Supporting Information

Itzkovitz *et al.* 10.1073/pnas.0712207105

SI Text

Analysis of Scaling Regimes for the Relation Between Minimal Number of Addresses and Network Size. In this section we analyze the two scaling regimes of the relation between the minimal number of addresses and the network size, and supply the critical network size above which the scaling becomes independent of network size. We start with Eq. **3** and divide by N^2:

$$\frac{D}{N} = \frac{\log_2 c}{N} + \left(\frac{c}{N}\right)^2 = \frac{1}{N}\left[\log_2 N + \log_2 \frac{c}{N}\right] + \left(\frac{c}{N}\right)^2$$

$$= f_1\left(\frac{c}{N}\right) + f_2\left(\frac{c}{N}\right) = f_3\left(\frac{c}{N}\right) \quad \textbf{[S1]}$$

We next examine D/N vs. c/N. Both of these scaled variables are bounded between 0 and 1:

The minimal number of addresses can be graphically obtained as the intersection of D/N with f_3. For small values of c the logarithm function f_1 predominates in its contribution to f_3 and the squared function f_2 can be neglected (solid blue line, Fig. S3). When D is larger than $\log_2 N$, f_2 predominates and the solution follows f_2 (dashed and dotted blue lines in Fig. S3). Thus, the critical network size, above which the minimal number of addresses becomes independent of network size is $N_{\text{crit}} = 2^D$. The dotted line in Fig. S3 depicts the solution for estimated parameters in the human brain. For the biologically relevant parameters the scaling of minimal number of addresses vs. network size follows a square root law.

Generalization of Scaling Results to Networks with Arbitrary Neighborhood Graphs. Our scaling results (Eq. **3**) do not explicitly require geometric constraints and can be generalized to other topologies, as long as the neighborhood of each node, that is the number of nodes with which it can potentially interact, D, is arbitrarily predefined. To test this we used two generalizations to our geometric network model. In the first one we generated networks in a similar manner to the random geometric networks, and then assigned each neuron the connectivity neighborhood and the connectivity pattern of an arbitrarily chosen different neuron. This model fits scenarios such as topographic mappings, in which long range axons connect to a small and geometrically close set of targets that are distant from the original neuron (1). In the second generalization, we generated arbitrary connectivity and neighborhood matrices, with the constraint that each neuron has, on average k outgoing edges, which are a subset of D arbitrarily chosen potential targets. This model also includes small world networks (2). In both cases we relaxed the symmetry of the neighborhood graph. We find that although the minimal number of addresses in these networks is higher than in the geometric networks, the square root scaling law of the minimal number of addresses with network size is invariant to these generalizations (Fig. S2).

Scaling Exponent Depends on Mean Connectivity. Whereas our analytical analysis predicts the scaling exponent of the number of addresses vs. network size to be 1/2, we find that the exponent is slightly influenced by the mean connectivity (Fig. S1). This can be understood by noting that the number of neuronal addresses required to encode the wiring of a fully connected neighborhood ($k = D$) and an empty network ($k \to 0$) approaches zero.

Controlling for Symmetries in the *C. elegans* Neuronal Network. In the main text we computed the minimal number of neuronal addresses required to wire the *C. elegans* neuronal network. To control for the prominent left-right and ventral-dorsal symmetries in the *C. elegans* neuronal network (3, 4) we repeated our calculations on a reduced network that includes only neurons from the right ventral parts whenever a homologous neuron from the left or dorsal parts existed. This was done by first eliminating all neurons whose names end with the letter L, when another neuron ending with the letter R exists. Of the remaining neurons we next eliminated any neuron with the ending DR, whenever a neuron with the same three-letter prefix and the ending VR or R existed. This resulted in a network with 170 neurons and 873 synapses.

We find that the real network can be encoded by 45 addresses, significantly less than the number of addresses required to encode the randomized networks with the same degree connectivity and geometric constraints ($P = 0.01$). The significantly smaller number of addresses required to encode the wiring of the *C. elegans* neuronal network is thus not only due to the left–right and ventral–dorsal symmetries in the network.

Minimal Number of Addresses Is an NP Complete Problem. Here, we show that the problem of finding the minimal number of addresses (denoted as *ADDRESS*) is NP-complete. Therefore no efficient algorithm exists that can solve this problem deterministically.

Theorem:

$3SAT \leq 16ADDRESS$

Proof:

We seek a mapping f from formulas in $3CNF$ to directed connectivity networks and corresponding non-directed neighborhood networks such that for every formula ϕ in $3CNF$, ϕ is satisfiable iff the minimal number of addresses of $f(\phi)$ is ≤ 16.

Formally, we need to find f that is easy to compute, such that:

$f : 3CNF \to \{G(V,E), G'(V,E')\}$

$\forall\ \varphi \in 3CNF$: $\varphi \in 3SAT \leftrightarrow f(\varphi) \in 16ADDRESS$

Given a formula ϕ in $3CNF$, f computes the directed network $G(V,E)$ and the nondirected neighborhood relation $G'(V,E')$. It starts by constructing the network G as follows:

First, it creates index nodes 1 to 16 and connects them as described in Fig. S4. Then, for each variable Xi in ϕ, it creates six literal nodes, denoted as: Xi, Xi', Xi'', $\neg Xi$, $\neg Xi'$, $\neg Xi''$ and connects them to each other and to nodes 1 to 3 as described in Fig. S4. Finally, for every clause C_i in ϕ, it creates a clause node, denoted as C_i, and connects it to three literal nodes as follows: first it connects it to the nontagged literal node representing the first literal in clause i, secondly it connects it to the tag literal node corresponding to the nontagged literal node representing the second literal in clause i, and finally, it connects it to the double tag literal node corresponding to the nontagged literal node representing the third literal in clause i. (see Fig. S4 for an example on two clauses: C_1 and C_2)

Notice that the order of the literals in the clauses affects the connections of the clause nodes to the literal nodes. The neighborhood network G' contains nondirected edges between all pairs of nodes which contain an edge in either direction in G, all pairs of index nodes and all pairs of literal nodes that represent a variable and its negation (the dashed brown line in Fig. S4 shows nodes that are neighbors according to G').

It is easy to see that f is easy to compute and it is complete (i.e., well defined for every formula in $3CNF$).

Direction 1:

Here, we show that if $\phi \in 3SAT$, then $f(\phi) \in 16ADDRESS$. $\phi \in 3SAT \rightarrow \phi$ is satisfiable, i.e., there is an assignment S that satisfies $\phi \rightarrow$ we can use S to assign addresses to $f(\phi)$ in the following way: first assign 16 different addresses to the index nodes as shown in Fig. S4. Then assign the red address of node 8 to every literal node Xi or $\neg Xi$ to which assignment S assigned a false value and a green address of node 5 to every literal node Xi or $\neg Xi$ to which assignment S assigned a true value. The reason that only one of these two addresses can be used is that index node 2 receives connections from both index nodes 5,8 and from all pairs of nontagged literal nodes.

Assignment S defines a consistent assignment of addresses for all of the literal nodes that are denoted as Xi or $\neg Xi$, because addresses 5 and 8 are disconnected among both the index nodes and the literal nodes.

Continue to assign addresses to the tagged and double tagged nodes according to S while maintaining consistency with the addresses of the index nodes. This means that the tagged nodes will be assigned either the address of node 4 or the address of node 9 and that the double-tagged nodes will be assigned the address of either node 6 or node 7. This is again dictated by the connection of index nodes 1 and 3, respectively. For instance: if node Xi was assigned the red address of node 8 then Xi' can only be assigned to the red address of node 9 and X'' can only be assigned to the red address of node 7. This also means that node $\neg Xi$ was assigned the green address of node 5 and $\neg Xi'$ can only be assigned to the green address of node 4 and Xi'' can only be assigned the green address of node 6.

Because S satisfies ϕ, it assigns a true value at least for one literal in every clause in $\phi \rightarrow$ there is no clause in which all of the literals are assigned the red address of node 8 $\rightarrow$ there is no clause for which C_i is connected to three red literal nodes (that were assigned the addresses of nodes 7 to 9) $\rightarrow$ it is possible to assign at least one of the addresses of nodes 10 to 16 to every C_i node $\rightarrow$ There is a consistent assignment of 16 addresses to all of the nodes in $f(\phi) \rightarrow f(\phi) \in 16ADDRESS$.

Direction 2:

We now show that if $f(\phi) \in 16ADDRESS$ then $\phi \in 3SAT$.

$f(\phi) \in 16ADDRESS \rightarrow$ There is a consistent assignment S that uses only 16 addresses for $f(\phi)$. In addition, $f(\phi) = \{G(V,E), G'(V,E')\}$ such that G contains a subgraph of 16 index nodes that are in neighborhood relation to each other in G'.

All of the index nodes must be assigned to different addresses (otherwise there will be inconsistency in the relation between the addresses) $\rightarrow$ S assigns all of the literal and clause nodes addresses that where also assigned to the index nodes.

Every literal node that is denoted by Xi or $\neg Xi$ must be assigned to the same address of node 5 or 8. In addition node Xi will not be assigned to the same address of node $\neg Xi \rightarrow$ we can construct an assignment S' that satisfies ϕ in the following way: every literal that was assigned by S to the same address of node 5 will be assigned by S' to the value true, and every literal that was assigned by S to the same address of node 8 will be assigned by S' to the value false.

There is no clause C_i in which all of the corresponding literals nodes are assigned the same address of node 8 by S because in this case there is no address in the existing 16 addresses that could be assigned by S to C_i in contradiction to the success of S in using only 16 addresses for $f(\phi) \rightarrow$ At least one literal was assigned a true value by S' in every clause in $\phi \rightarrow S'$ satisfies $\phi \rightarrow \phi \in 3SAT$

Theorem:

$ADDRESS \in$ NP-complete

Proof:

$ADDRESS \in$ NP (We can guess a solution and check in polynomial time if it is admissible)

$3SAT \in$ NP-complete

$3SAT \leq 16ADDRESS$.

$16ADDRESS \leq ADDRESS$ (By finding the minimal address of G and G', we solve the question if the minimal number of addresses is ≤ 16.)

1. Udin SB, Fawcett JW (1988) Formation of topographic maps. *Annu Rev Neurosci* 11:289–327.
2. Watts DJ, Strogatz SH (1998) Collective dynamics of ''small-world'' networks. *Nature* 393:440–2.
3. White JG, Southgate E, Thomson JN, Brenner S (1986) The structure of the nervous-system of the nematode *Caenorhabditis-elegans. Philos Trans R Soc London Ser B* 314:1–340.
4. Durbin, R (1987) *Studies on the Development and Organisation of the Nervous System of Caenorhabditis elegans.* (Cambridge Univ Press, Cambridge, UK).

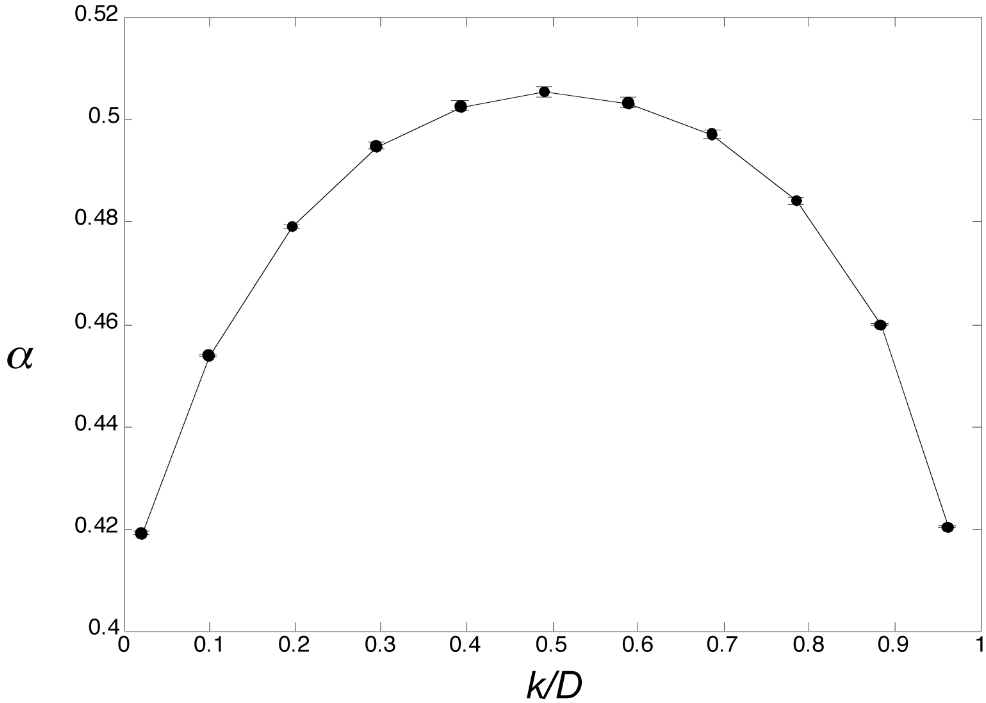

Fig. S1. Scaling exponent α depends on the mean connectivity k. The scaling exponent is obtained by fitting the formula $c = aN^{\alpha}$, where c is the minimal number of addresses and N is the network size (N was varied over three orders of magnitude). The highest scaling exponent, approximately 1/2, is obtained at intermediate values where each node can connects to about half of its neighborhood nodes ($k/D = 1/2$). For both low and high connectivity the number of addresses required to wire geometric networks becomes smaller.

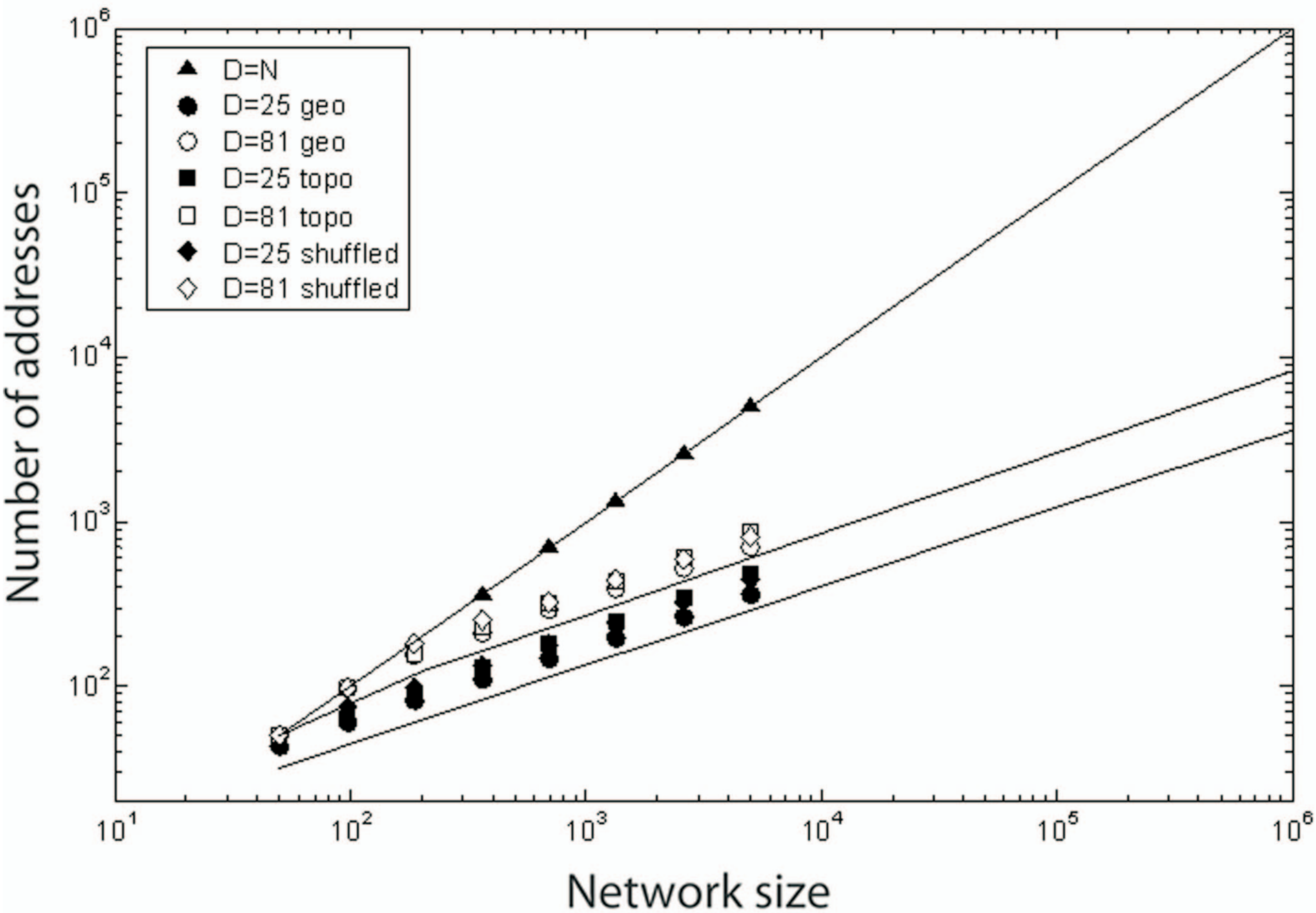

Fig. S2. Sublinear scaling of minimal number of addresses vs. network size holds for more generalized network models. Geo denotes the geometric networks, topo denotes the first generalization model, and shuffled denotes the second generalization model. All networks are one-dimensional and have a mean connectivity $k = 5$.

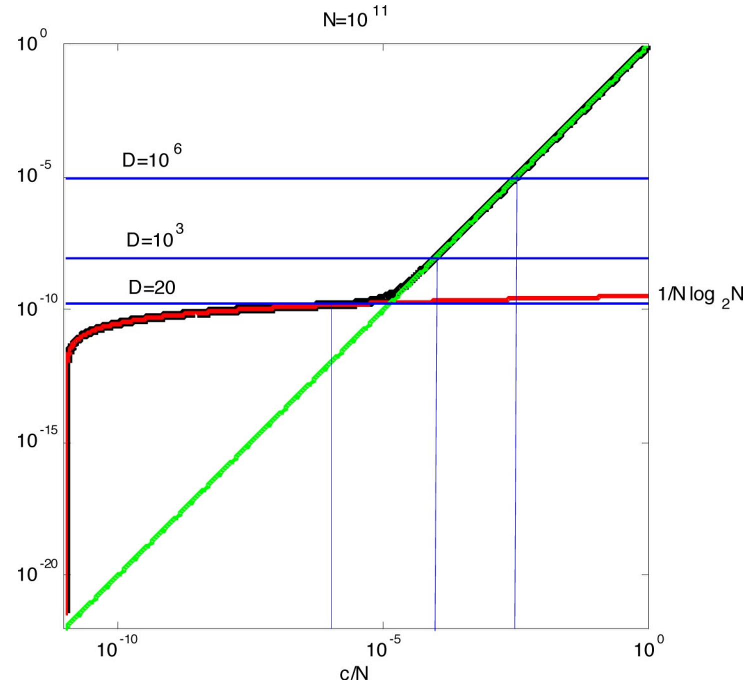

Fig. S3. Graphical solution for the minimal number of addresses for different network sizes N and neighborhood sizes D. f_1 is in dashed red, f_2 in dashed green, and f_3 in bold black. Vertical blue lines denote the solution, obtained from the intersection of f_3 and D/N. Solid line is the solution for $D = 20$, dashed line for $D = 10^3$, and dotted line for $D = 10^6$. Note that even for moderately high D the quadratic term in c/N (the green curve f_2) dominates in the solution. Note the logarithmic scales.

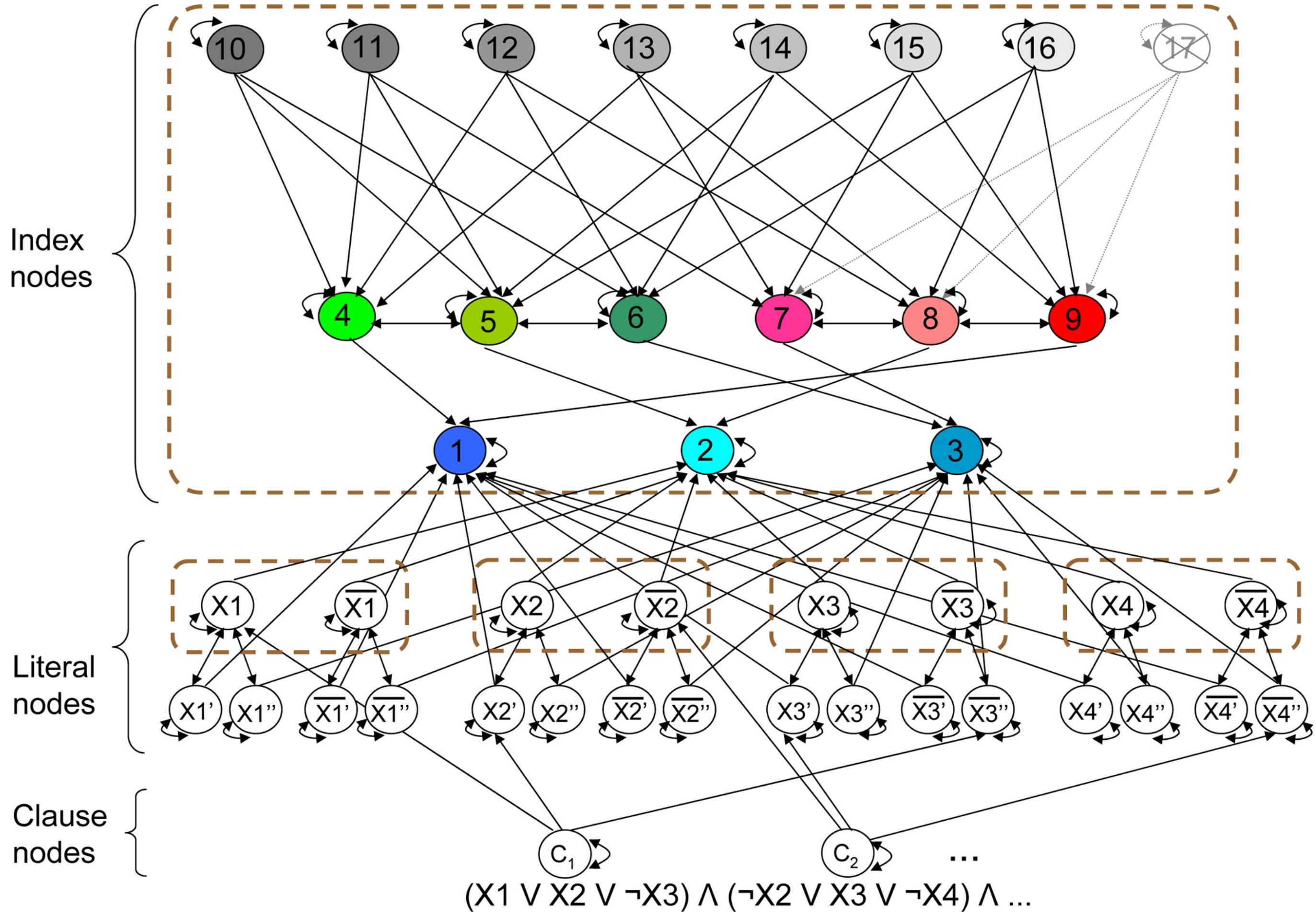

Fig. S4. Reduction from 3*SAT* to 16*ADDRESSES*. Node 17 and its connections are drawn for clarity purpose only (to stress the forbidden configuration) and are not part of the graph.