

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

NOTES

Make it smaller (1/3 to 1/4)

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

NOTES

I see the Introduction has a good set of components to it. I highly recommend that you try to break it down into paragraphs. Each paragraph should have a topic sentence. And that the overall structure of the Intro should be relatively straightforward: 1. Overall broad motivation for this project. Starting broad but probably ending more specifically with the question of interest. 2. Related work, what is known in this area. What has been done so far. 3. What challenge remains open and what are you proposing to do in this paper. 4. Organization of the paper. What should the reader expect in what follows?

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

NOTES

The Model/Method needs an initial paragraph that describes the section more broadly and says more about how it will be organized. From reading the Method, it is really hard to understand what you are doing. Try to Zoom out first and explain the larger goal first. Then explain the algorithm broadly. Then walk through details one by one. Use steps and orderings to help you at some point: There are 5 key steps in the process of achieving Y. First, we do this. Then, this. Finally, this.

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> (Witvliet et al., 2021)

Model

UPDATE WHEN ADDING RANDOM DIRECTION
SEEM I developed a generative model which spatially embeds edges as well as nodes. This model will 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 centrality 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

Initial Model

To answer the question of whether spatially embedding edges is an important factor in the formation of the *C. elegans* Neural Network (CENN), we compared networks from SEEM with the CENN. In order to gauge how close our model is to the *C. elegans* networks, we also compared these networks with RDDAM and random networks (Erdos-Renyi Networks / ERN). To rule out development as a factor, we made comparisons to both a single L1 connectome (0 hours) and a set of 3 adult (L5) connectomes (INSERT REFERENCE). We measured these generative network models from a set of 1000.

NOTE: Is it clear that we are averaging across 1000 graphs?

Network Statistics Comparing these networks can be done in several different ways. As an initial comparison, we

chose to use network statistics to compare essential aspects of these graphs.

We chose to compare the average clustering coefficient, average edge distance, and average total number of bidirectional links. These network statistics were chosen as they were used by Itzhack & Louzoun (2010) in their paper on the RDDAM.

We plotted the results of the network measures of all 1000 instances of each model (see fig 1) and compared them with the results of these measures on the CENN. For the newborn (L1), we compared it to a single graph, but, for the adults (L5), this was three graphs.

Looking at the results, we did not notice a significant difference between the results of L1 and L5 networks. One difference of interest was that the networks of Witvliet et al. (2021) had noticeably different clustering coefficients and total number of bidirectional links when compared with the enhanced connectome of White et al. (1984) (NOTE: N2U). It is difficult to determine what might be causing this discrepancy, whether it is a result of different environmental conditions or if this is indicative of flaws in recording methods of some or all of the connectomes. Rather than averaging the results of the connectomes, we chose to show all three individually (see fig 1).

		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

Clustering Coefficient: RDDAM was closest to *C. elegans* in this statistic. This reinforces the idea that spatial proximity plays an important role in how neurons form connections and how they cluster in the way they do. It is surprising that our model did not have a stronger clustering coefficient given that closer neurites do not need to cross as large of a distance to potentially connect. This might account for why our model does slightly better than a random network. Although this difference is not that great.

Edge Distance: The average edge distance found that our model was the closest to *C. elegans* in this regard. It is important to note that our random networks were not that far behind. RDDAM had much shorter edges on average than any of the other networks. Given that it prioritizes proximity

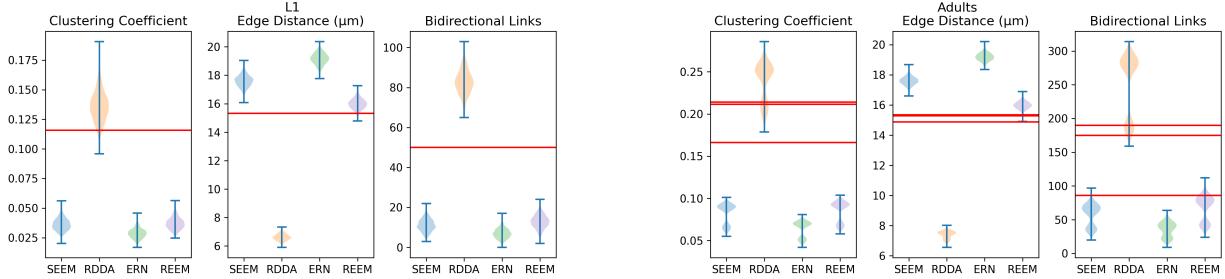


Figure 1: INSERT HERE

when making connections, this result is unsurprising. Given that our model was most similar to *C. elegans* in edge distance, this suggests that spatial embedding of neurites and soma alone might explain this characteristic of the network. As explained in the last section, neurites that start close to each other would be more likely to cross paths than those that start far away from each other.

Bidirectional Links: The number of bidirectional links in *C. elegans* do not appear to be similar to any of the graphs resultant of these presented models. Given that our model creates graphs which are random subgraphs of a bidirectional set of all potential connections, directionality does not play a role in deciding connections. It having a higher number of bidirectional links can simply be explained by this reduced pool of connections to pick from. One is more likely to flip two heads with a coin than one is to roll two 1's with a die, given the same number of flips/rolls. By the same logic, a marginal increase in the number of bidirectional links likely is not a result of any novel network property. The RDDAM overrepresents bidirectional links. Given its preference for short connections, an even smaller subset of possible connections are most likely to be made, which are direction agnostic. This would explain why RDDAM has so many bidirectional links. The stochasticity of the model would also explain the larger range of values. None of these models sufficiently explain the amount of bidirectional links exhibited by *C. elegans*.

What does this tell us? Looking at these statistics, it appears that RDDAM well characterizes the highly clustered neuronal networks of *C. elegans*. The edge distance of *C. elegans* appears to be only slightly shorter than random, which is likely explained by the spatial embedding of the network. The bidirectionality of *C. elegans* could not be explained by any network. It is notable that for all of these measures the results of our model are similar to random networks. This might be a result of how instance networks are created through random selection. It might also indicate that this network is not selective enough. As it stands, there is not a definitive explanation for these results.

Distributions While these statistics give us an idea of how some of the attributes of these networks in describing the underlying structure of the CENN, it does not give us a picture of how the edges in any of these networks are distributed. We had two questions: 1) What is the structural makeup of these networks? (Are they scale-free, small-world, etc.)? 2) What is the physical makeup of these networks?

To answer these questions, we compared the degree distributions of these networks to examine their network structure and the distance distributions of these networks to examine their physical structure (see fig 2). To get a better idea of how similar these distributions were, we also measured their Wasserstein Distances from the *C. elegans* distribution. For the adults, these distributions were averaged.

Degree Distributions: We found that our model and RD-DAM to have degree distributions closest to *C. elegans*, with RDDAM being slightly closer. This is the first result which differentiates our model from random networks. These two models result in networks which appear to have more variability in degree than random networks. Since both rely on spatial proximity of nodes, which are not evenly distributed, it would be expected for their degree distributions to favor nodes which are closer to many others. This could explain why we see these results. It is notable that the degree distributions of CENN have much longer tails. This means that *C. elegans* connectomes are more centralized than any of the other networks being compared.

Distance Distributions: The edge distances of each model show an issue with RDDAM: it favoring close connections. We see that our model is most like *C. elegans* in this distribution, the same as what we found when comparing average edge distributions. This indicates that *C. elegans* frontal ganglia is not particularly distance dependent as indicated in Itzhack & Louzoun (2010). Given the similarity in the distributions, spatial embedding alone could explain for the edge distance distribution of *C. elegans*.

Bidirectional Edges: Given the results of comparing numbers of bidirectional edges in these networks, we were curious as to what these bidirectional connections looked like in *C. elegans*. Did they result from a preference for short connections, were they random, or were they distributed in a

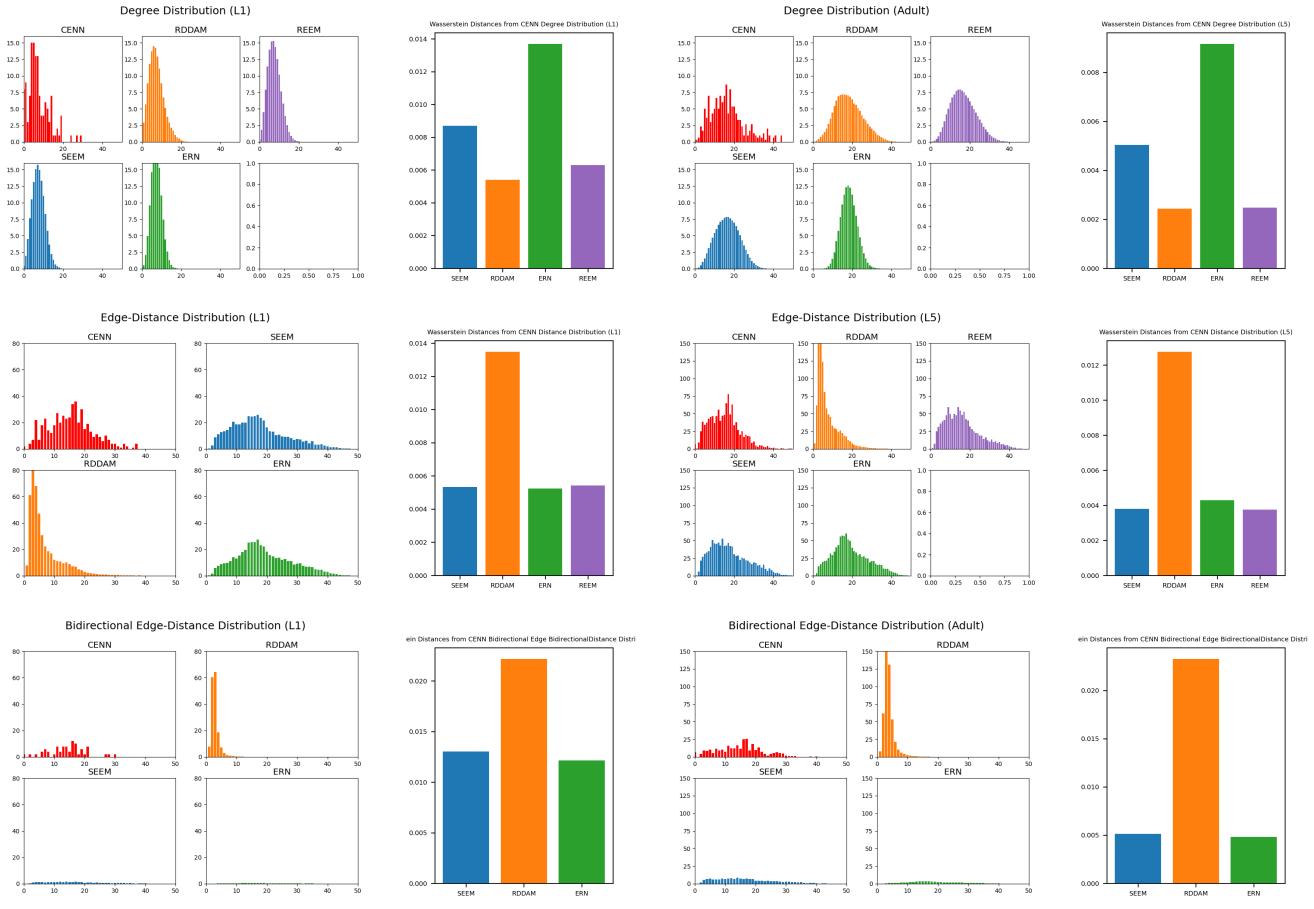


Figure 2: Degree distributions, plotting the number of nodes (y-axis) with a particular number of connections (x-axis). **NOTE:** Going to remove the purple sections from these graphs, use line graphs so I can overlap all distributions, add more caption, add letter labels, and will likely separate and improve the bottom two graphs.

completely different way from any of our models. To answer this, we compared the edge distance distributions of all bidirectional edges in each graph (see fig 2). From the results, the bidirectionality of *C. elegans* frontal ganglia appears to be mostly random with the random graphs being closest to *C. elegans* and our model being a close second. RDDAM, preferring short connections, resulting in a very different looking distribution. Despite the number of bidirectional edges being between the amount shown in RDDAM and our model, it appears that these connections are random with respect to spatial distance. It is important to note that the number of bidirectional edges is still much higher than one would expect in a random graph.

Spatial Embedding

Overlaps

Breaking Down the Model

What aspects of SEEM lead to a model that in some ways is more similar to the CENN? This model's input parameters are the location of the neurons, the expected number of connections, the location of the Nerve Ring (VALUE), and Epsilon (the maximum distance two neurites must have to form a potential connection). The first two of these parameters come from empirical data so they will not be adjusted. Adjustment of the location of the Nerve Ring cannot be meaningfully adjusted as there are only a small number of neurons which closely lie on this border, making such changes insignificantly impactful to the overall shape of the network. Rather than adjusting the nerve ring location, adjusting the algorithm for determining neurite direction may prove as more useful in parsing out the model's relevant parameters.

THIS IS WHERE WE TALK ABOUT THE RANDOM DIRECTION VERSION OF SEEM.

Conclusion

While significant experimental progress is made to understand the developmental process of *C. elegans*, a complementary modeling approach allows us to test what assumptions in our thinking are more and less important for the process.

To date, there's only been one model of the synaptogenesis of the *C. elegans* connectome that has aimed at predicting the connections between its neurons. In their work, they based their predictions on the distance between neurons alone.

In this work, we build on that work and test one important hypothesis: whether neurites directed in space is predictive of the neural structure of *C. elegans*.

Future Directions

Recent mappings of *C. elegans* connectomes have shown how variable it is (2). We are currently comparing the connectomes from various adults and at various developmental stages to potentially learn new heuristics about mechanisms to include and test using these computational models, accounting for the underlying variability.

Data and Code Availability

All scripts and files used to generate all figures are available at <https://github.com/Zach-Attach/SEEM.git>

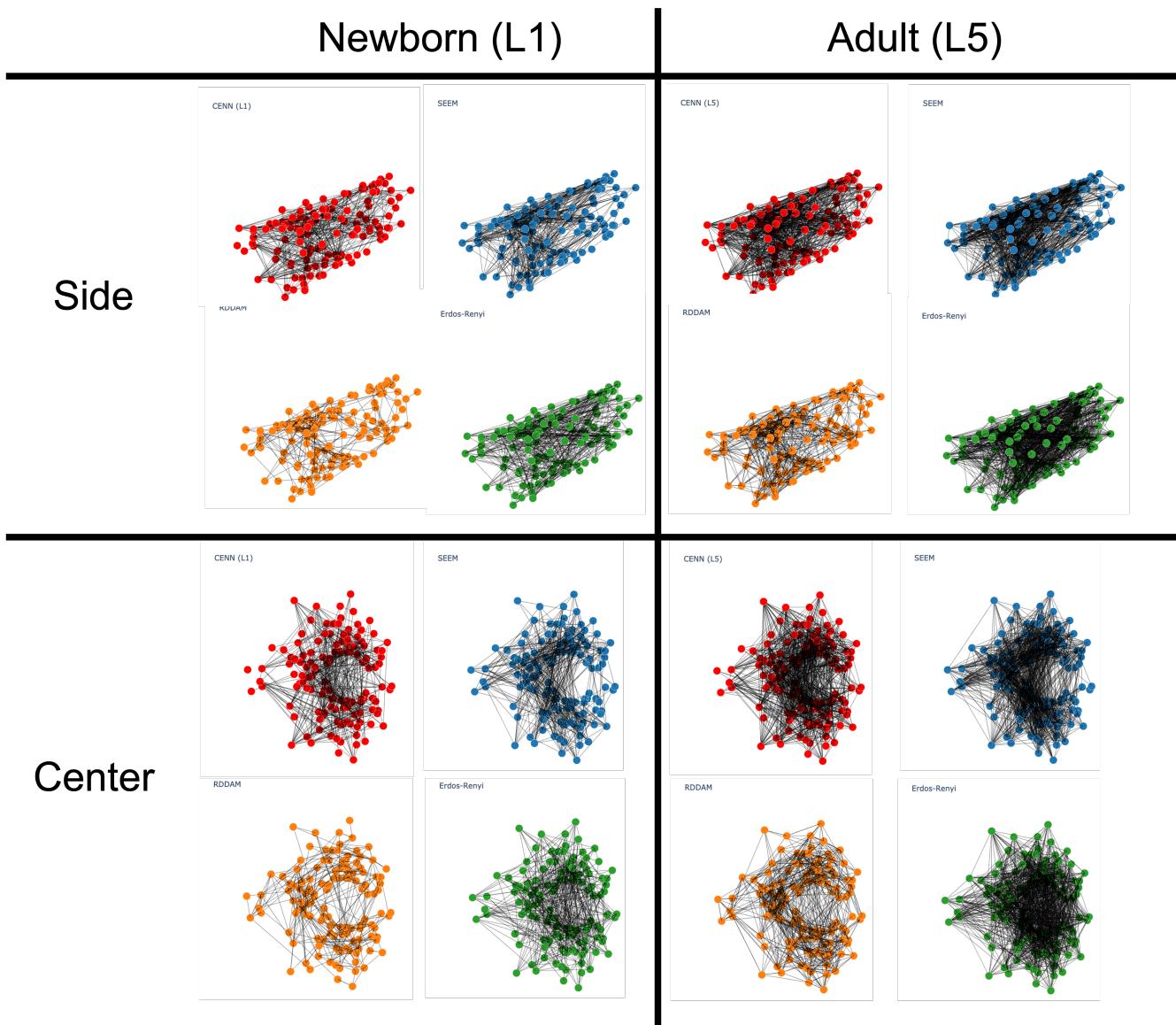


Figure 3: INSERT HERE