

Network Motifs in the *C.elegans* Neural Network and Their Dynamic Analysis

Yannan Zheng*

Department of Physics, Massachusetts Institute of Technology

(Dated: May 14, 2009)

Network motifs are defined as subgroups of interconnection pattern that recur at frequencies much higher than those found in randomized networks. This work applied the motif detection software developed by U.Alon to the neural network of *C.elegans*[1]. For the detected significant motifs, mathematical modeling and computational simulation are performed, and possible functions of them are inferred from the dynamics.

PACS numbers:

I. INTRODUCTION

Many models have been proposed to describe the topological and dynamic features of complex networks. These include the Erdős-Rényi random-graph model[2][3], percolation theory[3][4], Watts-Strogatz model[5], and Barabási-Albert model[7]. The neural network of the nematode *Caenorhabditis elegans* was shown to have a ‘small-world’ character, for that the characteristic path length of the network increases logarithmically with the number of vertices n . In addition, it has a clustering coefficient much larger than random networks and the connectivity distribution of the network has a fast exponentially decaying tail[5][8]. These characters agree with Watts-Strogatz model in a crude approximation.

To further understand the design principle of the *C.elegans* neural network, we need to go beyond these global statistical features. One way to disentangle this complex wiring diagram is to decompose it into ‘network motifs’. In this work, I applied the motif detector software[1] to *C.elegans* neuronal network. The dynamics of significant 3- and 4- node motifs are simulated. Generalizations from these simple network motifs and their functions are inferred. The simulation in this paper follows the work of [9],[10], and [11].

II. ALGORITHM

Uri Alon and his group at Weizmann Institute of Science developed a software, named mfinder1.2 for network motifs detection [1]. The basic idea of the algorithm is as following[9][10][12][13].

The network is represented as a connectivity matrix, M , and $M_{i,j} = 1$ represents a directed edge from node j to node i . $M_{i,j} = 0$ otherwise.

For stringent comparison, the randomized networks has to preserve the same incoming, outgoing degree per node as the real network. Two ways has been employed to generate such random networks. One is to use

a Markov-chain algorithm, starting with the real network and repeatedly swapping randomly chosen pairs of connections. The other is an *ab initio* algorithm, starting from an empty connectivity matrix and repetitively generate nonzero elements according to the weights $p_i = R_i / \sum R_i$ and $q_j = C_j / \sum C_j$, and update weights after each edge generation. Both algorithms do not allow multiple edges between nodes or self-connection.

To detect all connected n -node subgraphs in the connectivity matrix, exhaustively enumeration is performed. The algorithm recursively searches for nonzero elements in the matrix, then loops through all connected elements until n -node subgraph is obtained. Isomorphic effect (multiple submatrices of M may correspond to one architecture) is considered to allow correct counting of the subgraphs.

The subgraph counting is performed on both the real network and every randomized network. And the score of a certain subgraph is a function of the three parameters: (1) the appearance of the subgraph in the real network compared to it in randomized networks, or the Z-score, defined as $Z_i = \frac{N_{real_i} - \langle N_{rand_i} \rangle}{std(N_{rand_i})}$; (2) The uniqueness U , defined as the number of times it appears in the real network with distinct sets of nodes; and (3) Subgraph concentrations C_i , defined as the ratio between the number of this type of n -node subgraph and the total number of n -node connected subgraphs in the network, $C_i = N_i / \sum_i N_i$. And the network motifs are ranked according to this score.

III. SIMULATION ON *C.ELEGANS* NEURAL NETWORK

A. *C.elegans* neural network description

The entire nervous system of the adult *C.elegans* is the sole example of a completely mapped neural network. It has been reconstructed from serial section electron micrographs by White and his colleagues[14]. The reconstruction yielded an excellent understanding of the connectivity pattern of synapses in *C.elegans*. The nervous system contains approximately 300 neurons, 2000 neuromuscular junctions, 5000 chemical synapses between neurons and

*Electronic address: ynzheng@mit.edu

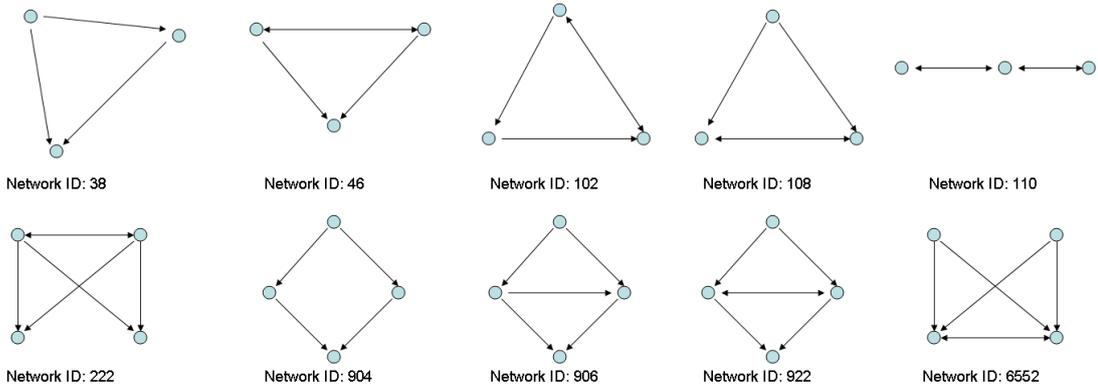


FIG. 1: The 10 most significant 3 and 4-node network motifs of *C.elegans* neural network. The motif id is labeled according to the ‘Motif Dictionary’ Uri Alon, *et al* compiled[12].

700 gap junctions, and the synaptic contacts is relatively invariant from animal to animal.

Here I use the raw data of [14] as employed in [5], which is available on [6]. Results are compared to [9], [10] and [11]. Note that the muscle cells were excluded from the data, and only those neurons with synapses connection > 5 were used in [9], [10] and [11].

In the neural network, the directed edge $X \rightarrow Y$ means neuron X has a synaptic connection to neuron Y. The classic rate equation

$$\tau \frac{dx_i}{dt} + x_i = f\left(\sum_{j=1}^N W_{ij}x_j - \theta_i\right) \quad (1)$$

describes the integrate-and-fire dynamics of the spiking neurons. It can be shown[15][16] that electric synapses, which has graded rather than spiking voltage potentials, can still be described by the same equation. Thus the rate model serves as the basic equation to describe the dynamics of neurons in the following sections. For simplicity, I use rectification function $f(u) = [u]^+ = \max\{u, 0\}$ as the activation function. This is a good approximation in the non-saturated region.

B. Motif Detection and Dynamic Analysis

1. Network Motif Detection

I applied mfinder1.2 to *C.elegans* neuronal networks. Since the exhaustive searching of subgraphs is extremely time consuming, it could take up to days to detect even 4-node motifs. Here I only use 100 randomized network for comparison. Therefore, P-value is ignored due to small number of control. To avoid detection of ‘pseudo’ 4-node motifs, which appear to be significant simply because it has a highly significant subpattern, metropolis algorithm[9] is also used. This algorithm preserved the same number of all 3-node subgraphs in addition to single node characteristics. As a cost, it is even more time

consuming, thus I only generated 20 randomized network using this algorithm. The final result of significant 4-node motif is a balance of the two algorithms. For detection of motifs larger than 4 nodes, in principle, I can use the efficient sampling algorithm mentioned in [17]. But still it is too expensive for personal computer. Thus I merely cite the results from [11], and support them by topological generalization of 3 and 4-node motifs.

Fig. 1 is the graphic representation of 10 most significant 3 and 4-node network motifs of *C.elegans* neural network. Table I gave their number of appearances, Z-score, uniqueness and subgraph concentration. Dynamics of these motifs will be discussed in the following section.

TABLE I: Statistics of detected motifs

Network ID	N_{real}	Z-score	Uniqueness	$C_i(\times 1000)$
38	1972	11.91	51	41.67
46	542	26	34	11.45
102	179	2	28	3.78
108	312	21.8	30	6.59
110	148	14.5	23	3.13
222	1665	84	20	1.19
904	5110	20	26	3.67
906	2110	17	24	1.51
922	2199	94	19	1.58
6552	803	45	18	0.58

2. Feedforward loop (FFL) is a network motif

Feedforward loop (FFL, id: 38) is the most common 3-node motif in the network. The Z score I got is even much higher than it is in [9]. One possible explanation is that FFL is commonly found in the signal processing system. I included the neuromuscular cells into the network, those are the effector of the sensory-response reflex arc. Inclusion of them might complete the previously open loop, and contribute to the appearance of FFL motif.

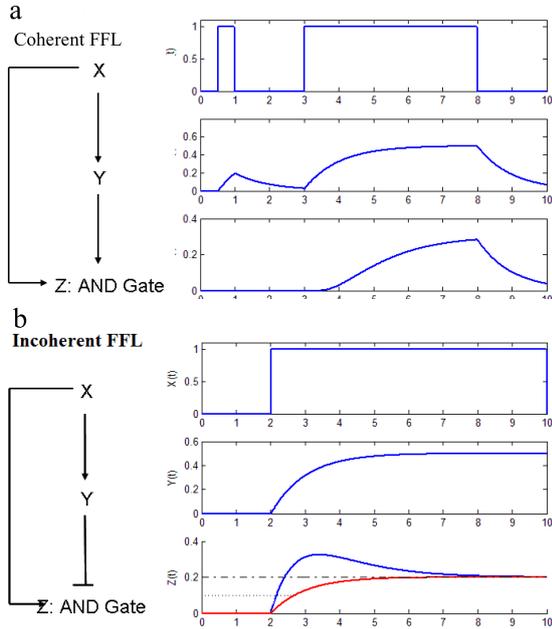


FIG. 2: The dynamic of FFLs. a: This Coherent FFL can act as a ‘sign-sensitive delay’ element and persistence detector. b: This Incoherent FFL can act as a pulse generator and response accelerator.

The popularity of FFL is highly related to the versatile functions it can perform. One type of Coherent FFL is shown in Fig. 2 (a), dynamic simulation shows that if Z has high threshold, it performs AND computation on X and Y signals. Thus it can reject transient activation signals from X and respond only to persistent signals, which activates Y, then activates Z. Meanwhile, the system allows a rapid shutdown. The delay of the ON step can be useful to filter out brief spurious pulses of signal, and allow robust performance in the noisy environment. The type of Incoherent FFL shown in Fig. 2 (b) is a response accelerator, the speed up of dynamics can be essential in response to crisis[13][18][19].

Interestingly, FFL is also common in transcription and signal transduction networks[13], though these biochemical networks operate on a very different spatial and temporal scales with respect to the neuronal network. This coincidence may reflect a convergence of evolution toward a similar goal, which is information processing on noisy signals using noisy components[9][19].

3. FFL generalize into multi-input FFL

Basic network motifs can also serve as building blocks of larger structure. Though FFL is common in both transcription and neural networks, their topological generalization are different. FFLs generalize into Multi-input FFLs in neural network whereas multi-output FFLs is common in transcription networks[11]. A graphic representation of multi-in and multi-out FFL is shown in

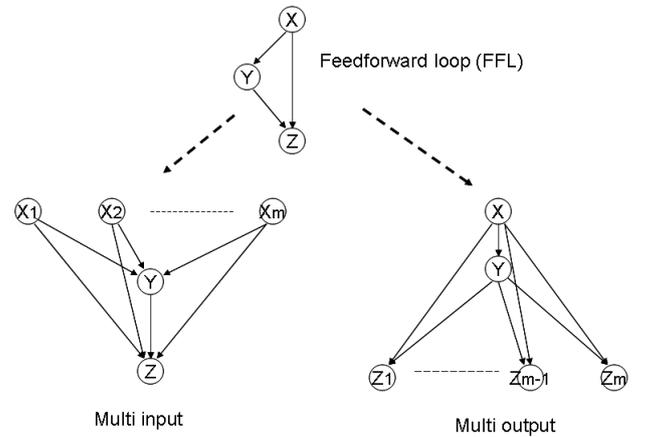


FIG. 3: Two topological generalization of simple network motif FFL. Adapted from [11]

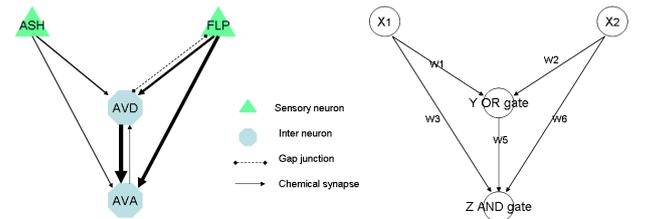


FIG. 4: One FFL generalization found in *C.elegans* locomotion neuronal circuit. The left picture is adapted from [20]

Fig. 3.

The multi-output FFL in transcription network is shown to perform detailed temporal regulation of gene expression[13]. To infer the function of multi-input FFL, here we analyze the dynamics of a two-input FFL, which is found in the *C.elegans* locomotion neuronal circuit[11], shown in Fig. 4. Here we chose the threshold of rectification function such that Y acts as an OR gate and Z acts like an AND gate. Synapse weight were chosen proportional to the number of synapses connecting two neurons[20]. And the dynamic of neurons are shown in Fig. 5.

From simulation, we can see that like simple FFL, the two-input FFL can also acts as a persistence detector with respect to its stimulus. In addition, it can also function as a coincidence detector, say, two transient input of X_1 and X_2 , which by itself cannot activate Z, can achieve activation if the two signals come concomitantly, due to the transient-memory storage capacity of Y[11].

4. Multi-layer perceptron (MLP) is a network motif

When examining patterns with four or more nodes, one finds that the most abundant network motifs are multi-layer perceptrons[19]. This can be inferred from Appendix A, which showed that the neuronal network

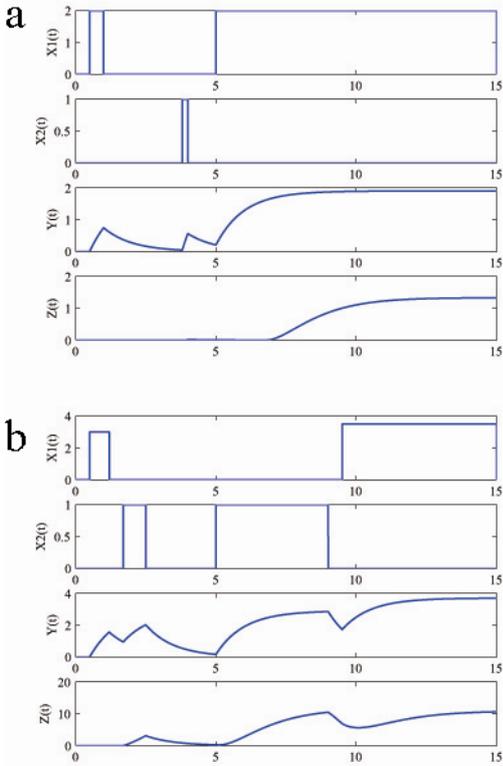


FIG. 5: Dynamics of two-input FFL. a: It can also acts as a persistence detector. b: It can also acts as a coincidence detector.

of the *C.elegans* is highly organized into 7 to 8 layers. It can also be supported by topological generalization of 3 and 4 node motifs. Motif 904, which has a diamond structure, is the basic building block of multi-layer perceptrons. Motif 46, 108, 222, 922 and 6552, all together constitute half of the top 10 most significant motifs, can generalize into MLP with symmetric lateral connections. motif 906 can be generalize into MLP with asymmetric lateral connection. I will first discuss the advantage of MLP, then discuss the possible functions of lateral connections.

One obvious reason for the *C.elegans* to use layered structure is that MLP has much powerful computation power than single layer perceptrons. In fact, it has been shown that any smooth function can be approximately computed by an MLP with just two layers of synapses and a sigmoidal activation function[15][16]. With a deeper network, MLP maybe able to realize computation with fewer neurons or synapses, but the computation could be more time consuming. Thus, the current layered structure of *C.elegans* must be some tradeoff between depth and size. It is interesting to note that in contrast to the deeper network of *C.elegans* neuronal circuits, *E.coli* transcription network has few long cascades[13]. One possible reason for this shallow architecture is the need for rapid response: it can take up to one cell-generation

time to pass a signal down each step of a cascade. Sensory networks that are designed to respond rapidly to external signals might therefore be limited in their use of long cascades[13][18]. But neuronal network works on the timescale of milliseconds to seconds, time is no longer a demanding constraint here. A more stringent requirement is the exact response to complex stimulus. Perceptrons can be either highly selective or highly invariant, depending on activation thresholds. By alternately performing convolution (conjunction) and sub-sampling (disjunction), MLP can potentially recognize and response to rather complicated signals[15][16].

Considering lateral connection, since I lack the detailed information of synapses connections. I will only mention briefly some possible functions of it. Lateral inhibition amplifies differences relative to commonalities[15][16], this can facilitate feature selectivity by setting up a dynamic threshold. Strong mutual inhibition as synaptic motif acts as a forbidden sets[15][16], and it is useful to implement computation involving constrains.

IV. CONCLUSION

In the current algorithm, all directed edges are treated with equal weights. In fact, interactions can be either excitory or inhibitory; there exists multiple types of interactions, e.g. chemical synapses and gap junctions; and the strengths of interaction can also be different. In the present studies, this information is added manually after motif detection, but it would be beneficial to incorporate this information at the beginning[4][12].

Self connection is forbidden in the current algorithm. However, autoregulatory is potentially important single-node network motif since it can provide key functions such as speeding response times and reducing noise[12][18].

In the generation of randomized networks, the probability of two neurons connecting is completely independent of their relative positions in the network. However, in reality, two neighboring neurons have a greater chance of forming a connection than two distant neurons. Without regarding these localized aggregation would misclassify a completely random but spatially clustered network as non-random and significant network motifs. In this case, a random geometric graph is more appropriate[22].

In conclusion, this project use a network motif software to detect network motifs in the *C.elegans* neuronal network. Mathematical modeling and computational simulation were performed to understand the function of the detected motifs. A better understanding of the network will depend on a finer measurement of weight for each edge and on improvement of motif-detecting algorithm. Finally, functions of network motifs inferred from theoretical analysis need to be tested with dynamic experiments in living cells.

APPENDIX A

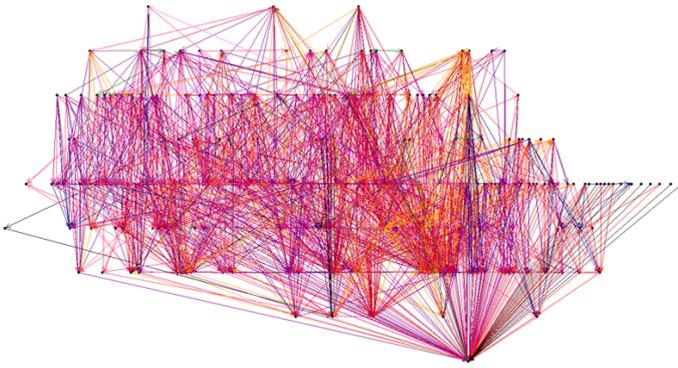


FIG. 6: Multi-Layer Perceptron structure of 3 and 4- node motifs of *C.elegans* neural network. Drawn using software Mdraw from [21].

-
- [1] Motif detector software mfinder1.2 is downloaded from <http://www.weizmann.ac.il/mcb/UriAlon/>.
- [2] Erdős, P., and A. Rényi, On random graphs. *Publ. Math. (Debrecen)* **6**, 290 (1959).
- [3] M. Kardar and L. Mirny, 8.592J Statistical Physics in Biology lecture notes. (2009)
- [4] Réka Albert and Albert-László Barabási, Statistical mechanics of complex networks. *Review of Modern Physics*, **74**, 47 (2002).
- [5] Watts, D. J. and Strogatz, D. H., Collective dynamics of ‘small-world’ networks. *Nature* **393**, 440 (1998).
- [6] *C.elegans* neuronal network raw data is downloaded from <http://cdg.columbia.edu/uploads/datasets/celegansrawdata> with a few revision.
- [7] Albert-László Barabási and Réka Albert, Emergence of Scaling in Random Networks, *Science* **286**, 509 (1999).
- [8] Amaral, L. A. N., M. Barthélémy, and P. Gopikrishnan, Classes of small-world networks. *PNAS*, **97**, 11149 (1999).
- [9] Milo, R. *et al*, Network motifs: simple building blocks of complex networks. *Science* **298**, 824 (2002).
- [10] Milo, R. *et al*, Superfamilies of designed and evolved networks. *Science* **303**, 1538 (2004).
- [11] Kashtan, N. *et al*, Topological generalizations of network motifs. *Phys. Rev. E* **70**, 031909 (2004).
- [12] The manual and motif dictionary of mfinder1.2 is downloaded from <http://www.weizmann.ac.il/mcb/UriAlon/>.
- [13] Shen-Orr, *et al*, Network motifs in the transcriptional regulation network of Escherichia coli. *Nature Genet*, **31**, 64 (2002).
- [14] White, J. G., Southgate, E., Thomson, J. N. and Brenner, S. *Philos. Trans. R. Soc. London B* 314, 1 (1986)
- [15] S. Seung, Classical neurodynamics (To be published).
- [16] S. Seung, 9.641 Intro. to Neural Networks lecture notes. (2009)
- [17] Kashtan, N. *et al*, Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs. *Bioinformatics* **20**, 1746 (2004).
- [18] Alon, U. Network motifs: theory and experimental approaches. *Nature Genetics* **8**, 450 (2007).
- [19] Alon, U., Introduction to Systems Biology: Design Principles Of Biological Circuits (CRC, Boca Raton, 2006).
- [20] The example of locomotion circuit of *C.elegans* is adapted from <http://ims.dse.ibaraki.ac.jp/research/Celegans.html>
- [21] Motif visualization software mdraw is downloaded from <http://www.weizmann.ac.il/mcb/UriAlon/>.
- [22] COMP572 Network Motifs lecture notes, spring (2008).