**THE EUROPEAN
PHYSICAL JOURNAL B**

Discovering motifs to fingerprint multi-layer networks: a case study on the connectome of C. Elegans

Deepak Sharma[1,a], Matthias Renz[1,b] , and Philipp Hövel[2,c]

[1] Department of Computer Science, Christian-Albrechts Universität zu Kiel, Christian-Albrechts-Platz 4, Kiel 24118, Germany
[2] Theoretical Physics and Center for Biophysics, Saarland University, Campus E2 6, Saarbrücken 66123, Germany

Abstract. Motif discovery is a powerful and insightful method to quantify network structures and explore their function. As a case study, we present a comprehensive analysis of regulatory motifs in the connectome of the model organism *Caenorhabditis elegans* (*C. elegans*). Leveraging the Efficient Subgraph Counting Algorithmic PackagE (ESCAPE) algorithm, we identify network motifs in the multi-layer nervous system of *C. elegans* and link them to functional circuits. We further investigate motif enrichment within signal pathways and benchmark our findings with random networks of similar size and link density. Our findings provide valuable insights into the organization of the nerve net of this well-documented organism and can be easily transferred to other species and disciplines alike.

1 Introduction

The nematode *Caenorhabditis elegans*, commonly known as *C. elegans*, has long been a prominent model organism in genetics and developmental biology research. Its relatively simple and well-defined anatomy, short life cycle, and fully sequenced genome make it an ideal subject for investigating fundamental biological processes [1,2]. Over the years, extensive studies on *C. elegans* have provided valuable insights into various aspects of cellular and molecular biology, contributing significantly to our understanding of eukaryotic biology [3–6].

In this study, we take a network science-driven angle [7] and focus on motif discovery in the nervous system of *C. elegans* to shed light on (i) the composition of the multi-layer connectome and (ii) the underlying regulatory circuits that govern the behavior of this model organism. The multi-layer nature arises from the presence of electrical (gap junction) and synaptic connections between the neurons. To achieve this, we adopt the *Efficient Subgraph Counting Algorithmic PackagE* (ESCAPE) algorithm [8], a powerful computational method that has proven effective in identifying

regulatory motifs in complex networks extracted from empirical data.

We focus on network motifs that emerge from sets of 4 or 5 nodes and their connecting links, such as triangles and cliques. Applying the ESCAPE algorithm to individual layers, the aggregated network, and the locomotory circuit, we compare the distribution of these small-scale, recurrent network structures to random networks of similar size and link density. This allows identification of a fingerprint of the structure of the nervous system. We hypothesize that the ESCAPE algorithm will uncover novel and functionally relevant motifs in *C. elegans* that play crucial roles for behaviors such as locomotion. By pursuing these objectives and testing our central hypothesis, we aim to provide valuable insights into the anatomical and functional landscape of the *C. elegans* connectome, exemplifying a universal approach to characterize network data in biology and beyond. This study holds the potential to advance our knowledge of networks in general and has implications in biological contexts in particular. Our approach is universally applicable and can be extended to network data (including directed links) from other sources and invites benchmarking against any network model. Insights into the motif distribution can also be used for artificial network generation for realistic *in silico* studies.

The rest of the manuscript is organized as follows: In Sect. 2, we elaborate on the workflow and data set considered. Section 3 presents our findings and highlights exemplary cases of significant and insignificant motif occurrences. Finally, we conclude in Sect. 4. We stress

---

[a] e-mail: des@informatik.uni-kiel.de
[b] e-mail: mr@informatik.uni-kiel.de
[c] e-mail: philipp.hoevel@uni-saarland.de (corresponding author)

Springer

that we focus on selected examples of motifs involving 3, 4, and 5 nodes in the main text and provide the full set of results in Appendices A–H

2 Methods and data

This section discusses the methodology applied to the *C. elegans* data for motif discovery. We start with a summary of the multi-layer network and then briefly review the ESCAPE framework [8].

2.1 Multi-layer network of C. elegans

The complexity of connectome structure in *C. elegans* extends beyond simple regulatory interactions, necessitating a comprehensive approach to capture the intricate relationships between anatomy and function. To achieve a more holistic understanding of the regulatory landscape, we consider a multi-layer network representation of *C. elegans*. This multi-layer network incorporates different layers of biological information, providing a powerful framework to investigate the interplay between neurons (network nodes), connections (network links), and their regulatory motifs (small, recurrent network structures). As discussed in the following, the multi-layer structure consists of (i) electrical connections and (ii) synapses that involve different neurotransmitters, monoamines, and peptides (see Fig. 1 for the different layers in a circular layout):

Acetylcholine Layer (ACh): The Acetylcholine (ACh) layer, the largest component of the network, encompasses approximately 33.4% of the total 3538 connections. It is distributed extensively across all neurons without showing a specific preference for any particular neuron type. In humans, ACh functions at skeletal neuromuscular junctions, interfaces between the vagus nerve and cardiac muscle fibers, as well as various locations within the central nervous system. Although ACh's role at neuromuscular junctions is well-documented, its precise function within the central nervous system remains a subject of incomplete understanding [9]. In the context of *C. elegans*, ACh is intricately involved in a multitude of behaviors, including but not limited to locomotion [10,11], egg-laying [12], feeding [13], and defecation [14].

Electrical Layer: Electrical transmission, comprising 29.1% of network connections, is the second-largest layer after ACh. Electrical synapses are universally present in nervous systems, allowing direct bidirectional flow of electrical current between neurons. This mode of transmission is exceptionally fast, virtually instantaneous, enabling communication without delay, a departure from typical chemical synapses. Electrical synapses are located where rapid coordination and synchronization of network activity are required. For example, certain brainstem neurons synchronize via electrical synapses for rhythmic breathing, and similar synchron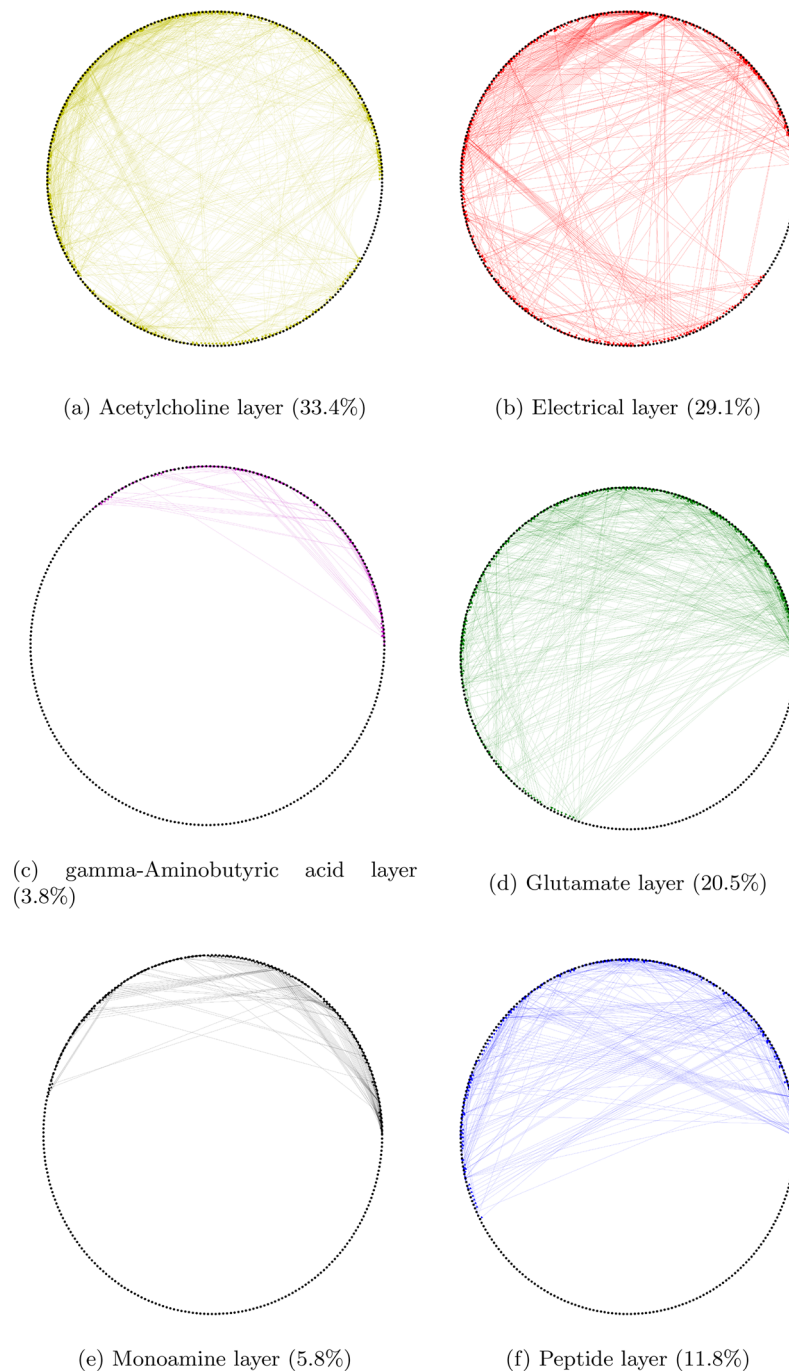ization occurs in interneuron populations in the cerebral cortex, thalamus, and cerebellum [9]. In *C. elegans*, electrical transmission plays a significant role in locomotion behavior and development [15,16].

Gamma-Aminobutyric acid Layer (GABA): The gamma-Aminobutyric acid (GABA) layer stands as the smallest component within the network, accounting for approximately 3.8% of the connections. Most of these connections are established by interneurons and motor neurons. GABA serves as the neurotransmitter for approximately one-third of the synapses in the brain and is predominantly found in interneurons within local circuits. Diverging from the excitatory nature of ACh and Glu, GABA exerts an inhibitory effect [9]. Notably, in *C. elegans*, GABA can also function as an excitatory transmitter, depending on the specific neuroreceptor involved. In its inhibitory role, GABA regulates head movements during foraging [17] and relaxes muscle cells during locomotion [18].

Glutamate Layer (Glu): The Glutamate (Glu) layer constitutes the third-largest component of the network, utilizing approximately 20.5% of its connections. In contrast to the ACh layer, this layer involves significantly fewer motor neurons. In the human brain, Glu plays a pivotal role as the primary neurotransmitter for normal brain function, with an estimated release from more than half of all brain synapses [9]. In the context of *C. elegans*, Glu contributes to various behavioral aspects, including foraging behavior [19], long-term memory processes [20], and spontaneous transitions from forward to backward movement, commonly referred to as reversals [21,22].

Monoamine Layer (MA): MA transmitters constitute approximately 5.8% of network connections, with no specific neuron-type preference although most motor neurons primarily receive postsynaptic signals. Among MA transmitters, dopamine accounts for about two-thirds. In humans, MA transmitters play a crucial role in regulating various brain functions, extending from central homeostatic processes to cognitive phenomena like attention. Dysfunctions in MA-mediated signal processing can lead to psychiatric disorders, exemplified by Parkinson's disease resulting from dopaminergic neuron degeneration [9]. In *C. elegans*, MA transmitters influence a range of behaviors, including egg-laying, pharyngeal pumping, locomotion, and learning, as detailed in [1]. For instance, dopamine modulates locomotion behavior and learning, facilitating the worms' responsiveness to environmental changes [23] and efficient exploration of new food sources [19]. Learning allows worms to adapt their behavior based on prior experiences, such as responding to non-localized mechanical stimuli like plate tapping by either moving backward or increasing forward locomotion, with repeated tapping leading to habituation and reduced reversal frequency [24].

Peptide Layer: Peptides account for approximately 11.8% of network connections, predominantly established by interneurons and sensory neurons, with most originating from the FMRFamide-like peptide (FLP)

(a) Acetylcholine layer (33.4%)

(b) Electrical layer (29.1%)

(c) gamma-Aminobutyric acid layer (3.8%)

(d) Glutamate layer (20.5%)

(e) Monoamine layer (5.8%)

(f) Peptide layer (11.8%)

Fig. 1 The *C. elegans* multi-layer network consists of nodes (circles) representing 279 neurons in a circular layout. In the aggregated network, there are 3538 distinct connections (lines) among the nodes. Panels (**a**)–(**f**) show the links of the different layers: Acetylcholine (yellow), electrical (red), gamma-Aminobutyric acid (magenta), Glutamate (green), Monoamine (black), and Peptide (blue) layers, respectively. The percentages refer to the proportion of the links in each layer

and peptide deformylase (PDF) families. Roughly 31% of these peptides act as neurotransmitters, while about 69% function as co-transmitters. In humans, peptides play roles in pain perception, emotional modulation, and complex responses to stress. Peptide co-transmission can modulate synaptic activity, impacting functions, such as food intake, metabolism, social

behavior, learning, and memory [9,25]. In *C. elegans*, peptides influence a wide array of behaviors, including locomotion, dauer-stage formation, egg-laying, sleep, learning, social behavior, mechano- and chemosensation. Notably, some FLP family peptides are involved in feeding behavior [26], while others, like nlp-22, regulate *C. elegans*' sleep-like state (lethargus) during larval

transitions[27]. Neuropeptides, constituting the third layer of information flow in neuronal communication alongside chemical and electrical transmission, are still not fully understood in *C. elegans*, with over 300 neuropeptides identified [28,29]. Similar to monoamines, neuropeptides form a small wired network but a vast wireless network[30], suggesting their potential involvement in all *C. elegans* behaviors.

In the subsequent sections, we present our findings based on this multi-layer network analysis, highlighting the discovery of regulatory motifs. This paves the way to understand their functional relevance, and the implications for understanding mechanisms governing the *C. elegans* nervous system. We anticipate that the insights gained from this comprehensive multi-layer network analysis will contribute significantly to the broader understanding of the *C. elegans* connectome and provide a valuable resource for the scientific community in further investigations that also invites the transfer of methodology to other model organisms and contexts.

2.2 Efficient subgraph counting algorithmic PackagE: ESCAPE

In a nutshell, the applied algorithmic framework ESCAPE revolves around breaking down complex patterns into smaller subpatterns and leveraging counts of these smaller patterns to obtain larger pattern counts. See Fig. 2 for a complete list of 3-, 4- and 5-node motif. The list also states their names, which are helpful for later identification. By employing this divide-and-conquer strategy, we can efficiently handle the enumeration and computation of subgraphs, even for large and complex networks. This approach allows us to avoid the excessive computational overhead associated with naively counting subgraphs. A key feature of the algorithm [8] is the exploitation of degree orientations within the network. By judiciously utilizing information about the degree distributions, we can further optimize runtime, resulting in faster computations. This optimization ensures that our algorithm can scale efficiently even for massive-sized networks.

2.3 Motif discovery algorithm

The enumeration of motifs, also called graphlets, particularly in the context of large-scale networks with millions of nodes and links, presents a formidable challenge due to the exponential growth in computational complexity. In the past, it was widely believed that subgraphs beyond three nodes were exceptionally difficult to enumerate, and conventional enumeration algorithms, which necessitate traversing each subgraph, were deemed impractical in terms of termination within a reasonable timeframe.

In response to the combinatorial explosion that hampers typical enumeration algorithms, the ESCAPE method was introduced by Pinar et al. [8]. This approach adopts a divide-and-conquer strategy, effec-

tively identifying substructures within each counting subgraph and partitioning them into smaller patterns. In other words, calculating all 5-node motifs counts requires enumerating only a small set of patterns. The corresponding formal framework *cuts* a pattern into smaller subpatterns, which is the main reason for its practical feasibility. A degree ordering applied to these subsets of fundamental patterns significantly cuts down the combinatorial expansion and results in a practical technique that is sufficient to count all 5-vertex patterns.

Despite its broad applicability, the method allows for tailored decomposition choices that facilitate the derivation of a set of formulas to compute the frequency of each subgraph efficiently. While the original paper delineates the resulting formulas for subgraphs up to size 5, diligent efforts can extend the application to larger sizes. For further details, we refer to Ref. [8].

To the best of our knowledge, ESCAPE stands as one of the most efficient algorithms available for counting undirected subgraphs and orbits up to size 5. Its strength lies in the ability to mitigate the exponential growth in complexity, enabling the enumeration of motifs even in sizable networks with millions of nodes and links. Through its divide-and-conquer paradigm and strategic substructure identification, ESCAPE opens new avenues for motif discovery, offering unprecedented opportunities to unravel recurring patterns within complex networks. Its scalability and performance make it a compelling choice for motif analysis in diverse domains, such as social networks, biological networks, and recommendation systems.

2.4 Proposed methodology

Next, we describe the stepwise methodology of the proposed work for *C. elegans* motif discovery. As schematically depicted in Fig. 3, the pipeline includes the following steps:

> **Step 1:** Input: Reading multi-layer network data and generation of random networks of the same size
> **Step 2:** Analysis: Application of the ESCAPE algorithm
> **Step 3:** Output: Motif distribution
> **Step 4:** Comparison: Benchmarking empirical against random networks
> **Step 5:** Comparison: Identification of significant motifs

The selection of random networks is one of many choices of network classes to generate artificial networks for benchmarking. Other prominent examples include regular, small-world, or scale-free networks, to name but a few. For the proper determination of a motif's subnetwork significance, we need comparison by an ensemble of a reference network class, e.g., random networks. Thus, the generation of this ensemble due to a given random network model is a necessary step of the algorithm. One of the popular random network models
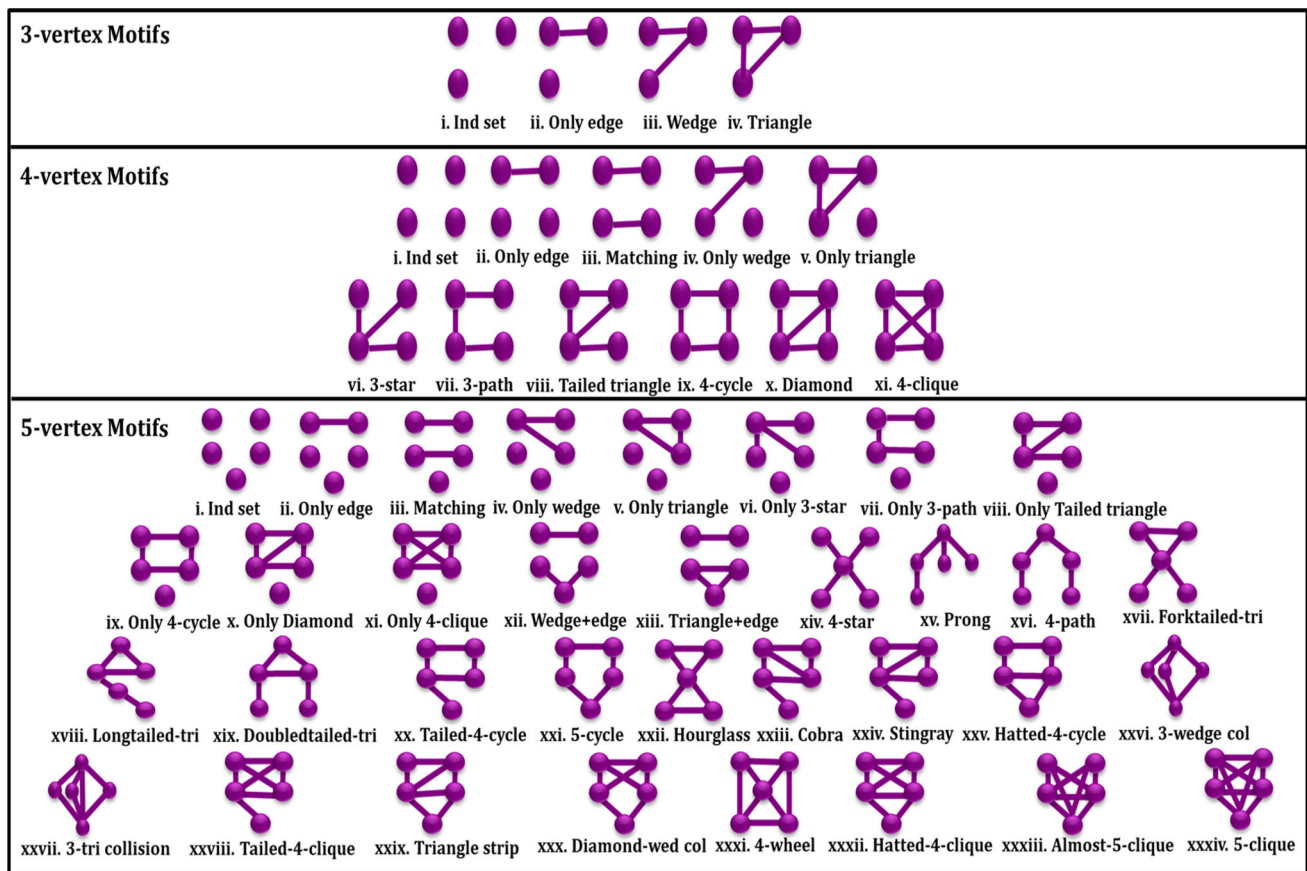
Fig. 2 Complete list of 3-, 4- and 5-node motifs including names for later identification
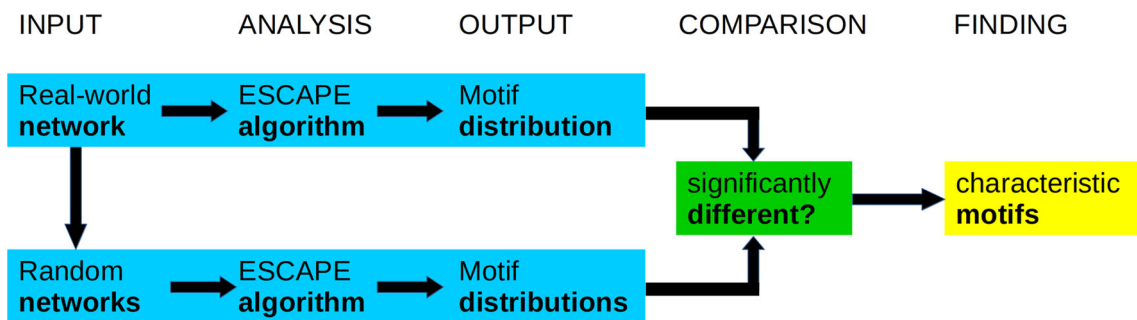

Fig. 3 Pipeline for the motif discovery

on which we also focused preserves the degree sequence of the original network. Note that there has been some research concerning the problem of subnetwork distribution within such networks for directed sparse random networks [31,32].

We follow a randomizing approach to generate reference networks of the same number of nodes and links. In that approach, similar to Milo's random model [31,32], switching operations are repeatedly applied on the randomly chosen vertices of the input network and their links, until the network is well randomized. By applying these switching operations, an ensemble of random networks is generated for comparing the real network to obtain the significance of each motif.

For the significance of the occurrence of each motif, statistical measures are introduced that lead us to the probable motifs in the input network: (i) *Frequency*: This is the simplest measurement for estimating the significance of a motif. The frequency is defined as the number of occurrences of a motif P in the network. (ii) *KS test*: We employ the Kolmogorov–Smirnov (KS) test to identify significant and insignificant motif frequencies by comparing the input network's motif frequency with those from 1000 generated random networks. First, the motif P frequency is counted in the input network. Then, the same motif is counted in each of the random networks, creating a distribution of motif frequencies for comparison. (iii) *p value*: The

Table 1 Data description: Number of nodes and links for the different layers, the aggregated network (All), and the locomotory circuit (LC)

	ACh	Electrical	GABA	Glu	MA	Peptides	All	LC
Nodes	258	253	102	197	131	162	279	83
Links	1145	1028	133	727	212	455	3700	435
Undirected links	1080	514	130	708	203	401	2287	435

Table 2 3-node non-induced motifs in *C. elegans* multi-layer network

	Motif	ACh	Electrical	GABA	Glu	MA	Peptides	All	LC
Ind set		2829056	2667126	171700	1254890	366145	695520	3580779	91881
Only link		276480	129014	13000	138060	26187	64160	633499	35235
Wedge		16848	3972	678	8035	1565	3028	56984	7056
Triangle		1083	170	18	376	28	125	4055	680

KS test compares the motif frequency distribution of the input network against the random networks' distributions, providing a p value to indicate significance. A low p value signifies that the motif frequency is significantly different from random, indicating the motif's significance, while a high p value indicates insignificance. Therefore, p values range from 0 to 1. The smaller the p value, the more significant the motif. Here, we consider a p value of 0.05, below which we declare the occurrence of a motif as significant, that is, characteristic.

3 Results and performance evaluation

This section provides the details of dataset and performance indicators used for the experiment analysis.

3.1 Data

The dataset [33] used in this study encompasses a comprehensive network representation of a biological system, capturing the interactions between various types of molecular entities. The data consists of nodes and links, representing different neuronal units and their connections within the biological system. The dataset contains a total of 279 nodes and 3,538 links. These nodes refer to neurons, each contributing to the overall functionality of the nervous system. The links in the network represent the connections or interactions between the neurons. A total of 3,538 links have been identified in the dataset across several layers (cf. Fig. 1), signifying the complex web of interactions within the biological system.

Table 1 provides an overview of the data, showcasing the different layers and the number of neurons involved as nodes in the network and the number of links that connect them. Each network layer corresponds to a specific neurotransmitter or signaling molecule involved in the biological system under investigation. The last two columns represent the total number of nodes in the aggregated network and the locomotory circuit, which is added as a functionally relevant subnetwork. In total, there are 3700 aggregated links, which amounts to 3,538 distinct connections between the neurons. Note that the electrical layer is—for reasons of neuro-physiology—bi-directional. All other layers are uni-directional. For the purpose of this study and to keep the combinatorics at a manageable level, we ignore directionality and treat all connections as undirected. The last row of Tab. 1 shows the number of links in the considered undirected network.

3.2 Network motifs

Next, we present the results of our analysis on the multi-layer network of *C. elegans* using the ESCAPE algorithm [8] for counting small subnetwork patterns. Specifically, we focus on motifs consisting of 3, 4, and 5 nodes and their prevalence across the different layers within the network. Additionally, we discuss the findings and implications of our motif counting analysis.

3.2.1 3-node motifs

Table 2 summarizes the number of occurrence of all possible motifs involving 3 nodes, from the *independent set* (3 nodes with no links) to the *triangle* (3-clique or

all-to-all connected set of 3 nodes). Obviously, the former is given be the binomial coefficient $\binom{N}{3}$, where N denotes the number of nodes in the layer (cf. Table 1). Evaluating the occurrences, it is important to note that a brute-force calculation of the motifs inevitably leads to double and multiple countings. For instance, a triangle automatically includes 3 motifs of *wedge* type. Here, network science/graph theory comes to our aid and provides the notion of non-induced and induced subnetworks.

As illustrated in Fig. 4, a *non-induced* subnetwork consists of a set of nodes, but not all possible edges that are present in the original network. In contrast, a motif is called *induced* if it contains the maximum set of links connecting its nodes in the original network. In the schematic example, the link between nodes 2 and 4 is missing in the non-induced subnetwork of nodes 1, 2, 3, and 4, while it is accounted for in the induced one. One can express the induced as a linear combination of the non-induced count and vice versa, that is, by multiplying the respective count vectors with a transformation matrix (or its inverse) that provides the combinatorics of lower-motif inclusions. For 3- and 4-node motifs, it is given by

$$
A_3 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad
A_4 = \begin{pmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
0 & 1 & 2 & 2 & 3 & 3 & 3 & 4 & 4 & 5 & 6 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 2 & 2 & 3 \\
0 & 0 & 0 & 1 & 3 & 3 & 2 & 5 & 4 & 8 & 12 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 2 & 4 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 2 & 4 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 4 & 6 & 12 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 4 & 12 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 3 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 6 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix},
$$
(1)

where the ordering of motifs corresponds to Fig. 2. For example, considering the last column of A_4, a *4-clique* gives rise to the following non-induced motifs: 6 *diamonds*, three *4 cycles*, 12 *tailed triangles* etc. For the case of connected 5-node motifs, that is, 21 out of the 34 motifs in Fig. 2, see the the 21×21 transformation matrix A_5 given in Appendix B of Ref. [8].

Therefore, it is helpful to focus on the occurrences of induced motifs, which is given in Table 3. That way, the count reflects subnetworks that are not part of a motif with the same set of nodes connected with more links. One can quantify their appearance via the ratio of induced to non-induced motifs. As shown in Table 4 that ratio is rather high, that is close to unity. For instance, considering the ACh layer, we find that 13599 *wedges* do not stem from *triangles* with one link left out. In fact, one can easily verify that the difference between the number of *wedges* of the induced to the non-induced case equals 3 times the number of triangles. In the following, we will consider only induced motifs. For 4- and 5-node motifs, which will be discussed in the subsequent sections, the tables for the non-induced case for

both the original and reference networks can be found in Appendices A,B,E and F.

3.2.2 4-node motifs

Figure 5 showcases the histogram of the counts of various 4-node motifs identified within the multi-layer network, classified by different molecular entity types, including ACh, Electrical, GABA, Glu, Monoamines, Peptides, and All (representing the total, aggregated network). Each motif type is represented by a corresponding graphical depiction. Note that the result of the independent set, only edge, and matching motifs are not shown. The respective numbers can be found in Appendix C.

To highlight some numbers: The motif *only edge* representing a wedge (an unconnected pair of nodes sharing a common neighbor) plus an independent node is prevalent across all molecular entity types, with counts ranging from 2856576 in the ACh layer to 52337 in the GABA layer. In the aggregated network, we observe a total count of more than 9 million *only wedge* motifs. The motifs *only triangle* (three nodes interconnected in a closed loop plus a single independent node) are also frequently observed across the network layers, with counts varying from 220488 in the ACh layer to 1415 in the GABA layer and an overall count of 823992 in the combined-layer network.

As the motif structure becomes more elaborate, the number of occurrences decreases. The *diamond* (two pairs of nodes connected by a central node) and *4-clique* (four fully interconnected nodes) motifs feature the least among all 4-node motifs. In fact, there is not a single occurrence of a *4-clique* in the GABA layer.

The ratios of the induced and non-induced 4-node motifs are given in Table 5 (See Appendix A for the absolute values of the non-induced motifs). We find that for sparser layers, such as Glu and MA, motifs with more links do not tend to be part of higher-order motifs, that is, they are induced. See, for instance, ratios of 0.92 (0.63) or 0.76 (0.25) for *3-star* and *4-cycle* motifs in the MA (ACh) layer.

The results of the occurrences of motifs in each layer compared to the aggregated network are summarized in Table 6. Here, one can see the difference between edges formed by electrical gap junctions and chemical synapses. Although the electrical layer is the second-largest layer, in terms of both number of nodes and edges, its contribution to the motif count is rather small. In other works, most of the motifs are found in the chemical layers. Furthermore, the ACh layer, while represented at almost 50% of the connections in the aggregated network (cf. Table 1) do not feature that prominently in the motif count. This means that many chemical links involve more than just a single neurotransmitter.

In order to identify the occurrences of significant motifs, we compare the motif distributions of the original network to 1000 random networks with the same number of nodes and edges and repeat the motif analy-
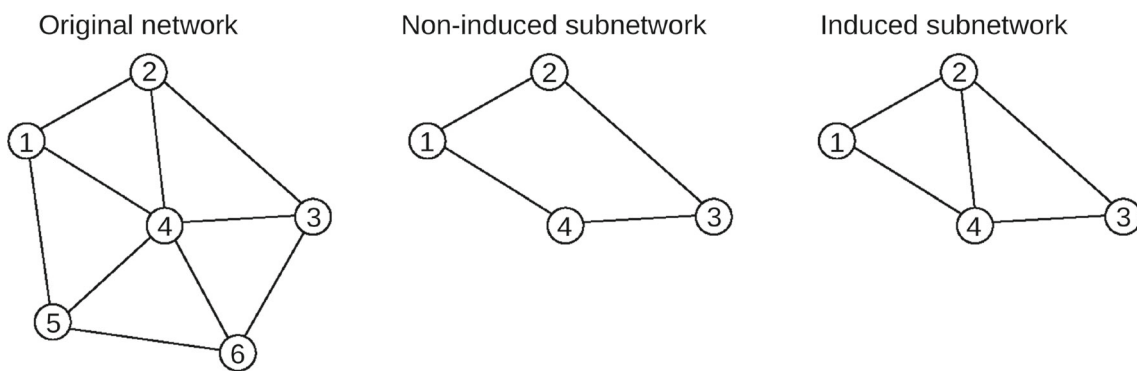
Fig. 4 Schematics of a 6-node network, a non-induced 4-node subnetwork, and an induced 4-node subnetwork. The latter consists of the maximum number of nodes and links taken from the original network

Table 3 3-node induced motifs in *C. elegans* multi-layer network

Motif	Motif	ACh	Electrical	GABA	Glu	MA	Peptides	All	LC
Ind set		2568341	2541914	159360	1124489	341495	634263	3000209	63022
Only link		246033	121580	11698	123118	23141	58479	531696	23163
Wedge		13599	3462	624	6907	1481	2653	44819	5016
Triangle		1083	170	18	376	28	125	4055	680

Table 4 Ratio of 3-node induced vs non-induced motifs in *C. elegans* multi-layer network

	Motif	ACh	Electrical	GABA	Glu	MA	Peptides	All	LC
Ind set		0.91	0.95	0.93	0.9	0.93	0.91	0.84	0.69
Only link		0.89	0.94	0.9	0.89	0.88	0.91	0.84	0.66
Wedge		0.81	0.87	0.92	0.86	0.95	0.88	0.79	0.71
Triangle		1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

sis. The resulting average counts of the different induced and non-induced 4-node motifs are given in Appendices B and D, respectively. As an exemplary illustration, Fig. 6 depicts the case of a significant motif (*tailed triangle*) in panel (a) and an insignificant motif (*3-path*) in panel (b). One can clearly see that the number of *tailed triangle* motifs (3570) lies outside the distribution calculated for the random networks, while the the number of *3-path* motifs (12473) can also be found in a substantial amount of random reference networks. Interestingly, we find only a few cases of insignificant

induced motifs. These are indicated by bold values in Tables 5, 6, and C5. Moreover, in almost all cases of significant motifs, the counts extracted from the original network is far away from those of the random reference networks.

3.2.3 5-node motifs

Next, we turn our attention to motifs that involve 5 nodes, and focus on those subnetwork structures that cannot be part of a 4-node motif with 1 additional dis-
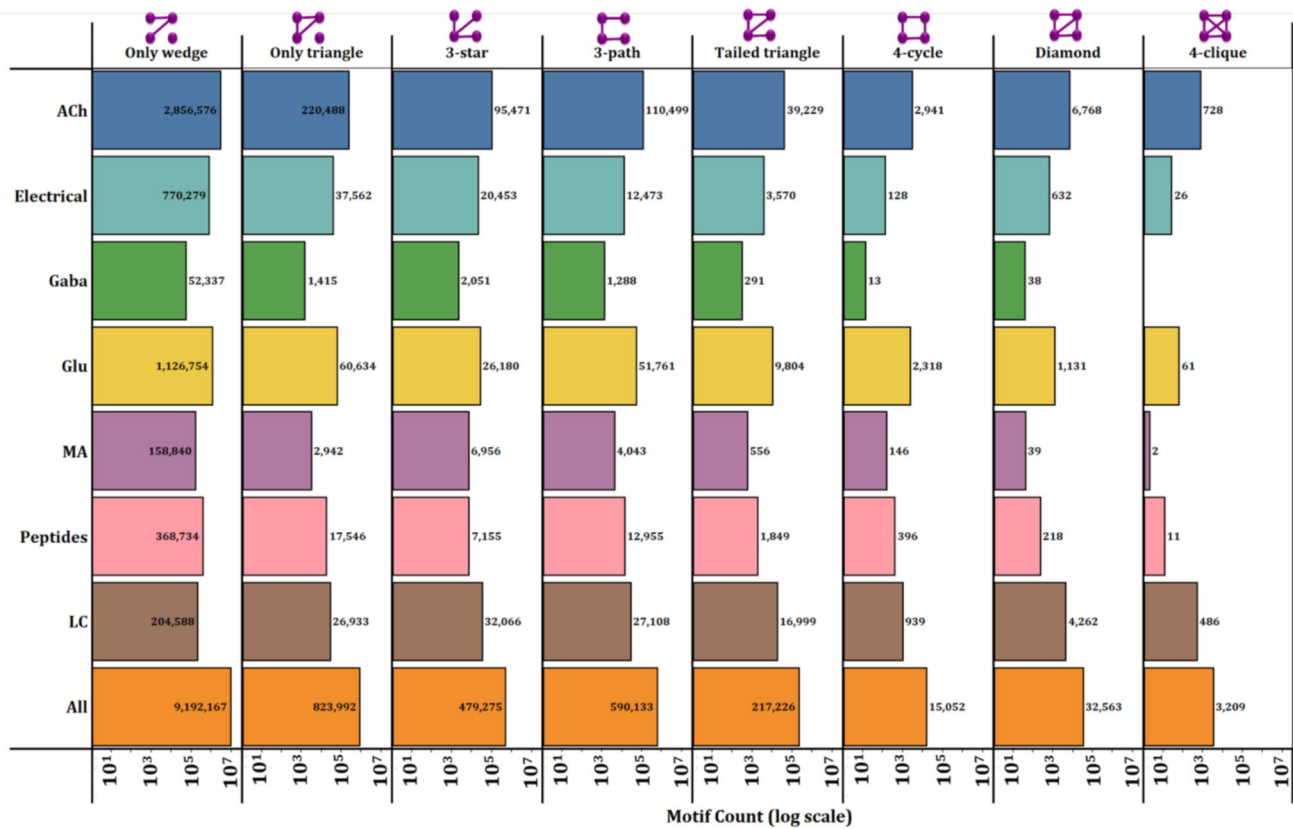
Fig. 5 Histogram of induced 4-node motifs in each layer. See Table 5 for a comparison to the non-induced counts given in Table A1
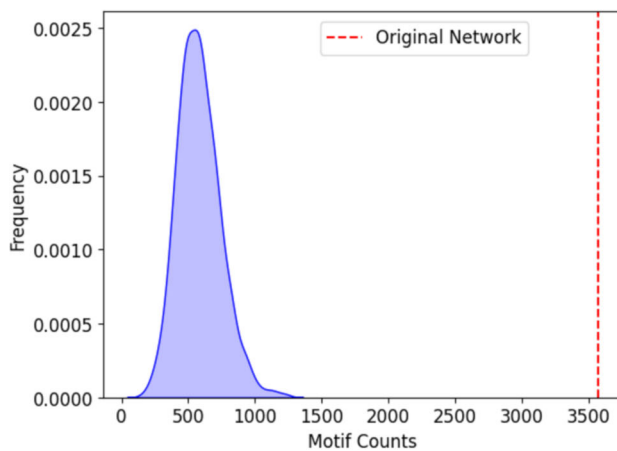
(a) Significance of tailed-triangle motif in the electrical layer as shown in Fig. 5 and Table C5.
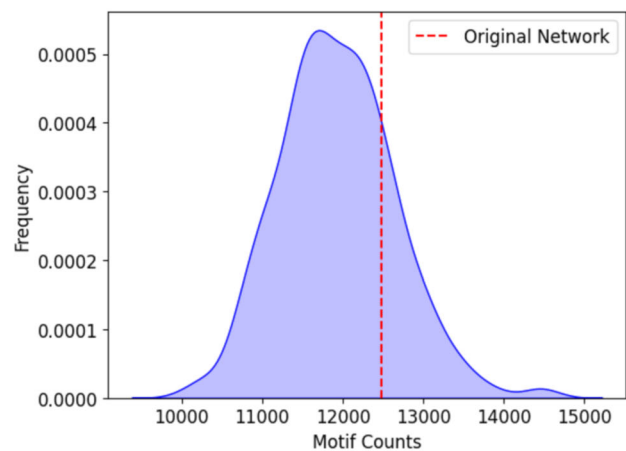
(b) Insignificance of 3-path motif in electrical layer as shown in Fig. 5 and Table C5.

Fig. 6 Examples for a significant and an insignificant 4-node induced motif in the electrical layer of original network (cf. Fig. 5) as compared to the random reference networks

Table 5 Ratios of 4-node induced vs non-induced motifs in *C. elegans* multi-layer network

	Motif	ACh	Electrical	GABA	Glu	MA	Peptides	All	LC
Ind set		0.83	0.91	0.86	0.81	0.87	0.83	0.71	0.50
Only link		0.77	0.88	0.79	0.77	0.77	0.81	0.67	0.41
Matching		0.70	0.86	0.78	0.72	0.74	0.79	0.64	0.36
Only wedge		0.66	0.78	0.78	0.72	0.79	0.77	0.58	0.36
Only triangle		0.80	0.88	0.79	0.83	0.82	0.88	0.74	0.50
3-star		0.63	0.81	0.85	0.68	0.92	0.75	0.62	0.54
3-path		0.44	**0.51**	**0.60**	0.59	0.67	0.66	0.45	0.28
Tailed triangle		0.52	0.56	0.66	0.65	0.76	0.65	0.56	0.43
4-cycle		0.25	0.15	**0.25**	0.64	0.76	0.61	0.26	0.14
Diamond		0.61	0.80	1.00	0.76	0.76	0.77	0.63	0.59
4-clique		1.00	1.00	-	1.00	1.00	1.00	1.00	1.00

The values in bold indicate that the discovered induced motif count is not significant

connected node. This includes 23 out of 34 possible 5-node motifs (cf. Figure 2). Note that we still include the *wedge+edge* and *triangle+edge* motif.

Table 7 summarizes the findings in terms of the percentage of 5-node induced motifs relative the aggregated network case. Again, we find that almost all motifs are significant in their occurrence and cannot be expected from random reference networks. It is noteworthy that except the *wedge+link* motif, the overwhelming majority of insignificant motif counts appear for very small fractions below 1 percent. In those cases, their existence is unlikely in random networks as well. This applies to the GABA and MA layers in particular because they involve the smallest number of links and nodes. The sparseness of these layers makes the formation of denser motifs less likely. The occurrence of such motifs is further reduced in the MA layer, whose nodes are more scattered across the network (cf. Figure 1(c) and (e)). The absolute values and numbers for non-induced 5-node motifs can be found in Appendices E to H for the sake of completeness, but are not displayed here to improve readability.

The exact numbers are also included in Fig. 7 in addition to the histograms. We find that in general, more complex motifs are present in lower numbers and are completely absent in small layers. At the same time, layers with more nodes and links have the tendency to allow for these motifs, e.g., *4-wheel*, *hatted-4-clique*, *almost-5-clique*, or *5-clique* (see bottom panel of Fig. 7), and the corresponding motifs remain significant to characterize the layer.

4 Conclusions

We have presented the application of a motif-discovery algorithm for network analysis. As a case study, the algorithm has been applied to the multi-layer connectome of *C. elegans*. We have considered motifs formed by 3, 4, or 5 nodes. Our analysis of the motif occurrences has yielded several important findings. First, the abundance of wedge motifs indicates the prevalence of shared regulators across different layers corresponding to different molecular entities (neurotransmitters), reflecting the existence of conserved regulatory mechanisms. Additionally, the presence of motifs with closed paths, such as triangles, 4 or 5 cycles and their vari-

Table 6 Percentage of 4-node induced motifs in *C. elegans* multi-layer network relative to ALL (aggregated network)

	Motif	ACh	Electrical	GABA	Glu	MA	Peptides
Ind set		85.29	86.57	2.1	28.03	5.85	13.17
Only link		46.24	23.98	0.87	17.51	2.18	6.98
Matching		23.98	6.7	0.37	10.55	0.85	3.72
Only wedge		31.08	8.38	0.57	12.26	1.73	4.01
Only triangle		26.76	4.56	0.17	7.36	0.36	2.13
3-star		19.92	4.27	0.43	5.46	1.45	1.49
3-path		18.72	**2.11**	**0.22**	8.77	0.69	2.20
Tailed triangle		18.06	1.64	0.13	4.51	0.26	0.85
4-cycle		19.54	0.85	**0.09**	15.4	0.97	2.63
Diamond		20.78	1.94	0.12	3.47	0.12	0.67
4-clique		22.69	0.81	**0.00**	1.9	0.06	0.34

The values in bold indicate that the discovered motif count is not significant

Table 7 Percentage of 5-node induced motifs in *C. elegans* multi-layer network relative to ALL (aggregated network)

	Motif	ACh	Electrical	GABA	Glu	MA	Peptides
Wedge+link		16.15	2.67	0.10	**5.80**	0.35	1.49
Triangle+link		13.08	1.43	0.03	3.33	0.08	0.78
4-star		12.38	2.33	0.1	1.41	0.50	0.28
Prong		12.55	1.01	0.08	3.60	0.31	0.66
4-path		11.41	**0.80**	**0.04**	5.33	0.21	0.93
Forktailed-tri		11.25	0.91	0.04	1.42	0.11	0.16

The values in bold indicate that the discovered motif count is not significant

Table 7 (continued)

Motif		ACh	Electrical	GABA	Glu	MA	Peptides
Longtailed-tri		10.15	0.48	0.02	2.86	0.05	0.39
Doubletailed-tri		12.77	0.46	0.04	2.51	0.12	0.27
Tailed-4-cycle		11.72	0.4	**0.02**	6.95	0.35	**0.82**
5-cycle		8.76	**0.33**	**0.01**	4.11	0.14	0.63
Hourglass		7.30	0.31	0.01	1.58	0.02	0.15
Cobra		13.39	0.41	0.02	2.22	**0.01**	0.24
Stingray		13.56	0.80	0.03	1.29	0.06	0.12
Hatted-4-cycle		9.55	0.16	**0.01**	3.72	0.08	**0.39**
3-wedge-col		18.5	**0.16**	**0.02**	20.50	0.97	**1.40**
3-tri-collision		15.83	1.96	0.04	0.72	0.02	0.09
Tailed-4-clique		15.66	0.28	**0.00**	0.95	**0.00**	0.07
Triangle-strip		12.06	0.31	0.01	1.31	**0.00**	0.13
Diamond-wed-col		13.78	0.15	0.01	5.88	0.04	0.51
4-wheel		16.19	0.21	0.01	3.11	**0.00**	0.22
Hatted-4-clique		15.89	0.46	**0.00**	0.56	0.00	0.08
Almost-5-clique		19.34	0.13	**0.00**	0.45	0.01	0.01
5-clique		23.11	**0.00**	**0.00**	0.06	**0.00**	**0.00**

ous variants with extensions to the core cycle, suggests the occurrence of regulatory feedback. Similarly, stars of spokes-and-hub motifs point to central regulators in the network.

To identify significant motifs, we have compared the original network to an ensemble of 1000 random networks of the same number of nodes and links, that is, same link density. For almost all layers and motif types, we find that the occurrence of motifs cannot be explained by a random network model. Notably, the 3-path and 4-path motifs stand out as an exception of insignificant motifs in the electrical and GABA layer, which are not suited to characterize the network at hand. Interestingly, the presence of 4 cycles and diamonds across multiple layers implies recurrent regulatory patterns and interconnected information path-
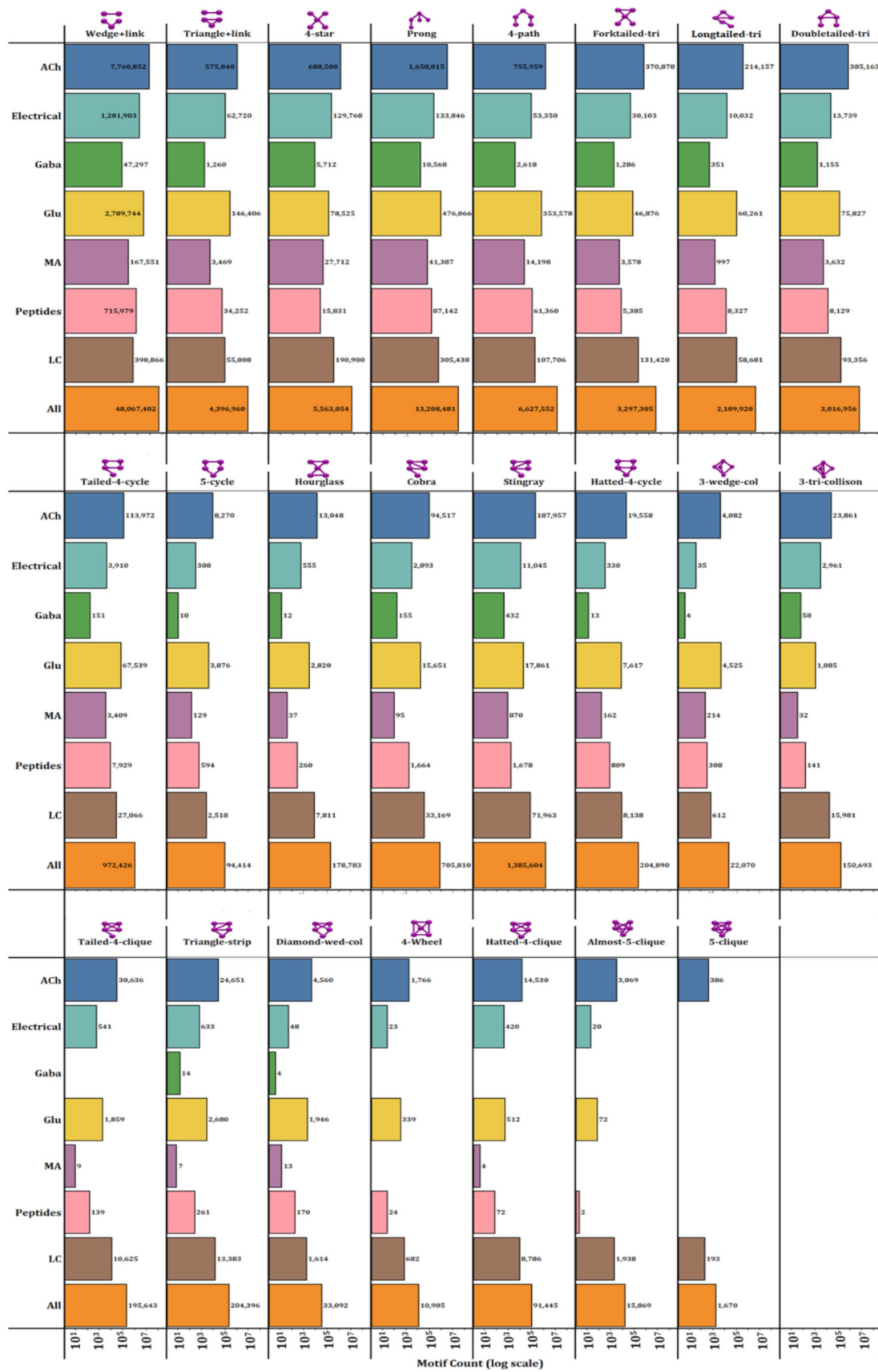
Fig. 7 Histogram of induced 5-node motifs in each layer. See Table G11 for a comparison to the non-induced counts given in Table E8

ways. Conversely, 4- and 5-clique motifs feature less prominently, which suggests that fully interconnected regulatory clusters are less prevalent in the network.

The proposed approach can be easily extended in several directions: (i) It invites the characterization of significant and insignificant motifs in other empirical networks from various backgrounds. (ii) We have chosen random networks as a reference case. The pipeline, however, allows to consider other classes of networks as well. That way, a network extracted from an empirical dataset can be fingerprinted and uniquely characterized by its composition of small network substructures. (iii) Information on the motif composition can also inform generative models to generate ensembles of realistic networks that can be used in simulations. (iv) We have focused on the occurrences of motifs in different layers. The general concept of these small network structures, however, can be extended to configuration where links are part of different layers. That way, information pathways across layers become accessible.

Author contributions

All authors designed the study, discussed the results, and wrote the manuscript. DS was the main programmer and coordinated the numerical analysis.

Data availability The neuron and connectivity data can be found on WormAtlas (http://www.wormatlas.org). The transmitter and receptor data are partly taken from B. Bentley et al. [30] and have been supplemented by own work. The dataset is available via the Kiel University Research Data Repository *opendata@uni-kiel*.

References

1. D.L. Chase, M.R. Koelle, Biogenic amine neurotransmitters in C. elegans. WormBook: The Online Review of C. elegans Biology (2007)

2. G. Yan, P.E. Vértes, E.K. Towlson, Y.L. Chew, D.S. Walker, W.R. Schafer, A.-L. Barabási, Network control principles predict neuron function in the Caenorhabditis elegans connectome. Nature **550**(7677), 519–523 (2017)

3. E.K. Towlson, P.E. Vértes, S.E. Ahnert, W.R. Schafer, E.T. Bullmore, The rich club of the C. elegans neuronal connectome. J. Neurosci. **33**(15), 6380–6387 (2013)

4. U. Dag, I. Nwabudike, D. Kang, M.A. Gomes, J. Kim, A.A. Atanas, E. Bueno, C. Estrem, S. Pugliese, Z. Wang et al., Dissecting the functional organization of the C. elegans serotonergic system at whole-brain scale. Cell **186**(12), 2574–2592 (2023)

5. E.K. Towlson, A.-L. Barabási, Synthetic ablations in the C. elegans nervous system. Netw. Neurosci. **4**(1), 200–216 (2020)

6. Y.L. Chew, D.S. Walker, E.K. Towlson, P.E. Vértes, G. Yan, A.-L. Barabási, W.R. Schafer, Recordings of Caenorhabditis elegans locomotor behaviour following targeted ablation of single motorneurons. Sci. Data **4**(1), 1–7 (2017)

7. D.L. Barabási, G. Bianconi, E. Bullmore, M. Burgess, S. Chung, T. Eliassi-Rad, D. George, I.A. Kovács, H. Makse, T.E. Nichols et al., Neuroscience needs network science. J. Neurosci. **43**(34), 5989–5995 (2023)

8. A. Pinar, C. Seshadhri, V. Vishal, Escape: Efficiently counting all 5-vertex subgraphs. In: Proceedings of the 26th International Conference on World Wide Web, pp. 1431–1440 (2017)

9. D. Purves, G.J. Augustine, D. Fitzpatrick, W. Hall, A.-S. LaMantia, L. White, *Neuroscience* (Oxford University Press, New York, 2019)

10. A.R. Winnier, J.Y.-J. Meir, J.M. Ross, N. Tavernarakis, M. Driscoll, T. Ishihara, I. Katsura, D.M. Miller, UNC-4/UNC-37-dependent repression of motor neuron-specific genes controls synaptic choice in Caenorhabditis elegans. Genes Dev. **13**(21), 2774–2786 (1999)

11. S. Hallam, E. Singer, D. Waring, Y. Jin, The C. elegans NeuroD homolog cnd-1 functions in multiple aspects of motor neuron fate specification. Development **127**(19), 4293–4301 (2000)

12. I.A. Bany, M.-Q. Dong, M.R. Koelle, Genetic and cellular basis for acetylcholine inhibition of Caenorhabditis elegans egg-laying behavior. J. Neurosci. **23**(22), 8060–8069 (2003)

13. D.M. Raizen, R. Lee, L. Avery, Interacting genes required for pharyngeal excitation by motor neuron MC in Caenorhabditis elegans. Genetics **141**(4), 1365–1382 (1995)

14. J.H. Thomas, Genetic analysis of defecation in Caenorhabditis elegans. Genetics **124**(4), 855–872 (1990)

15. D.H. Hall, Gap junctions in C. elegans: their roles in behavior and development. Dev. Neurobiol. **77**(5), 587–596 (2017)

16. K.T. Simonsen, D.G. Moerman, C.C. Naus, Gap junctions in C. elegans. Front. Physiol. **5**, 40 (2014)

17. J.G. White, E. Southgate, J.N. Thomson, S. Brenner et al., The structure of the nervous system of the nema-

tode Caenorhabditis elegans. Philosophical Trans. R. Soc. Lond. B **314**(1165), 1–340 (1986)

18. S.L. McIntire, E. Jorgensen, H.R. Horvitz, Genes required for GABA function in Caenorhabditis elegans. Nature **364**(6435), 334–337 (1993)

19. T. Hills, P.J. Brockie, A.V. Maricq, Dopamine and glutamate control area-restricted search behavior in Caenorhabditis elegans. J. Neurosci. **24**(5), 1217–1225 (2004)

20. J.K. Rose, K.R. Kaun, S.H. Chen, C.H. Rankin, GLR-1, a non-NMDA glutamate receptor homolog, is critical for long-term memory in Caenorhabditis elegans. J. Neurosci. **23**(29), 9595–9599 (2003)

21. Y. Zheng, P.J. Brockie, J.E. Mellem, D.M. Madsen, A.V. Maricq, Neuronal control of locomotion in C. elegans is modified by a dominant mutation in the GLR-1 ionotropic glutamate receptor. Neuron **24**(2), 347–361 (1999)

22. P.J. Brockie, J.E. Mellem, T. Hills, D.M. Madsen, A.V. Maricq, The C. elegans glutamate receptor subunit NMR-1 is required for slow NMDA-activated currents that regulate reversal frequency during locomotion. Neuron **31**(4), 617–630 (2001)

23. E.R. Sawin, R. Ranganathan, H.R. Horvitz, C. elegans locomotory rate is modulated by the environment through a dopaminergic pathway and by experience through a serotonergic pathway. Neuron **26**(3), 619–631 (2000)

24. J.K. Rose, C.H. Rankin, Analyses of habituation in Caenorhabditis elegans. Learn. Memory **8**(2), 63–69 (2001)

25. A.F. Russo, Overview of neuropeptides: Awakening the senses? Headache **57**, 37–46 (2017)

26. C. Li, K. Kim, *Neuropeptides* (The Online Review of C. elegans Biology. WormBook, WormBook, 2018)

27. M. Nelson, N. Trojanowski, J. George-Raizen, C. Smith, C.-C. Yu, C. Fang-Yen, D. Raizen, The neuropeptide NLP-22 regulates a sleep-like state in Caenorhabditis elegans. Nat. Commun. **4**(1), 2846 (2013)

28. N. Ringstad, Neuromodulation: the fevered mind of the worm. Curr. Biol. **27**(8), 315–317 (2017)

29. Y.L. Chew, L.J. Grundy, A.E. Brown, I. Beets, W.R. Schafer, Neuropeptides encoded by nlp-49 modulate locomotion, arousal and egg-laying behaviours in Caenorhabditis elegans via the receptor SEB-3. Philosophical Trans. R. Soc. B **373**(1758), 20170368 (2018)

30. B. Bentley, R. Branicky, C.L. Barnes, Y.L. Chew, E. Yemini, E.T. Bullmore, P.E. Vértes, W.R. Schafer, The multilayer connectome of Caenorhabditis elegans. PLOS Comput. Biol. **12**(12), 1005283 (2016)

31. S. Maslov, K. Sneppen, Specificity and stability in topology of protein networks. Science **296**(5569), 910–913 (2002)

32. R. Milo, N. Kashtan, S. Itzkovitz, M.E. Newman, U. Alon, On the uniform generation of random graphs with prescribed degree sequences. arXiv preprint cond-mat/0312028 (2003)

33. T. Maertens, E. Schöll, J. Ruiz, P. Hövel, Multilayer network analysis of C. elegans: Looking into the locomotory circuitry. Neurocomputing **427**, 238–261 (2021)