Investigating Motif Significance in the *Drosophilidae drosophila*

Chris Miller Taggart Bonham Sean McGowan

March 11, 2020

1 Introduction

Previous work has investigated patterns of local connectivity (network motifs) on biological graphs, such as signal-transduction networks (which help to control the development of organisms) and neuron connectomes (a structural map of the connections between neurons in the brain) [6, 7]. A network motif is a directed subgraph, typically consisting of a few nodes. In a neuron connectome, motifs correspond to repeated neuronal elements that process information (ie, repeated connection patterns between k neurons, where k is significantly less than the size of the network). Motifs in biological networks are often split into structural and functional motifs. Structural motifs are based exclusively on physical connections, while functional motifs also consider actual firing patterns and connection strength. Though the mapping between structural and functional motifs is often unclear, specific motifs are thought to have functional purposes, such as the feed-forward loop (Figure 1 - triad 9), which has been shown to perform certain signal processing tasks, such as regulatory functions [5].

Some have examined the presence of structural motifs in partially-mapped connectomes by investigating their Triad Significance Profiles [11]. At a high-level, the Triad Significance Profile shows the relative presence of different structural motifs in a network. The connectome of *C. elegans* was experimentally found to belong to the same biological information-processing super-family, share the same Triad Significance Profile, as signal-transduction cellular networks in the *Drosophila* [6]. This super-family is characterized by the presence of two-node feedbacks that regulate or are regulated by a third

node. As defined by Figure 1, these networks show triads 9 (030T), 12 (120D), and 13 (120U) with positive Triad Significance Profiles (a measure of normalized significance), and triads 4 (021D), 5 (021U), 7 (111D), and 8 (111U) with negative TSPs [6].

Further studies have shown that TSPs on partial hippocampal mappings of humans display similar distributions [9]. However, they show higher TSPs on triad 15 (210) and 16 (30D), the transitive triplet interactions. These results are in line with recent studies that have shown that less stable, feedback-containing motifs are more present in organisms that are capable of more complex function. [8] Further, many current neuroscientific theories correlate evolutionary development with greater neuronal connection [10]: humans have greater neuronal inner-connectivity that of the *C. elegans* .

Given that the *Drosophila* connectome is orders of magnitude larger than that of the *C. elegans* as well as evolutionarily more developed, we seek to investigate the differences between their triad significance profiles. We hypothesize that the connectome of the *Drosophila* will exhibit greater presences of more complex network motifs. Specifically, we expect to see higher levels in the Triad Significance Profile of 3-motif in the 120 range as well as 210 and 300.

2 Dataset

Organisms contain billions of neurons and synapses; thus, mapping neuronal connections has to be restricted in scope. In humans, the largest neuronal datasets comprise just sub-regions of functional components of the brain. For this analysis, we focus on the two largest full connectome mappings, the *C. elegans* , a sea-worm, with just 302 nodes (neurons) and 6,393 edges (synapses), and the *Drosophila* with approximately 25,000 neurons and 20 million synapses. Data limitations for both datasets prevented analysis by functional region. This is because neural circuitry for *C. elegans* is not divided into functional regions, and regional data is given by synapse, not by neuron, for *Drosophila* . For this investigation, we analyze the differences of the triad significance profiles between their connectomes [3, 12].

3 Methodology

As noted by [7], since motif frequency is inherently related to network degree, to interpret results we compared them to the random configuration model (a random graph generated with the same degree distribution as the original graph). This allows us to exclude degree distribution as the cause of motifs being unexpectedly common or uncommon. We use the NetworkX directed_configuration_model functionality to produce these random graphs. While prior works such as [7] have used a large numbers of random graphs, we found that the variance of motif frequency in our random graphs was quite low. Due to the time complexity of analyzing motif frequency on large graphs such as the *Drosophila* connectome, we chose to generate three random graphs for each real graph.

We utilized the motif detection algorithm implemented by NetworkX and introduced by [1]. This algorithm is $O(E)$, which enables sub-quadratic analysis for relatively sparse networks such as the *Drosophila* connectome and its random graphs (approximately six hours of analysis per graph). For speed and uninterrupted computation, we evaluated the full *Drosophila* graph using a virtual machine hosted by Google Cloud.

After generating motif frequency for each true graph and a set of random graphs, we normalize the frequency following the process outlined in [6]. For normalization, we first calculate a Z-score for each motif, given by the following:

$$z_{motif} = \frac{m_{true} - \overline{m_{rand}}}{\sigma(m_{rand})}$$

We then normalize the Z-score by dividing by the maximum value across all motifs for a given graph to produce values in the range $[-1, 1]$.

4 Results & Discussion

4.1 Comparison to C. Elegans and Simulated Rat Somatosensory Cortex

We analyze 3-motif presence in *Drosophila* and compare to a previously analyzed datasets: the *C. elegans* connectome, which we re-analyze for methodological consistency. We also include a qualitative comparison (due to lack

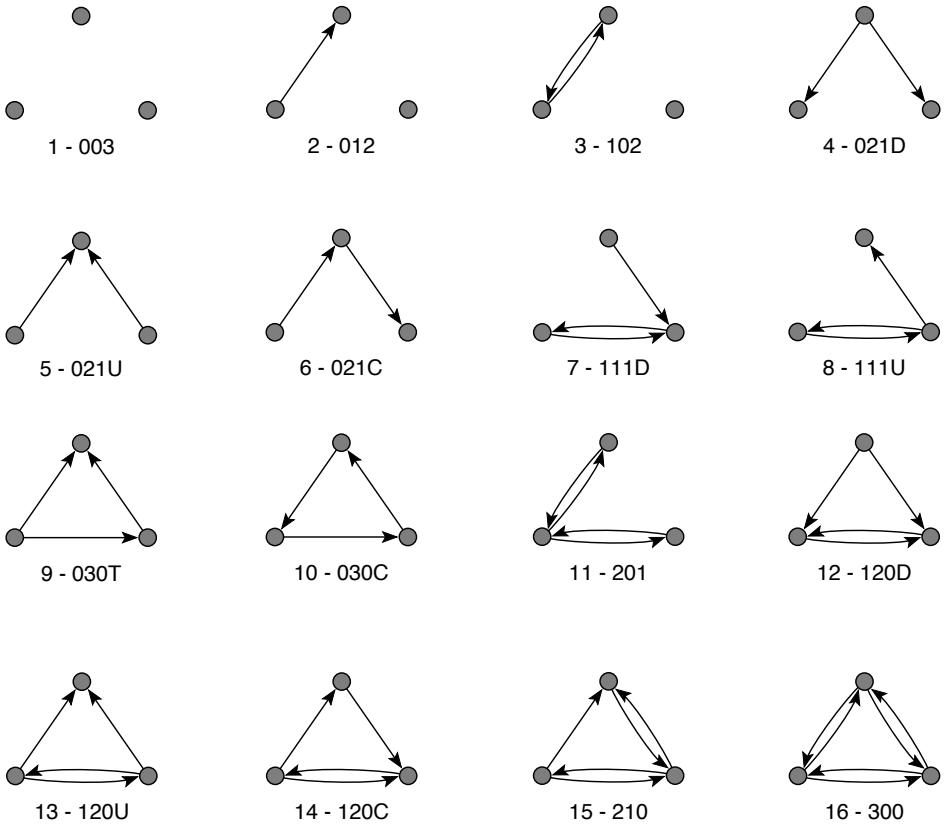

Figure 1: Possible 3-motifs, as defined by [1]

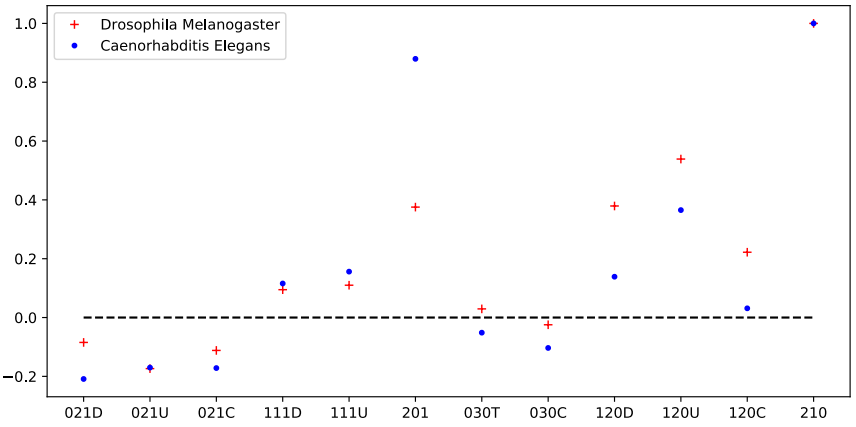

Figure 2: Motif frequency Z-score in *Drosophila* , compared to *C. elegans* for 3-motifs

of availability of the dataset) to a motif analysis of a simulated rat cortex as described in [4]. Gal et al. created the rat data by combining results from electron microscopy on a small section of a live brain with anatomical data, including cell-types and relative densities at cortical layers. The authors then developed a biologically constrained algorithm to generate a mammalian neocortical microconnectome with approximately 31,000 neurons and 36 million synapses.

The results shown in Figure 2 show striking differences between the *C. elegans* and *Drosophila* connectomes. *Drosophila* exhibits significantly more incidence of 120 motifs (120D, 120U, and 120C). These motifs are feedback loops, meaning outputs from one node will return to it after passing through other nodes.

Such structures have relatively low structural stability (ability to inherently reach a stable state after a perturbation) as defined by [8]. This means that they are evolutionarily selected against without strong functionality since producing a stable network with these motifs requires secondary inhibitory structures to avoid exploding or disappearing signals [8].

This results fits our hypothesis that the brain of *Drosophila* , which is more complex than that of *C. elegans* , requires more complex levels of interconnection. Note that this does not occur for all feedback loops—for example,

the simple three node feedback loop structure, 030C, is underrepresented in both networks. We speculate that this is due to the lack of additional information introduced by the loop structure (ie, additional intra-node processing does not occur via other connections). Our results are supported by the simulated data for the rat cortex in [4], which also shows peaks for the 120 type motifs.

Intriguingly, we also found that the basic feed-forward structure, 030T, was not as significant as expected, though it was markedly more important in *Drosophila* than in *C. elegans* . The feed-forward structure was suggested by [7] to be common in *C. elegans* compared to the random graph. However, we find that the motif occurs far less frequently in the true graph than the random graph for *C. elegans* (for which it occurs approximately 29% as often in the true graph). This discrepancy may be due to different methods of generating the random graphs for comparison, or due to dataset differences (we use a newer, corrected version of the *C. elegans* connectome). Though the feed-forward motif occurs approximately 47% more often in the true graph of *Drosophila* than in the random graphs, this increase is relatively small compared to other motifs, which suggests that this motif is less important than previously suggested for neural circuitry. As one counterpoint, the analysis in [4] of simulated rat brain circuitry also suggests that this motif is common. We posit that our analysis may imply that this motif is less common in simple circuitry (such as *C. elegans*) and becomes more common as complexity increases (hence increased presence in *Drosophila* and even higher presence in the rat brain). More research into intermediate complexity connectomes is necessary to evaluate the accuracy of this claim.

Meanwhile, we note that *C. elegans* has a far higher incidence of 3-motif 201, in which two neurons have bidirectional synapses with the same neuron but are not themselves connected. We suggest this can be attributed to the simple monopolar or bipolar morphology of the worm [2]. Since the majority of the worm's somatic system is comprised of non-branching neurons, we note the relative frequency of the 3-motif 201 as the transmission systems that run the length of the worm body are likely built on this neuronal motif.

Intriguingly, we also found parallels to the web motifs analyzed in [7]. For example, the uplinked mutual dyad (3-motif 120U) is classified as intrinsic to the WWW network. In the internet context, the motif typically implies multiple nodes linking to the same "authority node" or hub. We hypothesize that in a neuronal context, it may perform more of an information collection function, accepting information from multiple sensory neurons or lower-level

processing neurons and processing it before determining whether or not to fire and pass information along for further processing. The increase in this in *Drosophila* may imply that greater quantities of sensory and intermediate processing neurons requires more levels of analysis and information synthesis.

5 Conclusion

We show that the *Drosophila* connectome exhibits striking differences in 3-motif frequencies compared to the *C. elegans* connectome. Specifically, it shows a reduction in basic linked chains such as the 201 motif, and an increase in more complex feedback structures such as the 120 series motifs. We suggest that this result is due to the increase in complexity for the *Drosophila* brain compared to the *C. elegans* brain. This result holds when compared to the 3-motif frequencies in the We also find results for the basic feed-forward loop which contradict prior research, but suggest that this motif occurs more frequently in more complex organisms such as *Drosophila* brain and the rat cortex when compared to the *C. elegans* brain. We suggest that future work investigate the comparative frequency of 4-motifs, which we avoided due to computational constraints. We also suggest that a comparative analysis of feed-forward loop presence across more variably complex organisms may yield answers about its function and frequency in nature.

References

[1] Vladimir Batagelj and Andrej Mrvar. A subquadratic triad census algorithm for large sparse networks with small maximum degree. *Social networks*, 23(3):237–243, 2001.

[2] Andrew D Chisholm, Harald Hutter, Yishi Jin, and William G Wadsworth. The genetics of axon guidance and axon regeneration in caenorhabditis elegans. *Genetics*, 204(3):849–882, 2016.

[3] Richard Michael Durbin. Studies on the development and organisation of the nervous system of caenorhabditis elegans, 1987.

[4] Eyal Gal, Michael London, Amir Globerson, Srikanth Ramaswamy, Michael W Reimann, Eilif Muller, Henry Markram, and Idan Segev. Rich cell-type-specific network topology in neocortical microcircuitry. *Nature neuroscience*, 20(7):1004, 2017.

[5] Shmoolik Mangan and Uri Alon. Structure and function of the feed-forward loop network motif. *Proceedings of the National Academy of Sciences*, 100(21):11980–11985, 2003.

[6] Ron Milo, Shalev Itzkovitz, Nadav Kashtan, Reuven Levitt, Shai Shen-Orr, Inbal Ayzenshtat, Michal Sheffer, and Uri Alon. Superfamilies of evolved and designed networks. *Science*, 303(5663):1538–1542, 2004.

[7] Ron Milo, Shai Shen-Orr, Shalev Itzkovitz, Nadav Kashtan, Dmitri Chklovskii, and Uri Alon. Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002.

[8] Robert J Prill, Pablo A Iglesias, and Andre Levchenko. Dynamic properties of network motifs contribute to biological network organization. *PLoS biology*, 3(11), 2005.

[9] Christopher L Rees, Diek W Wheeler, David J Hamilton, Charise M White, Alexander O Komendantov, and Giorgio A Ascoli. Graph theoretic and motif analyses of the hippocampal neuron type potential connectome. *Eneuro*, 3(6), 2016.

[10] Michael W Reimann, Max Nolte, Martina Scolamiero, Katharine Turner, Rodrigo Perin, Giuseppe Chindemi, Paweł Dłotko, Ran Levi,

Kathryn Hess, and Henry Markram. Cliques of neurons bound into cavities provide a missing link between structure and function. *Frontiers in computational neuroscience*, 11:48, 2017.

[11] Olaf Sporns. The human connectome: linking structure and function in the human brain. In *Diffusion MRI*, pages 401–428. Elsevier, 2014.

[12] C Shan Xu, Michal Januszewski, Zhiyuan Lu, Shin-ya Takemura, Kenneth Hayworth, Gary Huang, Kazunori Shinomiya, Jeremy Maitin-Shepard, David Ackerman, Stuart Berg, et al. A connectome of the adult drosophila central brain. *BioRxiv*, 2020.