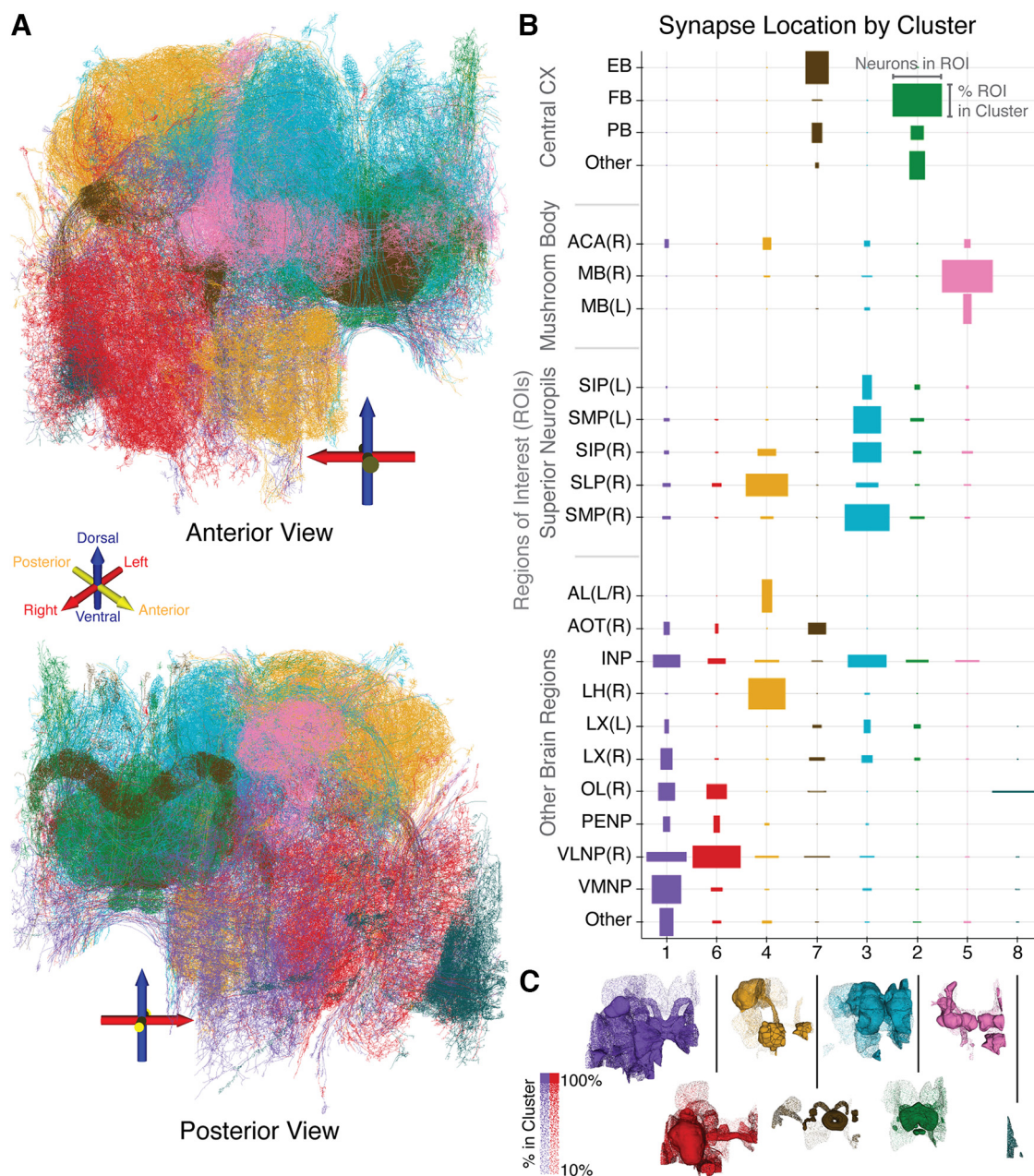
Figure 5. Overview of modularity and anatomy in the Hemibrain. At the coarsest scale modularity, $Q_g(0)$ is maximized by partitioning the Hemibrain network into 8 clusters. **A**, Morphologic rendering of neurons color-coded by cluster identity. Shown is a sample of 120 neurons from each cluster (except Cluster 8, which consists of only 88 neurons). **B**, Comparison between community structure and anatomy. Brain regions identified by anatomists partition the Hemibrain into disjoint volumes, listed on the y axis. Names are those assigned in the Hamibrain dataset, with L/R specifying left and right, respectively. Box height is the fraction of the volume on the y axis contained in the cluster on the x axis. The heights in each row sum to 1. The width of a box corresponds to the fraction of neurons in the cluster with synapses in the brain region given on the y axis. Individual neurons may have synapses in many brain regions, so the widths in one column may sum to >1. **C**, Volumetric renderings of the brain regions contained in each cluster, with shading indicating only partial containment.

Clustering automatically reveals layering in the fan-shaped body

Having identified the FB as a single community at $\chi = 0$, we next investigated the finer structure of this community by increasing χ. The FB splits into several subclusters, which correspond to previously identified anatomic layers.

At the coarsest resolution, the FB comprises a single cluster (Cluster 2, Fig. 5C). This cluster consists of 2391 cells, of which 2315 (97%) have synapses in the FB; this represents 90% of the 2570 cells in the Hemibrain volume with synapses in the FB. This cluster remained coherent at higher values of χ: Clusters containing at least one of these 2391 cells found using $\chi > 0$

were, on average, 95% contained in the original cluster (Fig. 6F). The subclusters we found are arranged from dorsal to ventral, with increasing resolution producing finer layering.

The FB is known to have a layered structure (Hanesch et al., 1989; Lin et al., 2013; Wolff et al., 2015), with different layers playing different functional roles in the brain (Donlea et al., 2011; Lin et al., 2013; Hulse et al., 2021; Kato et al., 2022). The exact number of reported layers in the FB varies, so we compared our results with the nine layers identified in the Hemibrain dataset (Scheffer et al., 2020). Increasing χ from 0 to 0.1 split the FB into four communities, which roughly correspond to layers 1, 2, 3-6, and 7-9. This is shown in Figure 7A by rendering each
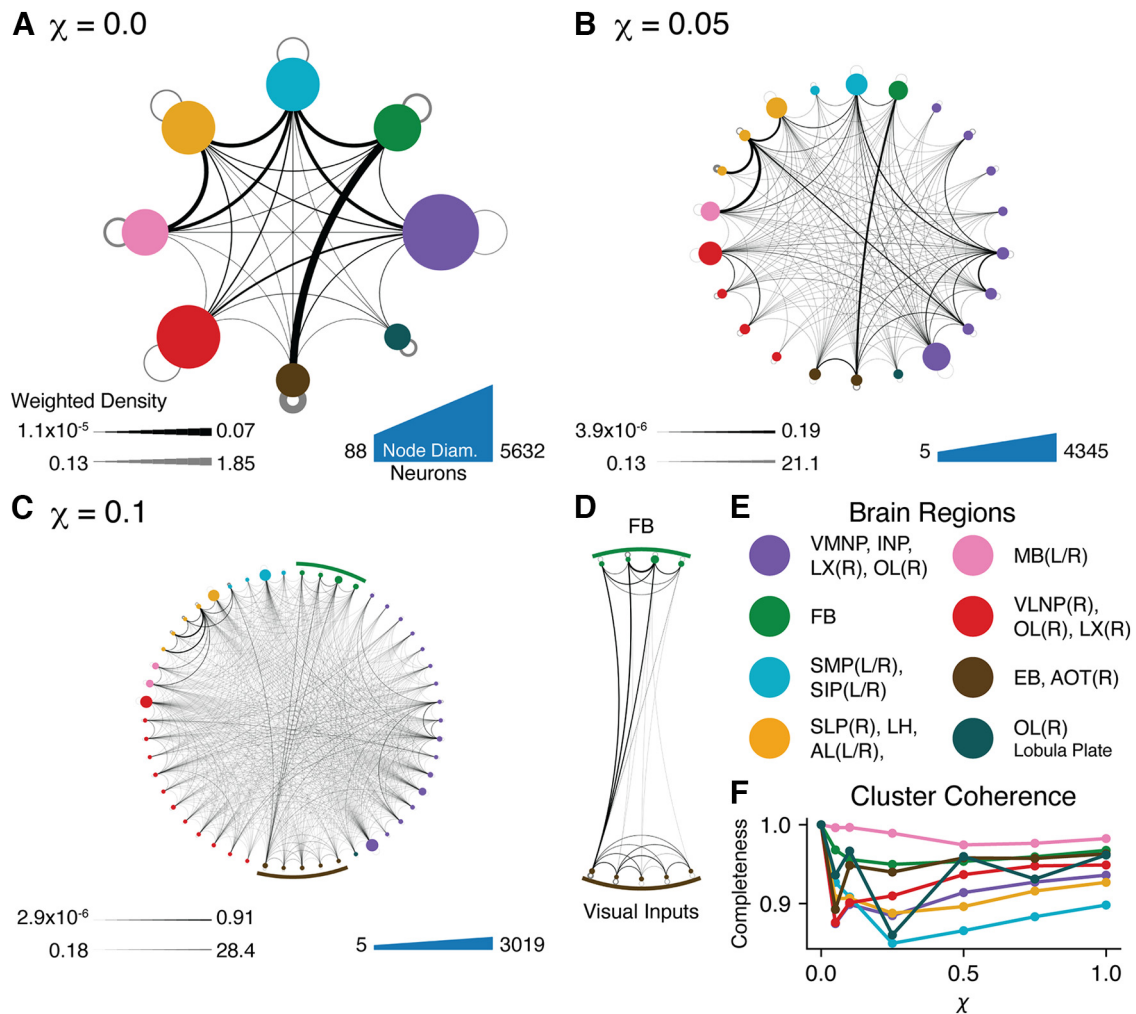
Figure 6. Reduced graphs obtained by representing communities by a single node, and collapsing all edges between nodes in different communities onto a single, weighted edge. Communities with <5 neurons are not shown. We reran the clustering algorithm with $\chi = 0.0, 0.05, 0.1$, producing results shown in ***A***, ***B***, and ***C***, respectively. Thickness of edges corresponds to weighted edge density, defined as the total weight of the edges connecting nodes in one cluster to nodes in another, divided by the number of possible edges between the two clusters. Gray border loops represent the connections between nodes within the cluster. The strength of within-community connections is orders of magnitude stronger than between communities. ***D***, The subclusters of the anterior visual inputs, ellipsoid body, and FB are shown in isolation to show the organization of the circuit connecting visual inputs to the ellipsoid body and the FB. ***E***, Brain regions represented by the different colors in this figure. ***F***, The coherence of each of the 8 clusters in ***A***. Coherence of Cluster C is defined as the fraction of each Cluster C′ found with $\chi' > 0$ contained in ***C***, averaged over the C′ which have nonempty intersection with C. If the clustering found by increasing χ was strictly hierarchical, all values would equal 1.0.

volume as a triangulated mesh, with the proportion of triangles of each color representing the partition of the synapses in the volume (see Materials and Methods). Increasing χ further separated these clusters along the dorsal-ventral axis. At $\chi = 0.5$, the FB separated into seven layers, combining the Hemibrain's layers 4 and 5 and layers 7 and 8 into single communities.

The strong separation between the dorsal and ventral halves may reflect their different functional roles: the ventral layers play a role in navigation, while the dorsal layers modulate arousal (Liu et al., 2006; Donlea et al., 2011; Kato et al., 2022). The parameter χ controls the sensitivity of the clustering to within-cluster density, so our findings show that the anatomically defined layers are also densely connected as networks. However, the individual layers themselves do not form single network communities. Rather, each layer is composed of a mix of densely connected networks that reside primarily in that layer (Fig. 7B).

Thus, our method automatically discovered the layered structure of the FB in an unsupervised way. Our results suggest that the layers are hierarchically organized: At the coarsest scale, the FB is split into four layers, which are further subdivided at

increasing resolution. Some of the clusters we identified relate to the columnar structure of the FB, which we revisit below. We next ask how the structures identified by our algorithm relate to cell types.

Common cell types form densely connected clusters
Understanding the importance of the identified structures and relationships between them becomes difficult as the number of clusters grows. In order to identify meaningful finer-scale networks, we combined our clustering results with cell type data attached to each node in the Hemibrain.

Most neurons in the Hemibrain (>90%) were previously assigned cell types based on morphology and brain region connectivity (Scheffer et al., 2020). In contrast, the clusters we found were defined using network connectivity alone, without the use of cell type information. We thus used the partitioning of cells by type in conjunction with clustering to uncover potential cell type-specific wiring principles. Unlikely conjunctions of cell types, such as clusters with a wide variety of cell types, or clusters containing all instances of a given cell type, indicate structures
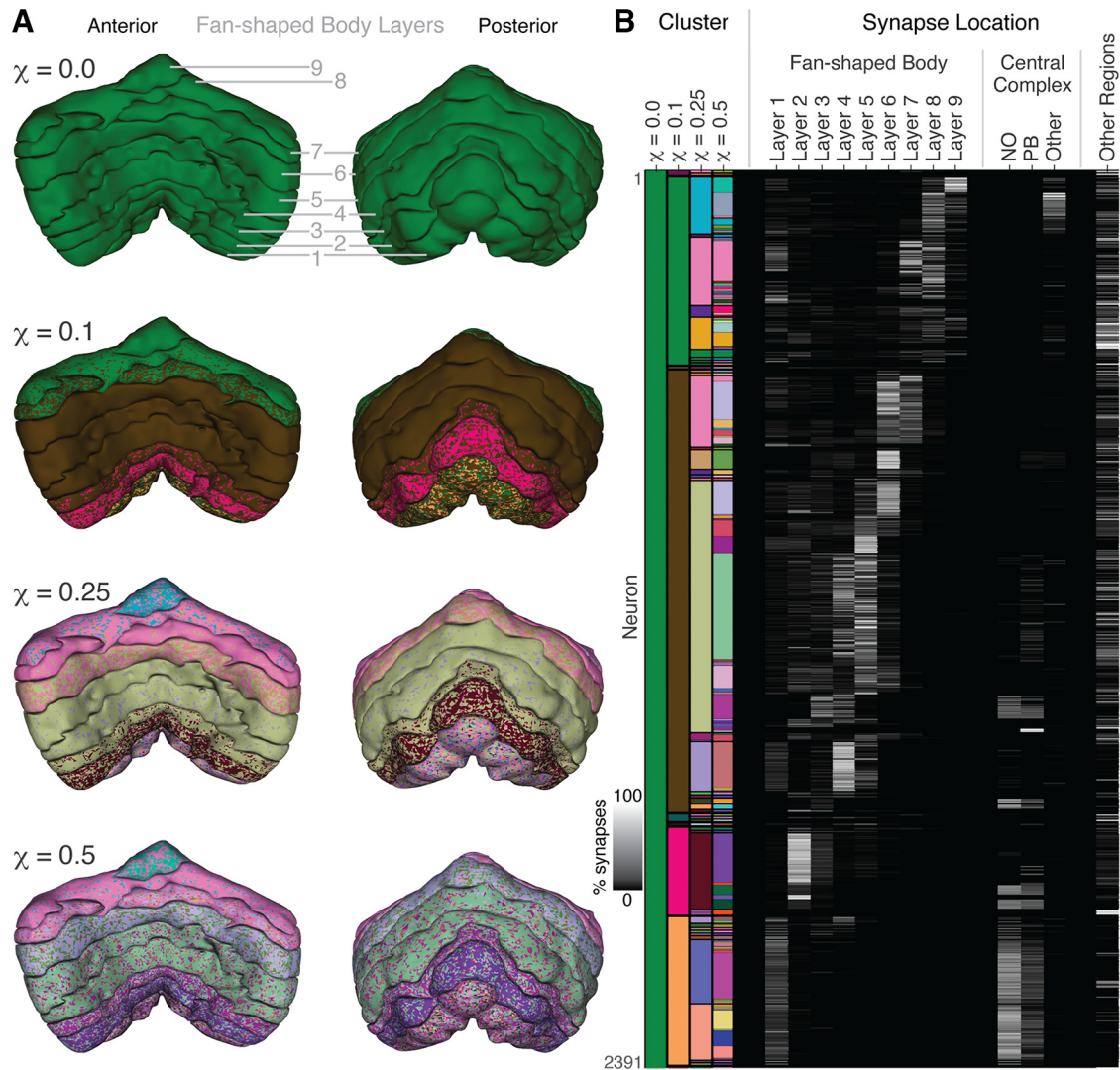
Figure 7. The layered, hierarchical structure of the FB is automatically revealed by varying the tuning parameter, χ. ***A***, The FB is partitioned into nine layers in the Hemibrain dataset, rendered here as 3D volumes. Each volume is rendered as a triangulated 3D mesh. The proportion of triangles of each color represents the cluster membership of the synapses in the volume (see Materials and Methods). Left column, Anterior view. Right column, Posterior view. At higher resolutions ($\chi > 0.1$), most layers are composed of multiple clusters, revealing structure within layers. This structure is reflected in the multicolor texture of each layer. Each of these individual clusters resides primarily in one or two layers, so different layers have different mixes of colors. ***B***, Synapse locations of the neurons in Cluster 2 ($\chi = 0$). Neurons on the *y* axis are sorted according to cluster membership at increasing values of χ; clusters are indicated by colors in the left columns labeled with values of χ. Repeated colors within one column represent the same cluster. Most clusters are nested hierarchically, so there are few repeated colors within each column. The remaining columns show neurons' synapse locations. For example, the dark pink cluster at $\chi = 0.1$ (near the bottom of the figure) resides primarily in layer 2, while some of its neurons have processes extending to the Noduli (NO) and PB.

that deserve further examination. A cluster with a wide variety of cell types could represent a functional circuit composed of diverse, but strongly interacting neurons. Alternatively, a very homogeneous cluster could reveal cell type-specific wiring principles. We first discuss the cell type distributions within larger clusters, identifying key functional circuits, and then consider how cells of particular types are partitioned by the clustering algorithm.

To quantify the diversity of cell types that constitute a cluster we used cluster heterogeneity, measured using Shannon entropy (for details, see Materials and Methods). Smaller clusters typically have lower entropy than larger clusters as they can contain fewer cell types (Fig. 8A). Increasing χ partitions the network into smaller clusters on average, and the average entropy across clusters decreases. However, we found that larger clusters tended to have significantly lower entropy compared with shuffled data (Fig. 8A, left column), obtained by permuting the cell types

between the cells. The largest clusters we found at higher values of χ have only a few bits of entropy, so they must be composed of many cells belonging to only a few cell types. This suggests that neuron types that are prevalent in the brain (i.e., types with tens or hundreds of exemplars) are most densely connected with other common cell types, much more so than would be seen in a randomly labeled network. Therefore, clusters are more uniform in their composition than expected by chance, with some being composed almost completely of one or a few cell types.

Cluster heterogeneity does not reveal the degree of homophily among the different cell types, that is, whether common types of cells are most densely connected to other cells of the same type, or whether they tend to share connections with different type cells. To answer this question, we introduced cell type completeness of a cluster, defined as the fraction of cells of a given type that belong to a cluster, averaged over the cell types in that cluster (for details, see Materials and Methods). A cluster with a
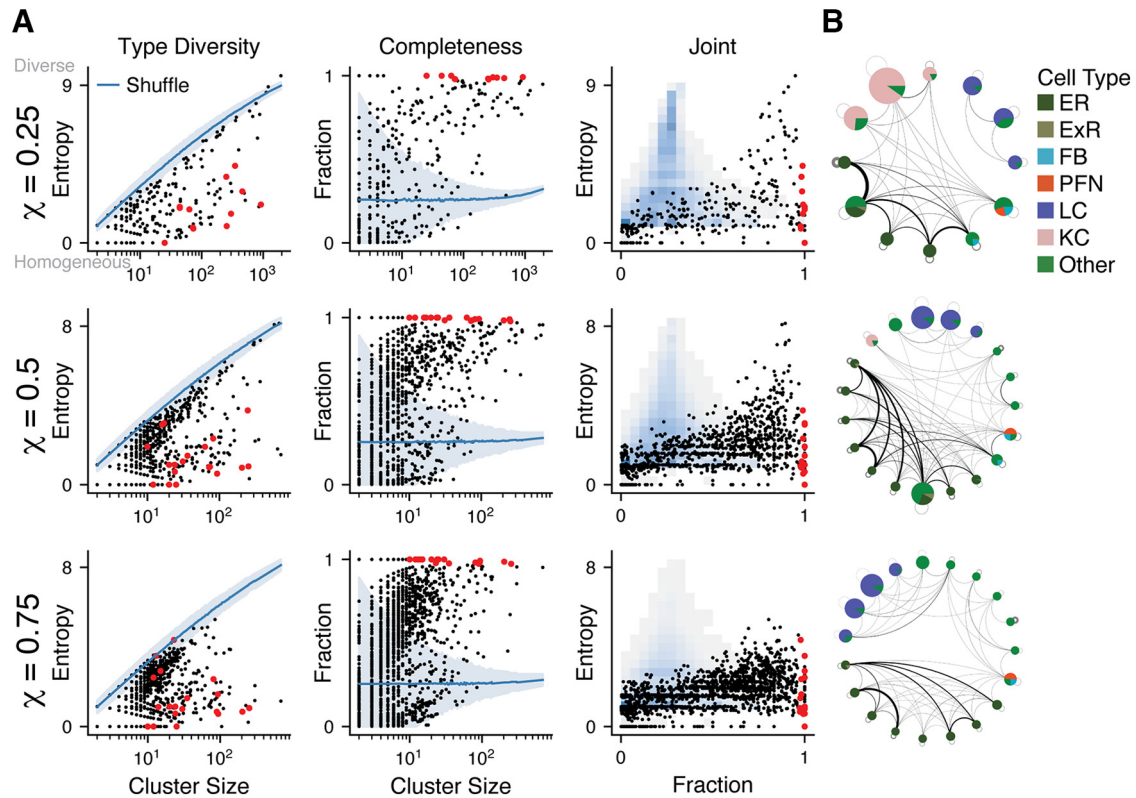
Figure 8. Cell type composition of clusters identified by maximizing $Q_g(\chi)$. ***A***, Heterogeneity and completeness of cell types in clusters. Each row shows clusters identified using a different value of χ. Left column, Heterogeneity (entropy) of cell type distributions within each cluster. Each black dot represents a cluster found using our method, with the corresponding cell type entropy plotted against cluster size. Large red dots indicate clusters shown in networks in ***B***. Blue line indicates the average entropy of clusters in the shuffled data ($n = 100$ shuffles), with the band indicating 3 SDs about the mean. Middle column, Cell type completion within each cluster. Right column, Heterogeneity against completeness for the same data. Black dots correspond to the actual data. Shaded squares represent a 2D histogram obtained from shuffled data. ***B***, Reduced networks of clusters with at least 97% cell type completion and at least 10 neurons. Node size is proportional to the number of neurons in the cluster. At different resolutions, ER cells and LC cells remain completely clustered.

completeness score of 1 may be composed of multiple cell types, but it contains all cells of those types. For each value of $\chi > 0$, we found an average completeness score of 0.4 (the 8 clusters found at $\chi = 0$ have an average completeness score of 0.8). Larger clusters tend to have higher completeness scores; excluding clusters with <10 neurons raises the mean completeness score to the range 0.5-0.7 and clusters with at least 100 neurons have a mean completeness score of 0.8-0.9 for $\chi > 0$ (Fig. 8A, middle column, Fig. 9). In other words, certain common cell types form strongly connected communities and these are automatically identified by our clustering method.

We found high-completeness clusters distributed throughout the brain, comprising a variety of cell types. The clusters with the highest completeness scores reside primarily in the mushroom body, in the inputs from the visual system, and in the central complex. The two most common cell types in the dataset are lobula columnar cells (LCs) and KCs, which combined represent 18.5% of the neurons in the Hemibrain. LCs form the bulk of the portion of the optic lobe in the reconstructed volume of the Hemibrain. There are many subtypes (Otsuna and Ito, 2006; Panser et al., 2016), the most prevalent of which are grouped together by our method across multiple resolution scales (Fig. 8B). KCs form the bulk of the mushroom body and the majority of the high-completeness clusters in that region. They are densely, but weakly connected (Takemura et al., 2017a), and thus do not completely cluster together at higher values of χ (Fig. 8B).

In the central complex, many of the ER cells which compose the head direction circuit in the ellipsoid body form cliques, all-
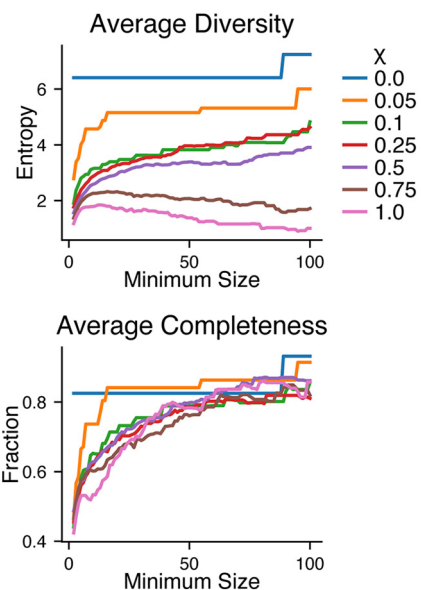
Figure 9. Cluster composition dependence on cluster size. Cell type diversity and completeness (see Materials and Methods), averaged across clusters. Shown here is the dependence of this average on the minimum size of clusters included. Cluster diversity decreases as χ increases, which reflects the smaller cluster size overall. Cluster completeness remains fairly constant regardless of χ. Excluding small clusters raises the average completeness.

to-all connections with cells of the same type (Scheffer, 2020; Scheffer et al., 2020; Fisher, 2022). These are automatically identified by our method (Fig. 8B, dark green). At the coarser resolutions, we find that nearly the entire central complex forms a small number of high-completeness clusters while at the finest scale, the communities we identified are composed of one or two types of ring cells (Fig. 8B, bottom).

Cluster heterogeneity and cluster completeness are complementary measures that summarize the cell type composition of clusters. Heterogeneity measures the diversity of cell types within the cluster, while completeness measures how strongly each type is represented. Plotting cluster heterogeneity against completeness organizes clusters into four quadrants allowing us to identify functional networks in a semiautomated fashion (Fig. 8A, right column). In the top right (high heterogeneity, high completeness) are clusters that are composed of nearly all cells of multiple types, potentially indicating important functional circuits. Clusters in the bottom right (low heterogeneity, high completeness) are communities composed primarily of a single cell type. In the bottom left (low heterogeneity, low completeness) are tightly linked subnetworks of a single cell type, or potentially cell subtypes. In the top left (high heterogeneity, low completeness) is noise — only the shuffled data appear here. Interesting high-completeness clusters are easiest to identify with this framework, as these are the clusters composed of nearly all the cells of one or more types. Clusters in the bottom left (low heterogeneity, low completeness) require a slightly different analysis, which we turn to next.

Clustering reveals cell type-specific wiring patterns
By definition, a high-completeness cluster contains most cells of certain types in the brain. However, the abundance of low-completeness clusters suggests that many cell types form multiple densely connected subclusters which may or may not be homogeneous. This may be a result of fine-scale organization within the cells of one type, such as submodules composed of cells of a single type, or repeated wiring patterns that could be revealed by similarly structured clusters composed of multiple cell types. Our approach can reveal these structures by partitioning cells of a given type into separate clusters with increasing values of χ. These separate clusters would have low completeness scores, as they would contain only a fraction of the cells of one type. Moreover, such clusters might only become apparent at high resolution when clusters are smaller and more densely connected. Thus, we sought to identify cell type-specific wiring patterns by examining the partitioning of cell types into clusters and the dependence of such partitioning on the resolution scale.

Such identification of wiring patterns requires us to quantify the cluster composition of cell types. That is, we are interested in finding multiple clusters that share cells of a given type to understand the wiring patterns of that cell type. We do so by leveraging the multiresolution nature of our modularity density measure, considering the progressive partitioning found by increasing the control parameter χ.

To this end, we considered measures analogous to homogeneity and completeness, now defined for cell types rather than clusters, and examined their dependence on χ. For a given cell type, the Fraction of Type is what fraction of that type is represented in each cluster, averaged over clusters which contain cells of that type (see Materials and Methods). For low values of χ, where the network is partitioned into a few large clusters, most cell types have a Fraction of Type score close to 1; for each cell type, all cells of that type are grouped together in one of the large clusters.

If Fraction of Type remains close to 1 across multiple values of χ, this suggests cells of that type form a single densely connected network that may or may not include cells of other types. Conversely, a decreasing Fraction of Type score indicates a cell type that is partitioned into densely connected subnetworks.

The complementary Fraction of Cluster score measures what fraction of each cluster is composed of cells of a given type, averaged over clusters which contain cells of that type. This is analogous to the completeness score for clusters, with the role of cell type and cluster identity reversed (see Materials and Methods). If Fraction of Cluster is close to 1, then all clusters that contain cells of a given type are predominantly composed of cells of that type. A low Fraction of Cluster score indicates that clusters containing cells of a given type also contain many cells of other types. For a fixed value of χ, Fraction of Cluster is not easily interpretable. However, considering how this score changes with varying χ, and moreover how Fraction of Type and Fraction of Cluster jointly vary, may reveal biologically relevant network structure.

As χ varies, the Fraction of Type and Fraction of Cluster vary, tracing out a curve (Fig. 10A). For most cell types, this curve starts in the top left corner (high Fraction of Type, low Fraction of Cluster) and moves to the right and down as χ increases (Fig. 11). Motion purely to the right indicates that our method is identifying progressively smaller networks containing all cells of the given type. Motion purely downward indicates a certain regularity of connectivity: While the cells of that type are partitioned into disjoint networks by our clustering method, the average composition of those networks remains unchanged. A cell type curve which moves purely rightward then purely downward thus might identify a wiring pattern involving multiple cell types. Computing the area swept out by the curve (Fig. 10A, gray shading) provides a simple heuristic for identifying cell types that are possibly hierarchically organized in this way. Many of the cell types with the largest swept area values are those that form high-completeness clusters. We therefore focus on the cell types that tend to be part of low-completeness clusters, that is, those types that are partitioned into subsets by the clustering algorithm.

Using this area score heuristic, hΔK and PFGs neurons stood out as cell types which potentially partition into type-specific wiring patterns (Fig. 10A). The FB has a grid-like layout with distinct layers and columns (Hanesch et al., 1989; Lin et al., 2013; Hulse et al., 2021). At a moderate resolution scale ($\chi = 0.5$), we find a single cluster consisting of 31 hΔK, 18 PFGs, 6 FB6A, 2 FB6D, 4 FB6M, and 2 ExR3 cells, which represents all cells of those types in the Hemibrain. For higher values of χ, the clusters containing the hΔK and PFG cells tile the middle layers of the FB, with each cluster composed of a small number of cells which arborize in two contralateral columns (Fig. 10B). PFG cells connect the PB and the FB; each PFG cell arborizes in one glomerulus of the PB and one dorsal columnar patch of the FB (Hulse et al., 2021). Each hΔK cell arborizes in two contralateral columns of the FB (Fig. 10D). Together, these two cell types form a roughly bipartite network with sparse, weak connectivity between cells of the same type (Fig. 10C). At the finest resolution scale we examined, corresponding to $\chi = 1$, each of these clusters consists of at most two PFG cells and at most three hΔK cells per PFG cell. One cluster contains an additional four neurons, the two ExR3 neurons and two of the FB6A neurons, which are strongly, reciprocally connected with all hΔK cells and all PFG cells (Fig. 10F). The PFG-hΔK-ExR3 circuit and its putative role in regulating sleep were previously described in Hulse et al. (2021). Our automated method recovers this
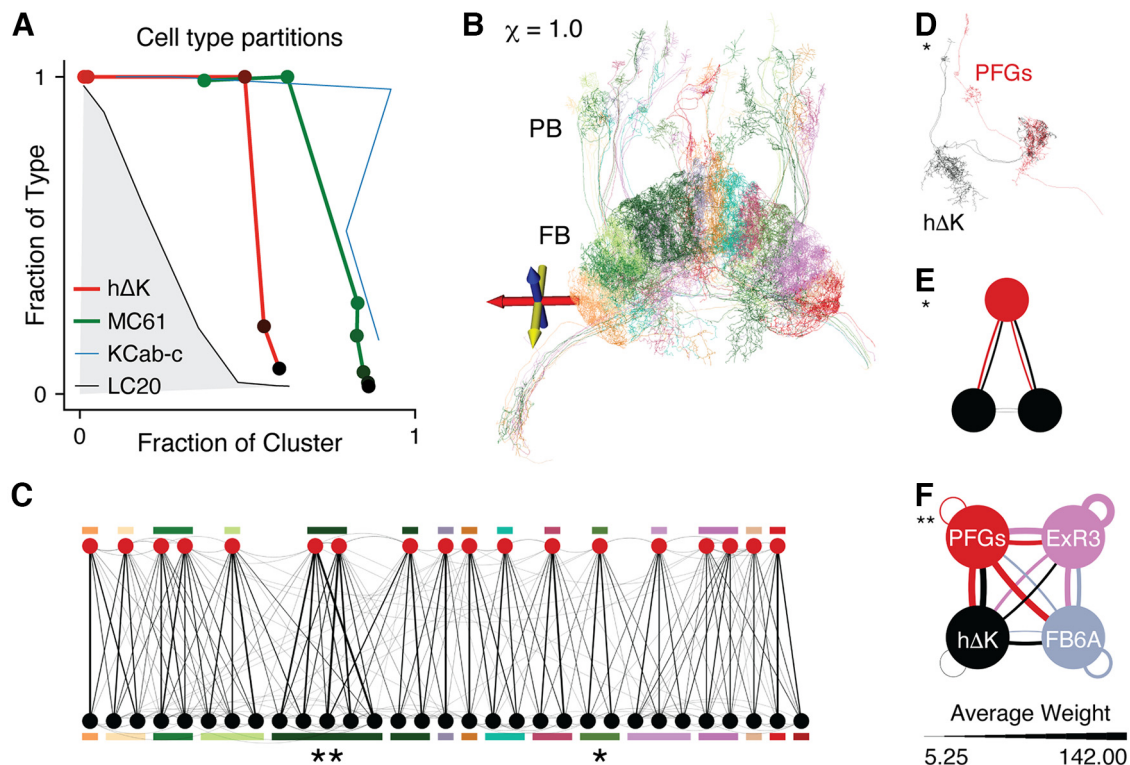
Figure 10. Algorithmically identified type-specific wiring patterns. **A**, For each cell type, the communities we find partition the cells of that type. This partition is summarized by the Fraction of Type and Fraction of Cluster scores. As the resolution scale parameter, χ, varies, each cell type traces out a curve in the Fraction of Type (y axis) versus Fraction of Cluster (x axis) plane. Shown here are these curves for four cell types: the cell type which sweeps out the largest area (KCab-c), a randomly selected type (LC20), and two highlighted curves which indicate the possible presence of a repeated type-specific wiring pattern (h$\triangle$K, MC61). Curves for all cell types with at least 10 exemplars are shown in Figure 11. Gray shading represents how the area score is computed for LC20. **B**, Clustering at $\chi =$ 1.0 reveals the columnar structure of the FB. Pictured are all clusters containing h$\triangle$K cells, which includes all h$\triangle$K neurons and all PFGs neurons. PFG cells connect the PB to the FB, while h$\triangle$K cells are local to the FB. Not shown are five cells which are part of these clusters (a left-right pair of ExR neurons and three FB neurons), but would otherwise obscure the figure. **C**, The undirected network of h$\triangle$K neurons and PFGs neurons in these clusters. The network is roughly bipartite, with strong connections between cells of different types, but only sparse and weak connections among cells of the same type. Each cluster is composed of one or two PFG cells and several h$\triangle$K cells. The 18 PFG cells (top row) are ordered according to their spatial arrangement in the FB. **D**, Morphologic rendering of the cells in the cluster marked *, colored by cell type. **E**, Directed network representation of cluster marked *, with edges colored by source cell. **F**, The cluster marked ** contains four additional neurons, two ExR3 neurons and two FB6A neurons. These neurons form reciprocal connections to all h$\triangle$K and PFG cells, in all clusters. Shown here is the cell type connectivity graph of the indicated cluster. Nodes represent cell types, and edge weights are given by the average connection strength between cells of the two types. Edges are colored according to the source node.
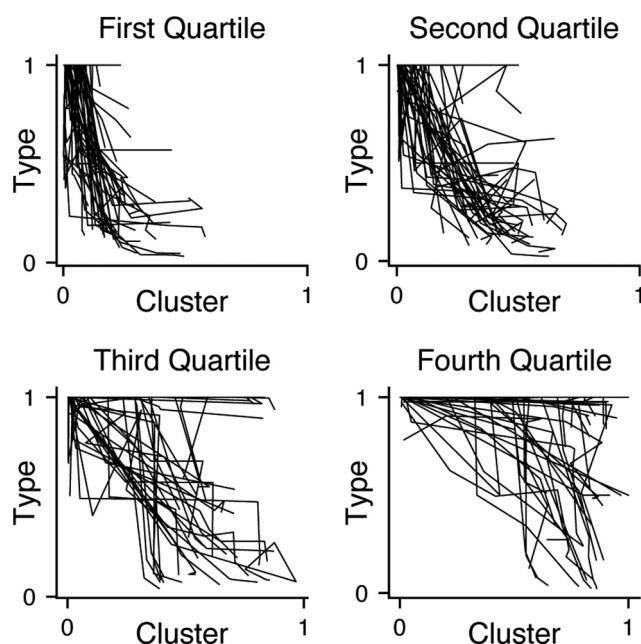

Figure 11. Fraction of Type–Fraction of Cluster curves for all cell types with at least 10 neurons in the Hemibrain, grouped by area score quartile.

structure at higher resolutions, and suggests a possible role for other neuron types (FB6D, FB6M) by their inclusion at moderate resolution scales.

Investigating another cell type identified using the area score heuristic revealed spatial and circuit organization in the inputs to the anterior visual pathway. The small clusters in the anterior visual pathway that emerge at higher values of χ (Fig. 6D, brown) are composed mostly of MC61 cells, which project visual information directly from the medulla to the central brain (Otsuna et al., 2014; Omoto et al., 2017). This algorithmically discovered partition corresponds to a spatial tiling pattern in the small unit of the anterior optic tubercle (Fig. 12A). The spatial tiling corresponds to a wiring pattern involving MC61 cells and tubercle-bulb cells (TuBu) (Fig. 12B). The clusters that are primarily composed of MC61 cells typically include one or two TuBu cells, each of which receives input from all MC61 cells in the cluster (Fig. 12C). TuBu cells innervate EB ring neurons, conveying visual information to the central complex in a retinotopically organized manner (Seelig and Jayaraman, 2013; Omoto et al., 2017). The striking tiling uncovered by our method provides a high-resolution view of findings previously described (Hulse et al., 2021) (Fig. 6). The network structure, with one tile per TuBu cell, implies a topographic organization of the MC61 dendritic arbors in the medulla, which is consistent with previous findings (Otsuna et al., 2014; Omoto et al., 2017). Unfortunately, MC61 cells are at the edge of the reconstructed
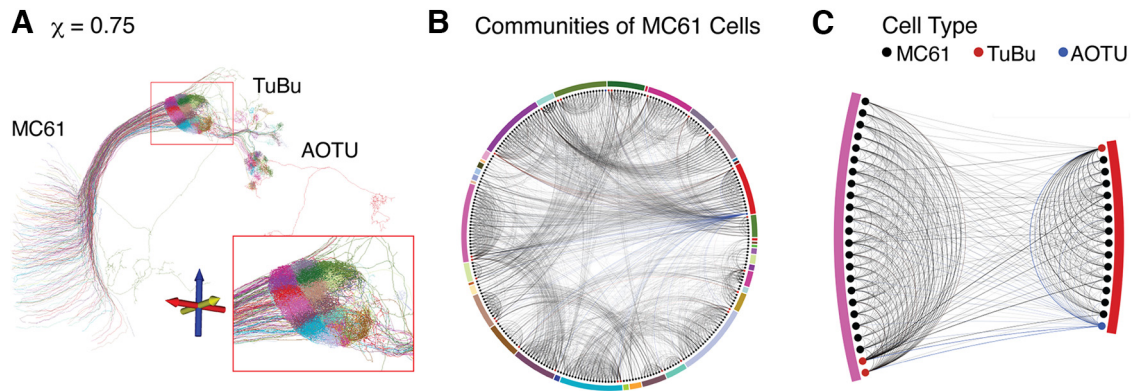
Figure 12. **A**, Clustering at $\chi = 0.75$ reveals a distinct spatial tiling pattern in MC61 cells and their outputs, with each tile receiving inputs from spatially congregated input fibers from the medulla. Shown are 293 cells, comprising all clusters which are at least 80% MC61 cells. This includes 258 of 346 MC61 cells (75%) in the Hemibrain dataset. **B**, The neurons in **A** form a highly organized network: Each cluster of MC61 cells converges on a common TuBu cell. Two of the clusters are shown in **C**, with edges colored by the presynaptic neuron.
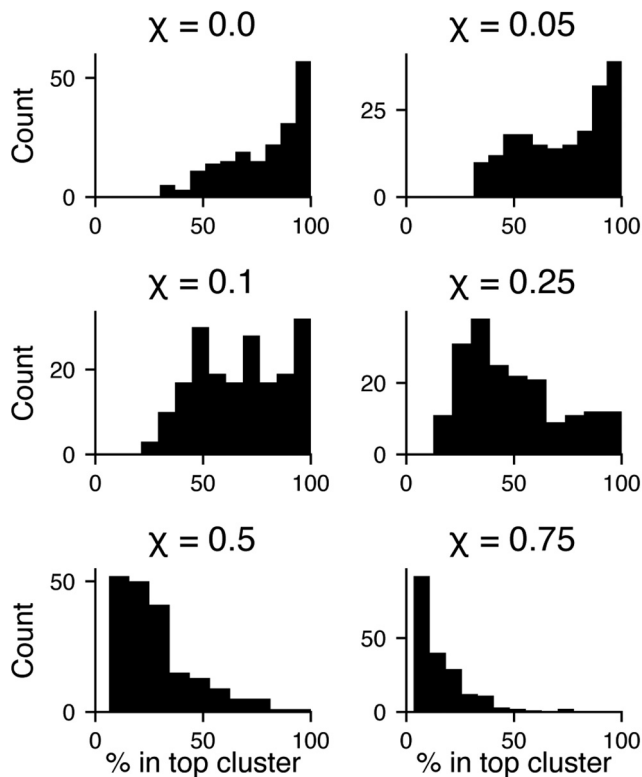

Figure 13. Cell body fiber (CBF) groups partition the neurons in the fly brain based on cell body location and clonal origin. For each CBF group, we computed what fraction of that group belonged to each cluster. Shown here is the distribution of the top fraction across groups. At lower resolutions, RenEEL clusters most neurons in one CBF group together, while at higher resolutions, these groups are split among several clusters.

volume and the medulla is not reconstructed, so we cannot describe the relationship between the spatial extent of MC61 cells in the medulla with their cluster identity.

Our method is thus capable of identifying fine, cell type-specific microcircuits, consisting of as few as three neurons. By comparing the partition of the network induced by cell type annotations to the partitions found by our method across several resolution scales, we can identify cell types that are progressively partitioned into functional modules. The intricate spatial topography revealed by the connectome's community structure in this way reveals striking fine-scale organization in the fly brain.

Discussion

We have shown that clustering by maximizing Generalized Modularity Density both recovers known anatomic structures and infers novel organizational principles from connectome data. Our approach allows us to uncover such structure automatically, unsupervised, solely from the network architecture. We also discover cell type-specific wiring patterns when we add node labels.

Our methods are scalable and can be applied to existing connectomes and to those that will be reconstructed in the future. Our community detection method is fully automated, but biological interpretation is required to fully appreciate the results. Including cell type data with clustering can help generate testable hypotheses about the organization of the network, hypotheses that can be tested using electrophysiology, imaging, genetics (Panser et al., 2016), and other methods. Thus, our methods are an important tool that can be used to help understand the organization and the function of neural populations and circuits.

There are multiple approaches to finding community structure in complex networks, each identifying communities with different attributes (Fortunato and Hric, 2016). Alternative clustering methods could account for link direction and use other information about the connectome. However, such methods would also require the development of efficient computational algorithms similar to the RenEEL algorithm we used. The results would also need to be validated using the known anatomy of the *Drosophila* brain. Maximizing Generalized Modularity Density identifies communities that are more densely connected than a random graph null model (Guo et al., 2019). We validated its use for finding connectome structure by analyzing the structure of the Larval Mushroom Body and comparing the results with those found previously with other methods (Chung et al., 2021). Our results identified a detailed hierarchical structure of nested communities consisting of heterogeneous cell types in the hemispheres that follows anatomy. The structure we found contrasts with that found using spectral embedding methods, for example, which generally identify single cell type clusters. Although our method of community detection finds structure that has an appealing, "straightforward" interpretation, it is not necessarily better than other methods. Rather, community structure found by different approaches should be considered complementary.

Brains inherit a degree of hierarchy and modularity during the course of development from neural stem cells (Molyneaux et al., 2007; K. Ito and Awasaki, 2008; Lai et al., 2008; Sawa, 2010). The fly brain, in particular, is composed of clonal units, densely
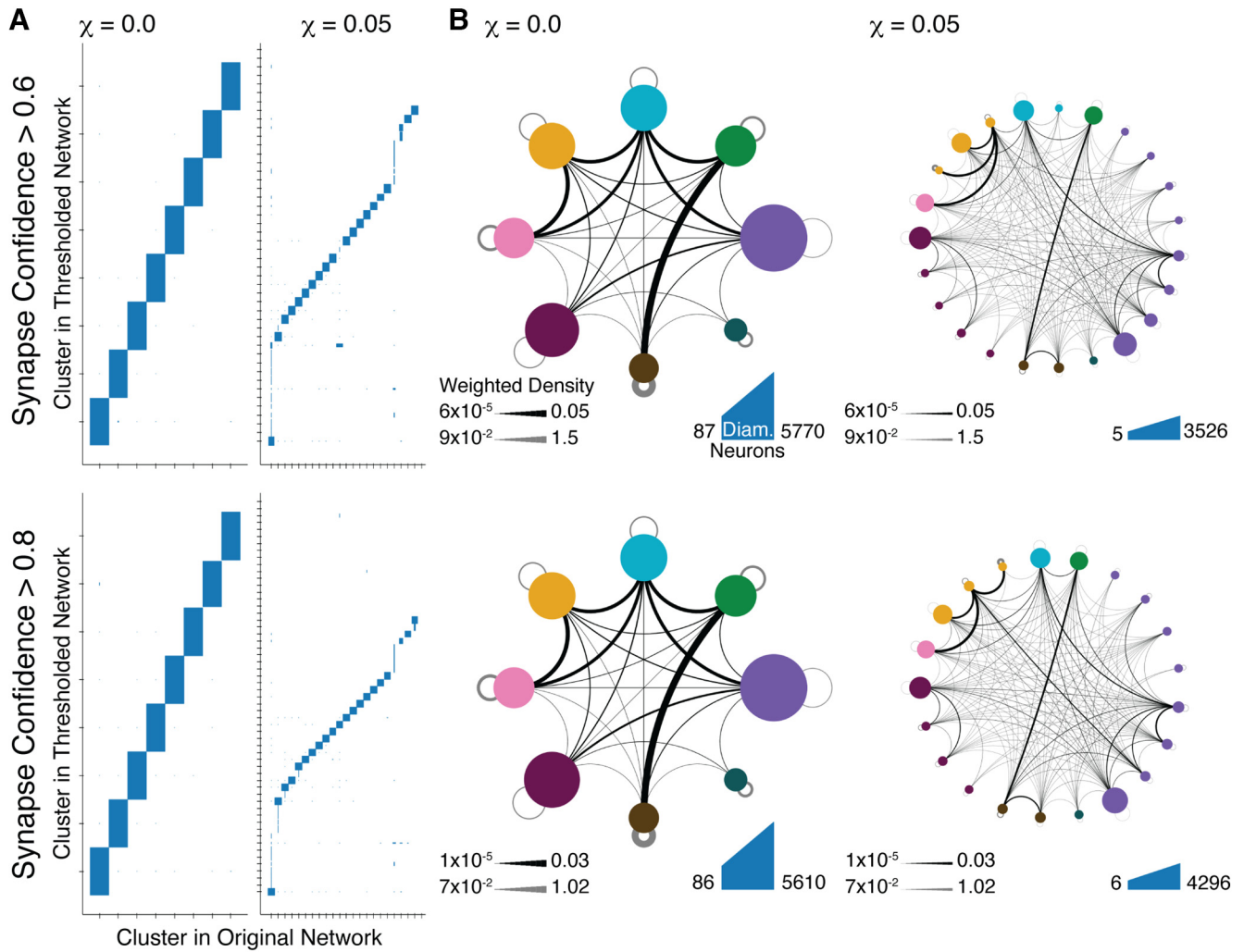
Figure 14. We perturbed the network to investigate the possible effects of reconstruction errors on our results. Community identification at the coarse scale did not significantly change when low-confidence synapses were dropped from the network. *A*, Comparison of the clusters found in the original network with those found in the perturbed network. Each box represents the neurons assigned to the cluster given on the *x* axis in the original network and the cluster given on the *y* axis in the perturbed network. Box width represents the fraction of the cluster in the original network; box height represents the fraction of the cluster in the new network. Clusters from the original network consisting of <5 neurons are not shown. At the lower values of χ presented here, the clusters found in the perturbed network are usually wholly contained in a cluster in the original network, shown by the full height of the boxes. *B*, The intercommunity connectivity is also fairly similar in the perturbed networks (compare to Figure 6*A,B*).

connected populations of cells derived from a single neural stem cell (Hartenstein et al., 2008; K. Ito and Awasaki, 2008). Since our approach partitions the network into densely connected clusters, many of the communities we find in the Hemibrain align closely with previously described clonal units (M. Ito et al., 2013; Omoto et al., 2017). A precise quantitative comparison to the systematic analysis by M. Ito et al. (2013), however, is difficult because we cannot perform a cell-by-cell alignment between their results and ours. Still, there are notable similarities between their Figure 1 (M. Ito et al., 2013) and our Figure 5: A visual inspection shows that many of the identified clonal units are strictly contained within the clusters we find at the coarsest scale parameter $\chi = 0$. Our method can thus be used to develop and refine hypotheses about the development of the brain. In particular, the subnetworks we identify by increasing the resolution of the clustering may identify structures which emerge at different stages of development.

In the Hemibrain, cell body fiber annotations group cells by the location of their cell body on the outer layer of the brain, which correlates with a cell's clonal origin (Scheffer et al., 2020).

At lower resolutions, cell body fibers tend to form subsets of single clusters, while at higher resolutions they tend to be split up (Fig. 13). In other species, clonally related neurons may give rise to a large structure, such as a cortical column (Costa and Hedin-Pereira, 2010), while having a fine-scale organization more similar to a bipartite or multipartite network (Cadwell et al., 2020).

By combining our analysis with cell type data, we were able to identify repeated microcircuits composed of as few as three neurons. This relied not only on our ability to overcome the resolution limit problem (e.g., by optimizing $Q_g(\chi)$ with any particular value of $\chi > 0$), but also on our ability to analyze the network at multiple different resolutions. For any fixed partition of the network into communities, a single community of three nodes could potentially be random noise. However, the small communities that we highlighted as emerging hierarchically appear to be real biological structures because they have reliable cell type composition.

The function of these structures in the fly brain network is currently unknown and will require additional experiments and more complete connectomes to uncover. The bipartite network

we described (Fig. 10) connects to many cell types and brain regions that play a role in vector calculations related to spatial navigation and sleep/wake regulation (Hulse et al., 2021), although their exact role in this circuit is unknown. The regular tiling pattern we describe in the anterior optic tubercle (Fig. 12) suggests a regular pattern of inputs from the medulla. However, the medulla and structures upstream from it are not included in the Hemibrain. The neural circuitry of these structures has been previously described, with many distinct wiring patterns connecting the ommatidia to downstream structures, including the medulla (Takemura et al., 2015). The regularity of wiring in these peripheral structures may serve a computational purpose that is conserved across space (Takemura et al., 2013, 2017b), and so the computational role of the tiling in anterior optic tubercle will require a more complete connectome to infer.

It is possible that some of our results are biased by reconstruction errors and the incompleteness of the dataset. For instance, Cluster 8 at $\chi = 0$ is at the edge of the reconstructed volume, and consists of many neurons which are only partially reconstructed (see Fig. 5). We choose to include the cluster in our analysis because it identifies the portion of the lobula plate that is reconstructed in the Hemibrain volume. Similarly, other clusters composed of cells that are only partially contained in the Hemibrain dataset could be affected by the incompleteness of the data. In the Hemibrain, automated synapse detection has an average precision of 0.8 and average recall of 0.8, although accuracy varies between brain regions (Scheffer et al., 2020). In order to determine whether our results are robust with respect to missing or misidentified synapses, we perturbed a comparable fraction of edges in numerical experiments, and observed no large changes in the identified communities, especially on the larger scale (Fig. 14). However, if reconstruction errors are systematic, they could change the finer structures detected using community detection methods. Further analysis could be done to estimate the impact of such errors, perhaps by perturbing the synapses in a way mimicking the errors known to occur in network reconstruction (Váša and Mišić, 2022). Our perturbation analysis also suggests that false negatives in identifying synapses can lead to a greater number of smaller clusters at higher resolutions (Fig. 2B); therefore, our ability to find true communities at these scales is limited by the accuracy of the data. Thus, our findings should only be interpreted in the positive: where we find certain kinds of structure, we are confident it is truly there; where we do not find structure, that may reflect the incompleteness of the data.

To help us identify communities of interest, we have made use of cell type data in addition to the architecture of the connectome. However, methods that directly incorporate such information could be used to identify network community structure. For instance, spatial information about the cells could be included to preferentially detect clustering between proximal or distal cells. Moreover, we could look at different modes of organization: A core–periphery analysis could identify cores of strongly interconnected cells, and peripheries of cells connected to the core (Rombach et al., 2014). Similarly, we could search for anti-communities of cells that are connected weakly to each other, but strongly to those in other anti-communities (M. Chen et al., 2014). An analogous bipartite or multipartite method that first distinguishes cells based on their type and then looks for structure within sets of cells of the same type would allow the direct inclusion of cell type information in the clustering, but would require the development of new generalized modularity measures and algorithmic methods.

Neurons do not interact only via synaptic connections, but can influence each other through gap junctions, expression of neuromodulators, and possibly ephaptic and other types of interactions. Moreover, in mammals, glia can play an important role in neuronal interactions. Currently available connectomes are based only on synapses between neurons, and thus offer only a partial picture of the interactions between neurons and subgroups. We view our results, and those of similar approaches, as an initial step in characterizing the structure of interactions in the brain, providing a picture that will both change and become more detailed as more complete data becomes available.

To interpret the results of the network analysis, it is important to consider and integrate additional information, such as cell type, spatial location, directionality of edges, and location within the reconstructed volume. Such tools are effective when used in conjunction with other analysis methods, and validated with synthetic and ground truth data. They can be used to generate concrete predictions which need to be independently verified. A slate of tuned and validated automated methods will be essential for progress, given the volume and complexity of new connectomics data.

References

Botta F, del Genio CI (2016) Finding network communities using modularity density. J Stat Mech 2016:123402.

Brandes U, Delling D, Gaertler M, Gorke R, Hoefer M, Nikoloski Z, Wagner D (2008) On modularity clustering. IEEE Trans Knowl Data Eng 20:172–188.

Bullmore E, Sporns O (2009) Complex brain networks: graph theoretical analysis of structural and functional systems. Nat Rev Neurosci 10:186–198.

Cadwell CR, et al. (2020) Cell type composition and circuit organization of clonally related excitatory neurons in the juvenile mouse neocortex. Elife 9:e52951.

Chen L, Yu Q, Chen B (2014) Anti-modularity and anti-community detecting in complex networks. Info Sci 275:293–313.

Chen M, Kuzmin K, Szymanski BK (2014) Community detection via maximization of modularity and its variants. IEEE Trans Comput Soc Syst 1:46–65.

Chen M, Nguyen T, Szymanski B (2013) A new metric for quality of network community structure. ASE Hum J 2:226–240.

Chen T, Singh P, Bassler KE (2018) Network community detection using modularity density measures. J Stat Mech 2018:053406.

Chung J, Bridgeford E, Arroyo J, Pedigo BD, Saad-Eldin A, Gopalakrishnan V, Xiang L, Priebe CE, Vogelstein JT (2021) Statistical connectomics. Annu Rev Stat Appl 8:463–492.

Clauset A, Newman ME, Moore C (2004) Finding community structure in very large networks. Phys Rev E Stat Nonlin Soft Matter Phys 70:066111.

Costa M, Hedin-Pereira C (2010) Does cell lineage in the developing cerebral cortex contribute to its columnar organization? Front Neuroanat 4:26.

Donlea JM, Thimgan MS, Suzuki Y, Gottschalk L, Shaw PJ (2011) Inducing sleep by remote control facilitates memory consolidation in *Drosophila*. Science 332:1571–1576.

Eichler K, Li F, Litwin-Kumar A, Park Y, Andrade I, Schneider-Mizell CM, Saumweber T, Huser A, Eschbach C, Gerber B, Fetter RD, Truman JW, Priebe CE, Abbott LF, Thum AS, Zlatic M, Cardona A (2017) The complete connectome of a learning and memory centre in an insect brain. Nature 548:175–182.

Fisher YE (2022) Flexible navigational computations in the *Drosophila* central complex. Curr Opin Neurobiol 73:102514.

Fortunato S, Barthélemy M (2007) Resolution limit in community detection. Proc Natl Acad Sci USA 104:36–41.

Fortunato S, Hric D (2016) Community detection in networks: a user guide. Phys Rep 659:1–44.

Guo J, Singh P, Bassler KE (2019) Reduced network extremal ensemble learning (RenEEL) scheme for community detection in complex networks. Sci Rep 9:14234.

Guo J, Singh P, Bassler KE (2023) Resolution limit revisited: community detection using generalized modularity density. J Phys Complex 4:025001.

Haberkern H, Jayaraman V (2016) Studying small brains to understand the building blocks of cognition. Curr Opin Neurobiol 37:59–65.

Hanesch U, Fischbach KF, Heisenberg M (1989) Neuronal architecture of the central complex in *Drosophila melanogaster*. Cell Tissue Res 257:343–366.

Hartenstein V, Spindler S, Pereanu W, Fung S (2008) The development of the *Drosophila* larval brain. In: Brain development in Drosophila melanogaster: advances in experimental medicine and biology (Technau GM, ed), pp 1–31. New York: Springer.

Hulse BK, Haberkern H, Franconville R, Turner-Evans D, Takemura SY, Wolff T, Noorman M, Dreher M, Dan C, Parekh R, Hermundstad AM, Rubin GM, Jayaraman V (2021) A connectome of the *Drosophila* central complex reveals network motifs suitable for flexible navigation and context-dependent action selection. Elife 10:e66039.

Ito K, Awasaki T (2008) Clonal unit architecture of the adult fly brain. In: Brain development in Drosophila melanogaster: advances in experimental medicine and biology (Technau GM, ed), pp 137–158. New York: Springer.

Ito K, Shinomiya K, Ito M, Armstrong JD, Boyan G, Hartenstein V, Harzsch S, Heisenberg M, Homberg U, Jenett A, Keshishian H, Restifo LL, Rössler W, Simpson JH, Strausfeld NJ, Strauss R, Vosshall LB, Insect Brain Name Working Group (2014) A systematic nomenclature for the insect brain. Neuron 81:755–765.

Ito M, Masuda N, Shinomiya K, Endo K, Ito K (2013) Systematic analysis of neural projections reveals clonal composition of the *Drosophila* brain. Curr Biol 23:644–655.

Jefferis GS, Hummel T (2006) Wiring specificity in the olfactory system. Semin Cell Dev Biol 17:50–65.

Kato YS, Tomita J, Kume K (2022) Interneurons of fan-shaped body promote arousal in *Drosophila*. PLoS One 17:e0277918.

Kim SS, Rouault H, Druckmann S, Jayaraman V (2017) Ring attractor dynamics in the *Drosophila* central brain. Science 356:849–853.

Lai SL, Awasaki T, Ito K, Lee T (2008) Clonal analysis of *Drosophila* antennal lobe neurons: diverse neuronal architectures in the lateral neuroblast lineage. Development 135:2883–2893.

Li F, et al. (2020) The connectome of the adult *Drosophila* mushroom body provides insights into function. Elife 9:e62576.

Lin CY, Chuang CC, Hua TE, Chen CC, Dickson BJ, Greenspan RJ, Chiang AS (2013) A comprehensive wiring diagram of the protocerebral bridge for visual information processing in the *Drosophila* brain. Cell Rep 3:1739–1753.

Liu G, Seiler H, Wen A, Zars T, Ito K, Wolf R, Heisenberg M, Liu L (2006) Distinct memory traces for two visual features in the *Drosophila* brain. Nature 439:551–556.

Molyneaux BJ, Arlotta P, Menezes JR, Macklis JD (2007) Neuronal subtype specification in the cerebral cortex. Nat Rev Neurosci 8:427–437.

Newman ME (2004) Fast algorithm for detecting community structure in networks. Phys Rev E Stat Nonlin Soft Matter Phys 69:066133.

Newman ME (2006) Modularity and community structure in networks. Proc Natl Acad Sci USA 103:8577–8582.

Omoto JJ, Keleş MF, Nguyen BC, Bolanos C, Lovick JK, Frye MA, Hartenstein V (2017) Visual input to the *Drosophila* central complex by

developmentally and functionally distinct neuronal populations. Curr Biol 27:1098–1110.

Otsuna H, Ito K (2006) Systematic analysis of the visual projection neurons of *Drosophila* melanogaster: I. Lobula-specific pathways. J Comp Neurol 497:928–958.

Otsuna H, Shinomiya K, Ito K (2014) Parallel neural pathways in higher visual centers of the *Drosophila* brain that mediate wavelength-specific behavior. Front Neural Circuits 8:8.

Panser K, Tirian L, Schulze F, Villalba S, Jefferis GS, Bühler K, Straw AD (2016) Automatic segmentation of *Drosophila* neural compartments using GAL4 expression data reveals novel visual pathways. Curr Biol 26:1943–1954.

Pereanu W, Kumar A, Jennett A, Reichert H, Hartenstein V (2010) Development-based compartmentalization of the *Drosophila* central brain. J Comp Neurol 518:2996–3023.

Rombach M, Porter MA, Fowler JH, Mucha PJ (2014) Core-periphery structure in networks. SIAM J Appl Math 74:167–190.

Sawa H (2010) Specification of neurons through asymmetric cell divisions. Curr Opin Neurobiol 20:44–49.

Scheffer LK (2020) Graph properties of the adult *Drosophila* central brain. bioRxiv 102061. https://doi.org/10.1101/2020.05.18.102061.

Scheffer LK, et al. (2020) A connectome and analysis of the adult *Drosophila* central brain. Elife 9:e57443.

Seelig JD, Jayaraman V (2013) Feature detection and orientation tuning in the *Drosophila* central complex. Nature 503:262–266.

Shapson-Coe A, et al. (2021) A connectomic study of a petascale fragment of human cerebral cortex. bioRxiv 446289. https://doi.org/10.1101/2021.05.29.446289.

Takemura S, et al. (2013) A visual motion detection circuit suggested by *Drosophila* connectomics. Nature 500:175–181.

Takemura S, et al. (2015) Synaptic circuits and their variations within different columns in the visual system of *Drosophila*. Proc Natl Acad Sci USA 112:13711–13716.

Takemura S, et al. (2017a) A connectome of a learning and memory center in the adult *Drosophila* brain. Elife 6:e26975.

Takemura S, Nern A, Chklovskii DB, Scheffer LK, Rubin GM, Meinertzhagen IA (2017b) The comprehensive connectome of a neural substrate for 'ON' motion detection in *Drosophila*. Elife 6:e24394.

Traag VA, Van Dooren P, Nesterov Y (2011) Narrow scope for resolution-limit-free community detection. Phys Rev E Stat Nonlin Soft Matter Phys 84:016114.

Turner GC, Bazhenov M, Laurent G (2008) Olfactory representations by *Drosophila* mushroom body neurons. J Neurophysiol 99:734–746.

Turner NL, Lee K, Lu R, Wu J, Ih D, Seung HS (2020) Synaptic partner assignment using attentional voxel association networks. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), pp. 1–5.

Turner NL, et al. (2022) Reconstruction of neocortex: organelles, compartments, cells, circuits, and activity. Cell 185:1082–1100.e24.

Váša F, Mišić B (2022) Null models in network neuroscience. Nat Rev Neurosci 23:493–504.

Vosshall LB, Stocker RF (2007) Molecular architecture of smell and taste in *Drosophila*. Annu Rev Neurosci 30:505–533.

Wolff T, Iyer NA, Rubin GM (2015) Neuroarchitecture and neuroanatomy of the *Drosophila* central complex: a Gal4-based dissection of protocerebral bridge neurons and circuits. J Comp Neurol 523:997–1037.