Search for computational modules in the C. elegans brain

Markus Reigl\textsuperscript{1}, Uri Alon\textsuperscript{2}, and Dmitri B. Chklovskii\textsuperscript{1}
\textsuperscript{1}Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724
\textsuperscript{2}Weizmann Institute, Israel

Does the C. elegans nervous system contain multi-neuron computational modules that perform stereotypical functions? We attempt to answer this question by searching for recurring multi-neuron inter-connectivity patterns in the C. elegans nervous system’s wiring diagram. Our statistical analysis reveals that some inter-connectivity patterns containing two, three and four neurons are significantly over-represented relative to the expectations based on the statistics of smaller inter-connectivity patterns. We do not find any over-represented five-neuron patterns. Over-represented patterns (or motifs) are candidates for computational modules that may perform stereotypical functions in the C. elegans nervous system. These modules need to be investigated further both anatomically and physiologically.
INTRODUCTION

There is little doubt that neurons are elementary building blocks of the nervous system (Cajal, 1999). It is less clear, however, whether multi-neuron modules (smaller than invertebrate ganglia or vertebrate nuclei and cortical columns) can be meaningfully defined, either anatomically (Peters and Sethares, 1996) or physiologically (Abeles and Gerstein, 1988). The existence of such multi-neuron modules would greatly simplify the description of the nervous system structure and function. An example of such simplification can be found in electrical engineering. An electronic circuit is often represented in terms of modules such as operational amplifiers, logical gates and memory registers rather than as a wiring diagram showing each transistor, resistor and diode. However, unlike electrical engineers who designed these modules themselves, neurobiologists did not design the brain, and evolution rarely leaves records of its experimentation. Therefore, if multi-neuron modules have indeed evolved they need to be discovered.

In this paper, we search for anatomically defined multi-neuron modules in the *C. elegans* nervous system. We choose *C. elegans* as a model organism because its wiring diagram is known entirely, including all 302 neurons and synapses between them (Achacoso and Yamamoto, 1992; Durbin, 1987; Hall and Russell, 1991; White et al., 1986; Wood, 1988). Our reasoning is that if a certain multi-neuron module performs some stereotypical function it may appear in the nervous system repeatedly (Milo et al., 2002; Shen-Orr et al., 2002). Therefore, search for multi-neuron connectivity patterns that appear more often than by “chance” (compared with the expectations as defined below) may yield these multi-neuron modules. Of course, there may be functionally important modules that appear infrequently and would be missed by our analysis. In the electronic circuit analogy, our approach would discover logical gates in a processor wiring diagram but not a rectifier in a power supply, which is essential but appears only once.

To search for *N*-neuron modules, we sort all *N*-neuron combinations into classes defined by their inter-connectivity pattern and count the number of combinations in each class. By comparing these counts with the mean counts from the random matrix ensemble, constructed based on our expectations, we detect significantly over-represented patterns, or motifs. In order to avoid assigning significance to a *N*-neuron pattern just because it contains *N−1*-neuron motifs we incorporate the *N−1*-neuron statistics into the expectations used to search for *N*-neuron motifs (Milo et al., 2002). To do this, we perform our search sequentially, by starting with doublets (or neuronal pairs, *N=2*) and then increasing the number *N* of neurons included in the pattern sequentially up to quintuplets (*N=5*).

We use three datasets of the *C. elegans* nervous system’s wiring diagram. Datasets 2 and 3 (Durbin, 1987) were obtained from serial sections electron microscopic reconstructions of two different animals (White et al., 1986). Dataset 1 (Achacoso and Yamamoto, 1992) combines data from these animals (White et al., 1986) with other reconstructions (Hall and Russell, 1991; Wood, 1988), for details see Methods. Although two given neurons may be connected by more than one synapse (multiplicity of connection), the small size of the dataset compels us to use the binary representation of these connections (connected or unconnected). In order to obtain binary connectivity matrices, we threshold the multiplicity of connections at various values Θ: Pairs having less than Θ synapses are considered unconnected while those having at least Θ synapses are considered connected. Such procedure is justified because more than a single synaptic contact may be necessary for an observable physiological effect of one neuron on another. Since we do not know the physiologically relevant count of synapses, we repeat our calculation for $1 \leq \Theta \leq 7$. 

2
RESULTS

1. Bi-directionally connected doublets (N=2) are over-represented. We classify all possible doublets (or pairs) of the C. elegans neurons into three classes: unconnected, uni-directionally connected, and bi-directionally connected, and compare the number of doublets in each class to that expected in a random network (Table 1). The random network ensemble consists of matrices that preserve the numbers of incoming and outgoing synapses for each neuron but not the identities of the partners (Maslov and Sneppen, 2002; Newman et al., 2001). The motivation behind this choice of the random matrix ensemble and the details of the algorithm are explained in Methods.

| Pattern: | 1 | 2 | 3 |
|----------|---|---|---|
| Actual # of doublets in C. elegans | $c_{2,i}(A)$ | 13,902 | 435 | 28 |
| Avg. # of doublets in random networks | $c_{2,i}(R)$ | 13,878.8 | 481.3 | 4.8 |
| Standard deviation | $\sigma(c_{2,i}(R))$ | 2.0 | 4.1 | 2.0 |
| Relative difference between A and R | $\Delta_{2,i}(A,R)$ | 0.002 | -0.1 | 4.8 |
| $P$ | $<=0.001$ | $<=0.001$ | $<=0.001$ |

Table 1: Doublet statistics in the C. elegans network, A, is significantly different from the random matrix ensemble {R}. Most importance is assigned to the bi-directionally connected doublet count because of the larger relative difference, $\Delta_{2,i}(A,R)$, defined in (1). Counts shown are for dataset 2, threshold $\Theta=3$, number of random matrices $n=1000$. Other datasets and thresholds give similar results.

Although the data in Table 1 indicate that the actual number of doublets in each class is significantly different from the mean of the random matrix counts, we think the most important finding is the over-representation of the reciprocally connected doublets (pattern #3), for two reasons. First, minor but inevitable mistakes in the datasets of the C. elegans wiring diagram can affect the number of doublets in each class. Therefore, only the differences greater than the typical error in the number of doublets are significant. Assuming that the error is proportional to the doublet count, we introduce a relative excess measure

$$\Delta_{2,i}(A,R) = \frac{c_{N,i}(A) - c_{N,i}(R)}{c_{N,i}(R)}, \quad 2 \leq N \leq 5, \quad i \text{ is the index of the } N\text{-neuron pattern}. \quad (1)$$

The high value of this measure for reciprocally connected neurons indicates the importance of this finding compared to the unconnected or uni-directional pattern. Second, if the total number of doublets in a network is kept the same then the over-representation of one class inevitably leads to the under-representation of another. Thus it is reasonable to assign most significance to the class that is relatively most over-represented. We repeat the above calculations for other datasets and threshold values and consistently find the significant over-representation of bi-directionally connected doublets (data not shown).

Can bilateral (left-right) symmetry of the C. elegans neuronal network account for the over-representation of the reciprocally connected doublets? Indeed, there are 99 bilaterally symmetric pairs of neurons in C. elegans. If connections between these pairs obeyed bilateral symmetry
then they could not be uni-directional, creating a bias in favor of bi-directional connections. To see whether this is the case, we calculate the percentage of bi-directional connected doublets, which consist of a bilateral neuron pair. We find that these percentages are small, 15%, 7.1%, and 5.5% in datasets 1, 2, and 3, respectively. Therefore, bilateral symmetry is not sufficient to explain the observed result. The over-representation of reciprocally connected doublets has been reported previously on the qualitative level (White et al., 1986) and explained (Durbin, 1987) as a consequence of correlation between adjacency and connectivity of neurons. Interestingly, an over-representation of bi-directionally connected doublets was also found for pyramidal neurons in layer 5 of mammalian neocortex (Markram et al., 1997; Sjostrom et al., 2001). This suggests that the over-represented motifs may be evolutionarily conserved.

2. **Several triplet classes (N=3) are over-represented.** We classify all connected triplets in the *C. elegans* wiring diagram into 13 classes and count the number of triplets in each class. We compare the actual number of triplets in each class to the null-hypothesis random matrix ensemble defined as follows. In order to include the observed over-representation of reciprocally connected doublets, we construct random networks that preserve the numbers of bi-directional and uni-directional connections for each neuron. Based on above arguments we plot the relative number of triplets in each class.

![Figure 1: Triplet counts in the actual network compared to a random matrix ensemble.](image)

For threshold $\Theta=2$ we find that several triplet counts are noticeably different from the mean of the random matrix ensemble, e.g. pattern #10, #12, and #14 in Figure 1. To ensure that these results are independent of the arbitrary threshold choice we repeat the calculation for different values of the threshold (within the biologically plausible range, $\Theta=1$ to 7), Figure 2.
Figure 2: Relative excess measure of triplet counts for different thresholds. Θ indicates an over-representation of three patterns. To detect over-representation of triplet patterns we plot the relative excess value $\Delta_{i,3}(A,R) = \left(\frac{c_{3,i}(A) - c_{3,i}(R)}{c_{3,i}(R)}\right)$ for various thresholds in datasets 2 and 3. Asterisk (*) indicates that the actual network value is outside the distribution calculated for the random matrices, i.e. the triplet count $c_{3,i}(A)$ in the actual network is greater than the count $c_{3,i}(R)$ for all random matrices $R$. All random matrices in the ensemble ($n=1000$) preserve the number of uni- and bi-directional connections for each neuron.

The significance of these findings may be characterized by the probability of obtaining a given number of triplets by chance within the chosen random matrix ensemble. For example, asterisks in Figure 2 indicate that this probability is less than 0.001. However, there are two problems with using this probability as significance measure. First, as mentioned in case of doublets, this measure is not robust against inevitable errors in the existing connectivity datasets. Second, because we search for over-representation in multiple classes simultaneously (i.e. perform multiple hypotheses testing) this probability under-estimates the true p-value.

To overcome these problems, we introduce a different significance measure, $P_{m}$, for every pattern and threshold. This measure represents the probability of finding a random matrix $R$ in our ensemble, in which at least one pattern $i$ has the same or greater relative excess value $\Delta$ as found for a given pattern in the actual network. This measure can be calculated by replacing $A$ in
the calculation of $\Delta N, i(A, R)$ with each of the random matrices from our ensemble and counting the number of matrices that have the same or greater excess value. By dividing this number of matrices by the total size of the random matrix ensemble, we estimate the multiple hypotheses testing corrected $p$-value $P_m$, Figure 3 (see Methods).

![Dataset: 2 (all connected neurons)](image)

![Dataset: 3 (all connected neurons)](image)

**Figure 3: Significance measure $P_m$ of triplet over-representation for different thresholds.** The multiple hypothesis testing corrected $p$-values for triplet patterns show a significant over-representation of patterns 10, 12 and 14 (datasets 2 and 3, $n=1000$). The significance measure $P_m$ is the probability that at least one triplet class in a random matrix has the same or greater relative excess value as the actual network. The triplet counts for high $\Theta$ in strongly connected patterns are often small or zero. To safeguard against fluctuations due to small counts, we do not display $P_m$ for a pattern and $\Theta$ if less than 5 triplets were found in the *C. elegans* network.

One of the most consistently over-represented motifs is the feedforward loop (triplet pattern #10), previously noticed in *C. elegans* (Hall and Russell, 1991) and other networks (Milo et al., 2002; Shen-Orr et al., 2002). Could some known feature of neuronal organization account for the observed over-representation of the feedforward loop? We consider two hypotheses and find that neither is sufficient to explain this observation:
i. The likelihood of connectivity between nearby neurons is not sufficient to account for over-representation of the feedforward loop. Since connectivity and adjacency are correlated in *C. elegans* and other nervous systems (White et al., 1986), one could argue the following. If two neurons have a common synaptic partner, then they are likely to be adjacent to that common partner, and hence to each other. If the two neurons are adjacent they are likely to be connected to each other. This argument can explain why strongly connected patterns (#10 to #16) are more likely to be over-represented than weakly connected patterns (#4 to #9). However, this argument would also predict an over-representation of the triplet classes #13 and #11 (the feedback loop), which is contrary to our observations, Figure 3. In fact, triplet class #11 is significantly under-represented in dataset 1 (data not shown). Therefore, the correlation between connectivity and adjacency by itself does not explain the feedforward loop over-representation.

![Dataset: 2 (only inter-neurons)](image)

![Dataset: 3 (only inter-neurons)](image)

Figure 4: Significance measure $P_m$ of triplet over-representation among inter-neurons shows that pattern #10 is significantly over-represented. Multiple hypothesis testing corrected $p$-values for the triplet over-representation in datasets 2 and 3 ($n=1000$). $P_m$ is calculated only if the triplet count in the actual network $c_{3,i}(A,R)\geq 5$. The figure shows that motif #10 is significantly over-represented within interneurons.

ii. The feedforward nature of the neuronal network is not sufficient to account for over-representation of the feedforward loop. If one views the *C. elegans* nervous system as a three-
layer feedforward network, where sensory neurons synapse mostly on interneurons, and interneurons synapse on other interneurons or motorneurons, this could explain the over-representation of the feedforward loop. We argue that this is not the case for two reasons. First, the feedforward loop is also over-represented among interneurons, Figure 4. Second, the three-layer model of the *C. elegans* nervous system is overly simplified. For example, there are feedback connections from interneurons to sensory neurons and from motorneurons to interneurons. To evaluate whether detected feedforward loops fit the three-layer feedforward network, we analyze the function of the neurons in these loops. About 40% of the detected feedforward loops either contain all neurons from the same functional group or at least one connection goes from a neuron in a lower layer to a neuron in a higher layer, Table 2. These loops do not fit into this three-layer model, undermining the hypothesis.

| Dataset | 1  | 2  | 3  | 4  | 5  |
|---------|----|----|----|----|----|
| 1       | 46 | 45 | 45 | 43 | 35 |
| 2       | 49 | 39 | 33 | 38 | 34 |
| 3       | 47 | 40 | 41 | 39 | 29 |

**Table 2:** Table shows the percentages of feedforward loops in which all three neurons belong to the same functional group or at least one of the three connections is made from an interneuron to a sensory neuron or from a motorneuron to an interneuron.

It is possible that the over-representation of the feedforward loop is a consequence of other factors or their combinations (such as feedforwardness and locality of connectivity combined). But even if these factors are found, the characterization of the network in terms of over-represented motifs remains valid. In transcription networks, the feedforward loop was proposed to carry out information processing functions such as filtering out fluctuations and responding only to persistent stimuli (Shen-Orr et al., 2002). Feedforward loop can also carry out other functions (Hall and Russell, 1991), depending on the polarity of synapses involved. These possibilities should be investigated experimentally.

In addition to the feedforward loop, we find that two other (both symmetric) patterns are consistently over-represented: pattern #12 and pattern #14 (Figure 3). These patterns weren’t identified as motifs in a previous work (Milo et al., 2002) because of the low absolute count at threshold $\Theta=5$ they used. Again, we ask whether this could be a consequence of the bilateral symmetry of the *C. elegans* nervous system. Indeed, the bilateral symmetry implies that pairs of bilaterally symmetric neurons are also connected symmetrically, meaning that triplets containing such a pair are likely to be symmetric. However, we find that the fraction of triplets #12 and #14 containing a bilaterally symmetric pair of neurons and an unpaired neuron is rather small (between 10% and 20% in datasets 2 and 3). This suggests that the bilateral symmetry of the nervous system is not sufficient to explain the over-representation of pattern #12 and #14.

3. **Several quadruplet classes (N=4) are over-represented.** We classify all connected quadruplets into 199 classes and count the number of quadruplets in each class. Then we compare the actual counts of quadruplets in each class to the mean counts of quadruplets in a random matrix ensemble. In this case, random matrices preserve the numbers of uni-directional and bi-directional connections for each neuron and, in addition, the numbers of triplets (see Methods). Because of the large number of quadruplet classes, we show results (Figure 5) only for patterns selected according to the following criteria: the multiple hypothesis testing corrected
significance values $P_m$ must be less than 0.1 for at least one threshold per pattern, while the number of quadruplets in the actual network $c_{4,i}(A)$ must be at least 5 (or the mean count of quadruplets must be at least 5 in the random matrix ensemble in case of under-represented outliers). The last two conditions exclude patterns that may appear as over-represented or under-represented due to very small quadruplet counts.

Figure 5: Significance measure $P_m$ of selected quadruplet over-representation for different thresholds. Multiple hypothesis testing corrected $p$-values for the quadruplet over-representation in datasets 2 and 3 ($n=1000$). The patterns shown satisfy the following selection criteria: There must be a significant value $P_m < 0.1$ for at least one $\Theta$ and the count $c_{4,i}(A)$ of this pattern in C. elegans must be at least 5. All random matrices in the ensemble ($n=1000$) preserve the number of uni- and bi-directional connections for each neuron as well as the count in all triplet classes for the whole network.

We find two consistently over-represented motifs: quadruplet patterns #45 and #51 (Milo et al., 2002). Can we explain this observation by some other known factor? We consider the following two hypotheses:

1. Bilateral symmetry of the nervous system is not sufficient to explain the over-representation of the two feedforward quadruplets. One could propose that symmetric patterns should be over-
represented because of the bilateral symmetry of the nervous system. We think that this argument by itself cannot explain the observed over-representation for two reasons. First, the fraction of feedforward quadruplets containing two bilaterally symmetric neuron pairs in motif 45 is rather small (less than 10% in dataset 2 and less than 14.3% in dataset 3). The percentage of bilaterally symmetric neuron pairs in motif 51 varies between 12% and 100% (data not shown). Second, many symmetric patterns are not over-represented, such as, for example, patterns 25, 30, 31, 35, 43, 44 and 65.

ii. Feedforward structure of the nervous system may partially explain the over-representation of the feedforward quadruplet. One could propose that the feedforward three-layer structure of the nervous system could account for this observation (see over-represented triplets). We find that 5% and 46% of the feedforward quadruplets do not fit into this proposition because either they contain a feedback connection or all neurons belong to the same layer (Table 3). After comparing these percentages to the relative excess values, Figure 6, we conclude that the feedforward structure may explain over-representation for some threshold values but not for others.

| motif i=45  | Theta = |
|-------------|---------|
| Dataset 1   | 1 28.5% 24.7% 26.9% 15.0% 14.3% |
| Dataset 2   | 2 37.1% 30.8% 29.4% 25.0% 22.6% |
| Dataset 3   | 3 34.4% 24.7% 15.6% 19.1% 14.3% |

| motif i=51  | Theta = |
|-------------|---------|
| Dataset 1   | 1 21.1% 18.3% 19.2% 5.2% 2.8% |
| Dataset 2   | 2 40.7% 37.4% 34.0% 45.5% 33.3% |
| Dataset 3   | 3 46.2% 47.5% 34.0% 34.6% 0.0% |

Table 3: Percentage of quadruplets in pattern #45 and #51, which do not fit into the three-layer feedforward network model. These quadruplets contain either all four neurons from the same layer or at least one connection from a motorneuron to an interneuron or from an interneuron to a sensory neuron.

It is possible that some other factors (in addition to feedforwardness) account for the reported quadruplet over-representation. Just as argued in case of triplets, discovering these factors would be complementary to the over-represented motif characterization.

4. We find no over-represented quintuplet classes (N=5). We classify all connected quintuplets into 9364 classes (out of 9608 patterns total) and count the actual number of quintuplets in each class. We compare these counts with the mean of the random matrix ensemble. In this case, the random matrices preserve the numbers of uni- and bi-directional connections for each neuron and, in addition, keep the numbers of all triplets and quadruplets in a 10% range of the actual network. We do not find any significantly over-represented quintuplets. This could happen for two reasons. First, there may be no significantly over-represented quintuplets with a given number of quadruplets. Second, the C. elegans dataset may be too small to detect statistical outliers.
DISCUSSION

By comparing counts of multi-neuron patterns in the *C. elegans* wiring diagram to the mean counts of the appropriate random matrix ensemble, we find several over-represented motifs. First, we find that bi-directionally connected doublets (out of three possible doublet classes) are over-represented, given the number of connections on each neuron is fixed. Second, several triplet classes (out of thirteen possible connected patterns) are over-represented, given the actual number of bi-directional (as well as uni-directional) connections for each neuron. Third, we find that several quadruplet classes (out of 199 connected patterns) are over-represented, given the numbers of triplets are preserved in addition to previously listed constraints. We find no over-represented quintuplet classes. The identified motifs are candidates for modules that may
perform stereotypical functions in the *C. elegans* nervous system, and they need to be investigated further.

Our motif search algorithm is different from previous attempts to find modules. For example, traditional clustering approaches look for the subsets of nodes, which are connected with their own subset stronger than with other subsets. In our algorithm, we consider all the connections within a pattern (unlike (Achacoso and Yamamoto, 1992), who considered only some connections within the pattern) but ignore the connections with neurons outside the pattern. One could question the expediency of ignoring multiple possible inputs to the neurons in a module since those inputs could influence the operation of that module. To counter this, we point out that the nervous system is capable of performing many different functions under different circumstances and neurons active in one case may be silent in another. Therefore, in any particular case, many of the anatomical inputs to the module may remain silent and can be safely ignored. This speculation may be verified experimentally by simultaneous monitoring of neuronal activity in different neurons. A second justification for looking at $N$-neuron patterns is that if there is a particularly recurring way to connect an external connection to a given motif, it would appear as an $N+1$-neuron motif. If, on the other hand, the motif is connected in many different ways in different instances, their significance will be washed out. Thus our approach may hierarchically detect modules with recurring input/output sites, growing them out of smaller patterns.

The strategy and algorithms we described in this paper can be applied to incompletely mapped networks because a highly significant pattern is also likely to be over-represented in a sub-network. However, the statistical power of our algorithm increases with the knowledge of the wiring diagram. Therefore it was natural to choose the *C. elegans* nervous system, which represents the biggest wiring diagram mapped entirely. Unfortunately, *C. elegans* has some disadvantages when it comes to the interpretation of the results: the polarity of a synapse (excitatory vs. inhibitory) in *C. elegans* is often unknown, electrophysiological investigations are still difficult in *C. elegans* (Lockery and Goodman, 1998), and the whole network contains only 302 neurons limiting the statistical power of the approach. Yet we hope that the recent technological developments (Kerr et al., 2000) will eliminate the first two disadvantages and allow functional analysis of the discovered modules. Moreover, we expect that our results have implications for understanding nervous system structure and function beyond *C. elegans*. The modules we identify in *C. elegans* may be evolutionarily conserved, and once identified can be searched for in other systems. Finally, the algorithm itself can be applied to other networks (Milo et al., 2002) once they become available.

As in any other theoretical analysis, we made several simplifications. In particular, we thresholded the multiplicity of connections, ignored the polarity of the synapses and the existence of gap junctions. Yet, our results are robust to the inclusion of these factors in the future because if an over-represented class is found, it will remain over-represented even if divided into smaller sub-classes. It would be interesting to see whether the inclusion of the above factors will reveal specific over-represent sub-classes.
METHODS

Representation of the networks

The C.elegans synaptic data was represented in adjacency matrices $A$, in which entry $A_{ij}$ is 1 if there is a connection from neuron $i$ to neuron $j$ and 0 otherwise. We assume a synaptic connection for threshold $\Theta$ if neuron $i$ makes at least $\Theta$ synapses onto neuron $j$. The order of neurons in this representation is not important for our calculations.

Detecting & counting patterns

We implemented two strategies for counting the number of triplets, quadruplets and quintuplets in a connectivity matrix. First, to get the count of all $N$-neuron patterns, we took all possible $N$-neuron subsets and characterized their connectivity. Second, in cases when we were only interested in connected patterns, i.e. there is a undirected connection between all $N$ neurons of this pattern, we took all possible $N$-neuron subsets out of the neighborhood of a neuron $a$. This neighborhood is defined by all neurons that can be reached from $a$, if the directed connectivity matrix is made undirected.

In both cases it is very crucial for run time of the algorithm to detect the pattern class from these connectivity sub-matrices as fast as possible. We realized that by designing a hash function, with maps all $N$-neuron sub-matrices to a unique function value and pre-calculating a lookup table with the corresponding pattern index for all these function values.

Creating Random Matrices

N=2. The random matrices we create preserve in-degree and out-degree of all neurons but change their connection partners. This choice of random matrices is motivated by the observation that the distribution of in-degrees and out-degrees in C. elegans is different from Poisson expected for the Erdős-Rényi random graph (Figure 7). Starting from an empty matrix the algorithm selects neurons in a random order and connects these neurons randomly to the necessary amount of other neurons with a degree (in the random matrix) which is still smaller than in the C. elegans network.

N=3. We keep the number of incoming and outgoing uni-directional connections as well as the number of reciprocal connections for each neuron the same. One of the implemented algorithms starts with an empty matrix, selects randomly a neuron and reconnects all outgoing connections to other neurons with an in degree smaller than in the C elegans network, all incoming connections to neurons with a too small out degree and all reciprocal connections to neurons with a unconnected reciprocal connection. We also implemented a second algorithm to verify the robustness of our results. This algorithm (Maslov and Sneppen, 2002; Newman et al., 2001) will randomly pick and swap 2 unidirectional or 2 bi-directional connections (a$\rightarrow$ b and c$\rightarrow$ d will changed to a$\rightarrow$ d and c$\rightarrow$ b).

N=4. For comparing the count of quadruplets, we construct random matrices that keep not only in degree and out degree of uni-directional and bi-directional connections for each neuron the same but also the count of the 16 different 3 neuron pattern in the whole matrix. Starting from a random matrix for $N=3$ as described above we used the Simulated Annealing algorithm by swapping two connections of the same type until the count for all triplets in the random matrix matches the real network. Since this swapping operation does not change the degrees of the
various connection types for the neuron we only had to include the triplet count into our energy function for the Simulated Annealing algorithm.

N=5. For the analysis of the quintuplets, we modified the Simulated Annealing algorithm to also match the count of all 4-neuron patterns to the real network. We could only find random matrices for which the relative difference between the count of each pattern in the random matrix and the real dataset was less than 10%.

**p-Value calculation/multi hypotheses testing correction**

Because we calculate the relative excess measure $\Delta$ for all patterns in parallel, there is an increased probability to find an over-represented pattern by chance (there are 13 patterns for $N=3$, 199 patterns for $N=4$ and 9364 patterns for $N=5$). We correct for that by calculation a $p$-value $P_m(X_\Delta)$ that reflects the probability that one random matrix $R_{k0}$ out of our random matrix ensemble $R$ will have at least one pattern $i$, which is over-represented by the same rate $X_\Delta$. To eliminate patterns with a very small count, we threshold $c_{N,i}(R_{k0})$.

For a given relative excess value $X_\Delta$ we perform the following procedure:

1. For all random matrices $R_{k0}, k_0 \in \{1, \ldots, n\}$ out of the ensemble we calculate the relative excess value $\Delta(R_{k0}, \overline{R_{k0}})$ between $R_{k0}$ and all other random matrices in this ensemble for each pattern $i$.

2. We then count the number of random matrices $R_{k0}$ for which this value $\Delta(R_{k0}, \overline{R_{k0}})$ is greater than $X_\Delta$ for at least one pattern $i$ and, to eliminate patterns with a very small count, the count $c_{N,i}(R_{k0})$ of this pattern in $R_{k0}$ is above a threshold (usually 5).

3. We derive $P_m(X_\Delta)$ by dividing the number of matrices we found in step 2 by the total number $n$ of random matrices.

In mathematical notation this can be written as

$$P_m(X_\Delta) = \frac{1}{n} \left\{ R_{k_0} \exists i : \Delta(R_{k_0}, \overline{R_{k_0}}) > X_\Delta \land c_{N,i}(R_{k_0}) \geq c_{\text{cutoff}} \right\}, \quad 2 \leq N \leq 5, \ R_{k_0} \in \{R\}.$$
Datasets/Datasources

We used data from two sources in our calculations:

1. (Achacoso and Yamamoto, 1992) assembled a database of neuronal connection from different sources. From this dataset we took all reported synaptic connections between 280 neurons. We did not use the connection data for the 4 classes WE, WI, WM, WN in this dataset. The data does not include data for the 20 pharyngeal neurons and also leaves the neurons CANL and CANR unconnected. We used this dataset mainly for control.

Corrections made to the data:
- We changed the class for neuron AUAR from CRI to CRS (makes AUAR an sensory neuron). Source: (White et al., 1986), symmetry to AUAL.
- We changed the class for neuron CEPDR from BLS to BRS (associates CEPDR with the right side). Source: [http://www.wormatlas.org](http://www.wormatlas.org), symmetry to CEPDL.

2. (Durbin, 1987) provides separate connectivity data for the different reconstructions JSH and N2U done by White et.al. We calculated our results for both of these reconstructions separately taking only the send and send_joint synaptic connection out of the Durbin’s data. We also deleted 11 non-neuronal cell or classes from the dataset: CEPshDR, CEPshVL, CEPshVR, GLRDL, GLRDR, GLRL, GLRR, GLRVL, GLRVR, hyp, mu_bod. The classification of the neurons into their function and their location was taken from dataset 1.

| Source                  | Dataset | Threshold Θ | Connected neurons | Total connections | Uni-directional connections | Reciprocal connections |
|-------------------------|---------|--------------|-------------------|-------------------|----------------------------|------------------------|
| Achacoso, Yamamoto      | 1       | 1            | 280               | 1973              | 1776                       | 197                    |
|                         |         | 2            | 275               | 1141              | 1079                       | 62                     |
|                         |         | 3            | 268               | 755               | 726                        | 29                     |
|                         |         | 4            | 250               | 552               | 531                        | 21                     |
|                         |         | 5            | 231               | 386               | 372                        | 14                     |
|                         |         | 6            | 203               | 295               | 288                        | 7                      |
|                         |         | 7            | 183               | 231               | 225                        | 6                      |
| Durbin JSH              | 2       | 1            | 179               | 1152              | 1011                       | 141                    |
|                         |         | 2            | 175               | 649               | 603                        | 46                     |
|                         |         | 3            | 170               | 463               | 435                        | 28                     |
|                         |         | 4            | 161               | 328               | 313                        | 15                     |
|                         |         | 5            | 140               | 226               | 216                        | 10                     |
|                         |         | 6            | 119               | 163               | 156                        | 7                      |
|                         |         | 7            | 106               | 124               | 119                        | 5                      |
| Durbin N2U              | 3       | 1            | 187               | 1288              | 1143                       | 145                    |
|                         |         | 2            | 181               | 728               | 685                        | 43                     |
|                         |         | 3            | 172               | 460               | 445                        | 15                     |
|                         |         | 4            | 162               | 323               | 313                        | 10                     |
|                         |         | 5            | 138               | 206               | 200                        | 6                      |
|                         |         | 6            | 119               | 143               | 138                        | 5                      |
|                         |         | 7            | 97                | 104               | 101                        | 3                      |

Table 4: Number of connected neurons and the count of the different connection types after applying thresholds to the 3 datasets.
REFERENCES

Abeles, M., and Gerstein, G. L. (1988). Detecting spatiotemporal firing patterns among simultaneously recorded single neurons. J Neurophysiol 60, 909-924.

Achacoso, T. B., and Yamamoto, W. S. (1992). Neuroanatomy of C. elegans for computation (Boca Raton, FL, CRC Press).

Cajal, S. R. y. (1999). Texture of the Nervous System of Man and the Vertebrates, Vol 1 (New-York, Springer).

Durbin, R. M. (1987) Studies on the Development and Organisation of the Nervous System of Caenorhabditis elegans, Ph.D>, Cambride University, Cambridge.

Hall, D. H., and Russell, R. L. (1991). The posterior nervous system of the nematode Caenorhabditis elegans: serial reconstruction of identified neurons and complete pattern of synaptic interactions. J Neurosci 11, 1-22.

Kerr, R., Lev-Ram, V., Baird, G., Vincent, P., Tsien, R. Y., and Schafer, W. R. (2000). Optical imaging of calcium transients in neurons and pharyngeal muscle of C. elegans. Neuron 26, 583-594.

Lockery, S. R., and Goodman, M. B. (1998). Tight-seal whole-cell patch clamping of Caenorhabditis elegans neurons. Methods Enzymol 293, 201-217.

Markram, H., Lubke, J., Frotscher, M., Roth, A., and Sakmann, B. (1997). Physiology and anatomy of synaptic connections between thick tufted pyramidal neurones in the developing rat neocortex. J Physiol 500 ( Pt 2), 409-440.

Maslov, S., and Sneppen, K. (2002). Specificity and stability in topology of protein networks. Science 296, 910-913.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network motifs: simple building blocks of complex networks. Science 298, 824-827.

Newman, M. E., Strogatz, S. H., and Watts, D. J. (2001). Random graphs with arbitrary degree distributions and their applications. Phys Rev E Stat Nonlin Soft Matter Phys 64, 026118.

Peters, A., and Sethares, C. (1996). Myelinated axons and the pyramidal cell modules in monkey primary visual cortex. J Comp Neurol 365, 232-255.

Shen-Orr, S. S., Milo, R., Mangan, S., and Alon, U. (2002). Network motifs in the transcriptional regulation network of Escherichia coli. Nat Genet 31, 64-68.

Sjostrom, P. J., Turrigiano, G. G., and Nelson, S. B. (2001). Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. Neuron 32, 1149-1164.

White, J. G., Southgate, E., Thomson, J. N., and Brenner, S. (1986). The structure of the nervous system of the nematode Caenorhabditis elegans. Philos Trans R Soc Lond B Biol Sci 314, 1-340.

Wood, W., ed. (1988). The nematode Caenorhabditis elegans (Cold Spring Harbor, NY, Cold Spring Harbor Laboratory).