Methodology of Network Connection Removal Reveals Connection and Node Impact and Function in C. Elegans Locomotion Neural Network For Guiding Effective Designs for Artificial Neural Networks

By Kathryn Le

Personal Section

I have always had aspirations for learning more about computer science and neuroscience. Since middle school, I have been fascinated with the brain and how it works. At the beginning of high school, I set out to learn python and, later, machine learning. With the growing prominence of technology and artificial intelligence, I was attracted to the many applications it would have in society. Especially, when it came to creating artificial neural networks modeled after the ones in our brain.

Hearing about my school's notable research program, I decided to join the three year math research track. After sending 50+ emails to multiple professors and research lab heads across NYC, I finally found a mentor in computational neuroscience. When I connected with my mentor, I was so thrilled that I immediately started reading many of his papers to gain a better understanding of the work done in his lab. On being recruited to the mentorship, I was ecstatic to learn more about the computational neuroscience research that the lab was currently conducting on the C. elegans' neural network, their symmetry (or asymmetry), and the individual role of each neuron and connection.

As I started to analyze a C. elegans' neural network, I was able to gain a better understanding of their brain and expand that knowledge to larger organisms such as humans. Though our understanding of the brain has been limited, by exploring the C. elegans' connectome, I hope to uncover some of the many mysteries of the brain and neural network.

When the pandemic hit New York City, many labs, colleges and schools shut down indefinitely. Because of the severity of the situation, I lost contact with my research mentor for a few months, putting a halt on my research endeavors while I adjusted to online learning.

Once I was able to start conducting research, I had weekly meetings with my mentor on Zoom to report my progress and ask questions. It was very difficult for me to come up with a specific research question and methodology at first even though I had already nailed down a topic. The research question I

ended up tackling was inspired by my mentor's smaller training projects where he taught me basic neuroscience concepts I needed to know and simulations on Matlab. While conducting my research, I had to learn linear algebra alongside learning a completely new programming language: Matlab. Since Matlab mainly deals with matrices, linear algebra was key to understanding and starting my project.

Before research, I was still unclear about how computer science connected to the real world and its applications to science and other fields. Through research, I learned how computer science can represent and simulate complicated biological processes like neuron firings in the brain. My interest in science blossomed as I was exposed to the endless possibilities of science and math in discovering the unknown. These unknowns drive me to continue asking questions and researching to find an answer.

I had an amazing research experience despite the fact that it was during the COVID-19 pandemic. I would 100% recommend other high school students to undertake a research project combining science and mathematics. To these high school students, I would advise them to find a mentor who conducts research in the field that they are interested in and ask their mentor questions. There was no way that I could have completed this research project without the help and guidance of my mentor. I would also advise them to take advantage of the opportunity to learn as much as they can and have fun! At times, I would run into challenges and problems that would make me want to give up but redirecting my motivation to the purpose of my research and why I was interested in the topic in the first place helped me make the best of this opportunity. One such challenge was coming up with a research question and a direction for my research project, which is why I would advise high school students interested in research to read many research papers that intrigue them and build their understanding on the topic. After understanding the basic concepts in the computational neuroscience field, I was able to ask better questions and uncover some unknowns mentioned in research papers.

Research Section:

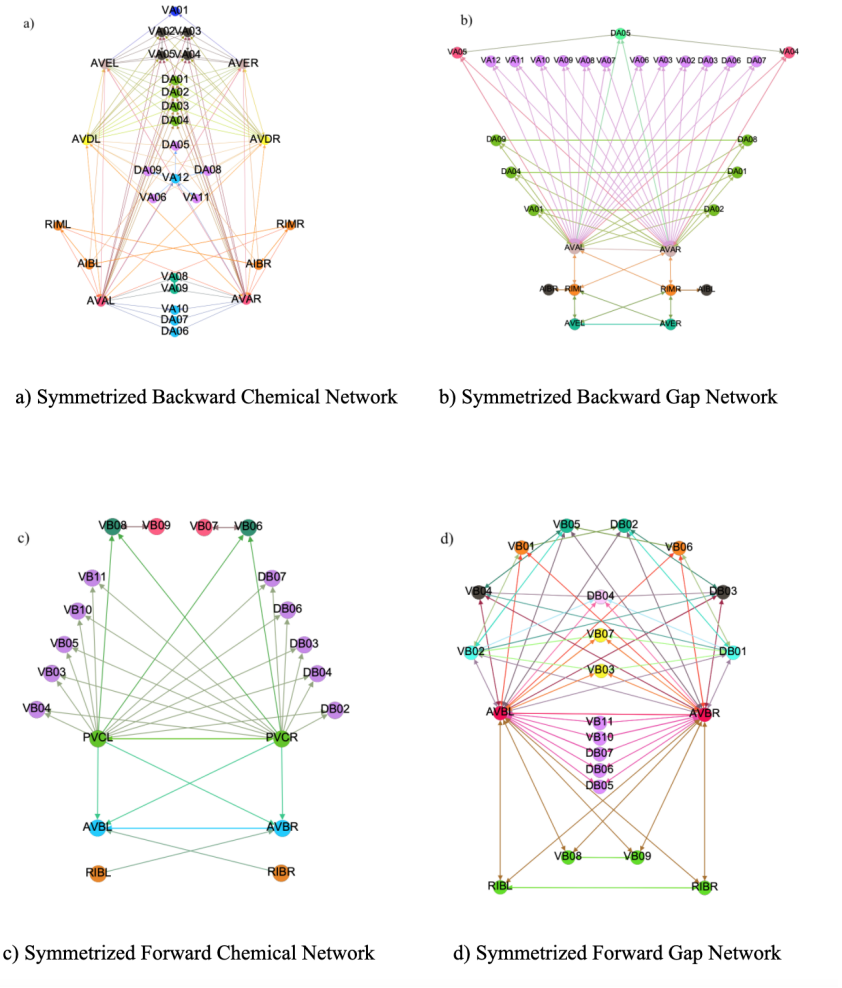
Introduction:

Human brains are way too complicated with billions of neurons and hundreds and even thousands of trillion connections that are still not completely understood. Because of this, studying a smaller "brain" permits one to better understand how the brain and the neural network influences the behavior of a creature. The C. elegans' connectome is the ideal network to research because of its simplicity (only consisting of 302 neurons and the fact that it has been completely mapped out. In this project, I strive to find the most important connections within different C. elegan sub-neural networks (chemical forward, gap junction forward, chemical backward, and gap-junction backward neural networks) using the symmetrized neural sub-networks. I randomly break the symmetry and use stochastic binary simulations to approximate its dynamics.

My study focuses mainly on the locomotion neural circuits of *C. elegans*. These neurons are categorized as forward or backward, controlling the forward and backward motion of the C. elegans, respectively. The forward and backward neural networks are further broken down into the gap-junction and chemical circuits where the gap junction circuit connections carry information that travels to and from the neurons connected (bi-directional) while the chemical synapse connections carry information that travels only in one direction (uni-directional). The locomotion circuit consists of two main functional classes of neurons called command interneurons and motor neurons. Command interneurons function as information processors as they receive input from sensory neurons (not studied) and pass on information and decisions to motor neurons or other interneurons.

My project includes using the symmetrized version of these circuits to observe the impact of stochastically removed connections on the symmetrized networks' dynamics. A neural network can be represented as a graph, which is a set of vertices connected through edges that can be bi-directional or uni-directional. A graph can be mathematically represented using a square matrix, also called the adjacency matrix. Using the discussion of permutation symmetry in "Symmetry group factorization reveals the structure-function relation in the neural connectome of Caenorhabditis elegans", the forward and backward networks from the Varshney C. elegans connectome were computationally symmetrized independently. I used the symmetrized networks instead of the experimentally mapped network because

the connectome variation can be up to 25% of the total number of neurons from worm to worm while the function of the connectome is basically unchanged. The symmetrized network acts as a generalization of many C. elegans neural networks that is required for this experiment.



a) Symmetrized Backward Chemical Network b) Symmetrized Backward Gap Network

c) Symmetrized Forward Chemical Network d) Symmetrized Forward Gap Network

Figure 1 | Graphs of symmetrized networks (Morone, F. & Makse, H. A., 2019)

The significance of randomly removing connections is to find the most significant neural connections in the network by comparing the dynamics of the original symmetrized network to the dynamics of the modified symmetrized network to determine which connections, when removed, cause the greatest difference in the dynamics of the neural network.

After symmetrizing the adjacency matrices, it was used to simulate the neuron dynamics of each sub-circuit independently. The algorithm used for simulation implements a Heaviside step function to

output 0 (low voltage) or 1 (high voltage) if the inputs reach a certain threshold to determine whether the neuron turns on or off (Triplett MA, Avitan L, Goodhill GJ., 2018). The Heaviside step function is shown in the Fig. 2 as Θ. It has an activation threshold, $\theta * InWeight$ (the percentage of input weights of a neuron), that the sum of the neuron's input weights (Wwhich is represented by an adjacency matrix) plus a stochastic term must be greater than θ to be able to turn on. β is a constant and $\curlyvee$(represented as vectors) are the neurons. The subscripts portray the type of neuron (i.e. F are forward neurons and B are backward neurons) and the superscripts portray whether the neurons are chemical (C) or gap junctions (G). The stochastic term that introduces noise into the simulation is *InWeight * Gaussian * Randi(-1, 0, 1)*. *Gaussian* represents the constant ideal mean and standard deviation values (mean: 0.4, std: 0.05). *Randi(-1, 0, 1)* outputs either -1, 0 or 1 with equal probability so that when it returns -1, the neurons are most likely to be turned off, 0, the neuron state is left to be determined by the neurons that send signals to it and 1, the neurons are most likely to be turned on. *InWeight* is the sum of all the chemical and gap input weights. The left side of the equation depicts the time step and time. Because we are looking at sub-circuits, The original equation is used for when the forward and backward chemical and gap junction circuits are interacting with one another, but because I was looking at sub-circuits, in order to conduct my investigation, the equations were modified to take this form:

$$x_i^F(t+1) = \Theta[(W_{FF}^G * X_F + W_{FF}^C * X_F + W_{FB}^G * X_B - W_{FB}^C * X_B - \theta * InWeight_F + InWeight_F * Gaussian * Randi(-1, 0, 1))]$$

$$x_i^B(t+1) = \Theta[(W_{BB}^G * X_B + W_{BB}^C * X_B + W_{BF}^G * X_F - W_{BF}^C * X_F - \theta * InWeight_B + InWeight_B * Gaussian * Randi(-1, 0, 1))]$$

Figure 2 | Binary Neuron Firing Simulation Equations

With the equations above, I was able to simulate the dynamics of the neural network. In order to solve the problem of which neural connections are most impactful, I constructed methods for removing neural connections from the network and calculating the impact of the removed connections on the network.

Methods:

With the symmetrized networks, I wanted to observe the impact of randomly removing neurons from each network. The networks were converted into a format that allowed me to work with them computationally. This format was converting the networks into adjacency matrices where the row neurons connected to the column neurons and the number in that position in the matrix represented a weight (0, 1) representing whether or not there is a connection between those neurons (0 indicates no connection and 1 indicates a connection). Since the weights of all the connections are converted to 1, the importance of the connections are determined by the number of edges to and from each neuron instead of the weights of the edges. This allows for more accurate results for which connections are most important because the scale of non-uniform weights can dramatically impact the results from the simulation and cause the results to be misinterpreted. Neurons from the chemical circuits are removed randomly and neurons to be removed from the gap junction circuits are selected randomly but removed symmetrically (i.e. if the connection AVAL to VA12 from backwards gap was randomly chosen, the connections AVAL to VA12 and VA12 to AVAL would be removed) because of the nature for gap junction connections to be bi-directional instead of uni-directional as in the chemical network. When removing connections, I also made sure that no connection was removed or selected for removal multiple times in all four networks. This method is both simple and efficient. It can also be applied to other organisms with less comprehensive data available than the C. elegans because the weights are only either 0 or 1 so the only data needed are the neurons and whether there is a connection between two neurons.

I then removed the same percentage of connections from each network since proportionally removing connections made for more accurate comparisons and analysis. For each circuit, I ran the neuron firing simulation a total of 200 times, where each time had 2000 time steps with a modified network that had 5%, 20% and 40% of total neural connections in each circuit randomly removed. Using these percentages, the backward chemical circuit, with a total of 119 connections, had 6, 24, and 48 connections randomly removed, respectively; the forward chemical circuit, with a total of 38 connections, had 2, 8, and 15 connections randomly removed, respectively; the backward gap circuit, with a total of

118 connections, had 6, 24, and 48 connections randomly removed, respectively; the forward gap circuit, with a total of 122 connections, had 6, 24, and 50 connections randomly removed, respectively. I analyzed the results by splitting the 200 simulations into two groups of 100 simulations. I then programmed a method for recording the removed connections in 10 of the 100 times the simulation was run where the dynamics of the modified network (using the Pearson's correlation) showed the greatest difference to the dynamics of the original circuit (this was done twice because there were a total of 200 simulations total for each network). Using the recorded removed connections from these top 10 simulations that showed the greatest difference between the modified network and the original network, I found the connections that were removed most frequently (the mode connections), indicating that these connections have a great impact on the dynamics of the network. I also modeled the frequency connections that were removed visually on a heatmap to show which connections possibly had the greatest effect on the original circuit.

I recorded the results and visually analyzed them using heatmaps and lists. For the heatmap, the rows and columns represent the neurons and each value/colored square represents the frequency that the connection was removed in the top 10 simulations for the first and second half of the 200 simulations. The list created was a combined list of all the connections that were removed most frequently in the top 10 simulations for both the first and second half of the 200 simulations. These heatmaps and lists were created for each experimental trial for each network (backward chemical, backward gap, forward chemical and forward gap).

Results:

To visualize my results, I created multiple graphs, heatmaps and other ways to visualize and understand the results. Connections in the heatmap are represented by small boxes of color on the heatmap where the neuron on the row (y-axis) is connected to the neuron on the column (x-axis). A higher frequency is indicated by a color higher up on the spectrum/color bar. Shown below are the connections for the backward chemical and gap circuits.

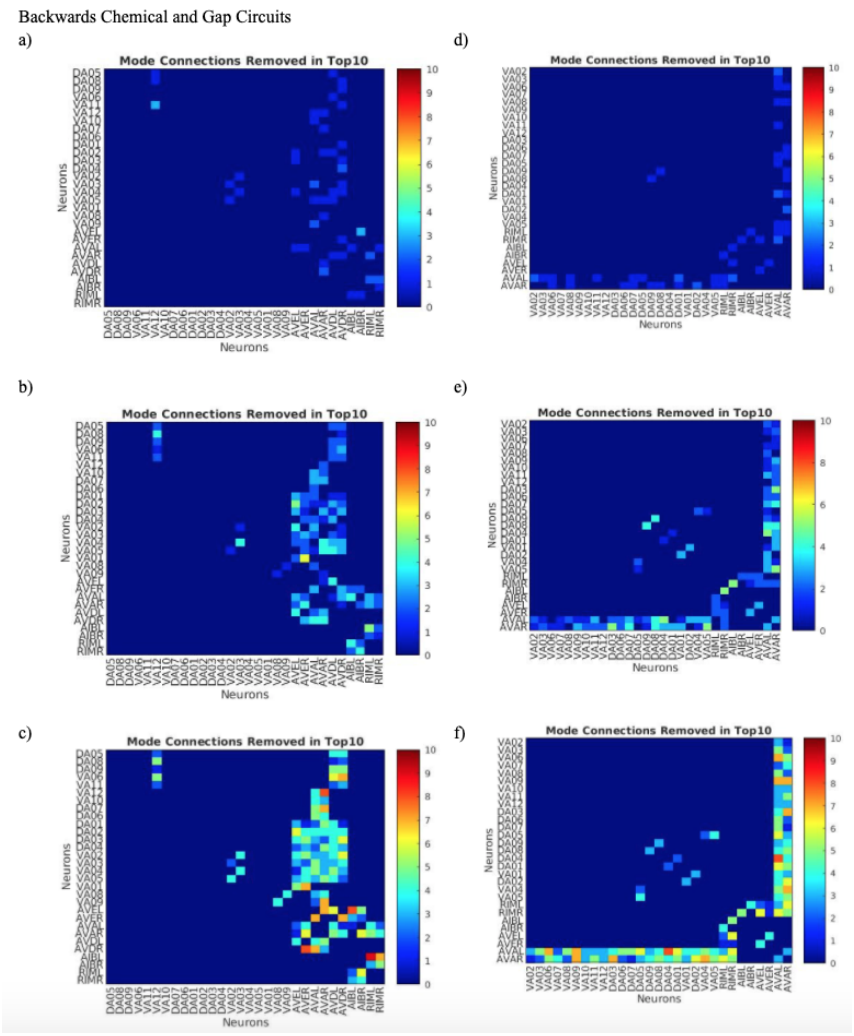


Figure 4 | Heatmaps of the frequency of neural connections removed in the top 10 simulations out of 100 simulations that showed the most difference in average Pearson correlation between the original network and the modified network for the backward networks in the first half of the total 200 simulations total.

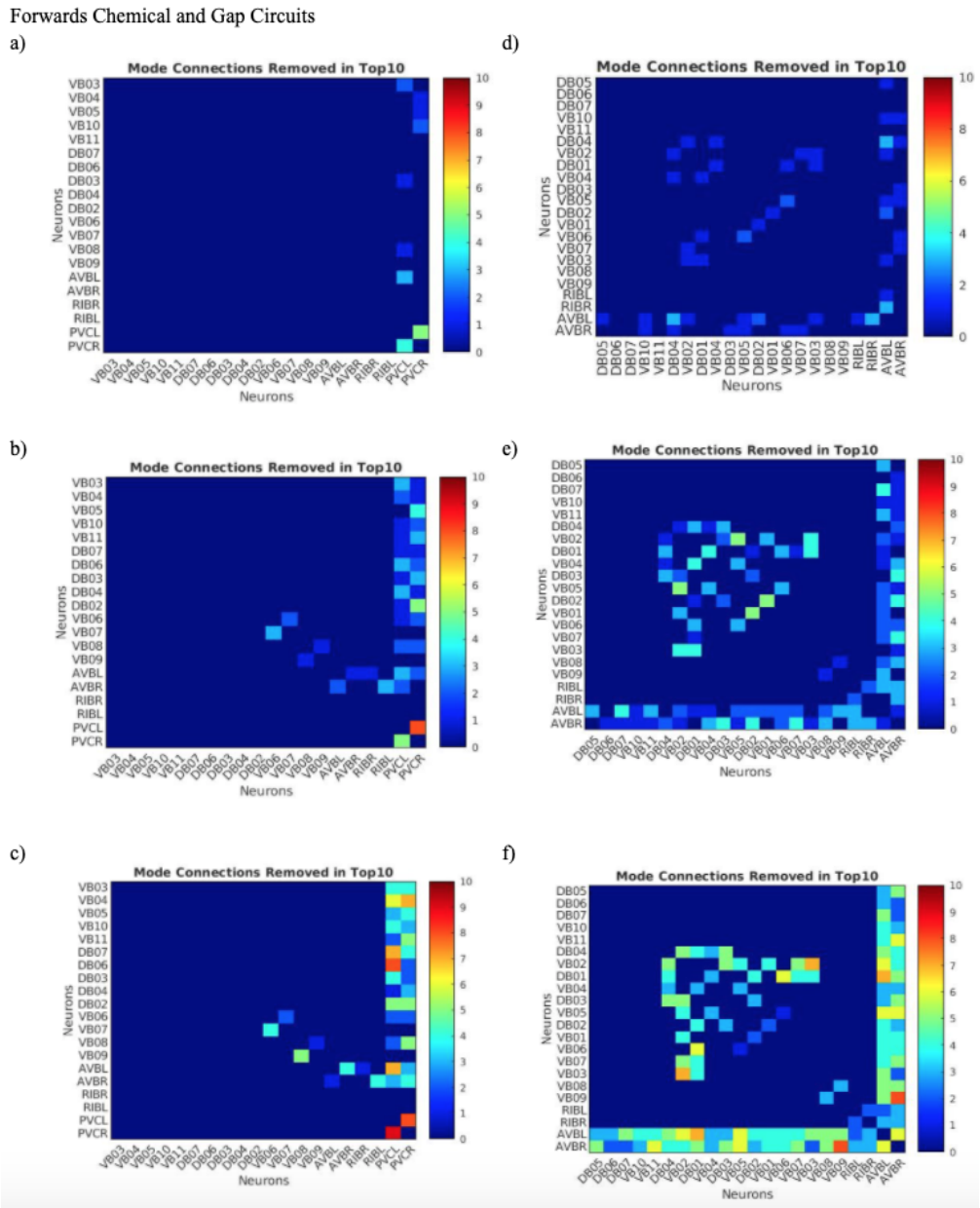


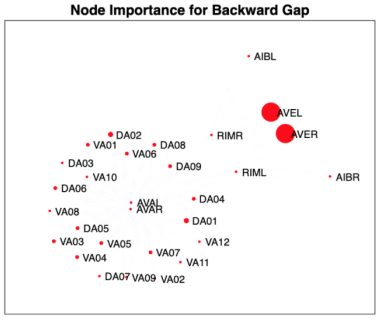
Figure 5 | Heatmaps of the frequency of neural connections removed in the top 10 simulations out of 100 simulations that showed the most difference in average Pearson correlation between the original network and the modified network for the forward networks in the first half of the total 200 simulations total.

To further understand the importance of each node in the subnetworks, I used the concept of centrality in graph theory and K-core to find the node importance to add to my results. Using the centrality function provided in Matlab, I graphed the nodes, labeled as the neurons they represent, with

the size of each node corresponding to their importance. The type of node centrality used were ‘degree’ for undirected networks and ‘indegree’ and ‘outdegree’ for directed graphs. These options in Matlab indicate that the node centrality will be calculated by the number of edges/connections connecting each node where a self-loop is counted. Since the chemical networks are directed (caused by its unidirectionality), I used a directed graph to represent their nodes and since the gap networks are undirected (caused by its bidirectionality), I used a graph to represent their nodes. For the undirected graphs, there were two different modes of determining node importance: In degree (degree of its inputs) and Out degree (degree of its outputs). There is only one mode for the directed graphs. The results of these graphs can be seen in Figure 7. Along with node importance, I also used the K-core filter on Gephi to determine the “layers” of each network and the degrees of each node. The last layer (or layer of the highest degree in the network) of each network can be seen in Figure 9.

Centrality for gap networks

a)



b)

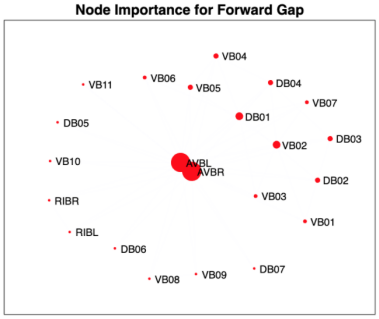


Figure 7 | Centrality graphs for gap junction networks.

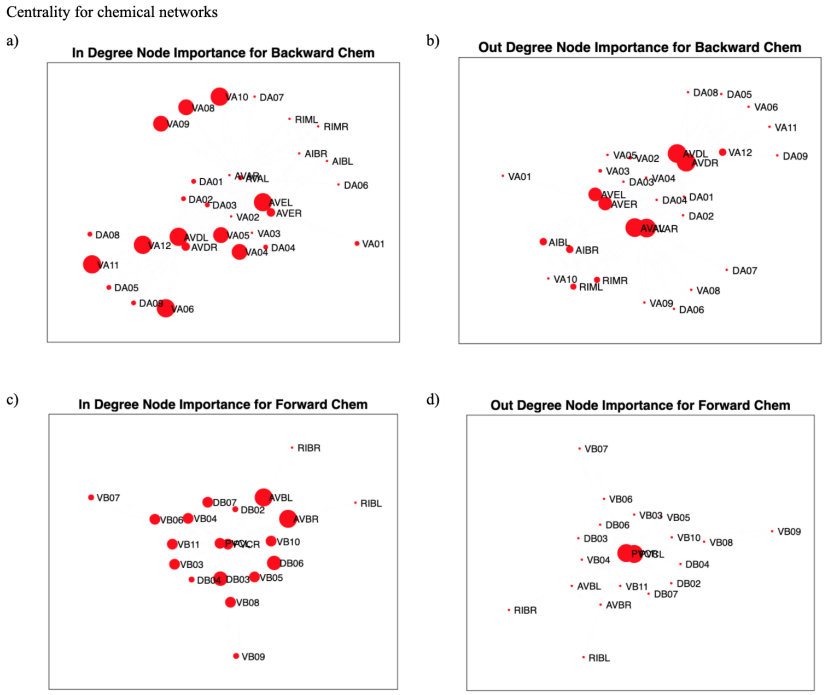


Figure 8 | Centrality graphs for chemical networks. a) Graph of node importance based on In degree for the backward chemical network.

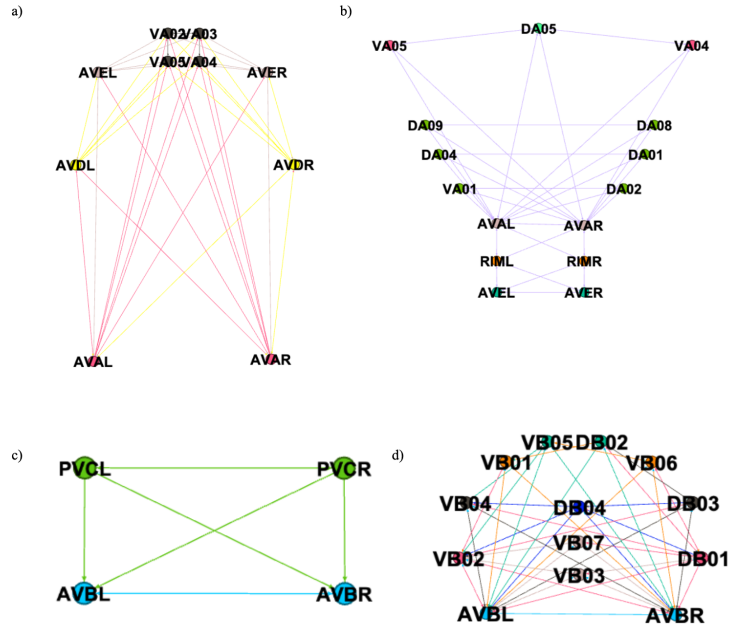


Figure 9 | K-core graphs for each network in its last layer. a) 7th and last K-core layer for the backward chemical network. b) 6th and last K-core layer for the backward gap network. c) 4th and last K-core layer for the forward chemical network. d) 8th and last K-core layer for the forward gap network.

Discussion:

With the results and graphs I found, I also wanted to observe important connections in each network and explain why they might be important through the structure of the network and types of neurons that make up the connection. For the forward chemical network it is easily observed that the most important connections are PVCL → PVCR and PVCR → PVCL because these connections had the greatest frequency. This might be explained by the fact that PVCL and PVCR are the main nodes connecting to the motor neurons and the communication between these neurons are crucial for the firing and symmetry of the circuit as a whole. In addition, these neurons are command neurons and control the firing of the motor neurons causing these neurons to have a great impact on the network. The connection between PVCL and PVCR might pass important information that affects the entire subnetwork. To confirm my analysis, I used the k-core analysis through the Gephi software, which shows the degree of nodes in the network. The k-core analysis for the forward chemical network reveals that the PVCL and PVCR nodes have a high input and output degree of 4.

The importance of connections for the forward gap network, however, is less obvious. Some general patterns that can be seen in the results for the forward gap network is that connections between the AVB neurons and the motor neurons were quite important and appeared frequently as impactful connections but the connections between the motor neurons were not as significant. In addition, the turquoise-colored neurons (VB02 and DB01) and the pink-colored neuron (DB04) as shown in Fig. 1 were significant neurons in which many connections involving these neurons were important. The k-core analysis for the forward gap network reveals that the last layer of nodes is with a degree of 8 with all of the motor neurons, AVBL and AVBR neurons present. This supports my analysis and results because

connections involving command neurons and motor neurons (especially AVBL, AVBR, DB04, VB02 and DB01) are important connections with a high degree of inputs and outputs.

The important connections in the backward chemical network are connections between command neurons and motor neurons only. The connections between motor neurons or between command neurons are not as important as those between command and motor neurons for the backward chemical network. This is seen in the results because connections involving AVE, AVD and AVA (left and right) neurons show up frequently as removed neurons in the top 10 simulations that had an average Pearson correlation that differed most from the average Pearson correlation of the original network. This is because it is important for the command neurons to be able to communicate with the motor neurons so that the organism can obtain the information necessary to ensure it moves correctly. Through the K-core analysis, this pattern is also seen with AVE, AVD and AVA neurons still present with a degree of 7. However, the motor neurons, VA02, VA03, VA04 and VA05 are also present which differs from the results collected. Similar to the forward chemical network, the backward chemical network is also a directed network so the k-core analysis may only be an approximation for this network.

Lastly, the backward gap network is similar to the backward chemical and forward gap networks in that the important connections are mainly between interneurons (e.g. RIMR → AIBL) and between interneurons and motor neurons (e.g. AVAR → VA01). As seen in Figure 6, most of the connections that have been removed in a simulation iteration that showed a high difference in average Pearson correlation from the original network are connections involving the interneurons, AVA, AVE and RIM. This is consistent with the findings from the other networks because command neurons and interneurons send important information to the motor neurons and receive important feedback from the motor neurons. Using the k-core filter, AVA, AVE and RIM were a part of the last layer (layer 6) along with some of the motor neurons that also frequently showed up in Figure 6 (DA01, DA02, DA04, DA05, DA08, DA09, VA01, VA04, and VA05).

To further my analysis, I used a graph theory concept, centrality of the graph to measure the importance of each node. As can be seen in Figure 7 and 8, the size of the nodes are proportional to the

degree of each node. Interestingly, for the chemical networks, the out degree graphs match the results collected (which show that the connections between PVCR and PVCL are important connections) closer than those of the in degree graphs. This makes sense, however, because command neurons are more frequently sending information out to motor neurons than receiving information. For the backward chemical network, the out degree graph shows that the AVE, AVD, and AVA neurons have high out degrees. For the forward chemical network, similar to the other findings in this paper, the out degree graph shows that PVCL and PVCR have significantly high out degrees, but in the in degree graph, AVBL and AVBR have the highest in degrees. The gap network graphs are also similar to the results collected in this paper because the command neurons AVE and AVB in the backward gap and forward gap networks, respectively, had the highest in and out degrees. These results show that the importance of the significant and impactful connections in a network correspond to the nodes that they connect. After seeing that the out degree graphs match the results collected more closely than the in degree graphs, it can be concluded that this methodology more accurately reflects which connections/nodes in a network have the highest out degree if the network is directed and which connections/nodes in a network have the highest degree if the network is undirected. This is because the methodology removes connections, which would affect the entire network more greatly if the connection removed connected nodes that had high out degrees and sent information throughout the network. However, if the connection removed connected nodes that had high in degrees, the node would only be missing one of its inputs and can still receive inputs from other nodes so that its impact on the entire network is not as significant.

Overall, I was able to determine the most important nodes and connections in a network (in terms of how the network reacts without these connections and the inputs and outputs of the nodes) using a novel methodology of equally weighting the symmetrized network and randomly removing connections to compare the average Pearson correlation of the modified network and the average Pearson correlation of the original network. In summary, connections that had a high impact on the dynamics of the network were mainly connections involving interneurons or command neurons. It was also found that in the process of determining the importance of certain connections in a network, the connections can provide

information about the nodes that they connect such as the level of their degree. This methodology and its results lead to new methods for studying the structure and function of C. elegans' neural network and even other organisms' neural networks. Further research can be done on the full network instead of just the locomotion component. This method can also be used for other organisms whose neural networks are not as well understood as the C. elegans' neural network. With these current discoveries and research, scientists can use the results from C. elegans' neural connectome to improve artificial neural networks by applying this understanding of a neural network's structure to artificial neural networks, bettering their performance and efficiency. Because artificial neural networks depend on knowledge of the actual neural networks, the more these artificial neural networks emulate biological neural networks the more efficient and independent they will become. With this piece, I can apply these discoveries to artificial neural networks and conduct further research on other important neural connections in C. elegans and even in humans.