An entropic measure of diverse specialisation highlights multifunctional neurons in annotated connectomes

Sung Soo Moon[1] and Sebastian E. Ahnert[1,2]

ssm47, sea31 @cam.ac.uk
[1]Department of Chemical Engineering and Biotechnology, University of Cambridge, Philippa Fawcett Drive, Cambridge CB3 0AS, United Kingdom
[2]The Alan Turing Institute, 96 Euston Road, London NW1 2DB, United Kingdom

March 19, 2025

Abstract

The creation and curation of synaptic-level neuronal networks, or connectomes, enables the study of the relationship between structure and function of these networks. Topological characteristics of neuronal networks have been studied extensively. Separately there have been considerable efforts to classify the morphology, cell types, and lineages of neurons. Here we introduce a network metric that combines topological analysis with node metadata. This entropic quantity measures the diversity of incoming or outgoing connections to a node in terms of the metadata distribution. We find that in *C. elegans*, the top-scoring neurons in terms of this metric have known functions that integrate and disseminate multimodal information involved in sensorimotor functions. In the nerve cord of *Drosophila melanogaster*, we find that the top-scoring neurons are located in the abdominal neuropil, where sensorimotor coordination is required for complex innate behaviour such as mating.

1 Introduction

A brain is often modelled as a network [Bassett and Sporns, 2017, Bullmore and Sporns, 2009], but how are brain networks organised? Neuronal resolution synaptic mapping of the brain, or connectomes, are networks of nodes and weighted, directed edges between them. While the *C. elegans* connectome [White et al., 1986] was a landmark data set in neuroanatomy, we have yet to formulate a general network-level understanding of the organisational principles of nervous systems. More recently, the adult *Drosophila melanogaster* brain [Zheng et al., 2018, Scheffer et al., 2020, Dorkenwald et al., 2024] and nerve cord [Venkatasubramanian and Mann, 2019, Takemura et al., 2024, Cheong et al., 2024, Marin et al., 2024, Court et al., 2020, Stürner et al., 2024], data sets have been imaged, segmented and reconstructed, and annotated. With high-throughput experimental techniques granting more metadata about the nature of the neurons [Harris et al., 2015, Janssens et al., 2022] and the synaptic connections [Buhmann et al., 2021, Eckstein et al., 2024, Bentley et al., 2016], connectome data sets are complex annotated systems and not simple undirected networks. Annotations are key supplements to our understanding of structure and function of brain networks, often related to dynamical properties of the system, and intrinsic to biological function [Bazinet et al., 2023a,b]. Extensions of network methodologies respecting annotation metadata in neuroscience is needed.

The application of network analysis to neuroscience has proved fruitful, establishing key interpretable features of brain networks, including small-worldness, modularity, hubs, rich

clubs and community structure [Bassett and Sporns, 2017, Bullmore and Sporns, 2009, Betzel et al., 2024, Towlson et al., 2013, Jarrell et al., 2012]. Many analyses focus on assigning an importance to a node in context of global topology, like centrality measures. However, for studying biological networks with specific cross-modular functional roles, a measure based on local topology may be more appropriate. For instance, network motifs have been shown to form the basis of many complex networks [Milo et al., 2002], and in brain networks have revealed a dependency on local building blocks of circuits for an emergence of a network that facilitates complex behaviour [Sporns and Kötter, 2004]. Topological network metrics have been considered in combination with metadata in some contexts, such as assortative mixing [Bazinet et al., 2023a], but previous studies on network motifs in single neuron resolution data (in *C. elegans* [Varshney et al., 2011, Bentley et al., 2016] and *Drosophila* [Lin et al., 2024, Hulse et al., 2021]) do not fully exploit the rich metadata available. Comprehensive brain networks are key to principles of development [Schlegel et al., 2017] and locally defined connectivity can illuminate the architectural blueprint of the brain [Vogelstein et al., 2019]. In this work, we introduce a method to explore local connectivity with annotations.

As the number of available connectomes continues to increase, we require more network methods that incorporate the rich complexity of annotations in order to link the topology of the network to its biological function. Here, we focus on node metadata to introduce a measure of the diversity of local connections and apply this to the *C. elegans* connectome and the *Drosophila* ventral nerve cord (VNC) connectome. We exploit several different annotation schemes available to us, from fine-grained labels such as cell types to broader categories such as lineage. While there is prior work characterizing diversity with respect to community annotations [Bertolero et al., 2017, Guimerà and Amaral, 2005, Guimerà and Nunes Amaral, 2005, Pedersen et al., 2020, Delaplace, 2025, Eagle et al., 2010] and information entropy in network science is not novel [Omar and Plapper, 2020, Bashiri et al., 2020], to our knowledge none characterizes the diversity of specialised connectivity in the local neighbourhood of a given node.

Our measure, the *s*pecialisation-diversity, is derived as follows: First, we construct each neuron's connectivity vector p, which expresses the proportion of ingoing or outgoing connections to annotated partners. The metadata we use are cell classifications or lineage annotations (see **Methods** for details). Then, for a set of connectivity vectors $\{p\}$, the specialisation-diversity is defined as: $\Delta_S = S(\bar{p}) - \bar{S}$ where $S = -\sum_i p_i ln(p_i)$, information entropy and $\bar{p}$ and $\bar{S}$ denote the means of p and S. A large Δ_S value is only possible when the vectors $\{p\}$ are individually specialised but in diverse categories (see figure 1A and **Methods**). Δ_S quantifies the level of diverse specialisation in connectivity with respect to the annotations present, and this is a property of the set of connectivity vectors. We consider two choices of vector sets for a given neuron that characterizes how diverse the integrative and distributive connectivity is. The first is calculated from the diversity of specialisation in the downstream outputs, and the other is calculated from the diversity of specialisation in the upstream inputs. More precisely, for a target neuron, we gather *out*going connectivity vectors of *downstream* partner neurons, and calculate a specialisation-diversity. The resulting Δ_S associated to this outwards branching motif we call the *distributive specialisation-diversity* and notate it as $\Delta_S^{D,out}$. Similarly, a given node's *upstream* partners' *in*coming vectors can be used to calculate a specialisation-diversity. We call this *integrative specialisation-diversity*, $\Delta_S^{U,in}$. These measures quantify the level of diversely specialised connectivity channels in a node's local vicinity (figure 1B), with respect to an annotation scheme we use to label the nodes. The annotations we used for the connectivity vectors were cell category in *C. elegans* and hemilineage in *Drosophila* (see **Methods** for details).

In this paper we interpret the diversity of local connectivity as characteristics of roles of

nodes in a connectome by using the specialisation-diversity measure. The top 10 highest-ranking neurons in summed integrative and distributive specialisation-diversities we call "entropic hubs". In *C. elegans* we find these to be linked to multimodal functions like roles in sensorimotor circuits, and in the neuropeptide signalling network responsible for social behaviour [Ripoll-Sánchez et al., 2023]. In *Drosophila*, our entropic hubs are concentrated in the abdominal region where higher-level innate behavioural programmes are controlled [Cheong et al., 2024, Pavlou et al., 2016].

2 Results

2.1 Cell classes occupy interpretable and distinct regions in specialisation-diversity distributions

We first verified whether specialisation-diversity was interpretable when considering broad cell classifications in the two connectomes by inspecting the distributions for integrative and distributive diversities (figure 2A for *Drosophila* VNC and 2D for *C. elegans*). In both distributions, neurons are broadly speaking similarly distributive and integrative, but nevertheless there are clear deviations that differ according to the cell classes. In figure 2B and 2E we show the distributions of the quantity $\Delta_S^{D,out} - \Delta_S^{U,in}$ which quantifies how much more distributive or integrative a given neuron is. We report varying biases between the cell types - sensory (orange) being most distributive, and the intrinsic/interneurons (purple) and motor neurons (green) being more integrative. In both *Drosophila* and *C. elegans*, sensory neurons with mostly downstream connections will propagate into the connectome for the processing of external sensory signals, and so are biased towards being distributive rather than integrative. Motor neurons, by contrast, will mostly form upstream connections that control their output, and therefore are more integrative than distributive. The ascending (red) and descending neurons (blue) in *Drosophila* lie in the distributions as we expect: descending neurons being more distributive and ascending more integrative. For intrinsic neurons in *Drosophila* and interneurons in *C. elegans* that scaffold the connectivity between the input and outputs of the system, we see a slight bias towards integrative diversities. The Pharynx (teal) in *C. elegans* is very sparsely connected to the rest of the network and therefore exhibits low distributive and integrative diversities.

2.2 Entropic hubs are not synaptic hubs

As the distributions of the broad categorizations of neurons yield results we can intuitively interpret, we now consider finer categorizations and individual neurons. Firstly we consider the highest-ranking neurons in $\Delta_S^{U,in} + \Delta_S^{D,out}$ value, which we call *entropic hubs*, (red squares in figure 2). In *Drosophila* the membership of cell classes of the ten highest-ranking entropic hubs are three descending neurons, one ascending neuron, five intrinsic neurons and one motor neuron and all these neurons are more distributive than integrative. These neurons have not been previously described in detail individually, but we note that all these neurons have soma positions in the abdominal neuropils, which are shown to contribute to complex mating behaviour [Pavlou et al., 2016, Marin et al., 2024]. In *C. elegans*, the entropic hubs play an active role in the neuropeptidergic network (see below). For both connectomes we find that entropic hubs are not synaptic hubs (figure 2C and F), meaning that a high number of synaptic partners does not result in a highly diverse set of upstream or downstream neighbour specialisations. This highlights that our method focuses on the specificity of connectivity rather than the volume of it. The full distributions of all the neurons and their metadata is made available (see **Methods**).

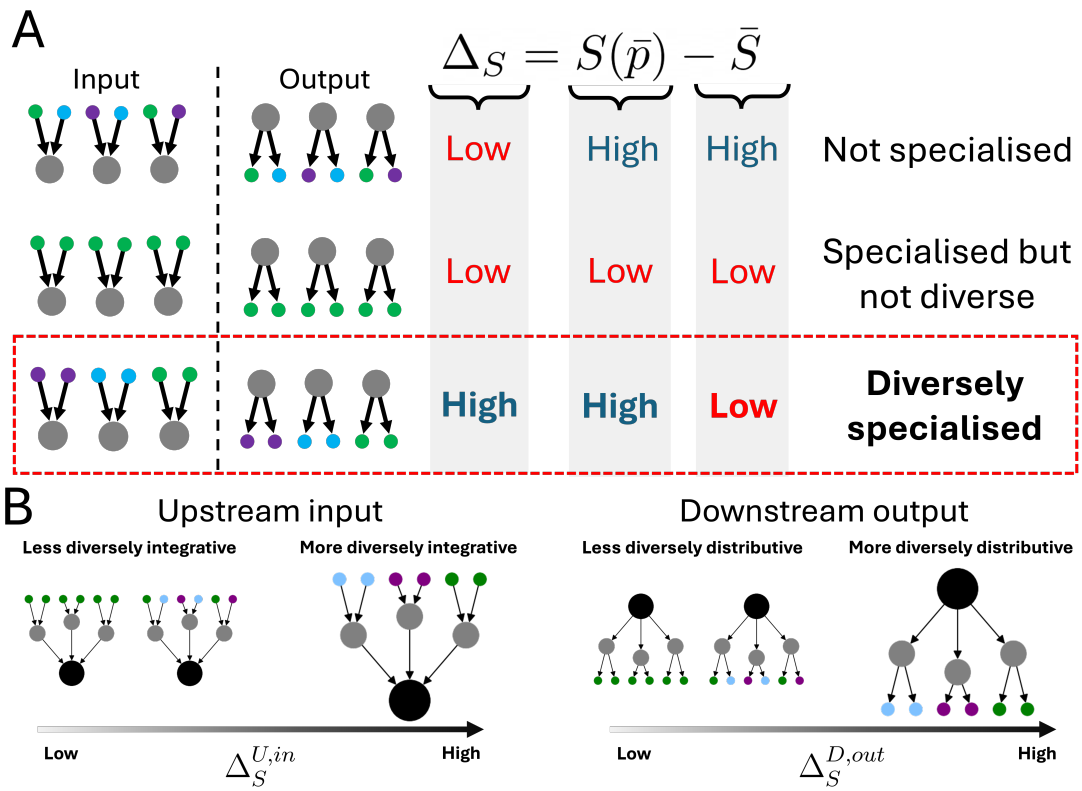

Figure 1: **A measure for quantifying diversely specialised connectivity vectors** **A**, Here we illustrate three scenarios concerning input or output connectivity vectors of the three grey nodes (left), where the coloured partners represent different annotation types. The value of entropy is low when the probability distribution is most specialised in one category and highest when it is most evenly distributed. In the first row, the individual vectors are not specialised, leading to a high average entropy, $\bar{S}$ and a high entropy of average probability vector $S(\bar{p})$. In the middle row, the three grey nodes are specialised but are all specialised in the same annotation type. Therefore, the average entropy is low and the entropy of the average vector is also low (in this case $\bar{S} = S(\bar{p}) = 0$). For the above reasons, the two cases yield a low value for Δ_S. In the last row, the grey nodes are also individually specialised in their vectors. In contrast to the previous case, the grey nodes are specialised in different types. Therefore, while the average input entropy is low again, the average input $\bar{p}$ is not specialised and so $S(\bar{p})$ is high. Thus, the difference, Δ_S is only high when the grey nodes are specialised diversely. We invite readers to the **Supplementary Information**, for a mathematical comment on the bounds of possible Δ_S values. **B**, We extend the previous argument to two motifs to calculate the specialisation-diversity, and illustrate the cases where the Δ_S values are low and high. Firstly, with the black node's upstream partners' input vectors we measure the integrative specialisation-diversity $\Delta_S^{U,in}$ (left). Secondly, the specialisation-diversity of the downstream partners' outputs measures the distributive specialisation-diversity $\Delta_S^{D,out}$ (right).

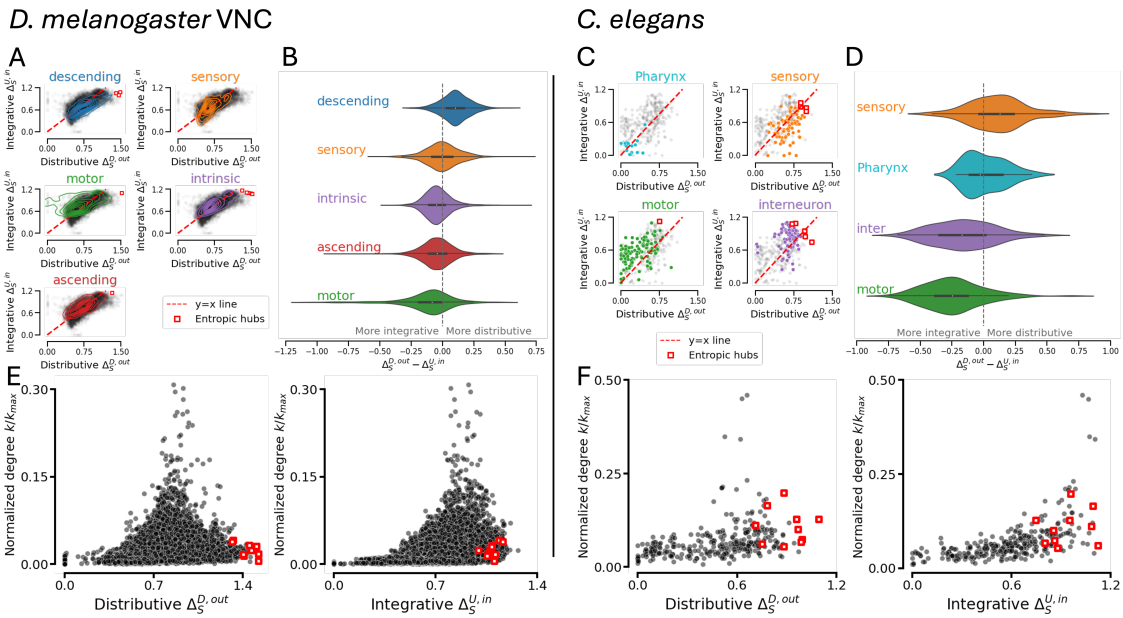

Figure 2: **Specialisation-diversity distributions are localised for different cell classes, and are distinct from synaptic degree**
The distributions of integrative vs distributive specialisation-diversity for *Drosophila* VNC neurons (**A**) and *C. elegans* (**C**) show localisation of the cell class populations in the plot, with transparent black markers showing the full distribution. In dotted red is the y=x line where the neuron is equally distributive and integrative. Red squares are entropic hubs: neurons that rank in the top 10 highest values of $\Delta_S^{D,out} + \Delta_S^{U,in}$. The distribution of distributive minus integrative specialisation-diversity for all neurons in the *Drosophila* VNC (**B**) and *C. elegans* (**E**), reveals that all cell classes have members that are strongly distributive and or integrative, but the majority are well balanced. The averages of the distributions are sorted from most distributive (compared to integrative) at the top to most integrative (compared to distributive) at the bottom. The cell classes are appropriately biased corresponding to their functional role. If the role is primarily projective into the system, then it is more distributive than integrative (descending, sensory). In *Drosophila* the sensory neurons are more balanced as the distribution mean lies around zero, suggesting that the average neuron will be equally integrative and distributive. If the role is primarily integrating diverse channels of information, then outputs in specific and specialised ways, the class is more integrative than distributive (motor, ascending). Intrinsic neurons in *Drosophila* and interneurons in *C. elegans* have a skewed mean towards integrativeness suggesting the function requires more integration of diverse connectivity than distribution. Subfigures **E**, **F** show that our measure is distinct of degree, the number of total partners of a node. In both animals, the entropic hubness cannot be fully explained by sheer number of synaptic partners, as neurons with the highest specialisation-diversity values occupy low (normalized) degree values.

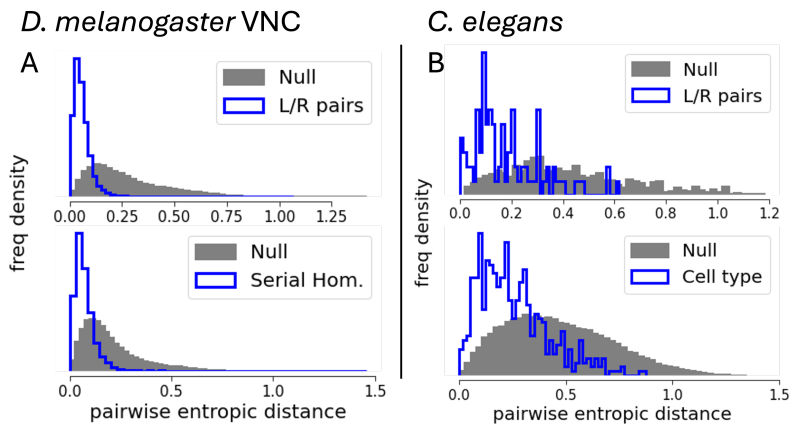

Figure 3: **Cell types, left-right pairs and serial homologues are statistically significant in their specialisation-diversity values**
The figure shows statistical significance of the specialisation-diversity values for different types of fine grained neuron categorisation. In all cases, we calculate pairwise Euclidean distances of the specialisation-diversity vector $[\Delta_S^{D,out}, \Delta_S^{U,in}]$, within a specific category and compare this to an appropriate randomised null model. We call this distance the pairwise entropic vector distance. In the VNC (**A**), left-right pairs (top) and serial homologues (bottom) are demonstrated to be statistically significant in their specialisation-diversity values. For left-right pairs, we construct a null model by randomly generating pairs of neurons that conserve the side of the soma, i.e. each randomised pair would be from opposite halves of the VNC. The serial homologue null model picks random groups of six neurons but conserves the leg neuropil membership, i.e. six random neurons from each leg neuropil. In all null models we preserve the connectivity vector and thus the entropy vector. We find qualitatively identical results for *C. elegans* (**B**) using analogous distance calculations using appropriate null models for statistical significance of left-right pairs (top) and fine grained cell types (118 types of which left-right pairs are a subset). The left-right null model again conserves the side membership, and the null model for cell type uses an all-by-all distance calculation (see main text). All null model results are aggregated over 10 realizations of random sampling. Two-sample Kolmogorov-Smirnoff tests yielded significant results for all distribution comparisons with their nulls ($p << 10^{-3}$).

2.3 Neurons within fine-grained annotation types are similar in terms of their specialisation-diversity values

The finest level of annotation in the connectome data sets are cell types. In *C. elegans* this comprises 118 cell types that are characterized and named according to their classification and position. In the *Drosophila* VNC, we have cell types that signify identical cell morphology or connectivity. Nervous systems exhibit a well described bilateral left-right symmetry observed in the connectome [Venkatasubramanian and Mann, 2019, Winding et al., 2023, Cheong et al., 2024, Marin et al., 2024]. In *Drosophila* VNC, due to the six-fold symmetry of the leg neuropils, we have six identical analogous neurons forming one *serial homologue* group [Marin et al., 2024, Haszprunar, 1992]. These neurons are annotated by the same label, but are not necessarily strongly connected, so do not form a modular structure within the connectome. Rather, these neurons are similar in morphology so have similar connectivity within each modular structure.

We asked if these finer groups of similar neurons were similar in terms of their specialisation-diversity. To answer this, we calculated pairwise distances within the fine groupings using the euclidean metric on the distributive and integrative diversities as a vector $[\Delta_S^{D,out}, \Delta_S^{U,in}]$,

then tested whether this distribution deviated from appropriate null models (figure 3). We find a statistically significant similarity of contralateral (left-right) pairs using a side-preserving null model of random samples of two neurons, one from each lateral side (top row, figure 3). We also find statistically significant similarity within serial homologue groups in *Drosophila* (figure 3A, bottom), where the null models are pairwise distributions drawn from sampling one neuron from each leg neuropil. Lastly, we test for pairwise entropic distance within within the 118 cell types in *C. elegans* [Hobert et al., 2016] (figure 3B) bottom), with a null model that considers the all-neuron-by-all-neuron pairwise entropic distance distribution. Our results show that alike neurons have similar values of integrative and distributive specialisation-diversities to a statistically significant degree according to a two-sample Kolmogorov-Smirnov test.

2.4 C. elegans entropic hubs have active roles in complex behaviour

Since all neurons in *C. elegans* are named and have been subject to (in some cases extensive) experimentation, we describe the entropic hubs and their active roles in complex behaviour. The top ranking neuron, `PVR` is a tail interneuron not well described in function. Recently, `PVR` has been identified as a neuropeptidergic network hub, important in the 'wireless' neuropeptide-GPCR signalling complementary to the synaptic and gap junction 'wired' network [Ripoll-Sánchez et al., 2023]. Neuron type `RMG` (`RMGL` and `RMGR`) are interneurons described in the literature as an essential neuropeptide expressor responsible for all types of social behaviour, whose complex behaviour depends on many modes of sensory cues and many diverse repertoire of outputs, with `RMG` being a gap junction hub for connecting 7 classes of neurons [Macosko et al., 2009]. Neuron `DVA` is a stretch sensory neuron for proprioception, and positively and negatively affects locomotion to be able to fine-tune motor neuron stimulation, from a single source. Its central role in regulating the sensory-motor integration explains the diverse downstream specialism [Li et al., 2006]. Neuron `CEP` is a dopaminergic touch sensory neuron in the nose essential for regulating learned bacterial avoidance, along with neurons `ADE` [Chou et al., 2022]. Neuron `URXR` is a body cavity oxygen sensory neuron, central in homeostatic integration of oxygen level fluctuation and metabolic processes. Its sensory signals stimulate body fat loss only when the body has sufficient reserves [Witham et al., 2016]. Neuron `RIGL` is a ring interneuron part of 4 pairs of neurons that mediates CO_2 response, operating downstream of `BAG` oxygen sensory neurons. Hunger causes the animal to be attracted to CO_2, and feeding reverses this, and neuron pairs `RIG` and `AIY` mediate this [Rengarajan et al., 2019]. Neuron `SMBDL` is a dorsal neck motor neuron that is a part of the `SMB` cell type that is crucial in sensorimotor transformation for governing klinotaxis, or stimulus provoked lateral locomotion, in particular, changes in salt concentration [Matsumoto et al., 2024].

3 Discussion

Our objective with this study was to incorporate biological annotations to enhance our understanding of functional roles in structural brain networks. To achieve this, we have introduced a novel measure based on information entropy that describes the level of diverse specialisation in local directed node neighbourhoods that are integrative and distributive. The specialisation-diversity considers aggregated, directed connectivity from next-nearest network neighbours, thus it is a localised measure for a given node that does not rely on global knowledge.

The application of our approach to the *C. elegans* and *Drosophila* connectomes, using cell classification annotations and hemilineage annotations respectively, yielded distributions of specialisation-diversity that reflect the biological functionalities of these neuron

classes. If a given target neuron is on any directed path of length two as the source or sink of that path, then these quantities can be calculated. Therefore, almost all neurons can be quantified by their proximity to specialised connectivity patterns. The capability of neurons to integrate and broadcast the many modes of diversely specialised information aligns with the roles of cell classes. However, all cell classes have outliers that are on either extremes, suggesting heterogeneity and hierarchy in diverse specialisation within class modalities in nervous systems. Interestingly, intrinsic neurons and interneurons that form the bulk of the connectivity bridging inputs and outputs of the network are more integrative than distributive on average. This asymmetry suggests that their connectivity role is to integrate diversely specialised channels of annotation types. Specific entropic hubs with highly integrative and distributive connectivity channels are neurons that are responsible complex sensorimotor behaviour in *C. elegans*, thus requiring close proximity to specialised channels in their inputs and outputs. We propose that entropic hubs in *Drosophila* are responsible for sensorimotor behaviour, and previous experimental work on the abdominal neuropil suggests an important role in innate behaviours like mating [Pavlou et al., 2016], and so we expect this result to be qualitatively conserved across different individuals.

While we can identify top entropic hubs that are most integrative and distributive, we are limited in terms of the conclusions we can draw from this in *Drosophila*. The VNC facilitates both reflexive and coordinated movement in the fly independent of brain signalling, the network topology is rich with functional circuitry that allows for complex behaviour [Takemura et al., 2024]. Secondary (post-embryonic) neurons of specific hemilineages have been shown to relate to specific motor functions [Harris et al., 2015], but in our analysis we found the specialisation-diversities to vary between hemilineages annotated with the same function, suggesting further sub-roles of requiring different levels of diversely specialised connectivity (see **Supplementary Information**). Our Δ_S indirectly measures the neurons that are proximal to many diversely specialised hemilineage connectivity channels responsible for coordinated movement like walking and flight. The entropic hubs in *Drosophila* are mostly primary (embryonic) neurons (7 characterised as such, and 3 not yet annotated). Primary neurons are often globally and centrally connected, and more likely be a member of rich club organisation [Marin et al., 2024], possibly facilitating the high integrative and distributive specialisation-diversities. Early birthtime neurons go through a developmental programme that persists and are "remodelled" during metamorphosis into the adult [Shepherd et al., 2016, Marin et al., 2024] and we speculate that its proximity to highly diversely specialised connectivity are artefacts of dynamic utility and versatility pivotal during development. The entropic hubs are associated with complex, often sensorimotor behaviour in *C. elegans*, with some neurons that were only recently described to be significant like `PVR`, a neuropeptide hub [Ripoll-Sánchez et al., 2023]. We also report that entropic hubs are not conventional hubs in terms of their degree in the synaptic network. In *C. elegans*, while some entropic hubs have high degrees compared to the rest of the network, there are several neurons that exhibit highly distributive and integrative diversities without being synaptic hubs. In *Drosophila*, the absence of correlation between specialisation-diversity and degree is clearer. We observe that while highly connected nodes could also be in close proximity to highly specialised and diverse connectivity channels, their high degree makes them more likely to be connected to the same specialisations, and non-specialised partners, which penalises their specialisation-diversity value. This is consistent with previous studies that showed that node degree was independent of a node's diversity of connection to disparate modules [Pedersen et al., 2020].

We have verified that the cell types, contralateral pairs and serial homologues are more similar to each other in specialisation-diversity values than randomised null models. Thus, similar neurons are also similarly diversely connected. This is expected as cell types are often defined to be an irreducible subunit of neuronal grouping [Dance, 2024,

Schlegel et al., 2024, Hobert et al., 2016]. Therefore, they must be functionally identical and so identically connected. The statistical significance of the pairwise entropic distances against various nulls of cell types and serial homologues contributes to providing comparative methodologies for cross-identifying and contrasting different connectome data sets [Schlegel et al., 2024, Stürner et al., 2024]. However, the specialisation-diversity values of functionally different neurons overlap considerably. While we are able to recover the broad cell classifications to be localised at low or high integrative or distributive specialisation-diversity values, it is insufficient to infer the specific finer grained categorisations from specialisation-diversity alone. Since we have reduced the network of annotated nodes into the two numbers quantifying the proximity to diversely specialised channels of connectivity, equivalently diverse connectivity may arise from different connectivity specialisation. This inverse inference problem we leave for future work, and approaches like graph neural networks [Corso et al., 2024] have demonstrated convincing evidence that node annotations supplement our inference of function in the underlying network. We also speculate more open questions about the nature of biological annotations complementing the network connectivity. Many different annotation schemes are available from coarse to fine-grained; node or edge metadata and their informative or predictive relationships with network structure are not fully understood [Bazinet et al., 2023a,b]. While our methodology can be extended to handle these, we do not yet know of which annotation schemes are most biologically and inferentially complementary for detecting specific functional attributes of the underlying system.

The connectomes we have used here also have limitations, for instance in terms of the incompleteness of annotations in *Drosophila*. Hemilineage annotations are inferred from morphology and connectivity and are incomplete for primary (embryonic) neurons which are functionally and morphologically distinct from secondary (post-embryonic) neurons. We ignore connectivity with unlabelled nodes in our methodology, but expect our conclusions to remain qualitatively robust since primary neurons arguably should be treated as a separate hemilineage subtype in our methodology because their functions and morphology are distinct to secondary neurons of the same lineage. Another limitation of the work is that our results are speculative and predictive rather than conclusive of function or activity which has been described in anatomical perspectives of neuroscience previously [Prinz et al., 2004]. Since our specialisation-diversity pools connectivity from all neurons that are reachable by or with a directed path of length two, we are compounding biological variability, or errors attributed to connectome imaging and reconstruction present in the data. Despite this, our analysis shows sufficient evidence that aggregating by connectivity produces statistically significant results to justify this study and so expect our conclusions be qualitatively similar with future data that arises from advances in the acquisition of connectomic data.

Data on the large-scale organisation of nervous systems has only recently become available to the neuroscientific community. Our primary aim was to provide an augmented perspective of network analysis that incorporates node annotations. To do this, we introduced an entropy-based measure, demonstrated its potential as an interpretable descriptor of neuronal function, and proposed roles for neurons not yet individually characterized experimentally. More generally specialisation-diversity is a local centrality measure in annotated networks, that can propose key nodes that are important for aggregating and disseminating diversely specialised channels of connectivity. While we have demonstrated our method using neuronal connectomes, it can therefore be applied to any complex network with node annotations.

4 Methods

4.1 Specialisation-diversity

To calculate the specialisation-diversity, we require normalized connectivity vectors describing the proportion of connectivity of a given neuron to each annotation label. We first establish the labels for all nodes we adhere to for connectivity aggregation. For given target node, we collect its partners whose labels are known. The frequency distribution of the node label types that are connected to the target node is then determined. We normalize by dividing by the total in or out strength of each frequency of node label, giving the discrete probability distribution of the in/out-connections to each of the labels. This removes biases of scale in the network, and addresses the in and out connections separately. The resulting vector has entries corresponding to proportion of contributions *in*-from or *out*-to each node label type.

Mathematically, let a_{iz} be the elements of an adjacency matrix, denoting the weight of a directed edge from neuron i to neuron z, with value zero if the edge does not exist. Let b_{zj} be another matrix element where $b_{zj} = 1$ if node z is labelled with annotation type x_j, (e.g. a specific hemilineage) and zero otherwise. Then the vector of annotation type contributions for the in and out connectivity becomes:

$$\nu_{ij}^{\text{out}} = \sum_z a_{iz} b_{zj}; \qquad \rho_{ij}^{\text{out}} = \frac{v_{ij}^{\text{out}}}{\sum_{j'} v_{ij'}^{\text{out}}}, \tag{1}$$

and the in-vector is the same, with a transposed adjacency matrix element (a_{zi}):

$$\nu_{ij}^{\text{in}} = \sum_z a_{zi} b_{zj}; \qquad \rho_{ij}^{\text{in}} = \frac{v_{ij}^{\text{in}}}{\sum_{j'} v_{ij'}^{\text{in}}}, \tag{2}$$

where the probability vectors (ρ) are normalized to sum to unity, and can write as $\mathbf{p}_i^{\alpha}$, where $\alpha \in [\text{in}, \text{out}]$ can denote an in-vector or out-vector in reference to node i. For clarity, let's drop α as the derivations are identical and discuss in and out directions later.

Notice that $\mathbf{p}_i$ is a probability distribution equivalent to $p^{(i)}(X)$ where X is a random variable that can take the values of the annotation types, $X \in \{x_1, x_2, ..., x_M\}$. We use the shorthand for probability for annotation type x_j as $p^{(i)}(X = x_j) = p_j^{(i)}$. Then, by using information entropy, we can measure specialisation in this particular probability distribution:

$$S^{(i)} = -\sum_{j=1}^{N} p_j^{(i)} \ln(p_j^{(i)}). \tag{3}$$

Entropy is maximized when all contributions p_j are equal, meaning the node is connected to all types of labels equally so it is not specialised. Entropy is minimal when the connectivity is concentrated rather than distributed, with the minimum entropy value being 0 where all connections are with one label type. This means that specialised nodes have low entropy.

Now, if we have many probability vectors, we can calculate how specialised they are individually and so compare how diversely specialised the probability vectors are. Let $P = \{p^{(1)}(X), p^{(2)}(X), ..., p^{(N)}(X)\}$ be a set of normalized connectivity probability distributions, where N is the number of distributions. Associated with the probability vectors, are each of their individual entropies: $S = \{s^{(1)}, s^{(2)}, ..., s^{(N)}\}$.

The average entropy is given by:

$$\bar{S} = \frac{1}{N} \sum_{i=1}^{N} s^{(i)}. \tag{4}$$

The average connectivity vector component $\bar{p}(X = x_j)$, written with a shorthand $\bar{p}_j$ is given by:

$$\bar{p}_j = \frac{1}{N}\sum_{i}^{N} p^{(i)}(X = x_j) \tag{5}$$

which can be demonstrated that it is also a probability distribution with the property $\sum_j \bar{p}_j = 1$ (see **Supplementary Information**). Thus, the entropy of the average vector can be calculated and meaningfully compared:

$$S(\bar{p}) = -\sum_{j=1}^{M} \bar{p}_j \ln(\bar{p}_j). \tag{6}$$

We are particularly interested in neurons are connected to other neurons with their own specific specialisms which are individually diverse. If a given neuron's connectivity is concentrated or specialised, then the entropy of this neuron will be low. If a group of neurons whose connectivity is specialised in the same node label types, then the average entropy, $\bar{S}$, will be low as well as the entropy of the average vector, $S(\bar{p})$. If a group of neurons whose connectivity is specialised in different node label types, the average entropy, $\bar{S}$, will be low, however, the average vector will not be specialised so the entropy of the average vector, $S(\bar{p})$, will be high. This is illustrated in figure 1.

If the average entropy is high, the group of nodes are unspecialised so the average vector must also be unspecialised and the entropy of the average vector is also high. With this, we observe that large differences between the average entropy and the entropy of the average vector mean that the group of neurons have specialisms in a variety of different node label types. If the entropy averages' differences are small, then it could mean that the group of neurons were not specialised to begin with, or that the group of neurons were all specialised in similar node label types (figure 1A, top two rows).

Using these measures, we demonstrate that a high difference in entropy of the average vector and the average entropy of each vector is only observed when the set of vectors are diversely specialised (figure 1A, bottom row). We arrive at the Δ_S, the *average entropies' differences* - the difference in the average entropy and the entropy of the average vector, which measures the *specialisation-diversity*:

$$\Delta_S = S(\bar{p}) - \bar{S}. \tag{7}$$

Entropy is a concave function, so negative entropy is a convex function. This means that negative entropy must obey Jensen's inequality for convex functions: $\mathbb{E}[\phi(X)] - \phi(\mathbb{E}(X)) \geq 0$, where $\mathbb{E}$ denotes the expected value and ϕ is a convex function on a probability distribution X. Therefore, substituting for $-S(p) = \phi(X)$, and $\bar{\cdot}$ for the expected value, $-\bar{S} + S(\bar{p}) \geq 0$, and so $\Delta_S \geq 0$ always. This is logically sound, as a set of probability vectors would not become more specialised than the individual constituent vectors through averaging. In other words, the entropy of the average vector would always be at least value of the average entropy of those vectors.

Notice that have two choices of the set of probability distributions P we wish to calculate the specialisation-diversity: which set of nodes $i = 0, ..., N$ to consider, and which direction the connectivity is aggregated ($\alpha \in [\text{in}, \text{out}]$ we dropped earlier). We are free to calculate a specialisation-diversity for any combinations of upstream or downstream partners; input or output vectors. We focus here only on the upstream inputs (integrative) and downstream outputs (distributive), and label them $\Delta_S^{U,in}$ and $\Delta_S^{D,out}$ respectively. In other words, for a given target node, we may collect its upstream partners' input vectors and we can calculate the specialisation-diversity associated to this (figure 1B, right). Secondly, we may use the downstream partners' output vectors of a given node to calculate the specialisation-diversity.

4.2 Drosophila VNC and C. elegans data sets

The *Drosophila* VNC connectome and associated annotations used in this paper is the Male Adult Nerve Cord (MANC) data available from https://neuprint.janelia.org/ publically available with a Google account (see Takemura et al. [2024]). The *C. elegans* connectome is a hermaphrodite data set from www.nemanode.org [Witvliet et al., 2021] by Varshney et al. [2011] (based on White et al. [1986]) and the associated meta data was compiled from Cook et al. [2019] available from https://www.wormatlas.org/neuronalwiring.html, and Ripoll-Sánchez et al. [2023].

Annotations that were used in this paper for the connectivity vectorisation for specialisation-diversity calculations were available from the data sources listed above. For *Drosophila*, we have hemilineage determined from light level data and morphology and connectivity clustering detailed in Marin et al. [2024]. The secondary neurons of hemilineages have been experimentally demonstrated to form behavioural phenotypes in independent of brain signalling [Harris et al., 2015]. Here, lineages are specified by neuronal precursor cells (ganglion mother cell) which then diverge via Notch signalling into two distinct phenotypes 'A' and 'B' which form the hemilineages. These hemilineages are the developmental units of organisation in *Drosophila* [Truman et al., 2010]. By calculating the specialisation-diversities with the hemilineage annotation scheme, we are interested in how diversely specialised are the connectivity channels that perform these functions. In the **Supplementary Information**, we investigate whether specialisation-diversity averaged for each hemilineage reveals neurons that are highly important as they are proximal for collecting information from upstream and dispersing information downstream of many different functions.

In *C. elegans* the annotations we use are cell classifications based on categorisations into groups by Cook et al. [2019], based on anatomy, chosen as it was a middle ground between the finest (118 cell types) versus the high-level cell classes (motor, sensory, interneuron etc.). Too fine types suffer from the curse of dimensionality, and sparse vectors are impacted by noise if it exists in the system. Too broad types will not distinguish the many neurons that are similar connectivity profiles. Exact determination of how optimal an annotation scheme is correlated to what interpretable property of the system is an open question.

4.3 Code availability

All files for `python` analysis for generating figures is documented and made available from https://github.com/sungsushi/spec_div. We include all data and processed data from this work in https://doi.org/10.5281/zenodo.15052465.

We utilise the `scipy.stats.ks_2samp` implementation of the two-sample Kolmogorov-Smirnov test.

Acknowledgments

We thank Gregory S. X. E. Jefferis, Philipp Schlegel, Tomke Stürner and Elizabeth Marin for discussions and providing access to data. We also thank William Schafer, Petra Vértes and Lidia Ripoll-Sánchez for discussions and access to data.

S.S.M. acknowledges the support of the Harding Distinguished Postgraduate Scholars Programme Leverage Scheme and the EPSRC DTP.

Competing Interests

Authors declare no competing interests.

References

H. Bashiri, H. Rahmani, V. Bashiri, D. Módos, and A. Bender. EMDIP: An Entropy Measure to Discover Important Proteins in PPI networks. *Computers in Biology and Medicine*, 120:103740, May 2020. ISSN 0010-4825. doi: 10.1016/j.compbiomed.2020.103740. URL https://www.sciencedirect.com/science/article/pii/S0010482520301207.

D. S. Bassett and O. Sporns. Network neuroscience. *Nature Neuroscience*, 20(3):353–364, Mar. 2017. ISSN 1546-1726. doi: 10.1038/nn.4502. URL https://www.nature.com/articles/nn.4502. Publisher: Nature Publishing Group.

V. Bazinet, J. Y. Hansen, and B. Misic. Towards a biologically annotated brain connectome. *Nature Reviews Neuroscience*, 24(12):747–760, Dec. 2023a. ISSN 1471-0048. doi: 10.1038/s41583-023-00752-3. URL https://www.nature.com/articles/s41583-023-00752-3. Publisher: Nature Publishing Group.

V. Bazinet, J. Y. Hansen, R. Vos de Wael, B. C. Bernhardt, M. P. van den Heuvel, and B. Misic. Assortative mixing in micro-architecturally annotated brain connectomes. *Nature Communications*, 14(1):2850, May 2023b. ISSN 2041-1723. doi: 10.1038/s41467-023-38585-4. URL https://www.nature.com/articles/s41467-023-38585-4. Publisher: Nature Publishing Group.

B. Bentley, R. Branicky, C. L. Barnes, Y. L. Chew, E. Yemini, E. T. Bullmore, P. E. Vértes, and W. R. Schafer. The Multilayer Connectome of Caenorhabditis elegans. *PLOS Computational Biology*, 12(12):e1005283, Dec. 2016. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1005283. URL https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005283. Publisher: Public Library of Science.

M. A. Bertolero, B. T. T. Yeo, and M. D'Esposito. The diverse club. *Nature Communications*, 8(1):1277, Nov. 2017. ISSN 2041-1723. doi: 10.1038/s41467-017-01189-w. URL https://www.nature.com/articles/s41467-017-01189-w. Number: 1 Publisher: Nature Publishing Group.

R. Betzel, M. G. Puxeddu, and C. Seguin. Hierarchical communities in the larval Drosophila connectome: Links to cellular annotations and network topology. *Proceedings of the National Academy of Sciences*, 121(38):e2320177121, Sept. 2024. doi: 10.1073/pnas.2320177121. URL https://www.pnas.org/doi/abs/10.1073/pnas.2320177121. Publisher: Proceedings of the National Academy of Sciences.

J. Buhmann, A. Sheridan, C. Malin-Mayor, P. Schlegel, S. Gerhard, T. Kazimiers, R. Krause, T. M. Nguyen, L. Heinrich, W.-C. A. Lee, R. Wilson, S. Saalfeld, G. S. X. E. Jefferis, D. D. Bock, S. C. Turaga, M. Cook, and J. Funke. Automatic detection of synaptic partners in a whole-brain Drosophila electron microscopy data set. *Nature Methods*, 18(7):771–774, July 2021. ISSN 1548-7105. doi: 10.1038/s41592-021-01183-7. URL https://www.nature.com/articles/s41592-021-01183-7. Publisher: Nature Publishing Group.

E. Bullmore and O. Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3):186–198, Mar. 2009. ISSN 1471-0048. doi: 10.1038/nrn2575. URL https://www.nature.com/articles/nrn2575. Publisher: Nature Publishing Group.

H. S. Cheong, K. Eichler, T. Stürner, S. K. Asinof, A. S. Champion, E. C. Marin, T. B. Oram, M. Sumathipala, L. Venkatasubramanian, S. Namiki, I. Siwanowicz, M. Costa,

S. Berg, J. F. P. Team, G. S. Jefferis, and G. M. Card. Transforming descending input into behavior: The organization of premotor circuits in the Drosophila Male Adult Nerve Cord connectome. *eLife*, 13, Mar. 2024. doi: 10.7554/eLife.96084.1. URL https://elifesciences.org/reviewed-preprints/96084. Publisher: eLife Sciences Publications Limited.

S.-H. Chou, Y.-J. Chen, C.-P. Liao, and C.-L. Pan. A role for dopamine in *C. elegans* avoidance behavior induced by mitochondrial stress. *Neuroscience Research*, 178:87–92, May 2022. ISSN 0168-0102. doi: 10.1016/j.neures.2022.01.005. URL https://www.sciencedirect.com/science/article/pii/S0168010222000050.

S. J. Cook, T. A. Jarrell, C. A. Brittin, Y. Wang, A. E. Bloniarz, M. A. Yakovlev, K. C. Q. Nguyen, L. T.-H. Tang, E. A. Bayer, J. S. Duerr, H. E. Bülow, O. Hobert, D. H. Hall, and S. W. Emmons. Whole-animal connectomes of both Caenorhabditis elegans sexes. *Nature*, 571(7763):63–71, July 2019. ISSN 1476-4687. doi: 10.1038/s41586-019-1352-7. URL https://www.nature.com/articles/s41586-019-1352-7. Publisher: Nature Publishing Group.

G. Corso, H. Stark, S. Jegelka, T. Jaakkola, and R. Barzilay. Graph neural networks. *Nature Reviews Methods Primers*, 4(1):1–13, Mar. 2024. ISSN 2662-8449. doi: 10.1038/s43586-024-00294-7. URL https://www.nature.com/articles/s43586-024-00294-7. Publisher: Nature Publishing Group.

R. Court, S. Namiki, J. D. Armstrong, J. Börner, G. Card, M. Costa, M. Dickinson, C. Duch, W. Korff, R. Mann, D. Merritt, R. K. Murphey, A. M. Seeds, T. Shirangi, J. H. Simpson, J. W. Truman, J. C. Tuthill, D. W. Williams, and D. Shepherd. A Systematic Nomenclature for the Drosophila Ventral Nerve Cord. *Neuron*, 107(6):1071–1079.e2, Sept. 2020. ISSN 0896-6273. doi: 10.1016/j.neuron.2020.08.005. URL https://www.cell.com/neuron/abstract/S0896-6273(20)30612-7. Publisher: Elsevier.

A. Dance. What is a cell type, really? The quest to categorize life's myriad forms. *Nature*, 633(8031):754–756, Sept. 2024. doi: 10.1038/d41586-024-03073-2. URL https://www.nature.com/articles/d41586-024-03073-2. Bandiera_abtest: a Cg_type: News Feature Publisher: Nature Publishing Group Subject_term: Cell biology, Bioinformatics, Molecular biology, Scientific community.

F. Delaplace. Entropic detection of chromatic community structures. *Journal of Complex Networks*, 13(1):cnae051, Feb. 2025. ISSN 2051-1329. doi: 10.1093/comnet/cnae051. URL https://doi.org/10.1093/comnet/cnae051.

S. Dorkenwald, A. Matsliah, A. R. Sterling, P. Schlegel, S.-c. Yu, C. E. McKellar, A. Lin, M. Costa, K. Eichler, Y. Yin, W. Silversmith, C. Schneider-Mizell, C. S. Jordan, D. Brittain, A. Halageri, K. Kuehner, O. Ogedengbe, R. Morey, J. Gager, K. Kruk, E. Perlman, R. Yang, D. Deutsch, D. Bland, M. Sorek, R. Lu, T. Macrina, K. Lee, J. A. Bae, S. Mu, B. Nehoran, E. Mitchell, S. Popovych, J. Wu, Z. Jia, M. A. Castro, N. Kemnitz, D. Ih, A. S. Bates, N. Eckstein, J. Funke, F. Collman, D. D. Bock, G. S. X. E. Jefferis, H. S. Seung, and M. Murthy. Neuronal wiring diagram of an adult brain. *Nature*, 634(8032):124–138, Oct. 2024. ISSN 1476-4687. doi: 10.1038/s41586-024-07558-y. URL https://www.nature.com/articles/s41586-024-07558-y. Publisher: Nature Publishing Group.

N. Eagle, M. Macy, and R. Claxton. Network Diversity and Economic Development. *Science*, 328(5981):1029–1031, May 2010. doi: 10.1126/science.1186605. URL https://www.science.org/doi/10.1126/science.1186605. Publisher: American Association for the Advancement of Science.

N. Eckstein, A. S. Bates, A. Champion, M. Du, Y. Yin, P. Schlegel, A. K.-Y. Lu, T. Rymer, S. Finley-May, T. Paterson, R. Parekh, S. Dorkenwald, A. Matsliah, S.-C. Yu, C. McKellar, A. Sterling, K. Eichler, M. Costa, S. Seung, M. Murthy, V. Hartenstein, G. S. X. E. Jefferis, and J. Funke. Neurotransmitter classification from electron microscopy images at synaptic sites in *Drosophila melanogaster*. *Cell*, 187(10): 2574–2594.e23, May 2024. ISSN 0092-8674. doi: 10.1016/j.cell.2024.03.016. URL https://www.sciencedirect.com/science/article/pii/S0092867424003076.

R. Guimerà and L. A. N. Amaral. Cartography of complex networks: modules and universal roles. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(02):P02001, Feb. 2005. ISSN 1742-5468. doi: 10.1088/1742-5468/2005/02/P02001. URL https://dx.doi.org/10.1088/1742-5468/2005/02/P02001.

R. Guimerà and L. A. Nunes Amaral. Functional cartography of complex metabolic networks. *Nature*, 433(7028):895–900, Feb. 2005. ISSN 1476-4687. doi: 10.1038/nature03288. URL https://www.nature.com/articles/nature03288. Number: 7028 Publisher: Nature Publishing Group.

R. M. Harris, B. D. Pfeiffer, G. M. Rubin, and J. W. Truman. Neuron hemilineages provide the functional ground plan for the Drosophila ventral nervous system. *eLife*, 4:e04493, July 2015. ISSN 2050-084X. doi: 10.7554/eLife.04493. URL https://doi.org/10.7554/eLife.04493. Publisher: eLife Sciences Publications, Ltd.

G. Haszprunar. The types of homology and their significance for evolutionary biology and phylogenetics. *Journal of Evolutionary Biology*, 5(1):13–24, Jan. 1992. ISSN 1010-061X, 1420-9101. doi: 10.1046/j.1420-9101.1992.5010013.x. URL https://academic.oup.com/jeb/article/5/1/13-24/7322356.

O. Hobert, L. Glenwinkel, and J. White. Revisiting Neuronal Cell Type Classification in *Caenorhabditis elegans*. *Current Biology*, 26(22):R1197–R1203, Nov. 2016. ISSN 0960-9822. doi: 10.1016/j.cub.2016.10.027. URL https://www.sciencedirect.com/science/article/pii/S096098221631212X.

B. K. Hulse, H. Haberkern, R. Franconville, D. Turner-Evans, S.-y. Takemura, T. Wolff, M. Noorman, M. Dreher, C. Dan, R. Parekh, A. M. Hermundstad, G. M. Rubin, and V. Jayaraman. A connectome of the Drosophila central complex reveals network motifs suitable for flexible navigation and context-dependent action selection. *eLife*, 10:e66039, Oct. 2021. ISSN 2050-084X. doi: 10.7554/eLife.66039. URL https://doi.org/10.7554/eLife.66039. Publisher: eLife Sciences Publications, Ltd.

J. Janssens, S. Aibar, I. I. Taskiran, J. N. Ismail, A. E. Gomez, G. Aughey, K. I. Spanier, F. V. De Rop, C. B. González-Blas, M. Dionne, K. Grimes, X. J. Quan, D. Papasokrati, G. Hulselmans, S. Makhzami, M. De Waegeneer, V. Christiaens, T. Southall, and S. Aerts. Decoding gene regulation in the fly brain. *Nature*, 601 (7894):630–636, Jan. 2022. ISSN 1476-4687. doi: 10.1038/s41586-021-04262-z. URL https://www.nature.com/articles/s41586-021-04262-z. Publisher: Nature Publishing Group.

T. A. Jarrell, Y. Wang, A. E. Bloniarz, C. A. Brittin, M. Xu, J. N. Thomson, D. G. Albertson, D. H. Hall, and S. W. Emmons. The Connectome of a Decision-Making Neural Network. *Science*, 337(6093):437–444, July 2012. doi: 10.1126/science.1221762. URL https://www.science.org/doi/10.1126/science.1221762. Publisher: American Association for the Advancement of Science.

W. Li, Z. Feng, P. W. Sternberg, and X. Z. Shawn Xu. A C. elegans stretch receptor neuron revealed by a mechanosensitive TRP channel homologue. *Nature*, 440(7084): 684–687, Mar. 2006. ISSN 1476-4687. doi: 10.1038/nature04538. URL https://www.nature.com/articles/nature04538. Publisher: Nature Publishing Group.

A. Lin, R. Yang, S. Dorkenwald, A. Matsliah, A. R. Sterling, P. Schlegel, S.-c. Yu, C. E. McKellar, M. Costa, K. Eichler, A. S. Bates, N. Eckstein, J. Funke, G. S. X. E. Jefferis, and M. Murthy. Network statistics of the whole-brain connectome of Drosophila. *Nature*, 634(8032):153–165, Oct. 2024. ISSN 1476-4687. doi: 10.1038/s41586-024-07968-y. URL https://www.nature.com/articles/s41586-024-07968-y. Publisher: Nature Publishing Group.

E. Z. Macosko, N. Pokala, E. H. Feinberg, S. H. Chalasani, R. A. Butcher, J. Clardy, and C. I. Bargmann. A hub-and-spoke circuit drives pheromone attraction and social behaviour in C. elegans. *Nature*, 458(7242):1171–1175, Apr. 2009. ISSN 1476-4687. doi: 10.1038/nature07886. URL https://www.nature.com/articles/nature07886. Publisher: Nature Publishing Group.

E. C. Marin, B. J. Morris, T. Stürner, A. S. Champion, D. Krzeminski, G. Badalamente, M. Gkantia, C. R. Dunne, K. Eichler, S.-y. Takemura, I. F. Tamimi, S. Fang, S. S. Moon, H. S. Cheong, F. Li, P. Schlegel, S. E. Ahnert, S. Berg, J. F. P. Team, G. M. Card, M. Costa, D. Shepherd, and G. S. Jefferis. Systematic annotation of a complete adult male Drosophila nerve cord connectome reveals principles of functional organisation. *eLife*, 13, July 2024. doi: 10.7554/eLife.97766.1. URL https://elifesciences.org/reviewed-preprints/97766. Publisher: eLife Sciences Publications Limited.

A. Matsumoto, Y. Toyoshima, C. Zhang, A. Isozaki, K. Goda, and Y. Iino. Neuronal sensorimotor integration guiding salt concentration navigation in Caenorhabditis elegans. *Proceedings of the National Academy of Sciences*, 121(5):e2310735121, Jan. 2024. doi: 10.1073/pnas.2310735121. URL https://www.pnas.org/doi/10.1073/pnas.2310735121. Publisher: Proceedings of the National Academy of Sciences.

R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network Motifs: Simple Building Blocks of Complex Networks. *Science*, 298(5594):824–827, Oct. 2002. doi: 10.1126/science.298.5594.824. URL https://www.science.org/doi/10.1126/science.298.5594.824. Publisher: American Association for the Advancement of Science.

Y. M. Omar and P. Plapper. A Survey of Information Entropy Metrics for Complex Networks. *Entropy*, 22(12):1417, Dec. 2020. ISSN 1099-4300. doi: 10.3390/e22121417. URL https://www.mdpi.com/1099-4300/22/12/1417. Number: 12 Publisher: Multidisciplinary Digital Publishing Institute.

H. J. Pavlou, A. C. Lin, M. C. Neville, T. Nojima, F. Diao, B. E. Chen, B. H. White, and S. F. Goodwin. Neural circuitry coordinating male copulation. *eLife*, 5:e20713, Nov. 2016. ISSN 2050-084X. doi: 10.7554/eLife.20713. URL https://doi.org/10.7554/eLife.20713. Publisher: eLife Sciences Publications, Ltd.

M. Pedersen, A. Omidvarnia, J. M. Shine, G. D. Jackson, and A. Zalesky. Reducing the influence of intramodular connectivity in participation coefficient. *Network Neuroscience*, 4(2):416–431, Apr. 2020. ISSN 2472-1751. doi: 10.1162/netn_a_00127. URL https://doi.org/10.1162/netn_a_00127.

A. A. Prinz, D. Bucher, and E. Marder. Similar network activity from disparate circuit parameters. *Nature Neuroscience*, 7(12):1345–1352, Dec. 2004. ISSN 1546-1726. doi:

10.1038/nn1352. URL https://www.nature.com/articles/nn1352. Publisher: Nature Publishing Group.

S. Rengarajan, K. A. Yankura, M. L. Guillermin, W. Fung, and E. A. Hallem. Feeding state sculpts a circuit for sensory valence in Caenorhabditis elegans. *Proceedings of the National Academy of Sciences*, 116(5):1776–1781, Jan. 2019. doi: 10.1073/pnas.1807454116. URL https://www.pnas.org/doi/full/10.1073/pnas.1807454116. Publisher: Proceedings of the National Academy of Sciences.

L. Ripoll-Sánchez, J. Watteyne, H. Sun, R. Fernandez, S. R. Taylor, A. Weinreb, B. L. Bentley, M. Hammarlund, D. M. Miller, O. Hobert, I. Beets, P. E. Vértes, and W. R. Schafer. The neuropeptidergic connectome of *C. elegans*. *Neuron*, 111(22):3570–3589.e5, Nov. 2023. ISSN 0896-6273. doi: 10.1016/j.neuron.2023.09.043. URL https://www.sciencedirect.com/science/article/pii/S0896627323007560.

L. K. Scheffer, C. S. Xu, M. Januszewski, Z. Lu, S.-y. Takemura, K. J. Hayworth, G. B. Huang, K. Shinomiya, J. Maitlin-Shepard, S. Berg, J. Clements, P. M. Hubbard, W. T. Katz, L. Umayam, T. Zhao, D. Ackerman, T. Blakely, J. Bogovic, T. Dolafi, D. Kainmueller, T. Kawase, K. A. Khairy, L. Leavitt, P. H. Li, L. Lindsey, N. Neubarth, D. J. Olbris, H. Otsuna, E. T. Trautman, M. Ito, A. S. Bates, J. Goldammer, T. Wolff, R. Svirskas, P. Schlegel, E. Neace, C. J. Knecht, C. X. Alvarado, D. A. Bailey, S. Ballinger, J. A. Borycz, B. S. Canino, N. Cheatham, M. Cook, M. Dreher, O. Duclos, B. Eubanks, K. Fairbanks, S. Finley, N. Forknall, A. Francis, G. P. Hopkins, E. M. Joyce, S. Kim, N. A. Kirk, J. Kovalyak, S. A. Lauchie, A. Lohff, C. Maldonado, E. A. Manley, S. McLin, C. Mooney, M. Ndama, O. Ogundeyi, N. Okeoma, C. Ordish, N. Padilla, C. M. Patrick, T. Paterson, E. E. Phillips, E. M. Phillips, N. Rampally, C. Ribeiro, M. K. Robertson, J. T. Rymer, S. M. Ryan, M. Sammons, A. K. Scott, A. L. Scott, A. Shinomiya, C. Smith, K. Smith, N. L. Smith, M. A. Sobeski, A. Suleiman, J. Swift, S. Takemura, I. Talebi, D. Tarnogorska, E. Tenshaw, T. Tokhi, J. J. Walsh, T. Yang, J. A. Horne, F. Li, R. Parekh, P. K. Rivlin, V. Jayaraman, M. Costa, G. S. Jefferis, K. Ito, S. Saalfeld, R. George, I. A. Meinertzhagen, G. M. Rubin, H. F. Hess, V. Jain, and S. M. Plaza. A connectome and analysis of the adult Drosophila central brain. *eLife*, 9:e57443, Sept. 2020. ISSN 2050-084X. doi: 10.7554/eLife.57443. URL https://doi.org/10.7554/eLife.57443. Publisher: eLife Sciences Publications, Ltd.

P. Schlegel, M. Costa, and G. S. Jefferis. Learning from connectomics on the fly. *Current Opinion in Insect Science*, 24:96–105, Dec. 2017. ISSN 2214-5745. doi: 10.1016/j.cois.2017.09.011. URL https://www.sciencedirect.com/science/article/pii/S2214574517301578.

P. Schlegel, Y. Yin, A. S. Bates, S. Dorkenwald, K. Eichler, P. Brooks, D. S. Han, M. Gkantia, M. dos Santos, E. J. Munnelly, G. Badalamente, L. Serratosa Capdevila, V. A. Sane, A. M. C. Fragniere, L. Kiassat, M. W. Pleijzier, T. Stürner, I. F. M. Tamimi, C. R. Dunne, I. Salgarella, A. Javier, S. Fang, E. Perlman, T. Kazimiers, S. R. Jagannathan, A. Matsliah, A. R. Sterling, S.-c. Yu, C. E. McKellar, M. Costa, H. S. Seung, M. Murthy, V. Hartenstein, D. D. Bock, and G. S. X. E. Jefferis. Whole-brain annotation and multi-connectome cell typing of Drosophila. *Nature*, 634 (8032):139–152, Oct. 2024. ISSN 1476-4687. doi: 10.1038/s41586-024-07686-5. URL https://www.nature.com/articles/s41586-024-07686-5. Publisher: Nature Publishing Group.

D. Shepherd, R. Harris, D. W. Williams, and J. W. Truman. Postembryonic lineages of the Drosophila ventral nervous system: Neuroglian expression reveals the adult hemilineage

associated fiber tracts in the adult thoracic neuromeres. *Journal of Comparative Neurology*, 524(13):2677–2695, Sept. 2016. ISSN 1096-9861. doi: 10.1002/cne.23988. URL https://onlinelibrary.wiley.com/doi/10.1002/cne.23988. Publisher: John Wiley & Sons, Ltd.

O. Sporns and R. Kötter. Motifs in Brain Networks. *PLOS Biology*, 2(11):e369, Oct. 2004. ISSN 1545-7885. doi: 10.1371/journal.pbio.0020369. URL https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.0020369. Publisher: Public Library of Science.

T. Stürner, P. Brooks, L. S. Capdevila, B. J. Morris, A. Javier, S. Fang, M. Gkantia, S. Cachero, I. R. Beckett, A. S. Champion, I. Moitra, A. Richards, F. Klemm, L. Kugel, S. Namiki, H. S. J. Cheong, J. Kovalyak, E. Tenshaw, R. Parekh, P. Schlegel, J. S. Phelps, B. Mark, S. Dorkenwald, A. S. Bates, A. Matsliah, S.-c. Yu, C. E. McKellar, A. Sterling, S. Seung, M. Murthy, J. Tuthill, W.-C. A. Lee, G. M. Card, M. Costa, G. S. X. E. Jefferis, and K. Eichler. Comparative connectomics of the descending and ascending neurons of the Drosophila nervous system: stereotypy and sexual dimorphism, June 2024. URL https://www.biorxiv.org/content/10.1101/2024.06.04.596633v2. Pages: 2024.06.04.596633 Section: New Results.

S.-y. Takemura, K. J. Hayworth, G. B. Huang, M. Januszewski, Z. Lu, E. C. Marin, S. Preibisch, C. S. Xu, J. Bogovic, A. S. Champion, H. S. Cheong, M. Costa, K. Eichler, W. Katz, C. Knecht, F. Li, B. J. Morris, C. Ordish, P. K. Rivlin, P. Schlegel, K. Shinomiya, T. Stürner, T. Zhao, G. Badalamente, D. Bailey, P. Brooks, B. S. Canino, J. Clements, M. Cook, O. Duclos, C. R. Dunne, K. Fairbanks, S. Fang, S. Finley-May, A. Francis, R. George, M. Gkantia, K. Harrington, G. P. Hopkins, J. Hsu, P. M. Hubbard, A. Javier, D. Kainmueller, W. Korff, J. Kovalyak, D. Krzemiński, S. A. Lauchie, A. Lohff, C. Maldonado, E. A. Manley, C. Mooney, E. Neace, M. Nichols, O. Ogundeyi, N. Okeoma, T. Paterson, E. Phillips, E. M. Phillips, C. Ribeiro, S. M. Ryan, J. T. Rymer, A. K. Scott, A. L. Scott, D. Shepherd, A. Shinomiya, C. Smith, N. Smith, A. Suleiman, S. Takemura, I. Talebi, I. F. Tamimi, E. T. Trautman, L. Umayam, J. J. Walsh, T. Yang, G. M. Rubin, L. K. Scheffer, J. Funke, S. Saalfeld, H. F. Hess, S. M. Plaza, G. M. Card, G. S. Jefferis, and S. Berg. A Connectome of the Male Drosophila Ventral Nerve Cord. *eLife*, 13, May 2024. doi: 10.7554/eLife.97769.1. URL https://elifesciences.org/reviewed-preprints/97769. Publisher: eLife Sciences Publications Limited.

E. K. Towlson, P. E. Vértes, S. E. Ahnert, W. R. Schafer, and E. T. Bullmore. The Rich Club of the C. elegans Neuronal Connectome. *Journal of Neuroscience*, 33(15): 6380–6387, Apr. 2013. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.3784-12.2013. URL https://www.jneurosci.org/content/33/15/6380. Publisher: Society for Neuroscience Section: Articles.

J. W. Truman, W. Moats, J. Altman, E. C. Marin, and D. W. Williams. Role of Notch signaling in establishing the hemilineages of secondary neurons in Drosophila melanogaster. *Development*, 137(1):53–61, Jan. 2010. ISSN 0950-1991. doi: 10.1242/dev.041749. URL https://doi.org/10.1242/dev.041749.

L. R. Varshney, B. L. Chen, E. Paniagua, D. H. Hall, and D. B. Chklovskii. Structural Properties of the Caenorhabditis elegans Neuronal Network. *PLOS Computational Biology*, 7(2):e1001066, Feb. 2011. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1001066. URL https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1001066. Publisher: Public Library of Science.

L. Venkatasubramanian and R. S. Mann. The development and assembly of the *Drosophila* adult ventral nerve cord. *Current Opinion in Neurobiology*, 56:135–143, June 2019. ISSN 0959-4388. doi: 10.1016/j.conb.2019.01.013. URL https://www.sciencedirect.com/science/article/pii/S0959438819300170.

J. T. Vogelstein, E. W. Bridgeford, B. D. Pedigo, J. Chung, K. Levin, B. Mensh, and C. E. Priebe. Connectal coding: discovering the structures linking cognitive phenotypes to individual histories. *Current Opinion in Neurobiology*, 55:199–212, Apr. 2019. ISSN 0959-4388. doi: 10.1016/j.conb.2019.04.005. URL https://www.sciencedirect.com/science/article/pii/S0959438818301430.

J. G. White, E. Southgate, J. N. Thomson, and S. Brenner. The structure of the nervous system of the nematode caenorhabditis elegans. *Philos Trans R Soc Lond B Biol Sci*, 314(1165):1–340, 1986.

M. Winding, B. D. Pedigo, C. L. Barnes, H. G. Patsolic, Y. Park, T. Kazimiers, A. Fushiki, I. V. Andrade, A. Khandelwal, J. Valdes-Aleman, F. Li, N. Randel, E. Barsotti, A. Correia, R. D. Fetter, V. Hartenstein, C. E. Priebe, J. T. Vogelstein, A. Cardona, and M. Zlatic. The connectome of an insect brain. *Science*, 379(6636):eadd9330, Mar. 2023. doi: 10.1126/science.add9330. URL https://www.science.org/doi/10.1126/science.add9330. Publisher: American Association for the Advancement of Science.

E. Witham, C. Comunian, H. Ratanpal, S. Skora, M. Zimmer, and S. Srinivasan. *C. elegans* Body Cavity Neurons Are Homeostatic Sensors that Integrate Fluctuations in Oxygen Availability and Internal Nutrient Reserves. *Cell Reports*, 14(7):1641–1654, Feb. 2016. ISSN 2211-1247. doi: 10.1016/j.celrep.2016.01.052. URL https://www.sciencedirect.com/science/article/pii/S2211124716300304.

D. Witvliet, B. Mulcahy, J. K. Mitchell, Y. Meirovitch, D. R. Berger, Y. Wu, Y. Liu, W. X. Koh, R. Parvathala, D. Holmyard, R. L. Schalek, N. Shavit, A. D. Chisholm, J. W. Lichtman, A. D. T. Samuel, and M. Zhen. Connectomes across development reveal principles of brain maturation. *Nature*, 596(7871):257–261, Aug. 2021. ISSN 1476-4687. doi: 10.1038/s41586-021-03778-8. URL https://www.nature.com/articles/s41586-021-03778-8. Publisher: Nature Publishing Group.

Z. Zheng, J. S. Lauritzen, E. Perlman, C. G. Robinson, M. Nichols, D. Milkie, O. Torrens, J. Price, C. B. Fisher, N. Sharifi, S. A. Calle-Schuler, L. Kmecova, I. J. Ali, B. Karsh, E. T. Trautman, J. A. Bogovic, P. Hanslovsky, G. S. X. E. Jefferis, M. Kazhdan, K. Khairy, S. Saalfeld, R. D. Fetter, and D. D. Bock. A Complete Electron Microscopy Volume of the Brain of Adult *Drosophila melanogaster*. *Cell*, 174 (3):730–743.e22, July 2018. ISSN 0092-8674. doi: 10.1016/j.cell.2018.06.019. URL https://www.sciencedirect.com/science/article/pii/S0092867418307876.

Supplementary Information

Proof that the category-wise mean of a set of discrete probability distributions is also a probability distribution

Let $P = \{p^{(1)}(X), p^{(2)}(X), ..., p^{(N)}(X)\}$ be a set of probability distributions, where N is the number of distributions and x is the discrete random variable $X \in \{x_1, x_2, ..., x_M\}$.

By definition, the sum of the ith probability distribution adds to one:

$$\sum_{j}^{M} p^{(i)}(X = x_j) = 1, \quad 0 \leq i \leq N, \quad 0 \leq j \leq M.$$

We take the category-wise mean of this set of probability distributions:

$$\bar{p}(X = x_j) = \frac{1}{N} \sum_{i}^{N} p^{(i)}(X = x_j)$$

which is the probability of x averaged over all the distributions. $\bar{p}$ must be positive and bounded by the interval $[0, 1]$ since the constituent probability distributions are also positive and bounded for all i, j.

To check that this distribution indeed does sum to unity:

$$\sum_{j}^{M} \bar{p}(X = x_j) = \sum_{j}^{M} \left(\frac{1}{N} \sum_{i}^{N} p^{(i)}(X = x_j) \right) = \frac{1}{N} \sum_{i}^{N} \left(\sum_{j}^{M} p^{(i)}(X = x_j) \right) = \frac{1}{N} \sum_{i}^{N} 1 = 1$$

by using the previous definition of the probability vectors. Therefore, the category-wise mean of a set of discrete probability distributions is also a probability distribution.

Alternate form for Δ_S

Here, we derive an alternate form for the specialisation-diversity which is helpful for implementation purposes and finding bounds on the metric.

From the definitions of $\bar{S}$, $S(\bar{p})$ and Δ_S in equations 4, 6 and 7:

$$\Delta_S = -\sum_{j=1}^{M} \bar{p}_j \ln(\bar{p}_j) - \frac{1}{N} \sum_{i=1}^{N} s^{(i)} = -\sum_{j=1}^{M} \left(\frac{1}{N} \sum_{l=1}^{N} p_j^{(l)} \right) \ln\left(\frac{1}{N} \sum_{q=1}^{N} p_j^{(q)} \right) + \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{M} p_k^{(i)} \ln(p_k^{(i)}) \tag{8}$$

The first term can be simplified as follows:

$$-\sum_{j=1}^{M} \left(\frac{1}{N} \sum_{l=1}^{N} p_j^{(l)} \right) \ln\left(\frac{1}{N} \sum_{q=1}^{N} p_j^{(q)} \right) = \frac{-1}{N} \sum_{l=1}^{N} \sum_{j=1}^{M} p_j^{(l)} \left[-\ln(N) + \ln\left(\sum_{q=1}^{N} p_j^{(q)} \right) \right] \tag{9}$$

$$= \frac{\ln(N)}{N} \sum_{l=1}^{N} \sum_{j=1}^{M} p_j^{(l)} - \frac{1}{N} \sum_{l=1}^{N} \sum_{j=1}^{M} p_j^{(l)} \ln\left(\sum_{q=1}^{N} p_j^{(q)} \right) = \ln(N) - \frac{1}{N} \sum_{l=1}^{N} \sum_{j=1}^{M} p_j^{(l)} \ln\left(\sum_{q=1}^{N} p_j^{(q)} \right) \tag{10}$$

where we invoke that the category-wise average of a probability distribution is itself a probability distribution and must sum to one.

Now, substituting this back to equation 8 :

$$\Delta_S = \ln(N) - \frac{1}{N} \sum_{l=1}^{N} \sum_{j=1}^{M} p_j^{(l)} \ln\left(\sum_{q=1}^{N} p_j^{(q)} \right) + \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{M} p_k^{(i)} \ln(p_k^{(i)}), \tag{11}$$

we may rewrite the sums over the labels and the sums over the vectors (i.e. setting $j = k$ and $i = l$)

$$\Delta_S = \ln(N) + \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M}\left[-p_j^{(i)}\ln\left(\sum_{q=1}^{N} p_j^{(q)}\right) + p_j^{(i)}\ln(p_j^{(i)})\right]. \tag{12}$$

Then, writing it as one logarithm:

$$\Delta_S = \ln(N) + \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} p_j^{(i)}\ln\left[\frac{p_j^{(i)}}{\sum_{q=1}^{N} p_j^{(q)}}\right]. \tag{13}$$

We consider $p_j^{(i)}$ to be the probability of label j given that the distribution is the ith vector, and can write as a conditional probability: $p(j|i)$. And so in this interpretation, the term

$$\frac{p_j^{(i)}}{\sum_{q=1}^{N} p_j^{(q)}} = p(i|j), \tag{14}$$

as it is the probability of choosing vector i given that we consider label j. The notation is consistent with conditional probability obeying Bayes' rule.

Finally, we arrive at:

$$\Delta_S = \ln(N) + \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} p(j|i)\ln(p(i|j)). \tag{15}$$

Bounds on Δ_S

The lower bound for specialisation-diversity is zero due to Jensen's inequality. For an upper bound, consider equation 15. The second term is always negative since $p(j|i)\ln(p(i|j)) \leq 0$ as $0 \leq p(i|j) \leq 1$. (We make use of the limit $\lim_{x\to 0} x\ln(x) \to 0$, and by definition in equation 14, $p(j|i) = 0$ if $p(i|j) = 0$.) Consider the case where we have fewer vectors than categories, $N < M$. Δ_S is maximized when $p(j|i)\ln(p(i|j)) = 0$ for all pairs of i, j. Physically, this would mean maximally specialised ($p(i|j) = p(j|i) = 1$) in different categories, and the remaining entries being zero ($p(i|j) = p(j|i) = 0$). And so the maximum value for Δ_S is $\ln(N)$. When there are more vectors than categories ($N > M$), we must have a degeneracy in specialism even in the most diversely specialised scenario, thus imposes a stricter bound $\Delta_S < \ln(N)$. This would contribute to degree hubs (large N compared to M, the number of annotation types) being penalized in their specialisation-diversity value if their partners are degenerately specialised.

Hemilineage functionality is reflected in specialisation-diversity distributions

From Harris et al. [2015], we know which *Drosophila* hemilineages (secondary neurons) activate when certain functions are observed. These are: changes in posture, uncoordinated leg movement, walking, wing wave, wing buzz and take-off, and many hemilineages were activiated in more than one functions. We binarized these relationships, and sought to ask are the complexities in functionality, which is summarized in table 1.

When we averaged the integrative and distributive specialisation-diversities we found that even with the 'same' functional were very different in specialisation-diversity values (figure 4), suggesting that the functional categorisation may offer itself to further sub-roles that require different levels of diversely specialised connectivity channels.

Table 1: Hemilineages and their link to motor functions in the VNC in the absence of brain signalling from Harris et al. [2015]

Function	**Hemilineages**
Posture	03B, 23B, 12B, 13B, 05B, 20A.22A, 08A, 09A
Uncoordinated leg movement (`uc_leg_movement`)	06A, 06B, 19A, 21A, 23B
Walking	18B, 12A, 01A, 10B
Wing wave	12A, 03B
Wing buzz	07B, 11A, 11B, 02A, 18B, 12A
Take-off	07B, 11A, 11B, 18B

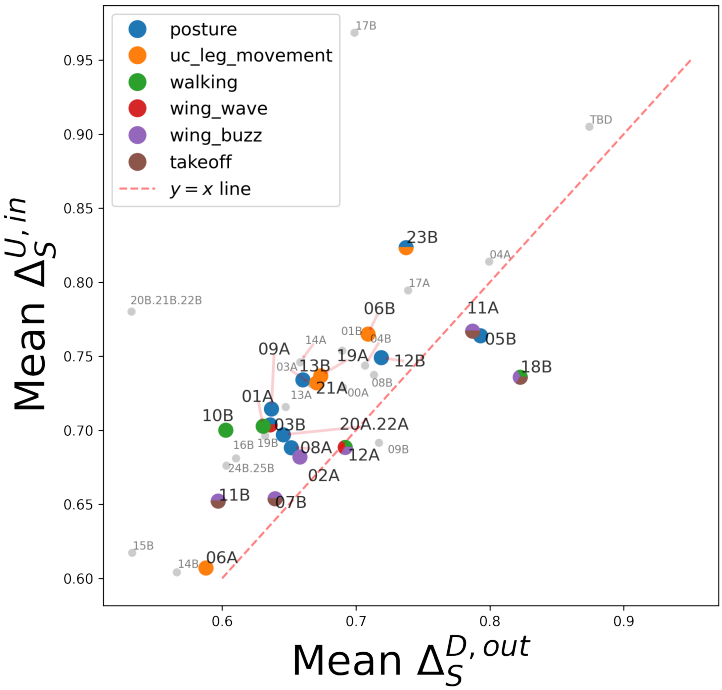

Figure 4: **Hemilineages with same function may have more specialised roles** Here, we plot the mean integrative vs distributive specialisation-diversities averaged across all secondary birthtime (post-embryonic) neurons of each hemilineage in the *Drosophila* VNC. We note each function by six colours in the legend, and split the marker colours when a hemilineage is associated to multiple functions. In grey are hemilineages not described in Harris et al. [2015]. The red dotted $y = x$ line is where points are equally distributive and integrative.