

# Modular Structures in Mutated Biological Neural Networks

By Charles Schumacher

17 November 2018

## Introduction

The *Caenorhabditis elegans* nematode is frequently used as a model organism in the field of biology due to the relatively simple nature of its structure and the fact that its neurons are almost identical to humans. It is one of a small pool of organisms which has had its entire genome sequenced and boasts the fact that it is the only multicellular organism to have its entire connectome (neural wiring diagram) mapped. The latter fact makes this of interest to neuroscientists and biologists who focus their research on the brain and its dynamical properties.

A recent theory generated in the realm of neuroscience and consciousness studies, integrated information theory, proposes a measure which aims to quantify characteristics of consciousness by examining the cause-effect power of a neural system relative to the cause-effect power of its highly synchronized subsystems. However, the theoretical measure in its pure form is intractable and requires examining all possible partitions of a system to find the one which minimally impacts the hierarchy of cause-effect relationships and preserves the most informationally independent components. To be exact, the number of computations required for a system of  $n$  vertices is given by the  $n^{th}$  bell number. The simple nervous system belonging to *C.elegans* only has 302 neurons, meaning that the number of partitions explored to calculate this measure exactly is given by the outstanding large  $B_{302} \approx 4.8 * 10^{457}$ . Attempts to make use of integrated information theory require the employment of various heuristics. A recent study which utilized such a modified approach was able to reliably distinguish between human patients who were either awake, sleeping (in dreaming or non-dreaming states), anesthetized, or comatose (in vegetative, minimally-conscious, or locked-in states), which suggests that approximation methods can be successfully applied [5]. Motivations for finding appropriate partitioning heuristics with the *C.elegans* connectome include the ability to correlate changes the nervous system structure (through study of mutated organisms) to changes in observed behaviors with real worm experiments.

Regardless of the domain being explored and the heuristic method applied, a consistent requirement for any IIT analysis is determining the partition which preserves relationships of highly synchronized components. When formulating a biological (or non-biological) structure of interest as a network, one approach to approximating this partition is by identifying community assignments which generate the highest modularity value for the full network. The current exploration into community detection and modularity uses only the *C.elegans* somatic nervous system (282 total neurons) because its other 20 neurons belong to a separate and autonomous pharyngeal nervous system which only impacts digestive functioning.

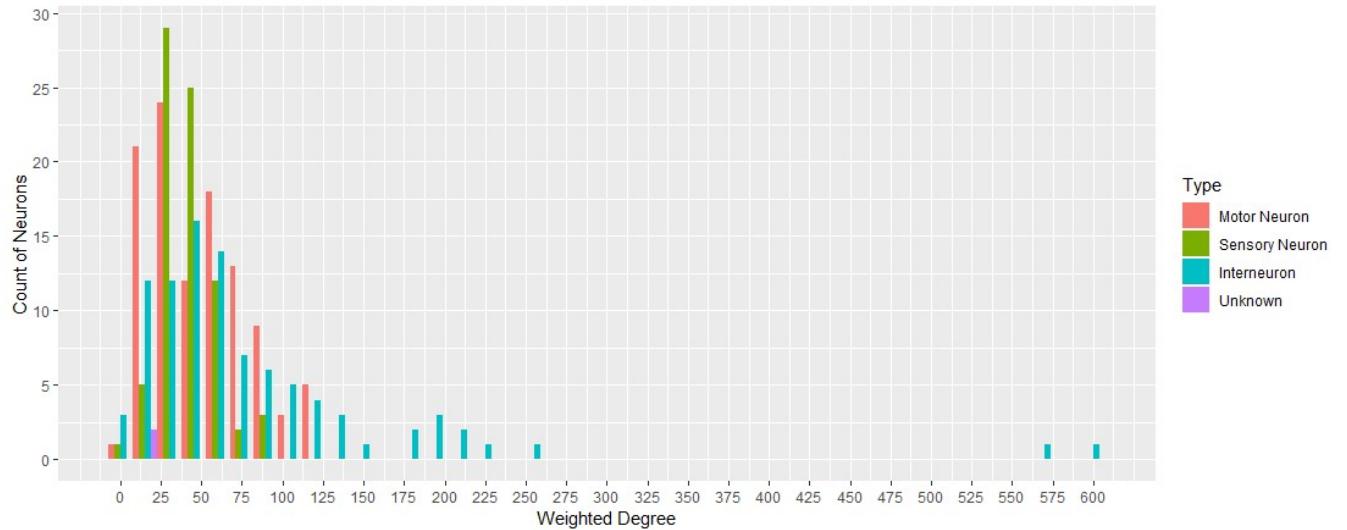
## Data Background

The data files for this investigation were collected from Openworm.org, an open-source project which collect information about *C.elegans* and aims to build a comprehensive computational model of the worm for use in research and simulation. Two files were collected and used to construct a directed and weighted network representation of the somatic nervous system:

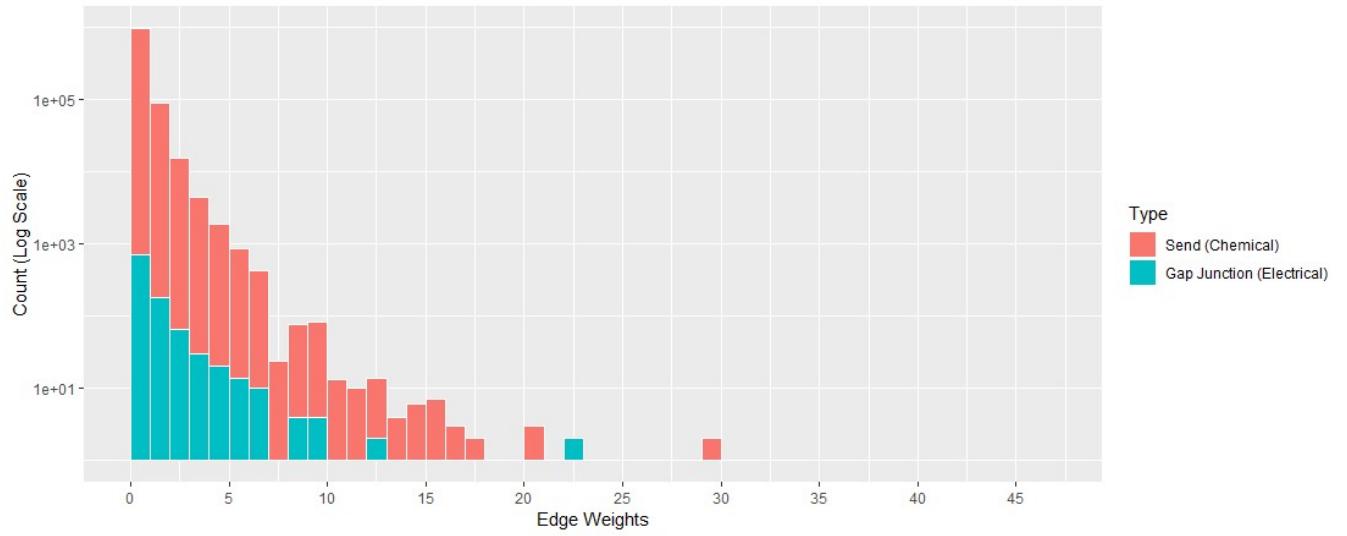
- An edge list describing the synaptic connections between neuron pairs. Attributes include:
  - Edge type
    - “Send” which describes a chemical synapse (unidirectional)
    - “Gap junction” which describes an electrical synapse (bidirectional)
  - Edge weight
    - Non-negative values which describe the number of synaptic join sites in a synaptic connection between two neurons. A higher number of join sites corresponds to a larger edge weight, and is considered to be a stronger connection where information is more likely to pass than a weaker synaptic connection
- A vertex list describing individual neuron names and functions from the somatic nervous system. Attributes include:
  - Primary neuron function
    - Interneuron
    - Sensory neuron
    - Motor neuron
    - Unknown function

Preprocessing included the removal of three neurons which had a given degree of zero, and one edge with a weight of zero, resulting in a total of 279 neuron-nodes (94 interneurons, 106 motor neurons, 77 sensory neurons, and 2 unknown) and 3604 edge-synapses. Documentation on the dataset did not include an explanation on why three of the nodes were disconnected or why there was an unweighted edge, but the resulting dataset after their removal was consistent with that which was used in other community detection research focused on *C.elegans* and thus deemed reasonable [8].

The connectome model with 279 neurons is used as a baseline model in subsequent analysis and is compared to connectome models which represent various neuron-pair ablated mutants: AIBR/AIBL removed (interneurons) and AIYR/AIYL removed (motor neurons). The weighted degree distribution by neuron type and edge weight distribution for the base model are shown in Figure 1 and Figure 2 respectively.



*Figure 1: Distribution of weighted degree of neurons in the base model representing the somatic nervous system*



*Figure 2: Distribution of synaptic-edge weights and types. Count of neurons is shown on a log scale due to most synapses having a low weight*

## Summary of Completed Work

### **Walktrap Community Detection in Directed, Weighted Connectomes**

Identifying highly synchronized communities as a preprocessing step for IIT-based analysis requires identifying maximally modular communities. Highly modular communities have a statistically surprising number of within-community connections relative to the

number that they have between-communities. Various community detection methods techniques are available which attempt to find communities in a network which maximize the modularity value (ex: Louvain's algorithm), however, these are typically limited to application in undirected networks. The frequent workaround for directed networks is unfortunately to throw away information about the direction of edges in the network and apply undirected community detection methods. In a brain network where the complex dynamics are not well understood, such as the one of interest in this paper, synaptic direction is crucial to having a reasonable understanding of neural communities that are dictated by physical structure. The traditional modularity metric is computed as following:

$$Q = \frac{1}{2m} \sum_{ij} [A_{ij} - \frac{k_i k_j}{2m}] \delta_{c_i c_j}$$

where  $A_{ij}$  is the element in an adjacency matrix,  $\delta_{c_i c_j}$  is the Kronecker delta which relates community assignments of nodes,  $c_i$  is the given community assignment for node i,  $k_i$  is the degree of node i, and m is the total number of edges in the network. However, in a directed and weighted network we want a modularity value which is high for statistically surprising configurations and which also factors in the edge weights. For that reason, our definition of modularity has taken simple modifications according to the work done by Leicht and Newman and by Sohn and is calculated in our analysis as:

$$Q = \frac{1}{m} \sum_{ij} [A_{ij} - \frac{s_i^{in} s_j^{out}}{m}] \delta_{c_i c_j}$$

This modified function is used in all subsequent analysis to accurately assess the strength of neural communities and account for synaptic strengths. The walktrap community detection measure functions by doing a random walk of specified distance from all nodes and determining where the walks end. Groups of nodes who have walks which get "trapped" amongst each other in a certain area are then placed into a community together. While this may not produce structurally as modular communities as other optimization methods, what this approach simulates is the probability of a signal generated by a particular node ending up within other nodes. While it is undoubtedly a dramatic simplification of brain dynamics, a walktrap in the context of a brain network can be thought of as identifying how a chemical or electrical signal might propagate through its neighboring structures.

To be consistent with additional related work [1] the walktrap method was used on the weighted and directed base and mutant models separately using a random walk of 6 steps from each neuron. The results from this test can be found in figure 3. Both of the ablated models found there to be 6 neural communities, which unsurprisingly consistent with the number found in the base model. Certain community structure remained identical in each of the models explored (for example: community #6 which is made up of 6 motor neurons (visuals in appendix A4-A6)) due to their physical location in the nervous system, while membership shifted more dramatically in others. The distribution of neuron membership in Base and AIY ablated models is much closer than in AIB ablated the largest community is split and joins with other community (see figures 6-8 which visualize the network and changes to membership. Appendix A1 and A2 include alternate visualizations of this and tables which denote the shifting membership).

The community in the base and AIB ablated models with secondarily maximal transitivity value (the constant community of 6 motor neurons has the highest transitivity) of 0.438 loses members in the AIY ablated model and has a transitivity value reduced to 0.434. Noteworthy is that the AIB model has a minimum transitivity value of 0.302 while the other two models have communities with a minimum of 0.228. The communities in base and AIY with the minimum transitivity are the largest ones with neuron membership over 100; the large community which was split for AIB contributed to higher minimum transitive community, and still produced a model with the equivalent overall transitivity (figure 5). While a high transitivity value may contribute to a greater probability of within-community connections, the relationship between overall model transitivity and distribution of community transitivities is not well described in the context of IIT.

Model	Community Number	Edge Density	Transitivity	Neuron Count	Diameter	Average Path Length
Base	1	0.102	0.228	105	8	2.59
Base	2	0.262	0.438	26	9	2.48
Base	3	0.141	0.354	43	8	2.74
Base	4	0.128	0.326	83	9	2.40
Base	5	0.188	0.348	16	13	1.54
Base	6	0.833	0.818	6	8	1.40
AIB Ablated	1	0.128	0.302	64	8	2.61
AIB Ablated	2	0.262	0.438	26	9	2.48
AIB Ablated	3	0.119	0.312	86	9	2.43
AIB Ablated	4	0.148	0.361	41	8	2.73
AIB Ablated	5	0.137	0.308	54	9	2.61
AIB Ablated	6	0.833	0.818	6	8	1.40
AIY Ablated	1	0.102	0.228	105	8	2.59
AIY Ablated	2	0.126	0.326	84	9	2.39
AIY Ablated	3	0.141	0.354	43	8	2.74
AIY Ablated	4	0.269	0.434	23	9	2.48
AIY Ablated	5	0.188	0.348	16	13	1.54
AIY Ablated	6	0.833	0.818	6	8	1.40

Figure 3: Metrics in detected communities using a weighted walktrap approach

Model	Modularity	Transitivity	Edge Density
Base	0.4643	0.2135	0.0465
AIB Ablated	0.4734	0.2127	0.0456
AIY Ablated	0.4586	0.2131	0.0465

Figure 4: Overall model metrics for our modified modularity value, transitivity, and density

It is additionally noteworthy the AIB ablated model actually produced a *higher* modularity value than the Base model. Despite boasting the highest modularity, the model actually produced a lower global transitivity and edge density for the whole network. To determine whether these were significant changes in modularity, 1000 random graphs were generated using the existing neuronal distribution and scrambling the existing synaptic-edges while preserving their weights. The average directed and weighted modularity value that was generated by randomly rewired structures using the C.elegans somatic components was normally distributed around 0.300 (Figure 5). This tells us that the modularity value in the C.elegans somatic connectome is indeed statistically significant since it exceeds all modularity values of random rewirings, and highlights that the modularity value in the AIB ablated connectome may be a statistically significant improvement upon the base connectome model. This is left for future work.

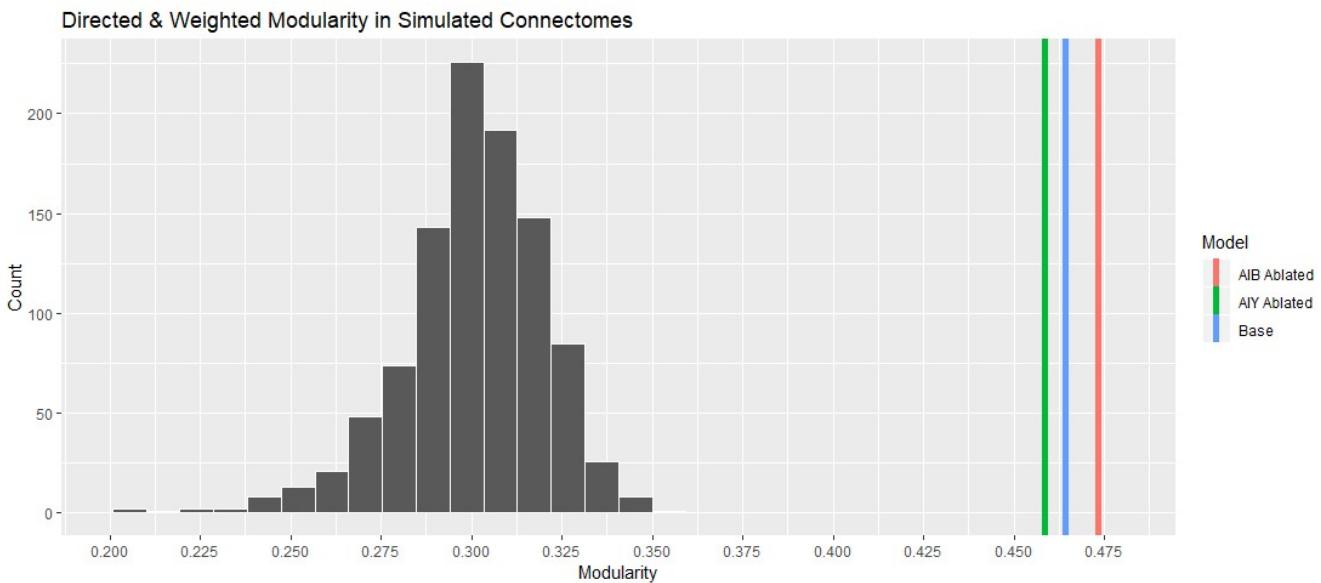


Figure 5: Comparing modularity values to randomly rewired networks. Red line=base model, yellow line=AIB ablated, and orange line=AIY ablated

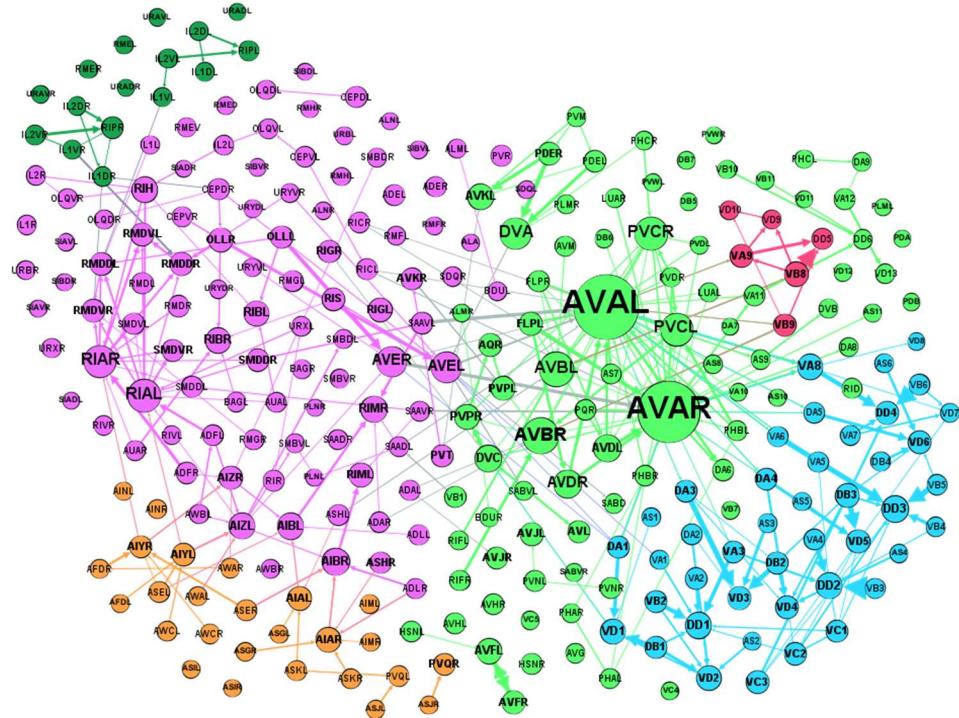


Figure 6: Base model. Colors denote community assignments, and node size indicates the weighted degree of the node. Edges with weights less than 5 have been filtered to reduce noise and highlight strong connections (isolated nodes in the visual only have weak synaptic connections)

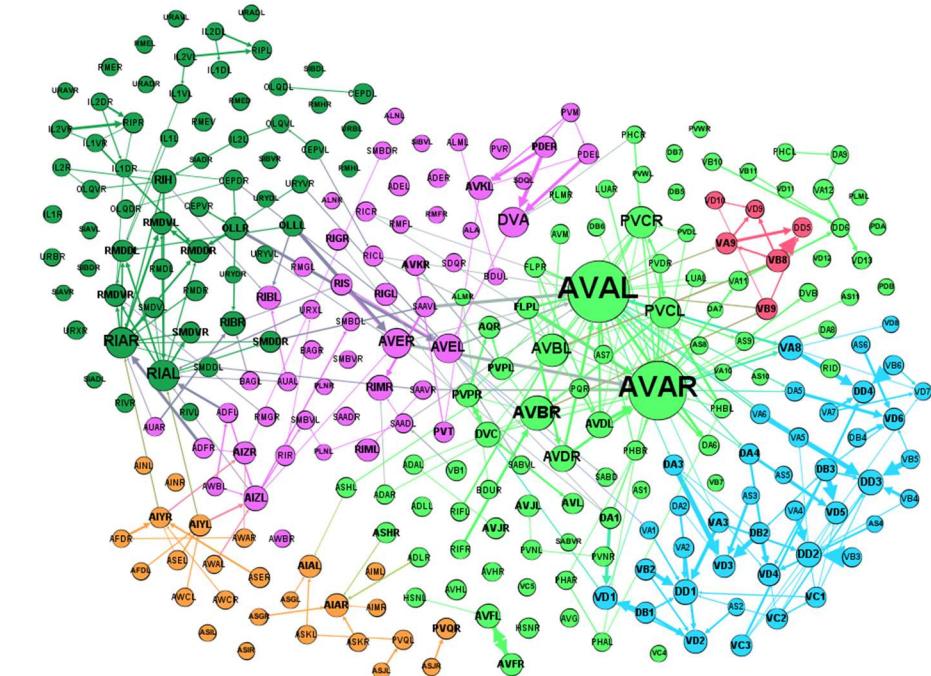


Figure 7: AIB Ablated. Note that the removal of the AIB pair causes walktrap membership to shift somewhat dramatically in the left region of the visual. The dark green nodes absorb a portion of the previously purple nodes.

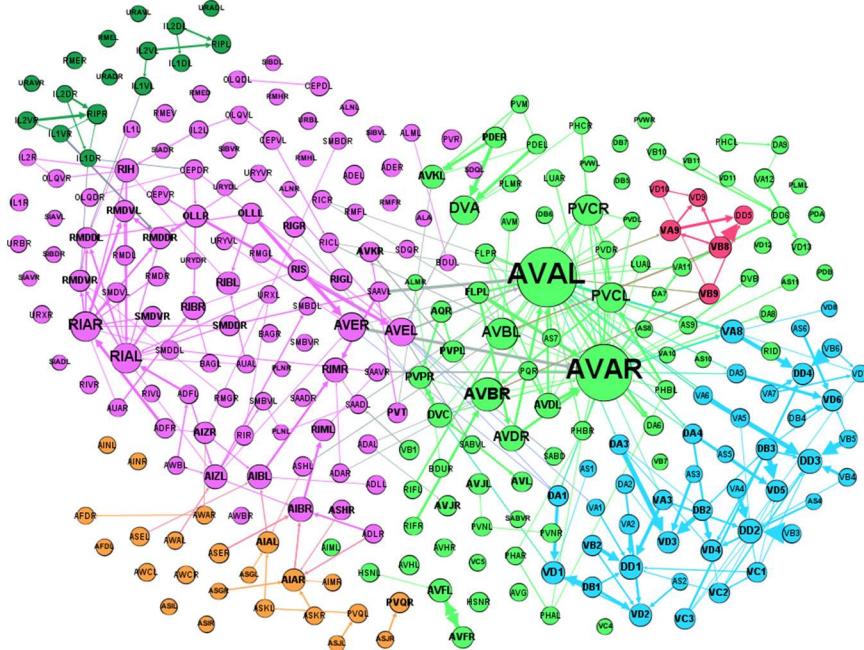


Figure 8: AIY Ablated. Note that the community structure is almost identical to that in the base model with the exception of absent AIYR/AIYL neurons and the changing membership of the AIML neuron

### Walktrap Analysis Conclusions

Removal of bilateral neuron pairs was expected to have some measurable impact on detected community structures since the overall neuron count in a such a small network could interrupt communication paths depending on how well connected the ablated neurons were (AIBR/AIBL have weighted degree of 137/127 and degree of 59/61, AIYR/AIYL have weighted degree of 70/77 and degree of 22/26). It is not surprising to see relatively subtle changes to the community membership, however, in the case of AIB ablated there was a meaningful shift in the membership between two communities. The maximum community size decreases, the minimum transitivity and edge density among the group of communities increased, and the overall transitivity and density for the model decreased while still showing an increase in network level modularity. This relationship between distribution of community level metrics and overall network metrics may be a useful to consider when analyzing other networks.

## Summary of Future Work

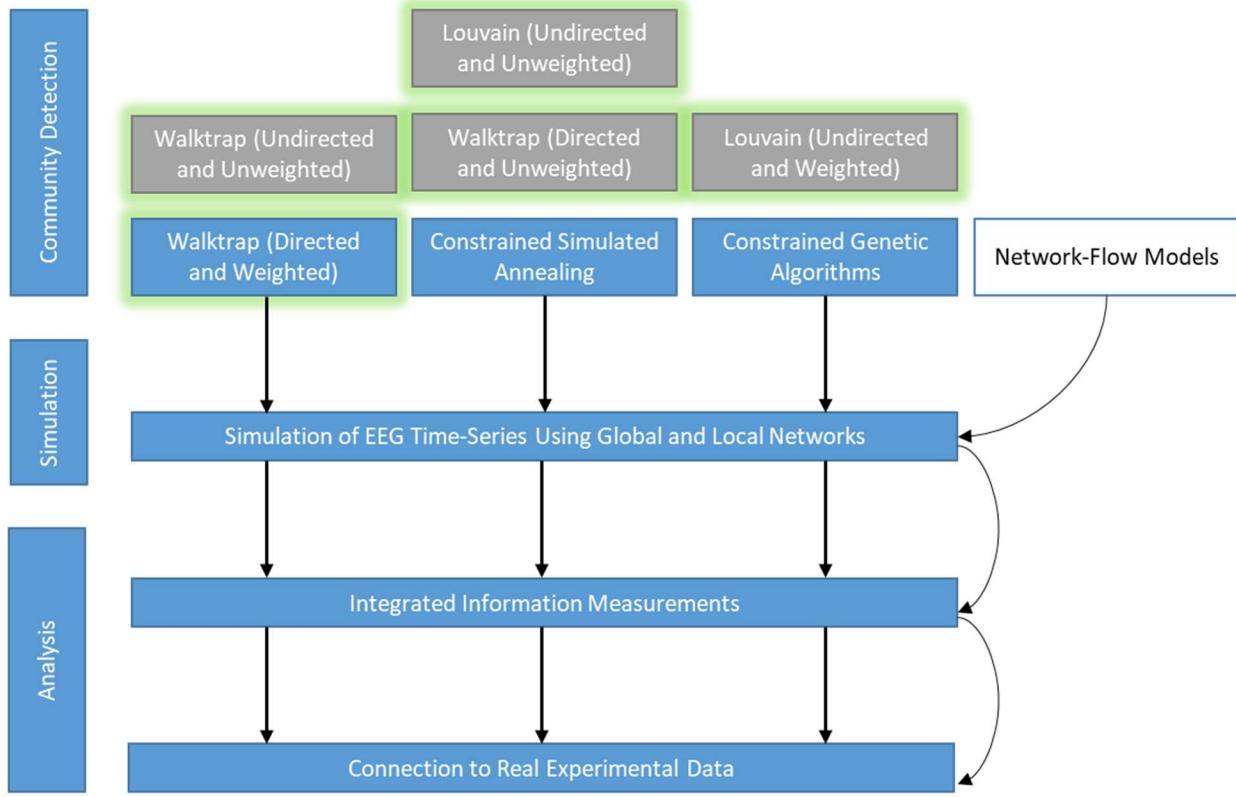


Figure 9: Future work pipeline. Green halo indicates completed work. Grey indicates results deemed irrelevant and not included in future work. White denotes potential work if time and resources permit.

### **Meta-Heuristic Community Detection in Directed, Weighted Graphs:**

Community detection is effectively a combinatorial optimization problem, thus, meta-heuristic approaches such as simulated annealing are considered to be more effective at identifying communities whose structure lends itself to highly synchronized components. Related work with an annealing approach using an identical dataset to the one in this paper can be found in [Sohn] and identifies the most modular communities in the *C.elegans* somatic connectome as having a value  $Q=0.490$  while adhering to the biological constraint the bilateral neuron pairs cannot be separate into different communities in order for the calculated communities to be biologically plausible (ex. AIBR and AIBL cannot be assigned to different communities even if it produces an increase in modularity) [8].

The work done by Sohn and colleagues considers this constrained form of optimal community detection and even takes it a step further to analyze the modularity values of individual discovered communities [8]. However, they do not consider the changes to biologically plausible community structures in *C.elegans* mutants, and at present there is no literature on that specific topic which this author is aware of. Efforts to replicate and improve upon their constrained simulated annealing community detection method is already underway and should be completed in a reasonably short time. Once this is complete and functioning effectively, there is excellent opportunity to expand on their analysis by exploring the changes in global modularity of mutant connectome structures; and using the natural somatic nervous system model as a base, investigating the interplaying relationship between global and local network metrics as modularity fluctuates.

An additional method to *improve* upon the constrained simulated annealing method for community detection is already under investigation and involves the adaption of the community assignment simulations to a genetic algorithm. The difference between this approach and simulated annealing is that the GA approach maintains and considers a pool of solutions which it iteratively selects from, “breeds” together, and mutates solutions over a series of generations to avoid getting caught in local maxima that simulated annealing may not be able to avoid in our data. Existing work in the area of genetic algorithms for community detection generally makes the assumption that the network is undirected and does not account for the kind of constraints we require [7].

The delta-energy function in the simulated annealing approach can be easily repurposed for a fitness function in a genetic algorithm approach and is described here. The original fast-unfolding community detection method described by Blondel suggests using the following equation to quickly and efficiently calculate the change in modularity for an undirected and unweighted graph [4]:

$$\Delta Q = \left[ \frac{\sum_{in} + k_{i,in}}{2m} - \left( \frac{\sum_{tot} + k_i}{2m} \right)^2 \right] - \left[ \frac{\sum_{in}}{2m} - \left( \frac{\sum_{tot}}{2m} \right)^2 - \left( \frac{k_i}{2m} \right)^2 \right]$$

where  $\sum_{in}$  is the sum of weights of the edges inside the community of interest C,  $\sum_{tot}$  is the sum of weights of the edges incident to nodes in C,  $k_i$  is the sum of weights of the edges incident to node i, and  $k_{i,in}$  is the sum of the weights of edges from i to nodes in C and m is the sum of the weights of all edges in the network. This equation simplifies to the following:

$$\Delta Q = \frac{k_{i,in}}{m} - \frac{\sum_{tot} * k_i}{m^2}$$

which describes the difference between the number of edges from node i to community C divided by the sum of all edges in the network, and the number of all edges going to C in the network multiplied by all edges incident to node i and divided by the square of the number of edges in the network. Being consistent with the modifications to our modularity equation already in use, we can easily transform this equation to the following which accounts for the direction and strength of edges [6][8]:

$$\Delta Q = \frac{s_{i,C}^{in} + s_{i,C}^{out}}{w} - \frac{\sum_{tot} * (s_i^{in} + s_i^{out})}{w^2}$$

where  $s_{i,C}^{in}$  describes the sum of the edge weights coming into node i to community C,  $s_{i,C}^{out}$  the sum of edge weights going from node i to community C,  $s_i^{in}$  the total weight of edges coming into node i,  $s_i^{out}$  the total weight of edges going out from node i, and w indicating the sum of all edge weights in the network. This equation is currently in use with the simulated annealing approach being developed and allows the approximation of changes to directed and weighted modularity using partial calculations rather than computing the Q function for the entire graph at every iteration. This modified delta Q function is entirely compatible with a genetic algorithm approach as a fitness function to approximate Q after each breeding or mutation step. The component that remains to be determined for a GA approach is how to optimally “breed” a series of solutions together since the community assignments aren’t independent; modification to one neuron’s assignment can impact the assignment of its adjacent neurons and would need to be accounted for with a modified solution-breeding algorithm or alternate method of progressing the pool of solutions over generations.

#### ***Computation of IIT Metrics with Connectome models:***

Heuristic methods have been developed which make it possible to approximate an integrated information measure in complex systems. As mentioned, an especially exciting application of this with humans has been the successful use of distinguishing between patients who were awake, asleep (dreaming vs non-dreaming), anesthetized, and comatose (vegetative vs minimally conscious vs locked in) [5]. This exciting work included the calculation of likely brain regions where measured signals originated, and used a modified form of an IIT measure to calculate what they call a PCI index (perturbational complexity index), which measures the information content of the brain’s response to magnetic stimulus.

A comparably crude application of a modified IIT measure has been calculated for C.elegans using a similar connectome model, but throwing away information about the synaptic direction and weights [1]. Additionally, this investigation only focused on the original connectome model and did not consider any changes to IIT values in the presence of neural ablations. It is an interesting opportunity to expand upon the limited research in this area by replicating their methodology with mutated connectomes, as well as improving upon it by factoring in the direction and weight of synaptic connections as well as potentially using information flow as well.

#### ***Analysis of Simulated Connectome Metrics and Real Observed Behaviors***

Ultimately, the aim of this exploration is to connect insights generated through simulation and theoretic measures of consciousness (which have some empirical validation) to observable behaviors by using analysis of data collected through the DePaul MedIX lab. Examples of the types of questions we want to answers are:

- Does the ablation of certain neurons or neuron pairs in C.elegans cause different changes to theoretical values of integration information in the modified connectome?
  - Are there some neurons which cause more dramatic changes than others? Are there any which cause an *increase* in the integration potential?
- As integrated information increases, do efficient behaviors of C.elegans increase?
- As integrated information decreases, do we observe more frequent or more dramatic behaviors deficits?

### ***Information Flow Models with Mutants***

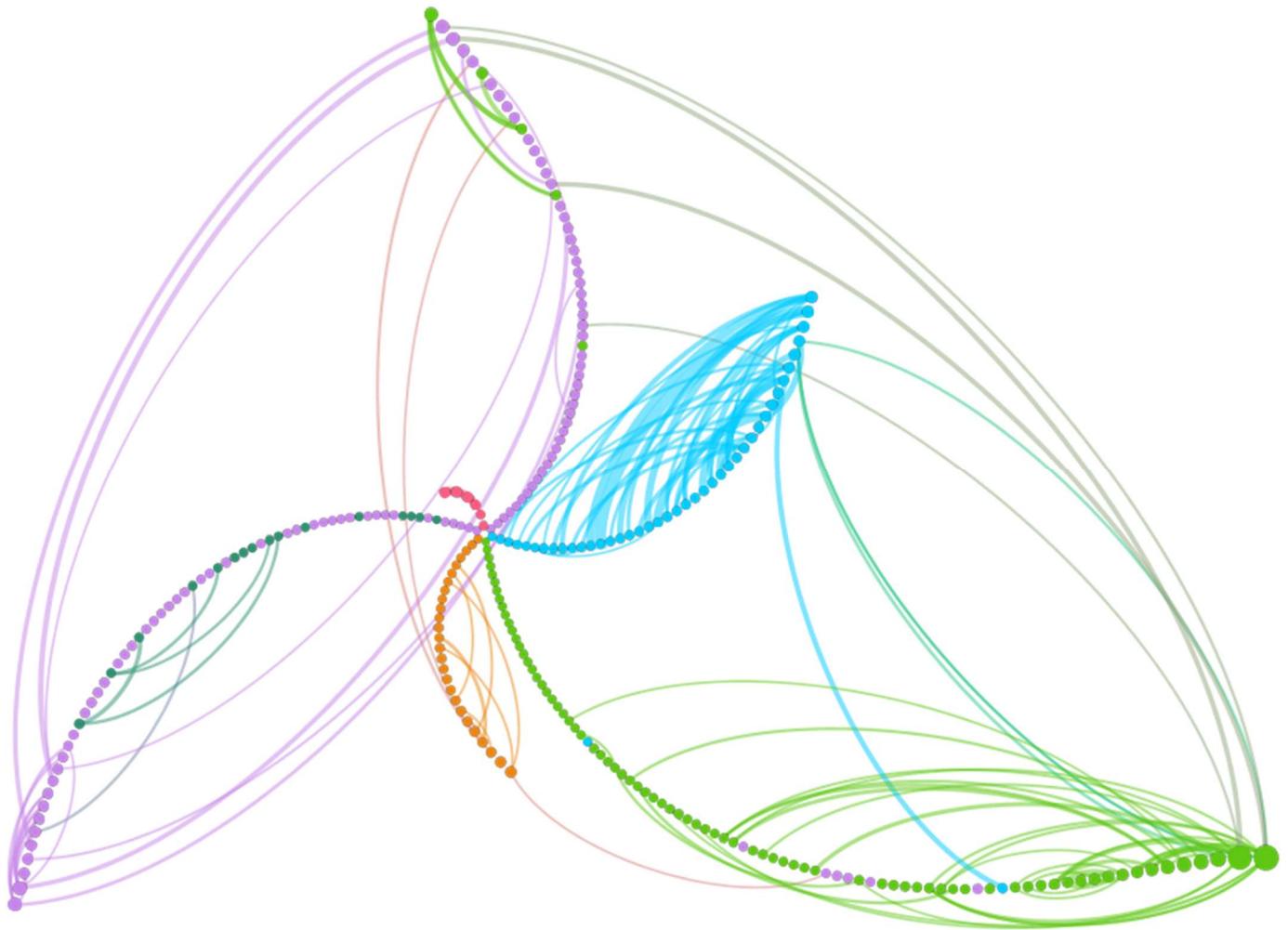
An alternate approach to understanding the C.elegans connectomes has been completed by Bacik and colleagues [2]. The techniques used by this team consider the full connectome (rather than the somatic nervous system alone) and focus on measurements from a information-flow perspective at varying scales of granularity. Rather than solely considering geometrical properties of the connectome structure that may promote or inhibit communication amongst highly synchronized components they attempt to connect these structural traits to the functional dynamics that they implement. The primary techniques used are Markov Stability for community detection at varying scales, and Role Based Similarity for extracting information about the directed and weighted information flows in order to classify neurons into different flow-role categories (neurons that have similar asymmetric patterns of incoming and outgoing networks flows at different scales are grouped together). Using these approaches, their team was also able to simulate “stimulus-response” experiments in modified networks and evaluate the flow-based changes to the global network in the presence of ablations (extensive variations of this were tested).

## References

1. Antonopoulos, C. G. (2016). Dynamical Complexity in the C.Elegans neural network. *The European Physical Journal Special Topics*.
2. Bacik, K. A., Schaub, M. T., Beguerisse-Diaz, M., Billeh, Y. N., & Barahona, M. (2016). Flow-Based Network Analysis of the Caenorhabditis elegans Connectome. *PLOS Computational Biology*.
3. Barrett, A. B., & Seth, A. K. (2011). Practical Measures of Integrated Information for Time-Series Data. *PLoS Computational Biology*, 7(1). doi:10.1371/journal.pcbi.1001052
4. Blondel, V. D., Guillaume, J., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10). doi:10.1088/1742-5468/2008/10/p10008
5. Casali, A., Gosseries, O., Rosanova, M., Boly, M., Sarasso, S., Casali, K., . . . Massimi, M. (2013). A Theoretically Based Index of Consciousness Independent of Sensory Processing and Behavior. *Science Translational Medicine*.
6. Leicht, E. A., & Newman, E. J. (n.d.). Community Structure in Directed Networks. *Phys Review Letters*.
7. Pizzuti, C. (2008). GA-Net: A Genetic Algorithm for Community Detection in Social Networks. *Parallel Problem Solving from Nature – PPSN X Lecture Notes in Computer Science*, 1081-1090. doi:10.1007/978-3-540-87700-4\_107
8. Sohn, Y., Choi, M., Ahn, Y., Lee, J., & Jeong, J. (2011). Topological Cluster Analysis Reveals the Systemic Organization of the Caenorhabditis elegans Connectome. *PLoS Computational Biology*, 7(5). doi:10.1371/journal.pcbi.1001139

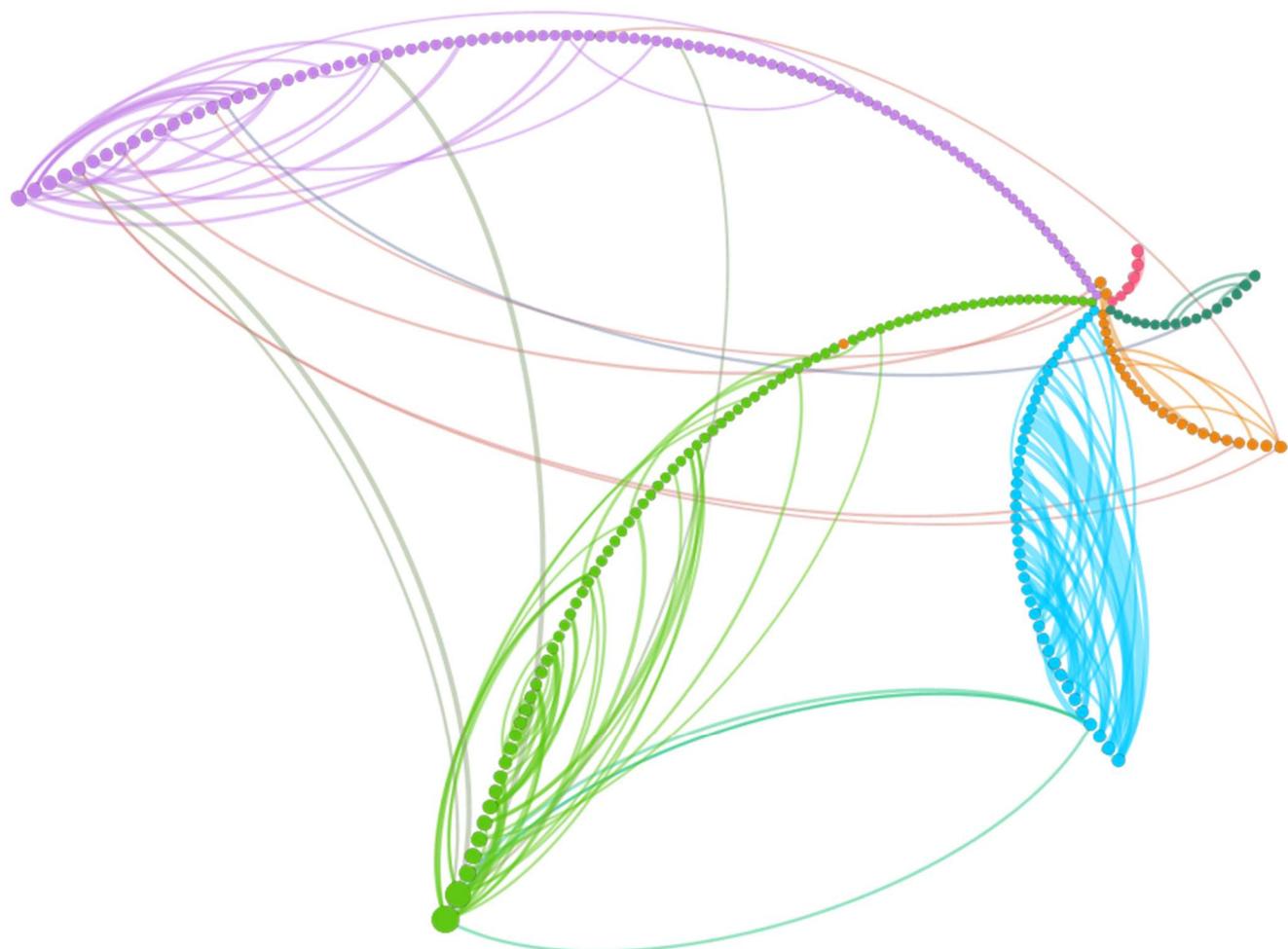
## Appendix

A1: Visualization of changing community membership from Base to AIB ablated. Colors of the nodes represent the community assignments in the natural Base model. Each branch represents the community membership of nodes after AIB ablation. Edges have been filtered to show only synapses with weights greater than 5 to reduce graphical noise and highlight strong connections



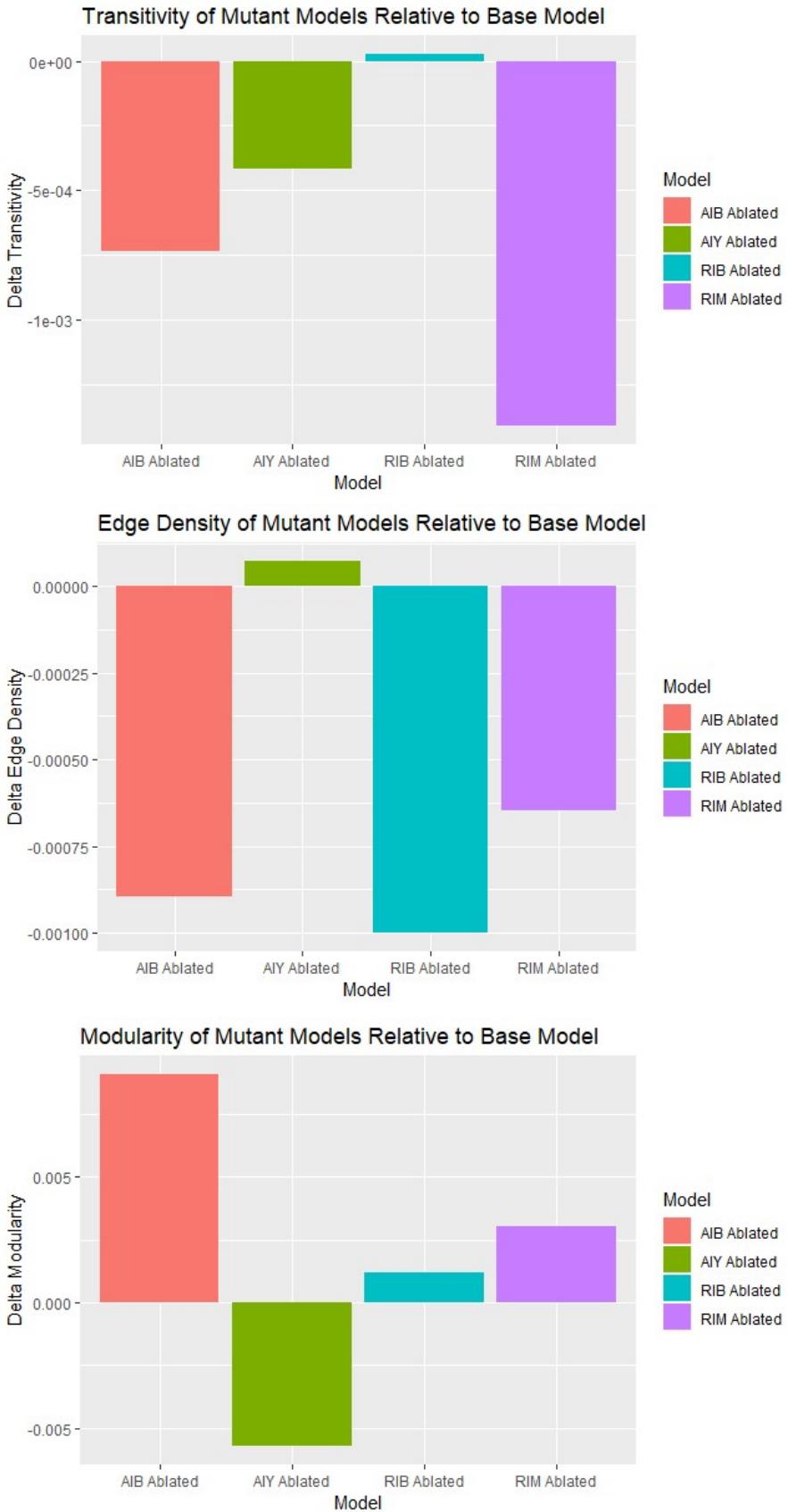
AIB Ablated Assignment	Base Assignment						
	Base Color:	Purple	Orange	Blue	Green	Dark Green	Pink
	Comm. #	1	2	3	4	5	6
Ablated	2	0	0	0	0	0	0
1	48	0	0	0	16	0	0
2	0	26	0	0	0	0	0
3	6	0	2	78	0	0	0
4	0	0	41	0	0	0	0
5	49	0	0	5	0	0	0
6	0	0	0	0	0	0	6

A2: Visualization of changing community membership from Base to AIB ablated. Colors of the nodes represent the community assignments in the natural Base model. Each branch represents the community membership of nodes after AIB ablation. Edges have been filtered to show only synapses with weights greater than 5 to reduce graphical noise and highlight strong connections

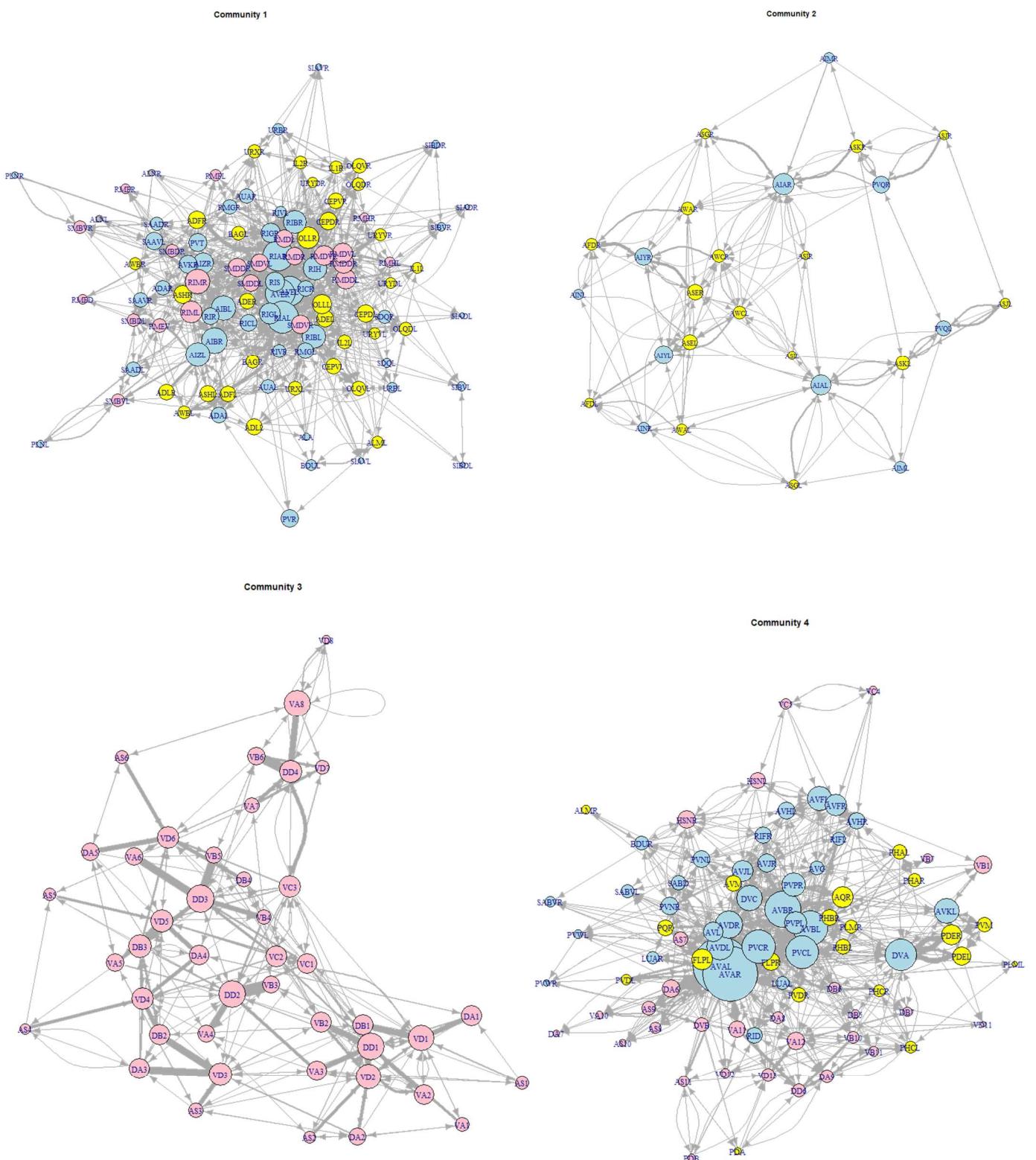


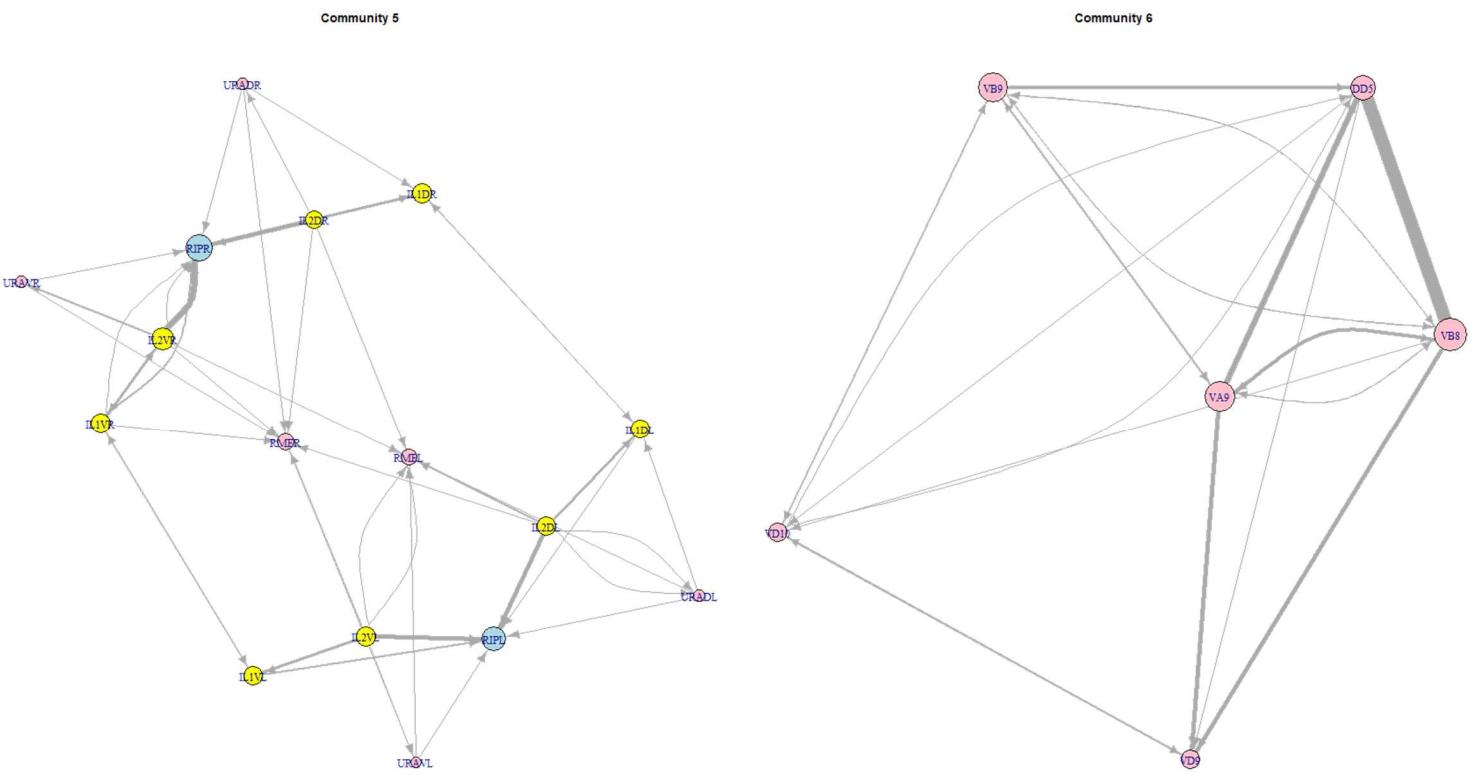
AIY Ablated Assignment	Base Assignment						
	Base Color:	Purple	Orange	Blue	Green	Dark Green	Pink
Comm. #	1	2	3	4	5	6	
Ablated	0	2	0	0	0	0	0
1	105	0	0	0	0	0	0
2	0	1	0	83	0	0	0
3	0	0	43	0	0	0	0
4	0	23	0	0	0	0	0
5	0	0	0	0	16	0	0
6	0	0	0	0	0	6	0

A3: Plots of structural metrics transitivity, edge density, and modularity (modified for direction and weights) including additional mutants RIB-ablated and RIM-ablated. While the difference from Base model are subtle, it is potentially noteworthy to note that AIY ablated which was the only mutation explored to show a global increase in transitivity and edge density also generated a decrease modularity.

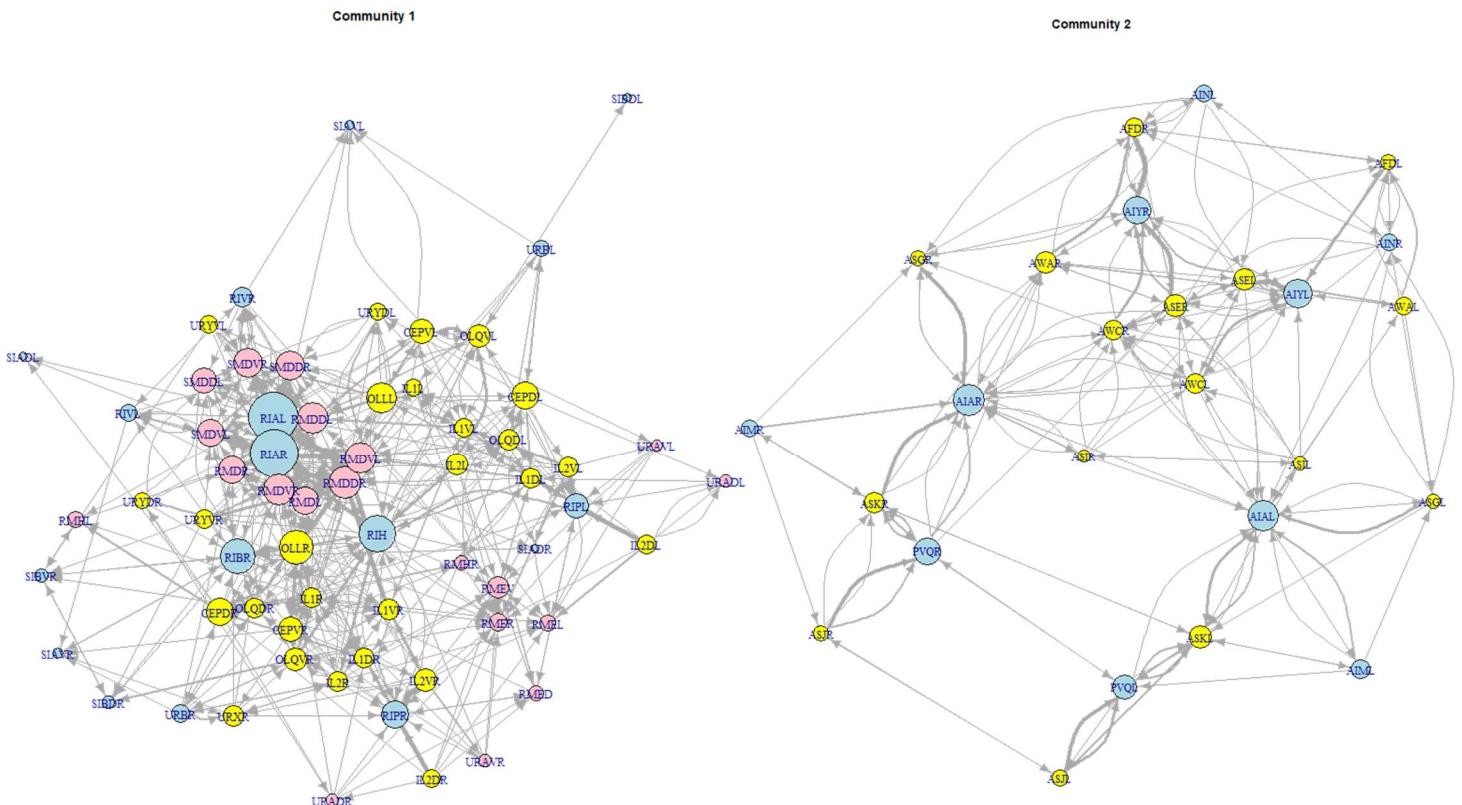


A4: Visualization of extracted sub-communities in C.elegans base model (directed, weighted walktrap approach):

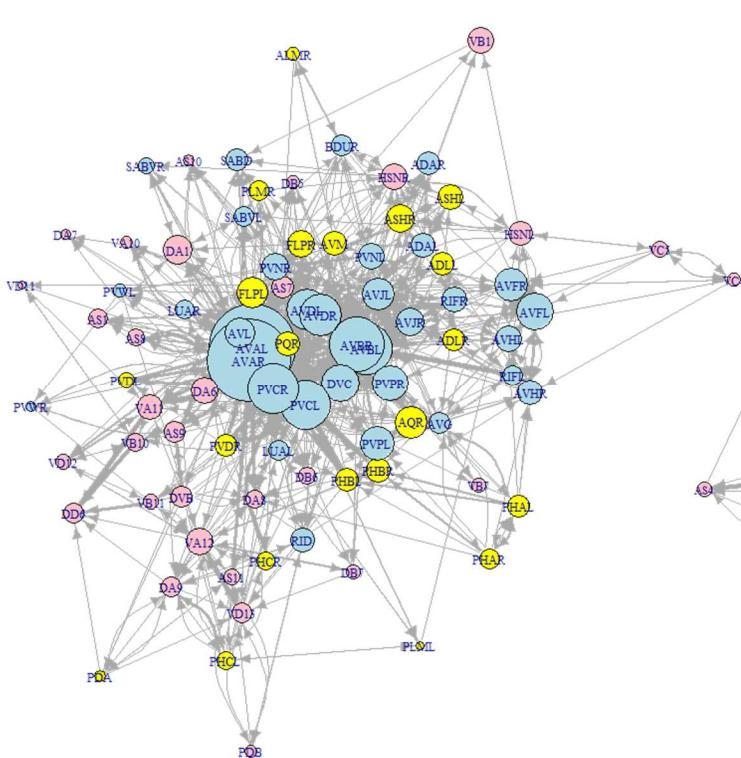




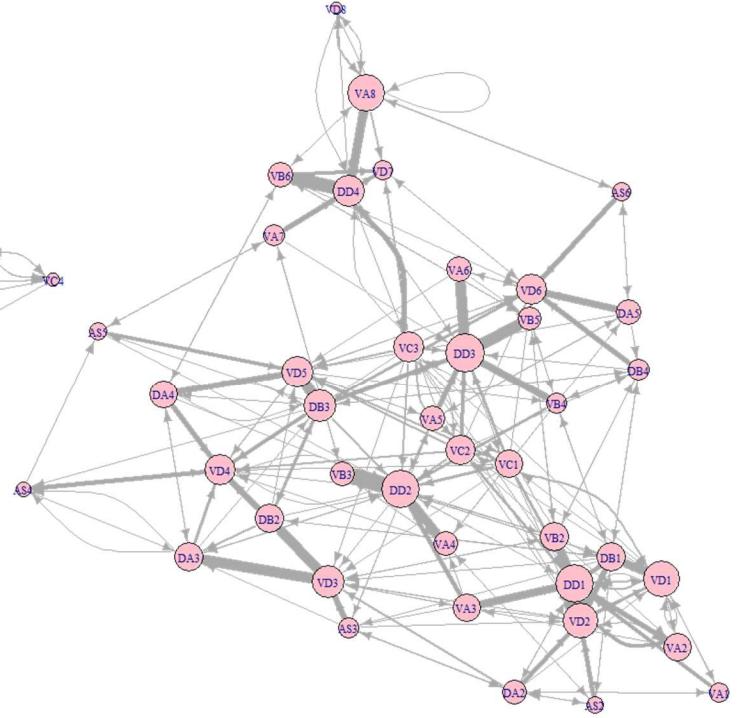
A5: Visualization of extracted sub-communities in C.elegans AIB Ablated model (directed, weighted walktrap approach)



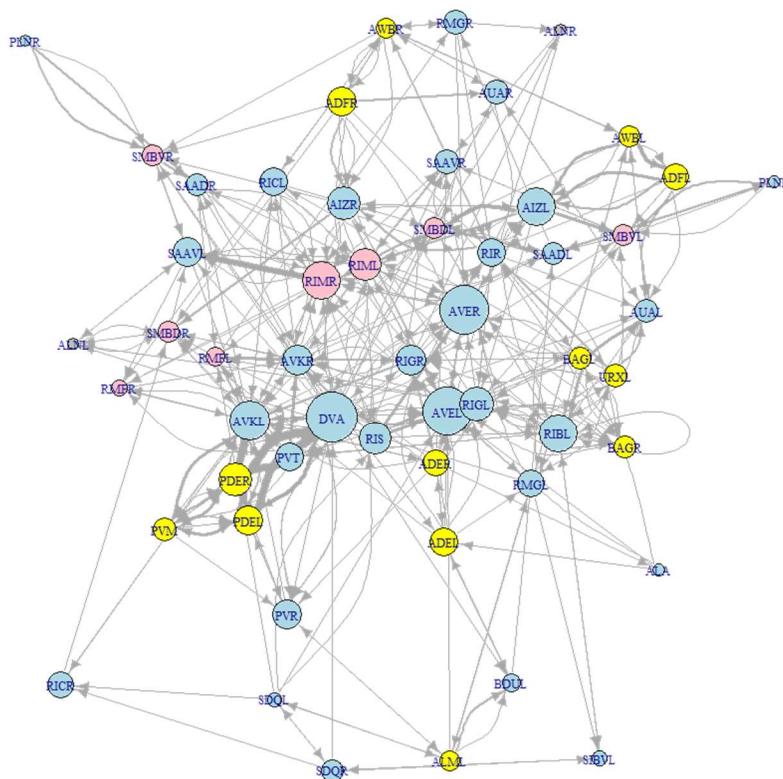
### Community 3



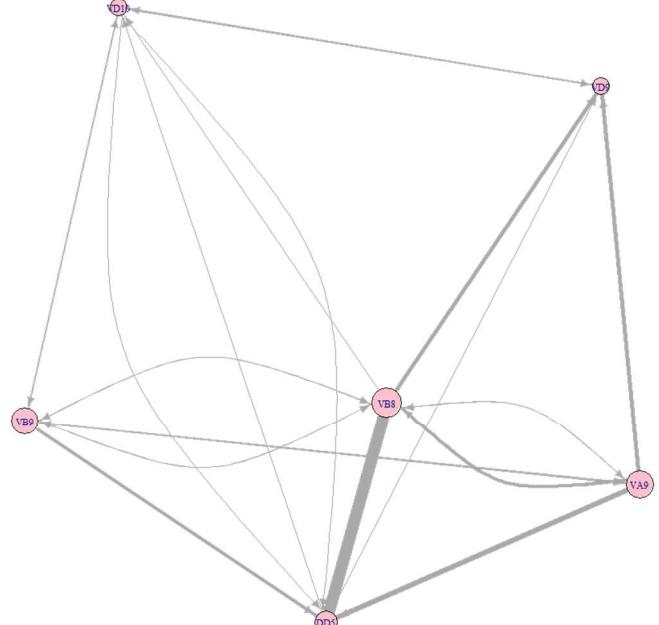
## Community 4



## Community 5



## Community 6



#### A6: Visualization of extracted sub-communities in C.elegans AIY Ablated model (directed, weighted walktrap approach)

