## Abstract

Graphical modelling of neural networks allows is useful for identifying topologically-related groups of neurons and thus providing clues to what neurons might serve a common function of the organism. The transparent nematode *c. Elegans* is of particular interest in the field of graphical modelling and analysis because it is the only organism for which the entire neural connectivity (connectome) is documented. However, despite the connectome having been mapped since 1986 (White et al 1986), most knowledge regarding the function of specific neurons has been obtained experimentally and no behaviorally predictive *in silico* model of *c. Elegans* neural function exists. This shortcoming has brought into question the ultimate utility in mapping the *c. Elegans* connectome. Here, we employ cutting edge spectral graph theory to identify key topological components on the roundworm neural network in the form of neuronal clusters. These clusters seem to correspond to specific animal behaviors critical to survival. We then identify highly-connected “hub neurons” that heavily innervate such clusters and perform *in silico* ablation experiments to test the importance of these master regulators for the overall network topology, i.e., well-defined clusters obtained through spectral analysis. Ultimately, we conclude that a select few of the 302 *c. Elegans* neurons are indispensable for the structural integrity of the network, challenging previous views that the *c. Elegans* neural network is scale-free and thus highly robust to failure.

## Background

In recent researches people have discovered the neuron circuits that corresponds to various behaviors of C. elegans, and identified functionalities of most of the neurons in C. elegans, though few still remain unknown. Our aim was to establish a general framework that could analyze the structure of neuron network, reproduce and even identify new subgraphs and important nodes within the neuron network with high functional significance.

A graph that models the neuron network consists of the set of vertices , which are neurons, and edges which are the synapses. In the *c. Elegans* central nervous system (CNS), all vertices have designated names and most have known or putative functions. In the connectome dataset ([www.wormatlas.org](http://www.wormatlas.org)), there are three basic types of edges.

Electrical synapses, also called gap junctions are formed by close contact between cells. In the nervous system, there exist gap junctions between neurons and between muscle cells. The electric synapses are generally considered symmetric: it allows information transfer between neurons in both directions. When reflected in the adjacency matrix of the neuron network graph, the subgraph consisting of all electrical synapses is symmetric.

Chemical synapses may occur between one presynaptic and one post-synaptic cell (a monad) or more than one post-synaptic partner (a polyad; two recipients make it a dyad and three recipients make it a triad).Chemical synapses are made en passant between neighboring processes where synaptic swellings are formed along the process shafts. (wormatlas) The subgraph consisting of all chemical synapses is generally asymmetric.

Finally, the neuromuscular junction is a special chemical synapse where neuron input to muscles occurs. Since this type of junction does not connect neurons to neurons, we generally exclude them in our connectome analysis unless otherwise stated.

Commonly, graphical models are built by defining a given edge weight based on the distance between the vertices that it joins. If two vertices are far apart, then the edge is deemed unimportant or weak. However, the spatial structure of C-elegans (FIG to come) does not readily reveal the connectivity of neurons. Distantly-connected neurons do not necessarily carry out distantly-related functions (REF). Consequently, to visualize the graph using spatial structure of neurons only reveals few information about significant clusters, hubs and paths. Spectral graph theory comes in to play, because it has both theoretical bases for graph clustering and partitioning, and experimentally out-performs traditional methods, such as Markov Chain Monte Carlo Methods (MCMC) (Sohn et al. 2011) , Expectation Maximization (EM) (Zhang et al. 2002) etc, in terms of efficiency and effectiveness (von Luxburg 2007).

By incorporating known connectivity between defined vertices, as well as specific classes of edges, here we attempt to apply spectral graph theory and novel hub analysis methods on the neural network of C. elegans, from which we show the usefulness of the technique and the insight it gives in understanding the graph model of C. elegans connectome.

## Previous results

There are numerous papers that discuss neural circuits present in C. elegans, and their correlation with its functional behaviors. Those are usually verified by removing particular neurons from C. elegans and recording behavioral changes. Even for this well-known organism, very few work has been done in computationally analyzing the network structure using a general framework for understanding the correlation between network and functionalities, and discovering/verifying subnetwork circuits.

Recent paper (Towlson et al. 2013) has addressed the issue of finding important neurons (rich club) in the network by analyzing graph parameters including degree, clustering coefficient and network motifs. However, it only provides information about these important neurons, with little systematic view on the organization of the entire network.

Topological clustering done by (Sohn et al. 2011) is so far the only comprehensive clustering analysis. It uses the method of simulated annealing to maximize intra-cluster synapses and minimize inter-cluster synapses. It belongs to the class of MCMC algorithms, which has the disadvantage of instability, i.e. there exist multiple convergent solutions for a network using different initial values. Additionally as shown in our results, the inherent property of C. elegans connectome makes it unrealistic to assign some important neurons, especially the hub neurons, to a fixed cluster deterministically.

In this paper, we aim to find an algorithm that is more computationally efficient, stable and reproducible that characterizes both the important hub neurons and the role of all neurons in the entire connectome network.

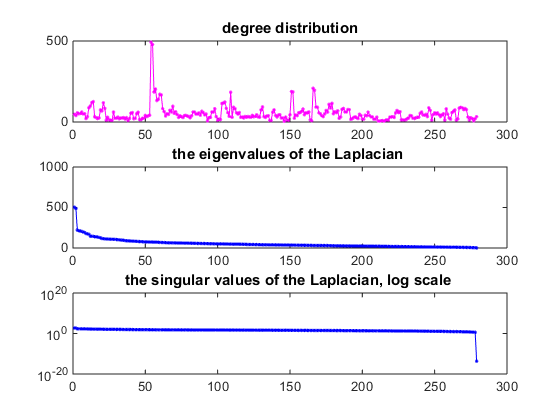
# Results

## Constructing a graph model of the c. Elegans central nervous system

To model the topology for the *c. Elegans* CNS, we included only connections that terminated within the CNS. These connections included electrical and chemical synapses, but not neuromuscular junctions (NMJs) which terminate onto muscle cells, acting as network “dead ends”. From dataset, we are able to construct the symmetric component of electrical synapses network, and the asymetric component of chemical synapses network. We represented them using sparse matrix form in MATLAB from which first few eigenvectors can be efficiency calculated using iterative methods. The edge weights are defined by the number of connections between a given pair of vertices.

For each individual neuron represented as vertices in the graph, we associate them with 3 general types: sensory neurons, which mainly respond to the stimuli of external environment; motor neurons, which have neuronmascular junctions to muscle cells and are in charge of physical activity of the work; and inter-neurons, which mediates between neurons with different types. The type information of vertices is represented by different colors in the visualization, that provides understanding of the general information flow of the neuron network.

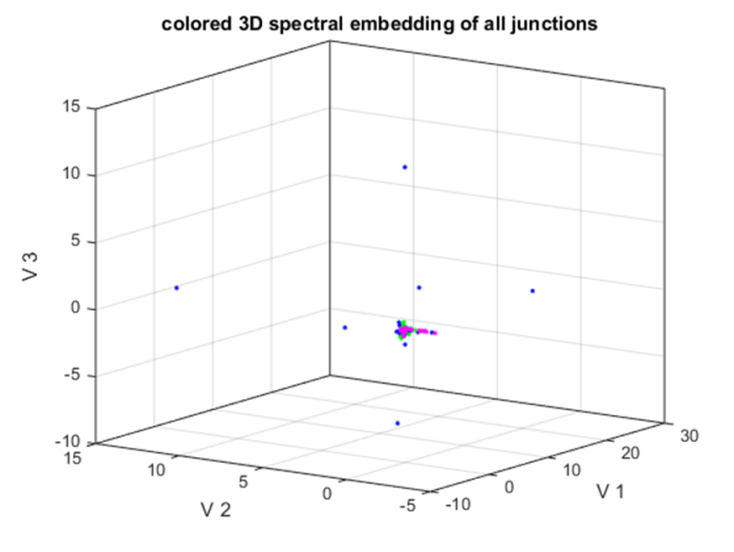
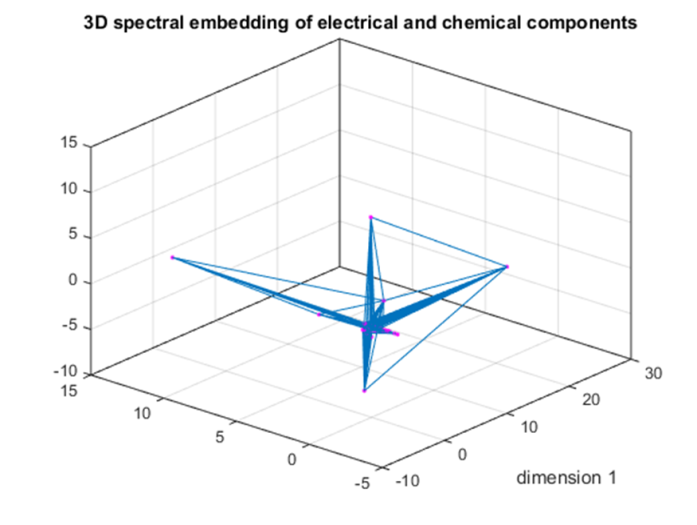
The degree of each vertex is first calculated, from which we generates the graph Laplacian and perform spectral analysis. As can be seen in Fig. 1, the eigenvalues decrease almost exponentially, except the last eigenvalue which is 0. There is only one eigenvector corresponding to 0 eigenvalue, implying that the neuron network consisting of 279 neurons is connected.



**Figure 1. Spectral analysis of the *c. Elegans* connectome.** (to be filled in)

## Identification of crucial neurons

In the spectral embedding of unnormalized graph Laplacian (Fig. 2), we can see that there are 3 points that are particularly far away from the central cluster of neurons, and 3 other points that are between the previous 3 points and the central cluster. The 3 neurons corresponding to the faraway 3 points are: AVAR, ABBR, RIAR; the intermediate 3 neurons are: RIAL, PVCL, DVA. All of which have common features of being inter-neurons with high connectivity. The hub analysis section addresses the significance of these neurons in more detail.



**Figure 2. Left**: 3D spectral embedding using the eigenvectors corresponding to the 2nd, 3rd, and 4th smallest eigenvalues of the un-normalized graph Laplacian incorporating both electrical and chemical junction network. Blue represents inter-neuron; green represents sensory neuron; magenta represents motor neuron. **Right**: The same embedding with graph connections shown. It is seen that DVA plays an important role in mediating between other hub neurons.

## Information Flow

The normalized version of graph Laplacian gives more interesting results in terms of the structure of connectome. From the embedding, we can see 4 arms that stem from a central cluster that mainly comprises of inter-neurons. The 4 arms have different compositions in terms of neuron types.

In general, the 4 arms represent 4 different set of functionalities of C. elegans. The motor and sensory neurons in each arm seldom communicate directly, but rather will go through an inter-neuron near the center of the embedding. For sensory and motor neurons in charge of similar functionalities, they communicate using a relatively shorter path by going through more peripheral inter-neurons located in the arm that they belong to in the embedding. However, sensory and motor neurons corresponding to different set of functionalities are located in different arms. They communicate with each other by going through a longer path by connecting to a more central command neuron located at the center of the embedding, and then travel to the other arm.

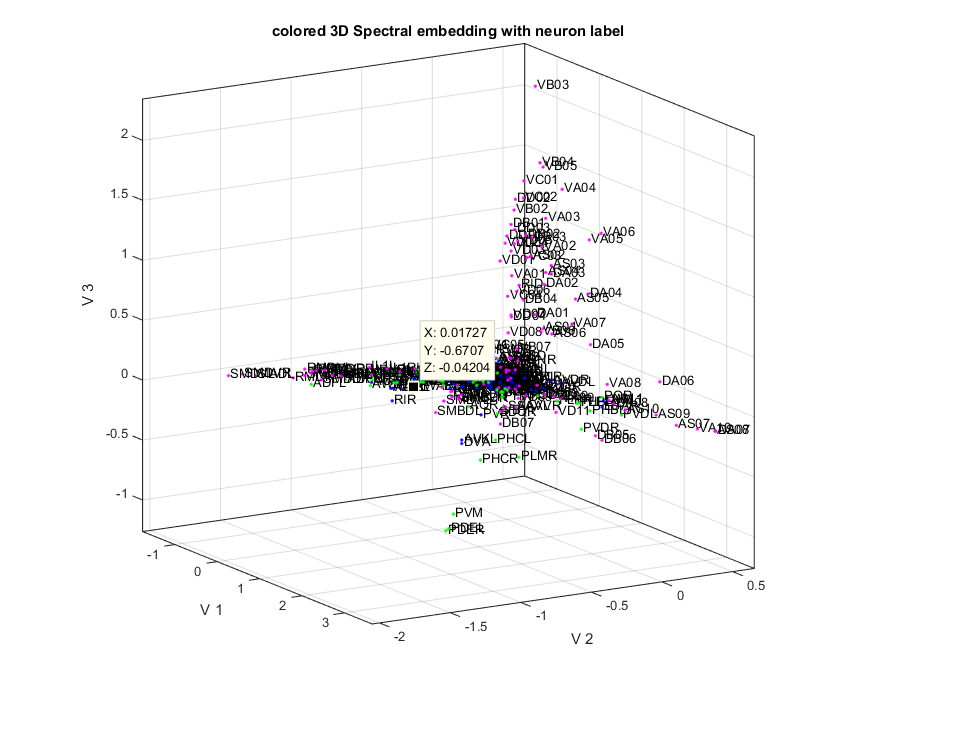
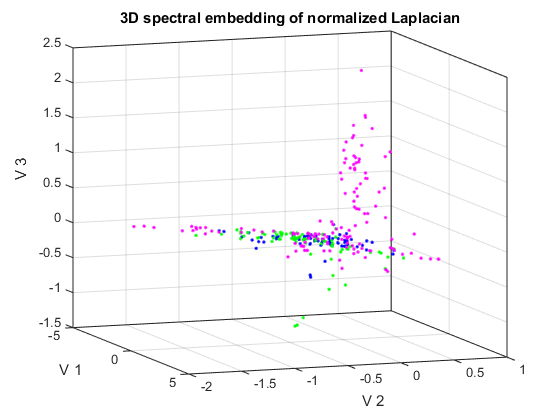
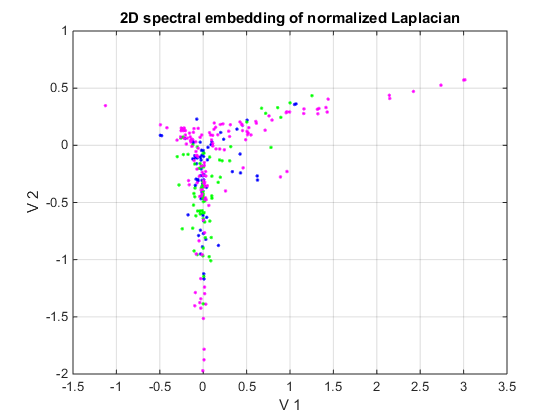
Top arm mainly consists of motor neurons. Examples of neurons in this arm are the DD neurons, which are responsible for sinusoidal body movement-locomotion, the DB neurons that are in charge of propagation of rhythmic activities along the body during forward locomotion, and the VA neurons which are also responsible for large locomotion. Almost no inter-neurons are present in this arm. The motor neurons receive their signal primarily through the central cluster of interneurons. In addition there are some neurons that act both as sensory and motor neurons in this arm. Eg. VB neurons are responsible for both locomotion and proprioception. This compensates for the efficiency of neuron activities without inter-neurons.

Bottom cluster consists mainly of sensory neurons. Examples of neurons in this arm include PDE which are responsible for the basal slowing response, PHE which are responsible for thermal sensing. The important interneuron in this arm is DVA, one of the less important hub neurons identified: it provides input to both the anterior and posterior touch circuits. Animals lacking these neurons respond to tap stimulus with diminished forward accelerations and reversals.

It is worth noting that DVA plays an important role in mediating between hubs. Each hub is responsible for a range of behaviors, and when one stimuli is sensed which requires response of neurons associated with other hubs, the hubs communicate through DVA. It is experimentally verified that DVA neurons are presynaptic to both the forward and backing interneurons, and provide input to both the anterior and posterior touch circuits. It regulates sensory-motor integration during C. elegans locomotion, where sensory responses are sent from sensory neurons to interneuron hubs RIA, and then sent to motor hubs PVC.

The left cluster has similar amount of sensory and motor neurons. The important hub neurons in this arm include RIA. They are the "second layer"interneurons in the process of integration of information from the outside world and the inner state of the animal, which then leads to a behavioral response.

Right arm has similar amount of sensory and motor neurons. The important hub neurons in this arm include PVC. It functions as command interneuron specifically for**forward locomotion,** driving forward movement of the animal along **with the hub neuron** [AVB](http://wormatlas.org/neurons/Individual%20Neurons/AVBframeset.html), which is located in the central cluster, but very close to this arm. The motor neurons in this arm consist of some VA neurons, and majority of AS neurons, which are responsible for locomotion. The sensory neurons in this arm include PVD which responds to hard touch and cold temperatures. It being in this arms explains for a need to have faster signal transfer from PVD to motor neurons, since generally the worm needs to respond quickly to such strong stimuli.



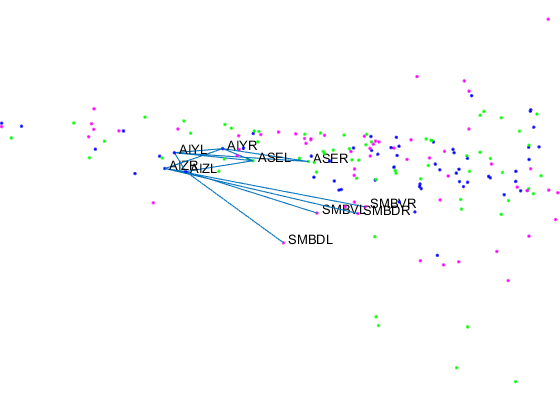
**Figure 3. Spectral cluster analysis of *c. Elegans* CNS. Top left**: 2D embedding of normalized graph Laplacian using the 1st and 2nd eigenvectors. The colors that represent neuron types are the same as before. **Top right**: 3D embedding of normalized graph Laplacian using the 1st, 2nd and 3rd eigenvectors. The central part mainly consists of interneurons. Four arms stretches from it, corresponding to different functional circuit categories. **Bottom left**: The same embedding with neuron name labels.

## Functional significance

Our spectral analysis proves to be a useful tool in exploring the structure of a neural network by providing hub information and the sub-modules of the network of the respective hubs.

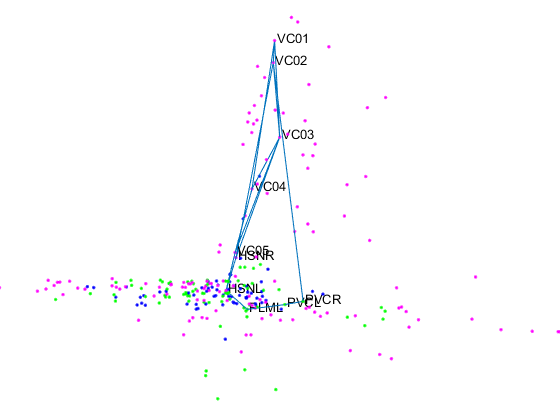
The minimum klinotaxis circuit identified (Izquierdo and Beer 2013), for example, is a typical circuit that resides in the left arm of the embedding. The paper identifies a minimum circuit by systematically searching the *C. elegans* connectome for pathways linking chemosensory neurons to neck motor neurons, and prune the resulting network based on both experimental considerations and several simplifying assumptions. The resulting circuit can be explained through our spectral cluster analysis. In fact, we can show that with known adjacency matrix and the motor/sensory neurons involved in a general circuit, it is possible to deduce the structure of the circuit sub-network given the spectral embedding.

Similar to most circuits, it contains sensory neurons that communicate with inter-neurons, which in turn stimulates the motor neurons. Notice that all neurons are located on the left arm, and the inter-neurons involved (AIY, AIZ) are embedded in the center of the left arm. In fact the important intern-neurons we discovered in multiple circuits on the left arm includes AIY, AIZ, RIA, all of which are embedded in the center of the left arm. Sensory and motor neurons on different parts of the arm communicate by going through these local hub neurons. In fact we have found that most circuits involving less motion intensive responses including klinotaxis, chemotaxis (Iino and Yoshida 2009), thermotaxis (Ma and Shen 2012), gentle locomotion (Bono and Villu Maricq 2005) involve neuron circuits that involves exclusively the neurons in the left arm.



Minimum klinotaxis circuit identified in silico and confirmed in vivo (Izquierdo and Beer 2013). This entire circuit is embedded in the left arm, mediated by local inter-neurons.

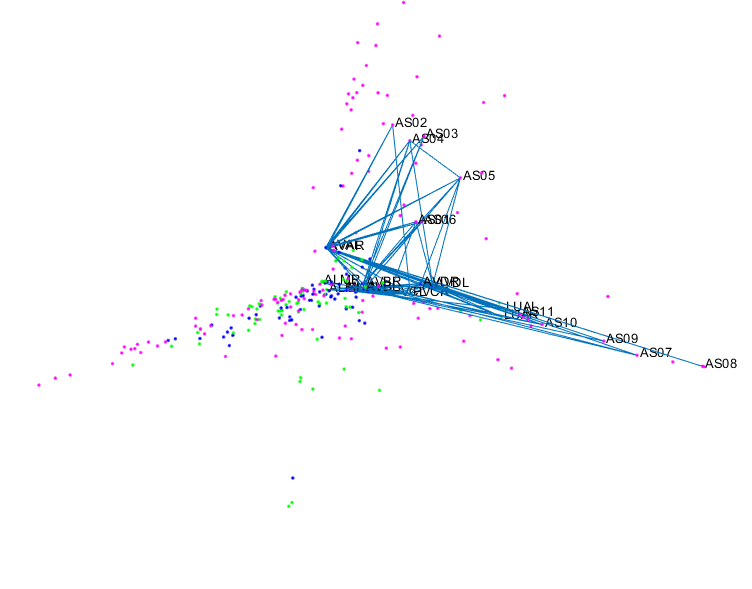
The top arm and right arm contain circuits that requires coordination of relatively large amount of motor neurons corresponding to complex physical responses, an example of which is the egg laying network in hermaphrodite C. elegans (Zhang et al. 2008). The circuit is embedded in the north arm. There are few sensory neurons involved in the circuit, compared to motor neurons. Signal is transduced in the network from the center cluster, and propage to motor neurons through the common command neurons (AVB), which acts as a hub for a wide variety of circuits.



Egg laying circuit identified in vivo (Zhang et al. 2008). The information flow in this circuit starts from the central cluster and propagates through minor inter-neurons (HSN) with specific functionality, to the motor neurons.

The touch circuit as elaborated in (Chalfie et al. 1985) is a good example for cross-arm communication. This circuit involves a wide range of neurons in both left arm, where sensory neurons are concentrated, and right arm where motor neurons are concentrated. Such circuit requires significant amount of coordination between a variety of sensory and mot

or neurons. The signal carrying information typically flows from sensory neuron in the left arm to the hub neurons within the arm, which in turn pass information to the command inter-neurons located in the central cluster. The command inter-neurons , are typically the hub neurons we have identified, and they pass information to the hub neuron in the right arm, which then relays the information to the motor neurons.



The touch circuit identified in vivo (Chalfie et al. 1985), containing a majority of motor neurons. It is a typical example of cross-arm communication that involves motion of the entire worm.

Sensory neurons

Left arm hubs

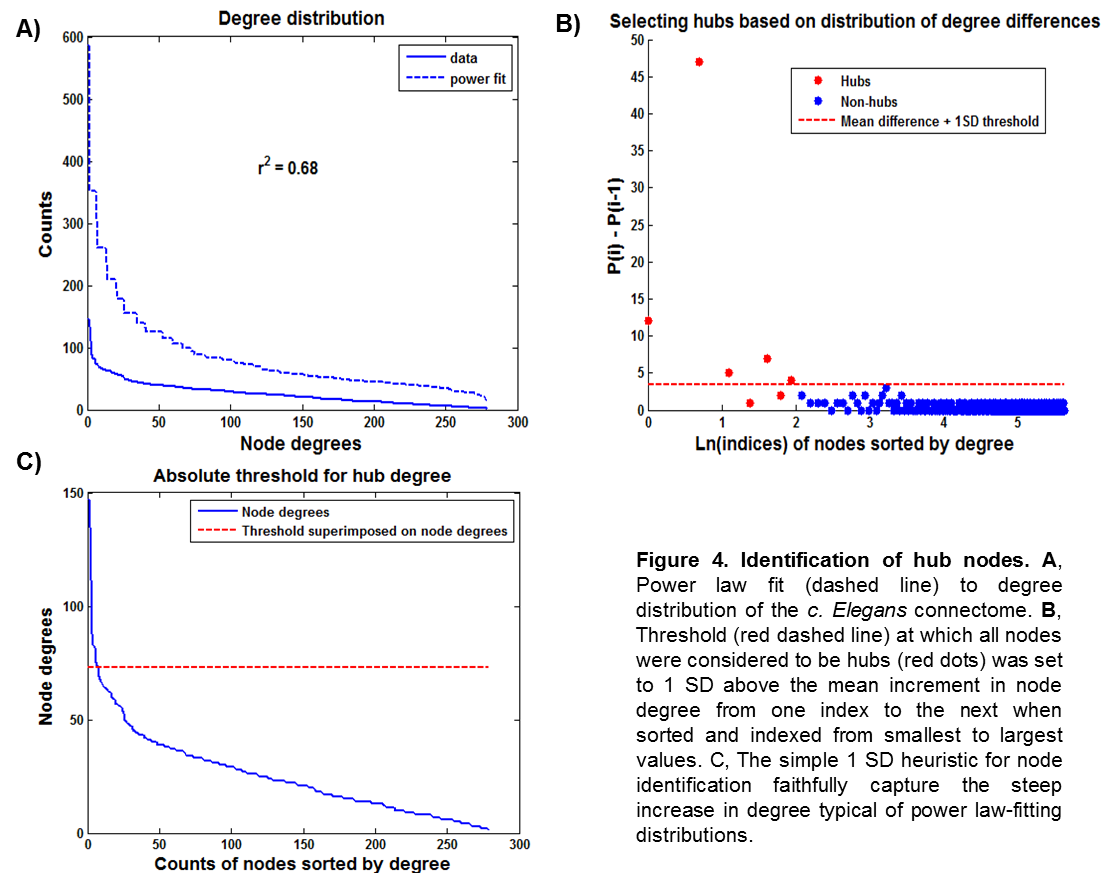
Right arm hubs

Central hubs

Motor neurons

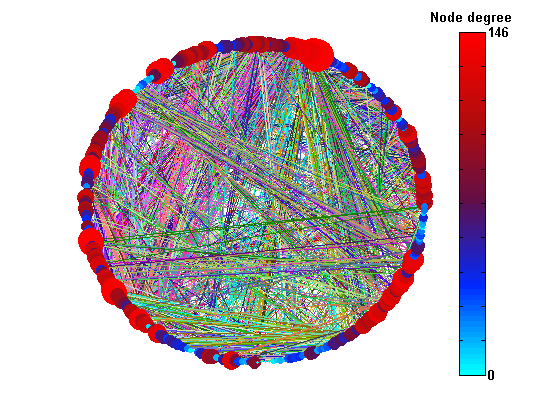
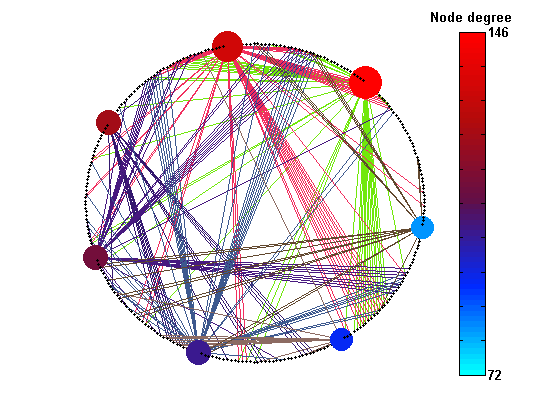
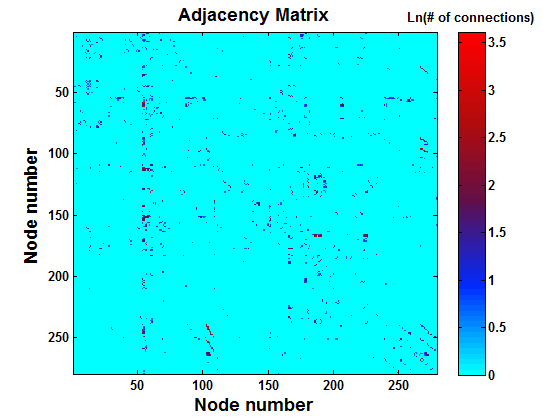
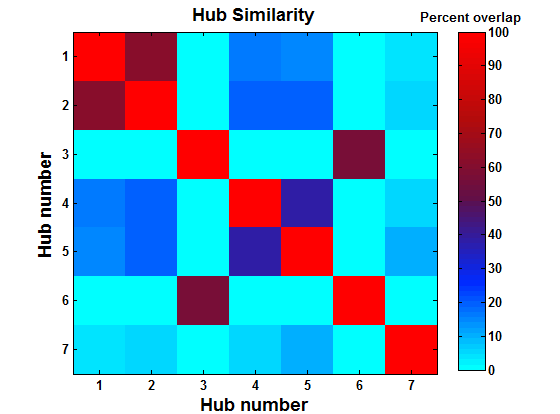
Cross-arm neuron circuit paradigm. There exists two layers of inter-neurons that mediates the information flow.

## Identification of “hub neurons” in the c. Elegans connectome

Although we found that certain groups of neurons preferentially linked to each other, it remained unclear whether some neurons were more extensively connected in the network than others. Hierarchical structure in degree of connectivity could allow small subsets of neurons to govern the activity of larger populations, possibly even specific clusters. It may be that any pair of neurons has an equal probability of being connected; this type of network is referred to as a random graph (ErdosRenyi1959). Previous work has shown that the node degree distribution of the *c. Elegans* connectome follows a power law. In other words, a small fraction of the nodes exhibit relatively high connectivity whereas the vast majority are equally and sparsely connected. (BarabasiAlbert1999). Indeed, when looking only at central synapses (chemical and electrical), we found the same phenomenon (Fig. 4A). 

A non-uniform degree distribution among the *c. Elegans* connectome is interesting because it implies that a select few neurons might be particularly important because they influence the activity of many others. Highly connected subsets of neurons have previously been identified in the *c. Elegans* connectome and referred to as “hubs” belonging to a “small rich club” (TowlsonBullmore2013). In this study, the degree of the hub neurons was the strongest factor that separated them from the rest of the network. Therefore, we employed a simple approach based solely on degree of primary (monosynaptic) connections to determine which neurons might serve as hubs in our data set. Additionally, we constrained our data set to include only those connections that were capable of propagating information through the central nervous system by excluding connections representing neuromuscular junctions (NMJs). NMJs are terminal synapses from neurons onto muscle cells and do not feed back into the network. We were thus left with chemical and electrical synapses.

We hypothesized that interneurons, typically inhibitory cells which serve to regulatethe activity of many neurons simultaneously, would be particularly likely to exist in the network topology as master regulators of network function. To identify those nodes in our network that were exceptionally well-connected to other nodes, we sorted nodes by degree and computed a difference vector containing the increment in node degree from each ith node relative to the ith-1 node. By setting a threshold at one standard deviation above the mean increment (Fig. 4B), we were able to index the beginning of the sharp upward tail of the node degree distribution (Fig. 4C). All nodes beyond this index in the sorted vector were classified as hub neurons (Fig. 4B).



**A)**

**D)**

**C)**

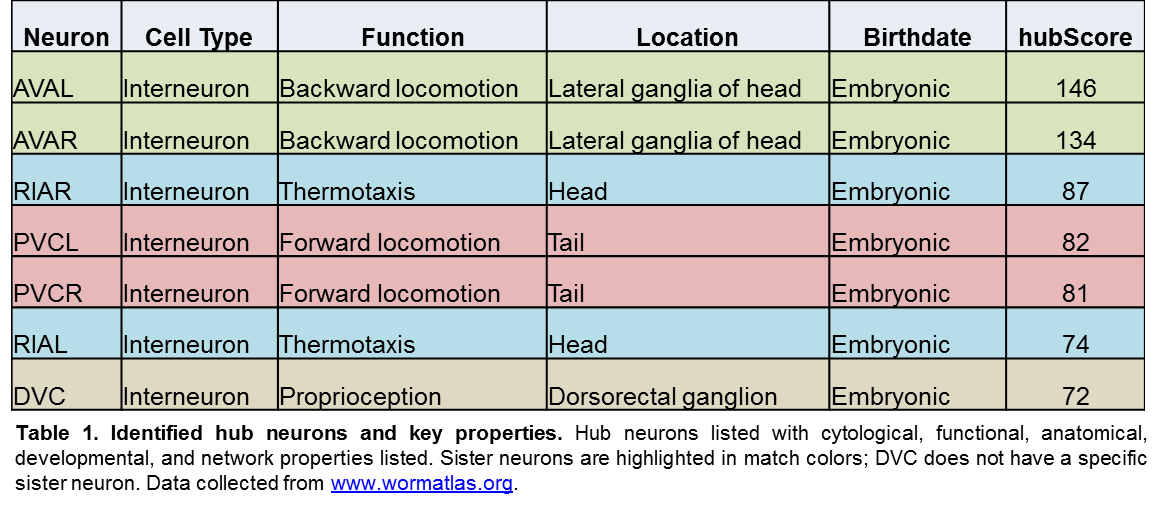
**B)**

**Figure 5. Visualization of connectome graph and hub node properties. A**, Heat map representation of adjacency matrix showing primary synaptic and electric connections between nodes. **B**, Connection diagram of matrix in A. Each circle represents a node, sized and heat mapped according to degree. Each line color represents a row in (A). **C**, Heat map matrix showing overlap between the targets of edges from each hub node. **D**, Connection diagram as in (B) showing only edges that arise from hub nodes. Non-hub nodes are present as small black dots; hubs are heat mapped according to degree.

As can be seen in Fig. 5A, the *c. Elegans* connectome is rather sparse. However, structure is apparent even in the adjacency matrix as solid rows or columns of higher values. These lines correspond to neurons that connect with much of the network. The existence of hub neurons becomes more apparent when examining a connection diagram where node size and color represent degree (Fig. 5B).

## Properties of identified hub neurons

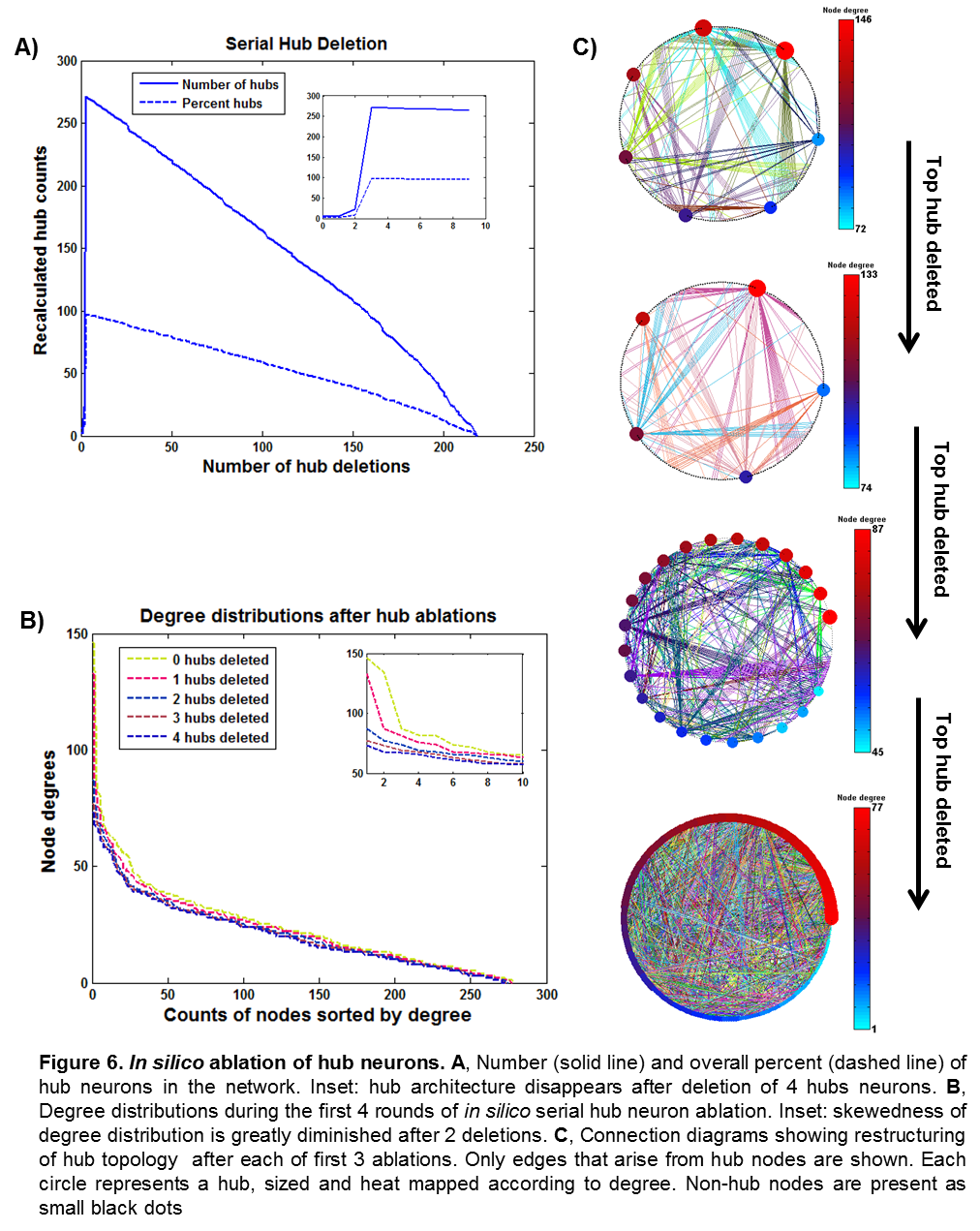
Interestingly, hubs identified as in Fig. 4B-C seem to share relatively few targets (Fig. 5C). This lack of overlap suggests that our identified hub neurons, while clearly being highly connected, may serve highly separate functions. Notable exceptions can be seen between hubs 1 & 2 (61.19% overlap) and 3 & 6 (57.14% overlap). Perhaps not surprisingly, these cells corresponded to “sister” cells AVAL & AVAR and RIAR & RIAL respectively (Table X). Sister neurons are those that have arisen from the same “parent” progenitor cell. Since neuronal wiring in *c. Elegans* is genetically determined, clonal relationship of these sister cells entails highly similar connectivity patterns (REF). The largely divergent efferent connectivity of identified hub neurons can be readily visualized in a connection diagram including only those connections originating from hubs (Fig. 5D).



Compartmentalized function within hub neurons could have interesting consequences for the robustness of the *c. Elegans* brain network. When there is a lack of redundancy in connections (i.e. when not every hub has a corresponding sister hub), any insult to that cell could profoundly damage network performance and animal survival. This effect would be mitigated if the function of hub neurons were distributed; in other words, if they diffusely connected to many targets involved in many behaviors. However, given that our hub neurons have been attributed specific functions ([www.wormatlas.org](http://www.wormatlas.org)), we predict that these cells are critical to the organism. Underscoring the importance of hub neurons that control distinct functions, all seven identified in our analysis are born embryonically rather than postembryonically.

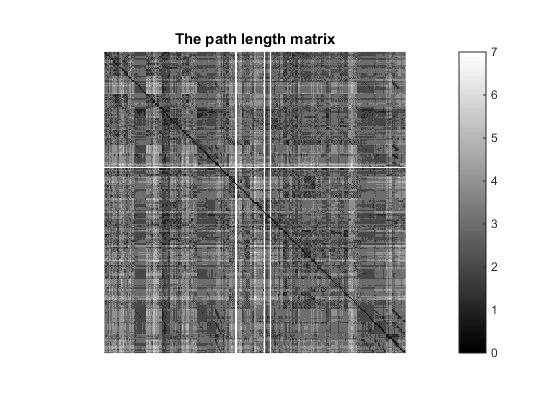
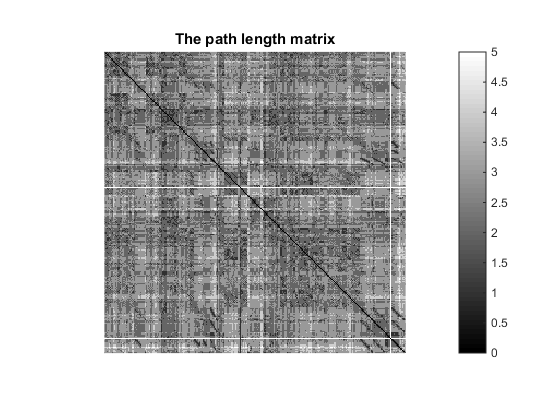
*Importance of Hub Neurons for Network Structure*

Since the degree distribution of the *c. Elegans* network follows a power law (REF, Fig. XA), it has been asserted that the network is “scale-free” (REF). By definition this is correct; however, a key property of scale-free networks, and what makes them of particular interest, is that they are highly robust to failure (CallawayWatts2000, CohenHavlin2000). Scale-free networks tend to be fault tolerant because if a major hub is removed, smaller hubs are waiting in line to take over. Notably, whether or not the *c. Elegans* connectome exhibits this fault tolerance has never been explored. To elucidate how the organism’s neural connectivity might benefit from its scale-free nature, iteratively ablated the strongest hub *in silico* and reanalyzed the network to determine hubs that still existed as well as any new hubs that took on the burden of the ablated hub.



Surprisingly, we found that the *c. Elegans* network of chemical and electrical synapses is highly vulnerable to the loss of only a few hub neurons. During our iterative ablation assay, any semblance of hub structure was lost almost immediately and nearly all nodes were erroneously identified as hubs (Fig. XA) due to the low variability in the degree distribution (Fig. XB). Indeed, plotting the degree distribution after each ablation revealed that, after just 2 hubs were removed, the previously reported “rich club” (TowlsonBullmore2013) was no longer readily distinguishable from the rest of the distribution (Fig. XB, inset). The dramatic effect of hub ablation on our data set can be more intuitively visualized in Fig. XC, where connection diagrams heatmapping only hub neurons are shown after deleting 0, 1, 2, and 3 hub neurons (top-to-bottom). The intolerance of the *c. Elegans* connectome network topology to single hub ablation is likely due to over-steepness in the degree distribution. This can be observed in Fig. XA, where the maximum likelihood estimation for our data is significantly shallower than the real data. If the degree distribution were denser near the highest values, then the network would likely respond better to the loss of single yet highly connected nodes.

Robustness of network determines how adaptive the organism is when some of its important neurons are impaired. In the following experiment, we investigated the effect of removing the 3 most important hub neurons (as identified in Fig. 6 and Table 1) from the network. We use a shortest paths matrix to store the shortest paths from one node to the other in the graph. stores the smallest number of hops needed to reach neuron from neuron . We observe a decrease in the percentage of nodes that can be reached, and 13.6% increase in the average path length of the network (Fig. 7, Table 2).

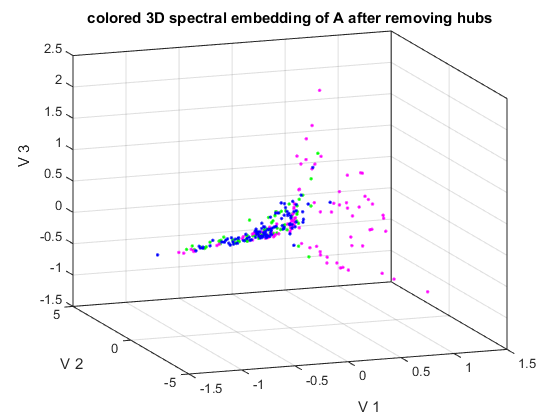


**Figure 7. Effect of *in silico* hub ablation on network path lengths. Left**: Path length matrix prior to ablation of the 3 vertices with highest hub scores. **Right**: Path length matrix post ablation.

|  |  |  |
| --- | --- | --- |
|  | Before removal | After removal |
| Avg path length | 2.5326 | 2.8762 |
| Max path length | 5 | 7 |
| Reachability percentage | 97.86% | 96.45 |

**Table 2. Effect of hub ablation on network performance.**

After removing hubs, the embedding shows less structure in terms of functional circuits. The interneurons tends to cluster together on the left side, while the motor neurons are mostly on the right side (Fig. 8). This suggest a significant dysfunction in the functional circuits, because the embedding suggests that most motor neurons are less connected topologically to the interneurons.



**Figure 8. 3D spectral embedding after *in silico* ablation of top 3 hub neurons.**

# Discussion

Here we present a framework for assessing functional classes of vertices in a biologically-based network and inferring their dependence on a subset of highly-connected vertices. In general, this was accomplished (1) identifying functional subgraphs of through graph spectral analysis, (2) identifying hub vertices within the network, and (3) reassessing cluster and hub composition of the network after one-by-one serial ablation of hubs *in silico* by reforming the adjacency matrix in the absence of those identified hubs. Through this approach, we identified specific neurons in the *c. Elegans* connectome that mediated the function of lower-degree vertices that were clustered together. Our clustering analysis successfully drew links between network topology and known neuronal function and our hub analysis exclusively identified interneurons as network hubs; both strong signs toward model validity. Therefore, we surmise that such a framework may be readily generalizable to other networks for which quality data sets are available.

Surprisingly, the *c. Elegans* network appeared to be unstable in the face of only a few key insults, as evidenced by the loss of tight spectral clustering and clear hub structure. However, an alternative viewpoint is that after deletion of the 3 most-connected vertices, the network became truly scale-free. In other words, the degree distribution was highly continuous from low to high values. This would imply that ablation past the top 3 hubs would not have had any major effect and so, beyond those key cells, the rest of the network is highly robust.

An important future direction should involve *in vivo* experiments in which our identified hubs are ablated; this show whether or not network topology completely determines robustness to vertex deletion. It is possible that factors other than connectivity contribute to network performance. Such factors could include diffuse action of neuromodulatory molecules as well as reactive axonal regeneration, synaptogenesis, or neurogenesis in response to injury. An acute *in vivo* ablation would illustrate network robustness to damage in real time, which may reveal information applicable to traumatic brain injury in humans, where the brain is forced to cope with a sudden and sever insult to the network. Alternatively, genetic deletion or impairment of a specific neuronal lineage would test the ability of the network to compensate through alternative wiring strategies. Understanding how the developing *c. Elegans* CNS is able to compensate for missing parts could improve our understanding of neurodevelopmental diseases such as Fragile X syndrome or autism spectrum disorder.

## Structural significance of spectral embedding

As can be seen in the above analysis, it is difficult to do a clear clustering of C-elegans neurons, primarily because (1) the hub analysis shows that neuron network lacks deep hierarchy: the hub neurons commonly acts globally (eg. AVA, AVB neurons) in the network, acting as the command neuron for a variety of behaviors. They cannot be effectively clustered. (2) It is difficult to define the size of each cluster since the functional neuron circuits have varying size depending on the complexity of behaviors.

Instead, we have proposed a new model based on information flow analysis. We obtained from spectral embedding the four-arm-with-central cluster topological structure. It can be seen from our investigation of functional significance that the 4 arms contain circuits of different types, ranging from sensory-based to motor-based. Within each arm, neurons utilize their own hub neurons to pass information, and seldom travel across arms.

We also found that the centrality of a neuron in the embedding is strongly positively related to the importance of the neuron as a hub in the network. The neurons in the central cluster are mostly interneurons in charge of relaying information between the sub-circuits in the 4 arms.

The framework describes the clustering information of less important neurons, according to which arm they are located in the embedding, but it also demonstrates the role of inter-neurons which are not clustered, but rather grouped together into a central cluster. They act as the intermediate relay station for large circuits involving coordination of neurons from multiple arms. Rather than assigning clustering labels deterministically, we embed them into a feature space where the closer a neuron is to one arm, the more closely related it is to the circuits within that arm.

## Connection with spatial locations

In general our graph embedding is independent of spatial structure. However given that the synaptic connections are more likely to exist between neurons that are close together, the connectivity also encodes spatial information, though to a less extent.

The distinction between the top arm and right arm is largely spatial-based. Notice that AS01-AS06 neurons are embedded into the top arm, while AS07-AS11 neurons are embedded into the right arm. The AS neurons span the entire body of the worm, with AS01-AS06 neurons positioned in the anterior of the worm body, while AS07-AS11 neurons positioned in the posterior of the worm body. Such spatial distinction results in different connectivity with inter-neurons, which causes them to be in different arms in the embedding. This demonstrates that neuron connectivity is not only dictated by the neurons’ role in the network, but also their spatial location in the network. Our framework thus incorporates both topological and geometric information through the embedding.

*Data Sources*

Connectome data as well as information regarding specific neurons was obtained from [www.wormatlas.org](http://www.wormatlas.org). To remove redundancies in our dataset, we analyzed only forward connections. To focus only on central synapses, we excluded NMJ connections.

*Hub Analysis*

All analyses and figures relating to hub neurons were performed and generated using custom MATLAB code which can be freely obtained from GitHub repository <https://github.com/RexYing/C-elegans-connectome.git>​.

Each node was assigned a “hubScore”, which simply reflected its degree within the network. Degrees were calculated by taking the row sum over the adjacency matrix corresponding to all chemical and electric synaptic connections.

# Supplement

**Methods**

*Spectral cluster analysis*

From dataset, we are able to construct the symmetric component of electrical synapses network, and the asymmetric component of chemical synapses network. We represented them using sparse matrix form in MATLAB from which first few eigenvectors can be efficiency calculated using iterative methods. The edge weights are defined by the synapse strength available in the dataset (need reference).

For each individual neuron represented as vertices in the graph, we associate them with 3 general types: sensory neurons, which mainly respond to the stimuli of external environment; motor neurons, which have neuronmascular junctions to muscle cells and are in charge of physical activity of the work; and inter-neurons, which mediates between neurons with different types. The type information of vertices is represented by different colors in the visualization, that provides understanding of the general information flow of the neuron network.

Spectral clustering is done by embedding the graph using eigenvectors of graph Laplacian as coordinates. When the eigenvectors corresponding to the smallest non-zero eigenvalues is chosen, the resulting embedding is an approximation to the Ratio Cut problem in graph theory (von Luxburg 2007).

The graph Laplacian is defined as:

Where is the diagonal matrix where , the degree of the ith vertex in the graph.

The Graph Laplacian satisfies the following property, as shown in the appendix, that for any vector of the same dimension as the column vectors of ,

Where is the adjacency matrix of the graph, and is the ith entry of the vector . It is called the quadratic form of a graph. In a special case, when is the normalized eigenvector corresponding to the smallest non-zero eigenvalue of , it follows that

The quadratic form reaches its minimum with the constraint that is perpendicular to the null-space of . Hence the spectral embedding that uses the eigenvectors corresponding to the smallest few non-zero eigenvalues optimizes for the total length of graph connections. If two neurons are closely connected through each other by strong synapses in very few steps, they are embedded close to each other in the resulting embedding.

When the graph is disconnected or almost disconnected (a neuron is connected to rest of the network with very few edges), the multiple eigenvectors correspond to the eigenvalue 0, and these eigenvectors essentially become connectivity vectors. Thus in practice we use the eigenvectors corresponding to the few non-zero eigenvalues of graph Laplacian.

The general procedure is to perform eigen-decomposition on the graph Laplacian generated from adjacency matrix of C-elegans neuron network, and choose eigenvectors corresponding to the smallest few eigenvalues as coordinates for each vertex. Plot the vertices using their respected coordinates. Finally, linear regression and K-means algorithm are performed on the resulting graph embedding, to obtain useful information about the structure of the connectome of C-elegans.