Spatial Embedding of Edges in a Synaptic Generative Model of *C. elegans*

Zachary Laborde and Eduardo J. Izquierdo

Cognitive Science Program, Indiana University Bloomington
Program in Neural Science, Indiana University Bloomington
zlaborde@iu.edu

Abstract

Despite so much research being put into neuronal networks, the development of their structure remains unclear. Synapses form because of many factors including spatial location, guidance molecules, and gene expression. It remains unclear how significant a role each of these factors plays in these networks' formation. With a relatively small and well-understood connectome, C. elegans is an ideal subject for this topic. Many unknowns and experimental challenges currently impede a detailed understanding of nervous system development. Computational modeling of these processes can begin to answer these questions, potentially leading to new insights that would otherwise be difficult to attain. Previous research has attempted this by stochastically assigning connections, weighted by distance. This model created networks that were found to be, on average, within one standard deviation of an adult C. elegans connectome in distance, connectivity, and the number of bidirectional links. The model did, however, result in networks with a very different clustering coefficient, which was 2 standard deviations away from the C. elegans connectome. We found that this model created networks with degree distributions more akin to random networks than to a C. elegans connectome. While the spatial location of the somas was accounted for, the spatial location of the neurites was not. Neurons in C. elegans typically have 1-2 processes. This physically constrains where synapses occur. Spatially embedding these processes may result in networks more like a C. elegans connectome than in the past. We created a new model which forms connections at intersections of artificial neurites, which pass through the densest region of the nervous system, the nerve ring. Because sensory/motor neurons often develop in a predetermined manner, testing was restricted to the frontal ganglia. The model's results were compared with the results of the previously mentioned model, random networks, and two C. elegans connectomes. The clustering coefficient, average distance, and total bidirectional links were all compared for consistency with previous research results. These metrics found the new model to be most like the C. elegans connectomes. The degree distributions of the resulting networks of this new model were closest to C. elegans. Mapping these networks in physical space revealed that the networks from this new model were visually dissimilar to those of C. elegans. Connections between neurons on opposite sides of the pharynx (the center) were underrepresented. The improvements made by this new model suggest spatial constraints of neural processes are important in the development of synaptic connections.

Introduction

Despite tremendous computational power, no person has been able to create an algorithm as capable as the human brain. With its approximately 86 billion neurons (Azevedo et al., 2009), the human brain is an organ so complex that charting its synaptic connections alone is a monumental task (Van Essen et al., 2012). Tremendous time and effort are being invested in the creation of a complete human connectome without a definitive explanation of how this data would explain how the human brain functions (Ceylan et al., 2022; Rheault et al., 2020). Accurate mapping of these connections in living brains alone prove difficult given modern technological limitations (Rheault et al., 2020). With a neuronal network that is approximately one percent of one percent the size of a human's (Azevedo et al., 2009; White et al., 1986), Caenorhabditis elegans is the only organism today with a completely mapped connectome (White et al., 1986). One might expect that the neural functions of C. elegans would be well understood. However, the dynamics and behaviors of the C. elegans neuronal network (CENN) have not been accurately replicated. To truly understand how a brain works, a neuronal model replicating the structure and dynamics of said brain is essential (Izquierdo, 2019). Synaptic connections are a fundamental part of microscale connectomes (Sporns et al., 2005), but many of these connections vary dramatically across individuals. Over half of all synaptic connections within an adult C. elegans are variable (Witvliet et al., 2021). While the topology of synaptic connections cannot predict the dynamics of a neuronal network, it does restrict what sort of dynamics are possible (Prinz et al., 2004). Any computational model of C. elegans that cannot account for all potential variabilities within the CENN would not be biologically accurate. The CENN exhibits consistent global attributes despite having so many variations (Pérez-Escudero & Polavieja, 2007). What combination of variable connections lead to biologically accurate CENN remains unclear (Witvliet et al., 2021). Any computational model of the CENN cannot be tested for complete accuracy without this information. Knowing what neuronal network topologies are possible for a neurotypical C.

elegans is necessary to accurately modeling the entire dynamical CENN. Manually mapping all possible topologies of C. elegans leads to a combinatorial explosion. With 302 neurons, there are 302302 or 9×10748 possible connection configurations for the CENN. This only becomes worse with more complex organisms such as Drosophila or humans. A trial-and-error approach is impractical in this context. Generative models are a promising solution to this problem of network combinatorial explosion. Generative network models create networks from a set of rules defining where and how connections are created (Betzel & Bassett, 2017). By focusing on how networks form and not on a complete list of nodes and edges, generative models are very scalable. These models are an effective solution for neuronal networks of any complexity. Several generative models attempt to accurately predict the network topology of C. elegans (Costa et al., 2007; Itzhack & Louzoun, 2010; Khajezade et al., 2019; Nicosia et al., 2013). The Random Distance Dependent Attachment Model (RDDAM) is one such model (Itzhack & Louzoun, 2010). It randomly forms connections with a probability that logarithmically decreases with an increase in spatial distance between two neurons. The RDDAM creates networks shown to resemble the CENN in average length of connections, average connectivity, and shortest path length distribution. Given its simple rules that align with observed neuronal organization in C. elegans (Pérez-Escudero & Polavieja, 2007), the RDDAM provides an excellent framework for future generative models of CENN connectivity to which to compare or improve. An essential aspect of the RDDAM is the spatial embedding of each neuron. Neurons exist in physical space, interacting chemically and electrically with their surroundings. What connections they make are a result of their location relative to other cells (Hentschel & Ooyen, 2000; Kaiser & Hilgetag, 2006; Pérez-Escudero & Polavieja, 2007). Spatial embedding is not only relevant for the somas but for the neurites as well. Neurites take up physical space within the body of C. elegans, limiting how many connections are physically possible within the organism. The neurons of C. elegans each have one or two neurites which rarely branch, forming most of their connections as "en passant" synapses (i.e. synapses that do not form at the tip of the axon) (Durbin, 1987; White et al., 1986). This pattern of neuronal connectivity is not accounted for in existing models of C. elegans neuronal connectivity (Costa et al., 2007; Itzhack & Louzoun, 2010; Khajezade et al., 2019; Nicosia et al., 2013). Constraining connectivity to intersections along the path of a neurite may result in a network which is more similar to CENN than the results of existing generative model. I hypothesize that a generative model which spatially embeds both nodes and edges will result in a network more similar to CENN than RDDAM.

Model and Methods

Data

The length and organization of neural processes in C. elegans are highly variable, particularly in the pharyngeal nervous system. To simplify this, this research only used neurons in the frontal ganglia, whose length in connections are less variable. All synaptic connections of 131 neurons in the adult C. elegans frontal ganglia were downloaded from www.dynamic-connectome.org/resources (Kaiser & Hilgetag, 2005; Kötter, 2004; McCormick et al., 2004). 3D positions of the neurons were taken from Skuhersky, Wu, Yemini, Boyden, & Tegmark (2021). Newborn (L1) (0 hours) connectome data was downloaded from https://nemanode.org (Witvlietet al., 2021)

Model

UPDATE WHEN ADDING RANDOM DIRECTION SEEM I developed a generative model which spatially embeds edges as well as nodes. This modelwill be referred to as the Spatially Embedded Edge Model (SEEM). SEEM iterates over every node. Each node locates the node closest to it that is on the opposite side of the nerve ring, a region which contains more than half of all neural processes (Altun, 2017). A line segment is drawn from the source node through the target node, terminating at the surface of the bounding box of the network. The bounding box is defined as the smallest 3D box that can hold all nodes in the network. These line segments are intended to represent a single neurite. After all "neurites" have been drawn, the overlaps between these line segments are recorded. Given these neurons have a diameter of 2-3 microns (Schafer, 2006), a distance greater than or equal to 1 micron between any two lines will be considered an overlap.All possible edges in the model have now been determined. To generate a network, a random selection of directed edges are selected from the set of all possible edges. The number of edges selected can be adjusted. For these experiments, I used as many edges as were in the comparable CENN. The code implementing this algorithm can be found at the link listed under *Data and Code Availability*.

Network Comparison Methods

The output of SEEM was compared to the CENN, RDDAM outputs, and Erdos-Renyi random graphs. Excluding CENN, each model was run for 1000 iterations, resulting in 1000 possible graphs. Each metric was run over every instance of the graph and the average value was recorded. To compare these networks, I used three measures used in Itzhack and Louzoun (2010). These were average clustering coefficient, average distance, and total bidirectional links. I attempted to include Average Connectivity as was done in the mentioned paper, but this proved difficult, so it was not included in this analysis. To better understand the connectedness of the graphs, I also included average closeness cen-

trality as a fourth comparison. Average degree distributions were also included for comparison. Finally, visual comparisons of randomly selected spatially embedded graphs were also incorporated to highlight patterns in the resulting networks.

Results

Network Statistics CHANGE

SEEM was found to be the set of networks closest to the CENN in average edge distance (see table 1 and fig 1). [INSERT STATS FOR STATISTICAL SIGNIFICANCE BETWEEN DISTRIBUTIONS]. RDDAM were closest to the CENN in average clustering coefficient and number of bidirectional links. [INSERT STATS FOR STATISTICAL SIGNIFICANCE BETWEEN DISTRIBUTIONS]. These results were not heavily dependent on which network was chosen.

		Clustering Coefficient	Edge Distance	Total Bidirectional Links
L1	SEEM	0.036307	17.593802	11.169000
	RDDA	0.136219	6.607809	82.841000
	ERN	0.028589	19.152152	6.943000
	CENN	0.115779	15.327489	50.000000
Adult	SEEM	0.090883	17.589113	68.313000
	RDDA	0.253409	7.550171	284.917000
	ERN	0.070523	19.193761	42.652000
	CENN	0.197319	15.170374	150.333333

Table 1: INSERT HERE

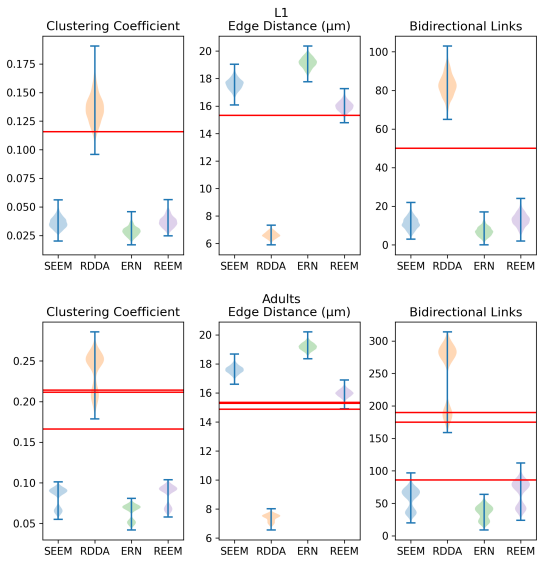

Figure 1: INSERT HERE

The degree distributions of RDDAM were also found to be most similar to the CENN with a Wasserstein Distance of 0.005 in L1 and 0.002 in adults (see fig 2). SEEM was close with a Wasserstein Distance of 0.006 in L1 and 0.003 in adults, making the distinction between the two insignificant. The degree distributions of Erdos-Renyi graphs appeared much less like the CENN with a Wasserstein Distance of 0.013 in L1 and 0.009 in adults. The average degree between all four graphs were not significantly different.

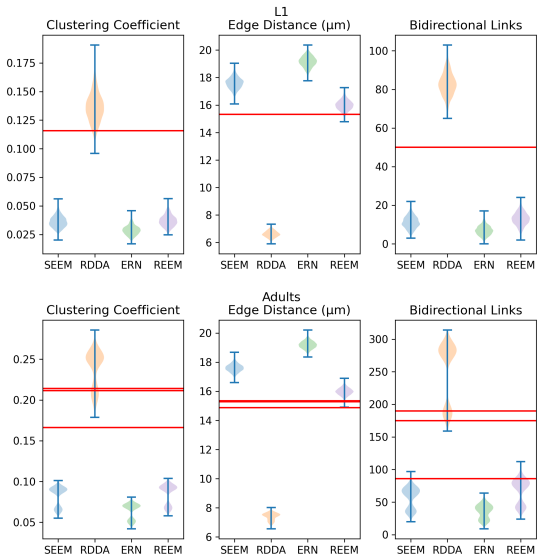

Figure 2: INSERT HERE

Comparing the distribution of edge distances shows a different picture. Here, the resulting graphsof SEEM were found to be closest to the CENN (see fig 3). The edge distance distributions in SEEM had a Wasserstein Distance of 0.004 in L1 and 0.003 in adults. The Erdos-Renyi graphs, with a Wasserstein Distance of 0.005 in L1 and 0.004 in adults, were even more similar to CENN than the RDDAM, with a Wasserstein Distance of 0.013 in L1 and adults. These results distinguish SEEM from RDDAM, showing the proclivity of RDDAM to form very short connections, something SEEM does not appear to do.

Spatially embedding these graphs using provided coordinates resulted in a particular connectivity pattern shared between the CENN and SEEM graphs (see Figure 4). [INSERT DATA ABOUT DISTRIBUTION OF CROSS-PHARYNX CONNECTIONS]

Discussion and Conclusions

What does it mean? Start off with a brief summary of the results and the main insights gained. Are there any things that came up during experiments and interpretation of the Results that warrant a further discussion. What are the limitations of your model. You are encouraged to be a good critic

of the work and express honestly its shortcomings. End in a forward-looking note. What could be done next? What do you see as the most immediate next steps that you would do (or that somebody else reading your paper might want to set out to examine).

How long should this document be? It doesn't matter for now. If you end up with a two-page report, that's fine. If you end up with a longer one, that's fine too. Keep in mind that the point of this report is to efficiently communicate your scientific findings in a way that makes your claims supported by clear evidence, your methods reproducible, and your results easily interpretable.

Additional reading? Finally, take some time to do some additional reading on the structure of scientific papers. These are two resources that I think might be useful (they are both hyperlinks, so should should be able to click on them): "Scientific Papers by Nature" and "The Structure, Format, Content, and Style of a Journal-Style Scientific Paper".