

THE SMALL WORLD OF THE NOBEL NEMATODE CAENORHABDITIS ELEGANS

J. W. Clark and A. T. Eggebrecht

Department of Physics, Washington University
St. Louis, Missouri 63130, USA

1. INTRODUCTION

Many complex networks, consisting of nodes connected by directed or undirected links, display a “small-world” character [1-4]. Like a randomly connected network, they exhibit short global separations between nodes, yet like a regular lattice, they show high local connectivity. Well-known examples include the web of human acquaintances (“six degrees of separation”), the network of movie actors linked by shared billing, the internet, and food webs in ecological systems. In this contribution, key concepts of the statistical physics of complex networks will be illustrated by studying the connectivity of the 302-neuron “brain” of the nematode *C. Elegans*, a tiny worm that has been the subject of a huge body of scientific work recognized in the 2002 Nobel Prize in Physiology or Medicine.

Our successes in understanding the behavior of the systems studied traditionally by condensed-matter physicists and many-body theorists can be largely attributed to the simplicity of the operant pairwise interactions between the constituents [4]. There is no ambiguity or irregularity in the topology of interactions; it is clear what is connected to (or affected by) what. In many cases, the strength of the interaction is uniquely determined by the ordinary physical 3-space distance between the constituents (together, perhaps, with their spin configurations). Indeed, the interactions are commonly local, in the usual 3-space sense. However, scientists are increasingly confronted with more complex systems – or, more specifically, complex networks of nodes or agents and their connections or interactions – in which ordinary physical distance loses its meaning, and the existence of an interaction between any two constituents becomes ambiguous. One faces situations (as in the airline system) where local or near-neighbor actions – whether in the conventional sense or more abstractly – can produce striking global effects, in a manner that depends sensitively on network structure. The last few years have seen a great surge of interest in the development of an effective statistical physics of complex networks as well as an improved mechanistic

understanding of their topological properties [2–4].

In Section 2, we focus on neural nets as a prime example of complex networks. We propose, as a primitive yet apt measure of the distance between two neurons, the smallest number of synaptic jumps needed to get from the one to the other. Following Ref. [1], we proceed to develop a theory of this measure for randomly connected neural nets. In Section 3, we present results for the average distance and distance distributions in two simple models of networks that deviate from the random paradigm and display varying degrees of local order. Section 4 enlarges our view and introduces, in addition to distance and its distribution, the further concepts of clustering coefficient and degree distribution as necessary ingredients of a statistical analysis of complex networks. Section 5 is devoted to a brief discussion of what it means to be a “small-world network” and offers numerous examples. In Section 6, we motivate the study of *C. Elegans* both as a “hydrogen atom” of multicellular animals and, through its nervous system, as a test-bed for understanding the small-world phenomenon. Section 7, the heart of our discourse, consists of a detailed analysis of the topological properties of a major portion of the nervous system of this noble worm, based on the theory and concepts provided in Sections 2 and 3 and employing connectivity data available on the world wide web. Finally, our analysis and results are compared with an earlier study [2] of the “brain” of *C. Elegans* that also served to verify the small-world nature of its topology.

2. DISTANCE IN RANDOM NEURAL NETWORKS

Consider a network of N neurons, regarded as structureless units that interact via synaptic junctions. A chemical synapse only works one way, with information flowing from the presynaptic neuron j to the postsynaptic neuron i ; thus, neuron j can have a direct effect in i without the reverse being true. In other words, neural interactions are generally not symmetric. (Newton’s third law is not obeyed in the nervous system: action and reaction are not equal!) The *connectivity* of such a neural network – its “wiring diagram” or “architecture” – is conveniently specified by a $N \times N$ matrix (V_{ij}) , where the matrix element $V_{ij} \in \{0, 1\}$ is nonzero if and only if unit j extends a connection to unit i .

We define the *distance* d_{ki} from generic neuron i to generic neuron k by the minimum number of synaptic junctions which information must traverse in going from i to k . The *distance matrix* (d_{ki}) is in general asymmetric.

Now imagine an ensemble of neural networks, each network being assembled randomly by the following algorithm.

- (a) Consider an arbitrary neuron j and potential target neuron i , also chosen arbitrarily; with probability m , connect j to i with a synapse directed from j to i . (Thus, i is a potential postsynaptic neuron of the presynaptic neuron j .)
- (b) Repeat for all j and then for all neurons i .

The parameter m , is called the *connection probability* or the *connection parameter* of the ensemble of networks. Such nets define a class of *unstructured random networks*. Each neuron has, on average, $M = Nm$ outputs and the same number of inputs. It

will be assumed that all distances in each sample net are finite. By this we mean that information can be transmitted from *any* neuron to *any* other neuron in the system via some sequence of neurons and across some sequence of synapses, while observing the fact that some (indeed, most) interneuronal junctions only work one way. This is the relevant case for our analysis, although it will not hold strictly for a biological nerve net because any functioning neural system must have inputs from, and outputs to, its surroundings in the organism or the world.

The above definitions and the coming analysis will actually apply just as well to the study of abstract networks composed of nodes connected by links. One may simply replace the specific terms “neuron” by “node” and “synaptic junction” by “directed link.” Thus one may consider an ensemble of randomly constructed graphs with oriented edges. One may also consider the simpler case of undirected edges, in which connections between nodes work both ways. The theory of random graphs was pioneered by Erdős [5]. For a standard text, see Bollabás [6]; for an excellent recent review, see Albert and Barbási [4].

Keeping with the case of directed random nets, let us focus on a particular neuron A and find, in the ensemble average, the number $n(d)$ of neurons lying a distance d from A. Denote by $q(d)$ the number of neurons at distances from A *greater* than d ; then

$$q(d) = N - \sum_{r=1}^d n(r). \quad (1)$$

On average, each neuron has Nm output lines, so

$$n(1) = Nm. \quad (2)$$

The number of neurons left over is $N - Nm$, so

$$q(1) = N(1 - m). \quad (3)$$

Next pick, arbitrarily, one of the neurons B which receives one of the Nm outputs from A. On average, this neuron will itself have output lines going to Nm neurons. However, some of these neurons reached by B are already reached after only one step from neuron A, the rest being at a distance $d = 2$ from A. At any rate, outputs from B reach a fraction m of the $q(1)$ neurons left over after one step. This leaves a smaller pool of unreached neurons numbering

$$q(1) - mq(1) = q(1)(1 - m) = N(1 - m)(1 - m). \quad (4)$$

But neuron A has Nm target neurons like B, each of which reduces the pool of left-over neurons by a factor $(1 - m)$. Hence there are

$$q(2) = q(1)(1 - m)^{Nm} = q(1)(1 - m)^{n(1)} \quad (5)$$

neurons unreached after two steps, making use of $Nm = n(1)$.

Continuing in this way, we may establish the following recursion relations for the function $n(d)$, where $d \geq 1$:

$$q(d + 1) = q(d)(1 - m)^{n(d)}, \quad (6a)$$

$$n(d + 1) = q(d) - q(d + 1) = q(d) \left[1 - (1 - m)^{n(d)} \right]. \quad (6b)$$

These equations are initiated by $n(1) = Nm$ and $q(1) = N(1 - m)$. We immediately verify that $n(1) = N$ and $q(d) = 0$ for all d if $m = 1$.

The probability distribution $\rho(d)$ of neuron-to-neuron distances, with normalization $\sum_{d=1}^{d_{\max}} \rho(d) = 1$, is formed simply as $n(d)/N$. The average interneuronal distance is then given by

$$\begin{aligned}
\langle d \rangle &= \sum_{d=1}^{d_{\max}} d n(d)/N \\
&= \{n(1) + 2n(2) + 3n(3) + 4n(4) + \dots\}/N \\
&= \{n(1) + 2[q(1) - q(2)] + 3[q(2) - q(3)] + \dots\}/N \\
&= \{n(1) + 2q(1) - 2q(2) + 3q(2) - 3q(3) + 4q(3) - \dots\}/N \\
&= \{Nm + 2N(1 - m) + q(2) + q(3) + \dots q(d) + \dots\}/N, \\
\langle d \rangle &= 2 - m + \sum_{d=2}^{d_{\max}} q(d)/N. \tag{7}
\end{aligned}$$

For fixed, finite m independent of N (implying that M scales like N^1), $\langle d \rangle$ saturates at $2 - m$ for large N . This follows since, when m is held constant, $q(d)$ remains arbitrarily small compared to N for $d \geq 2$ at large enough N .

In general, it could be that M in $Nm = M$ scales like N^p instead of N^1 . The case $p = 0$ means that M stays fixed at some finite value while N grows. In this case the mean distance $\langle d \rangle$ cannot saturate asymptotically, since the minimum number of steps needed to cover the net is N/M and the average distance must be greater. The lack of saturation may also be seen from the series representation of $\langle d \rangle$ on observing that the first recursion relation implies $q(d + 1) = q(d) = N$ for $N \rightarrow \infty$. With $q(1) = N(1 - m) = N$, all terms of the summation in Eq. (7) then contribute an equal finite amount. This second scenario (i.e., $M \sim N^0$) might correspond to considering an ever greater sample of cortical tissue, where the number of outputs (or inputs) per neuron remains roughly constant at something like 100–1000.

Figures 1 and 2 illustrate the behavior of the average distance $\langle d \rangle$ determined from Eq. (7) and the recursion relations (6) for an ensemble of randomly assembled networks, respectively as a function of the connection probability m and as a function of the average number $M = Nm$ of outputs per neuron. Figure 1 clearly demonstrates the saturation of $\langle d \rangle$ for large network sizes, in the case that m is independent of N .

3. STRUCTURED AND UNSTRUCTURED NETWORKS

Suppose we know the *connection matrix* (V_{ij}) for a particular network, which might be a member of an ensemble constructed randomly according to the algorithm of the last section, or might be a regular network with orderly connections, or might be something in between. In other words, consider a network with an arbitrary pattern of directed connections. The distribution of distances in such a network can be found from the fact that if the distance from neuron l to neuron k is d , then the kl th element of the d th power of (V_{ij}) will be nonzero, while the kl th element of all lower powers will vanish.

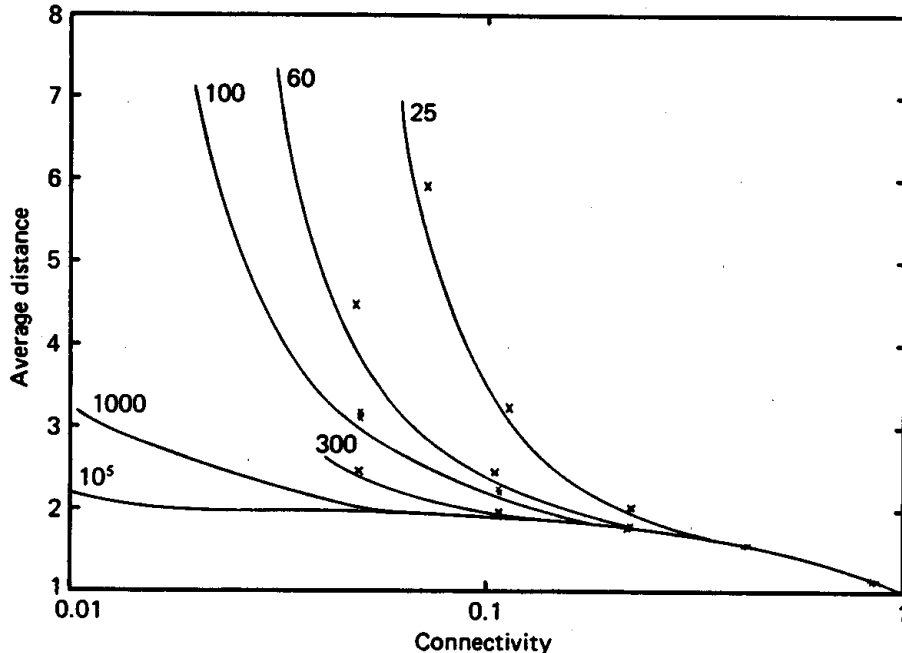


Figure 1. Average distance $\langle d \rangle$ as a function of connection probability m in an ensemble of randomly assembled networks for various sizes N . The solid curves are calculated from Eq. (7), while the crosses show actual mean distances in selected sample networks.

Thus, we have a simple recipe:

Compute successive powers of the matrix (V_{ij}) and, by counting the number of elements that change from zero to a nonzero value at each stage, construct the distance distribution $\rho(d)$ of the sample net.

Let us now consider two models for introducing order into network structure.

Collection of Weakly-Connected Netlets. Divide the set of N neurons into g groups (called netlets), each containing N/g neurons. Make *intranetlet* connections randomly, just as before, with connectivity parameter m_{int} , and form *internetlet* connections randomly with connectivity parameter $m_{\text{ext}} < m_{\text{int}}$. The ratio $l = m_{\text{int}}/m_{\text{ext}}$ serves as a measure of *local emphasis* or *clustering*. In simulations [1], the choices $N = 100$, $g = 4$, and $l = 5, 10$, and 20 have been considered.

Ring Structure. Having made an arbitrary initial labeling of the neurons, $i = 1, \dots, N$, form the connectivity matrix according to these rules:

- (i) Choose an integer L between 0 and $N-1$, and a number m satisfying $0 < m \leq 1$.
- (ii) For a chosen presynaptic neuron i and postsynaptic neuron j , the matrix element V_{ij} is permitted to differ from zero *only if* $i - j \leq L$. (Hence the integral parameter L limits the number of neighbors.)
- (iii) If the condition in (ii) is met, let j send a connection to i with probability m .
- (iv) Repeat for all $N(N-1)$ pairs ij . In simulations [1], the choices $N = 100$, L from 3 to 20, and m from 0.2 to 0.8 have been considered.

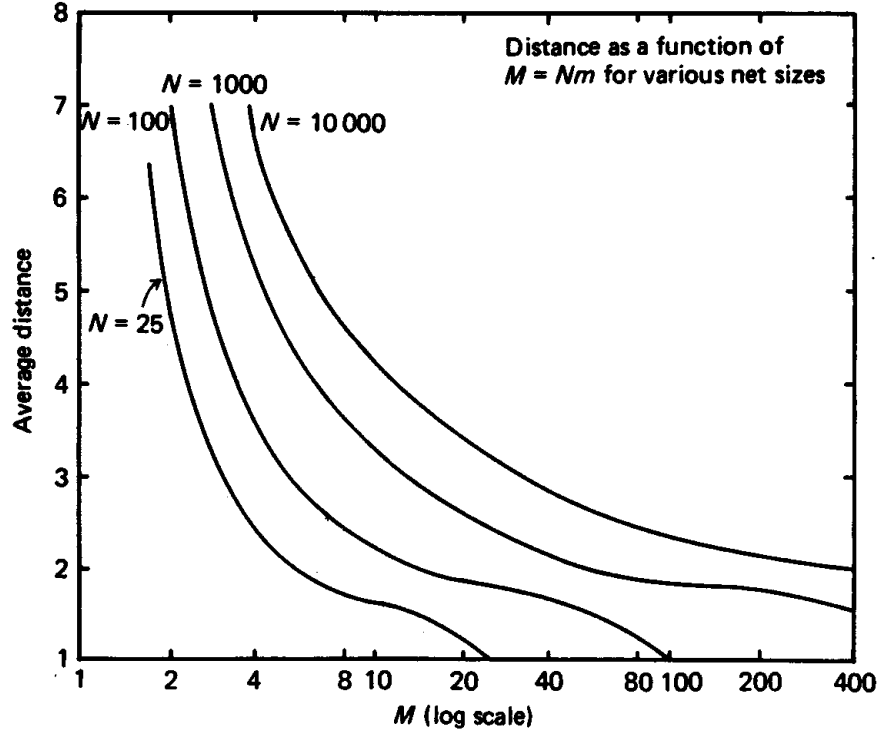


Figure 2. Average distance $\langle d \rangle$ as a function of the mean number $M = Nm$ of outputs per neuron in an ensemble of randomly assembled networks for various sizes N .

Some results for the distance distribution and average distance of structured nets of the above types are shown in Figs. 3 and 4, in comparison with results (sample or ensemble) for random networks. The most salient conclusion to be drawn from these results is that the average distance values obtained for the structured nets lie remarkably close to the curve derived for the corresponding ensemble of random nets, even for cases in which there is significant local emphasis (as measured by l or inversely by L). This feature is characteristic of systems that have come to be known as “small-world” networks.

A second observation is that the results for sample random nets, while showing satisfactory agreement with the statistical analysis of Section 2, nevertheless do evince appreciable deviations from the idealized, ensembled-averaged predictions for the distance distribution and average distance, owing to statistical fluctuations that increase with decreasing N (and fixed m) and with decreasing m (for fixed N). The fluctuations are not serious enough to obscure the small-world character of the networks studied, unless N or m is quite small. (More quantitative information on the effects of fluctuations may be found in Ref. [1].)

4. PROPERTIES OF COMPLEX NETWORKS

Let us continue to broaden our view and identify certain defining characteristics of complex networks, of whatever nature, made up of nodes or vertices connected by links or edges. We may consider networks with undirected edges, in which case a link

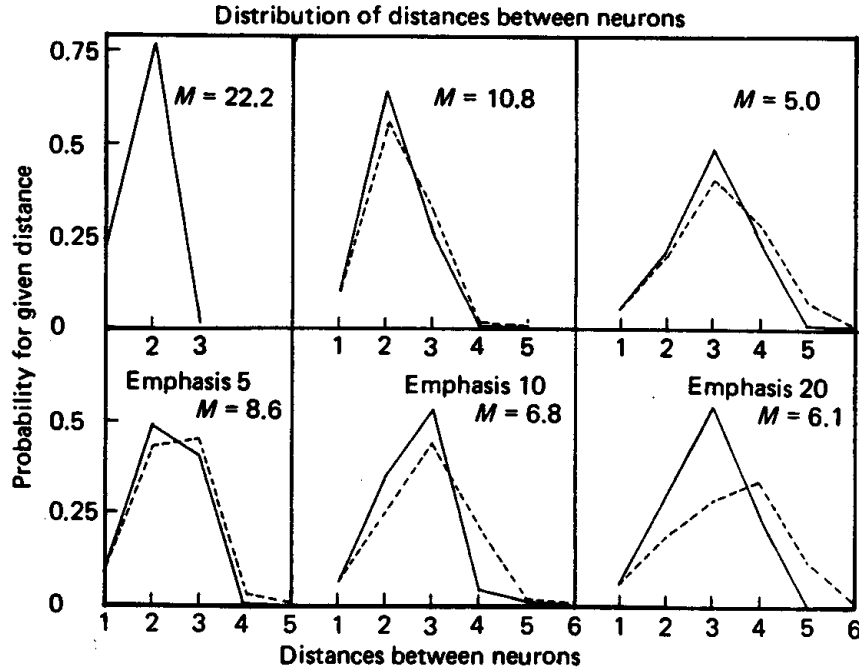


Figure 3. Dashed curves in upper triptych: actual distance distributions in three sample random networks of size $N = 100$ and respective connection probabilities $m = 0.222$, 0.108 , and 0.050 . Dashed curves in lower triptych: actual distance distributions in three sample networks of size $N = 100$ and respective average connection probabilities $m = 0.086$, 0.068 , and 0.061 , formed into $g = 4$ equal-sized netlet substructures with the indicated values of local emphasis l . Solid curves: ensemble-averaged statistical predictions for randomly assembled networks, determined from the recursion relations (6). The average number M of outputs per neuron is given for each case. The dashed and solid curves are practically coincident in the random case with $M = 22.2$.

between node j and node i establishes a one-step path from i to j as well as from j to i . This is the more common case studied in the literature. Alternatively, we may consider networks with directed, or oriented, edges; a connection directed from j to i does not imply the existence of a one-step path from i to j . This is the case relevant to neural networks. Chemical synapses are indeed directional connections; however, gap junctions (electrical synapses), which work both ways, are also present in neural systems. A gap junction between neurons i and j may be regarded either as a single undirected link or as a pair of reciprocal directed connections $j \rightarrow i$ and $i \rightarrow j$. The latter interpretation is the appropriate choice within our treatment of the model neural networks considered in the preceding sections, as well as for the nervous system of *C. Elegans*.

Def. The distance d_{ij} between two nodes j and i is the number of links or edges along the shortest path connecting them, with proper observance of the orientation of edges in the directed case.

Def. The average distance $\langle d \rangle$ is the mean of d_{ij} over all pairs ij (noting that ij and ji are regarded as different pairs in the case of directed links).

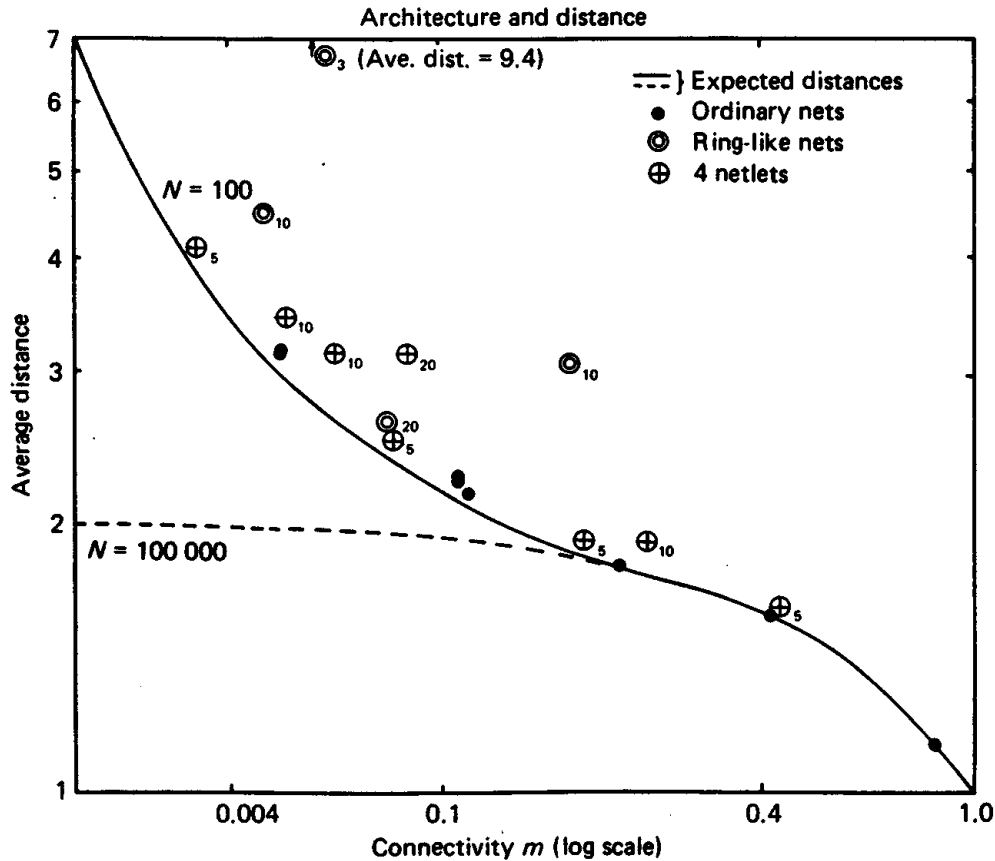


Figure 4. Average distances in randomly constructed sample nets (“ordinary nets”) and in structured networks of the two types considered (group of weakly connected netlets; ring networks) are compared with ensemble-averaged distances predicted by the formulas (6)-(7) for the purely random case (“expected distances”), across the full range of the connectivity parameter m . The solid curve and the symbols for sample networks correspond to $N = 100$; the dashed curve, calculated for $N = 10^5$, is included for reference. The numbers attached to the symbols for the overtly structured networks denote the local emphasis measure L or l , as appropriate.

Def. The tendency of the nodes in a network to form clusters or cliques is frequently measured by the *clustering coefficient* C , also known as the “fraction of transitive triples.” To find C , count the number of cases in which two nodes i and j that are neighbors of one another (i.e., one step apart) have a chosen node k as common neighbor. Divide by the number of directly-connected pairs i, j , thereby forming the clustering coefficient specific to node k . Then average over all k to obtain the overall clustering coefficient C . By choice (i.e., by definition), the direction of a link is irrelevant in forming the $\langle i, j, k \rangle$ triangles. For *random* networks, C coincides with the connection probability m (ignoring fluctuations).

Def. For the case of undirected edges, the *degree* of a node is the number of edges attached to it. In the case of directed edges, we define the *input degree* [output

degree] of a node as the number of incoming [outgoing] connections to [from] the node.

Def. The *degree distribution* is defined by the probability $P(\mu)$ that a randomly chosen node i has exactly μ edges (undirected case). Extension to the case of directed edges is transparent.

For a random net of N nodes with connection probability m , the degree distribution is binomial,

$$P(\mu_i = \mu) = \binom{N-1}{\mu} m^\mu (1-m)^{N-1-\mu}. \quad (8)$$

The distribution therefore becomes Poisson in the limiting regime where N is asymptotically large and Nm remains finite.

In 1998, Watts and Strogatz [2] introduced a one-parameter model of complex networks, which, by tuning one parameter, allows one to study the transition from complete order to complete randomness. Their strategy was to start with order and then randomize. The specific algorithm they adopted runs as follows:

- (i) Begin with a regular lattice of ring structure containing N nodes, each node being connected to its nearest K neighbors on the the ring. It is assumed that $N \gg K \gg \ln N \gg 1$ so as to ensure that the network is sparse yet always connected.
- (ii) Select a node on the ring and consider the link that connects it to its nearest neighbor, reckoned clockwise around the ring. With probability p , detach that link from the neighbor and reconnect it to a node chosen uniformly from all the nodes in the ring, subject to the provision that self-connections and duplicate connections are forbidden. Repeat this process $N-1$ times, working clockwise around the ring. In this way, some $pNK/2$ long-range connections replace the same number of short-range ones.

This model is strongly reminiscent of the ring model introduced in Ref. [1] and described in Section 3. However, there are some important differences. For one thing, the authors of Ref. [1] adopted a strategy opposite that of Watts and Strogatz, proceeding instead from randomness to order. For another thing, Watts and Strogatz do not assign directions to edges, whereas oriented links are intrinsic to the work of Ref. [1] as well as the present study.

5. WHAT IS A “SMALL-WORLD” NETWORK?

The *Nature* article [2] of Watts and Strogatz stimulated world-wide interest in the “small-world” phenomenon and a rapidly proliferating literature on this subject [3,4]. These authors found that realizations of their model could retain significant and discernible order, yet display “small-world” behavior in the sense that the average distance between nodes remains small, as in a random network. (We note that the studies of Ref. [1], carried out a decade earlier, gave evidence for the same phenomenon.)

Divorcing the discussion from any particular model, a small-world network is in general characterized topologically by (a) short separations between arbitrary nodes, as in the case of a random graph, together with (b) a high degree of local connec-

tivity, as in a regular lattice. These seemingly contradictory properties are reflected quantitatively in a relatively small average distance $\langle d \rangle$ and a relatively large clustering coefficient C (or large local emphasis). The degree distribution can show marked differences from a Binomial or Poisson distribution (e.g., a slower tail); on the other hand it can exhibit near-random behavior in some examples.

Actually, one of the first to draw attention to the small-world phenomenon in the modern era was the social psychologist Stanley Milgram [7], whose remarkable experiment revealing the short mean distance in the web of human acquaintances was the origin of the pithy adage “six degrees of separation,” also used as the title of a popular Broadway play by John Guare. Other examples include [4]: the world-wide web (site level); the internet (domain level); the network of movie actors linked by shared billing; co-authorship of scientific papers; genealogical networks; the substrate and reaction graphs of *E. Coli*; food webs in ecological systems; co-occurrence of words in text; networks of synonyms; the power grid of the U.S.; and the nervous system of *C. Elegans*. It seems reasonable to add the human brain to the list, but the analysis of its topological properties presents an awesome task at this stage of our knowledge.

6. C. ELEGANS: WHY STUDY THIS LOWLY WORM?

Let us turn now to an application of the concepts and analysis of Sections 2–4 to a real biological neural network.

First, what is *C. Elegans*? It is a tiny animal belonging to the *Phylum Nematoda*, which consists of the roundworms and threadworms, these being smooth-skinned, unsegmented worms having a long body of cylindrical shape, tapered at its ends. The Phylum includes worms that are free-living or parasitic, and worms that live in the water or on land. Our worm, *C. Elegans*, is not a parasite; it lives freely in the soil and feeds off bacteria. It is of no economic importance to mankind, and it does us no harm, but its conceptual importance is large. (This and other descriptive material on *C. Elegans* is drawn from a variety of sources on the world wide web [8-15], as well as the monograph of Achacoso and Yamamoto [16].)

The organism is only about 1 mm in length. It can be cultivated in a petri dish sprinkled with bacteria. Conveniently, it is transparent, so one can view its organs through a microscope. Its lifespan is normally only 2–3 weeks. (A curious exception is the *dauer phase*: the worm goes into a dormant larval state when food is scarce; with its mouth sealed shut, it can live for months and then revive.) There are two sexes – no, not the usual male and female, but rather a male (which is very rare, accounting for only one birth in a thousand) and a hermaphrodite. The hermaphrodite has both eggs and sperm and can reproduce either on its own or by mating with a male. Something like 300 baby worms are born after each sexual encounter (solo or pairwise). The reproductive system takes up much of the volume of the animal.

For the purposes of study, it is a remarkable and important fact that all hermaphroditic individuals are effectively identical, cell by cell. This version of the worm has 959 somatic cells, of which 302 – all identified, studied, and labeled – are neurons,

these being accompanied by 56 glial support cells. (In case you are curious, the male has a larger nervous system, with 381 neurons supplemented by 92 glial cells; this in spite of the fact that it is much easier for the male to find an hermaphrodite than vice versa.)

All very amusing, but why would biologists want to study this lowly worm with such intensity? The answer is that *C. Elegans* may be the simplest and most primitive organism that shares many of the essential biological characteristics of the human animal. For one thing, it is conceived as a single cell and undergoes a complicated process of development, embryonic cleavage and morphogenesis being followed by growth to the adult form. Second, it possesses a nervous system including a “brain” (identified with the circumpharyngeal nerve ring). It even shows adaptive behavior attributable to a simple learning mechanism. As already mentioned above, it manufactures both sperm and eggs, and it can reproduce by mating. After birth, the worm gradually ages, eventually weakens, and dies.

While exhibiting all these familiar phenomena, *C. Elegans* is only about 1 mm long and can be cultivated and studied like a microorganism. As a biological preparation, it offers a splendid compromise between complexity and tractability. Very significantly, it is the first organism for which a complete genomic sequencing was achieved (specifically, by a team including Washington University researchers). The importance of *C. Elegans* to biology was celebrated in the 2002 Nobel award in Physiology or Medicine to Sydney Brenner, H. Robert Horvitz, and John E. Sulston [10,11]. It was Brenner [11] who championed *C. Elegans* as a fundamental prototype for multicellular organisms.

Our interest is obviously focused on the nervous system of *C. Elegans*, especially the nature of its connectivity. The neural structures in the worm foreshadow those of higher animals, both in configuration and function. One finds a collection of sense organs in the head that respond to taste, smell, touch, and temperature. The creature has no eyes, but does show a slight response to light. There are 81 muscle cells stimulated by some subset of the 302 neurons making up the nervous system. An elegant swimming motion is achieved through the coordinated action of four longitudinal bands of muscle paired sub-dorsally and sub-ventrally (“dorsal” refers to the back or upper side of an animal, “ventral” to its belly).

Although it has only a small number of neurons, *C. Elegans* possesses a rich collection of neuron types. As intimated above, and according to “The Mind of the Worm” [13],

Each neuron has a unique combination of properties, such as morphology, connectivity and position, so that every neuron may be given a unique label. Groups of neurons that differ from each other only in position have been assigned to classes. There are 118 classes that have been made using these criteria, the class sizes ranging from 1 to 13.

The 302 neurons interact at some 5000 chemical synapses (based on asymmetric transfer of information through transmission of neurochemicals from presynaptic terminals to receptor sites on the postsynaptic cell), at more than 600 gap junctions (where the transmission is electrical and presumed to be symmetric), and at about 2000 neuromuscular junctions (where muscle actions are stimulated by motor neurons). Most gap junctions are found between neurons of the same class; this contrasts

with the case of chemical synapses, which rarely join members of the same class. The system of neurons is subserved by 56 non-excitabile glial cells that provide support while separating nerve cells from each other and from non-neural body cells.

Broadly speaking, neurons may be divided into three categories: (i) *sensory neurons* (providing input to the nervous system), (ii) *interneurons* (which modulate the flow of information internally), and (iii) *motor neurons* (which contact muscle cells and generate action corresponding to the computations that have been performed by the system based on its input). These three types account respectively for (roughly) one-third, one half, and one-quarter of the excitable neural cells.

The basic body plan of the organism involves two concentric tubes: an outer tube containing the cuticle, hypodermis, neurons, and muscle cells, and an inner tube comprised of a pseudocoelom space filled with intestines and gonads. Structurally, the nervous system divides into two parts, the pharyngeal nervous system (consisting of 20 neurons for which the connectivity data is not yet fully available) and the central nervous system (consisting of the remaining 282, whose connectivity is well mapped). The pharyngeal nervous system belongs to the pharynx, a neuromuscular pump in the mouth of the worm that is responsible for feeding behavior. It is largely independent of the central nervous system and therefore will be largely irrelevant to our investigation. (A pair of cells in the central nervous system, labeled CANL and CANR, do form connections with the pharyngeal system.)

The population of neurons in the central nervous system is organized into nine distinct ganglia, which contain afferent collectors and efferent senders of information. Between the ganglia are cells with relatively few connections, thought to act as logical gates [16]. Each ganglion is a “hub” for processing information relevant to a certain function or functions. Some collect data from many sources and convey output to only a few sources; and vice versa. The nerve tracks run in parallel “process bundles” with little branching. (Here, “process” is a biological term synonymous with “axon,” the extended fiber integral to a nerve cell that functions as a cable for electrical signal transmission.) The primary process bundle, called the ventral cord, is analogous to our spinal cord. Others include the nerve ring and the dorsal and lateral processes. The nerve bundles run antiposteriorly from the nerve ring and/or circumferentially. Synapses occur “en passant” in regions of dense neuropile. On average, each neuron is connected to about half of its *physical* neighbors; of course, it may also form connections with neurons at considerable physical distances.

For many purposes, it is correct to say that the neural connectivity (the “wiring diagram”) of *C. Elegans* has been completely mapped out experimentally. Basically, we know who talks to whom (i.e., which neuron communicates with which), and often we know through how many channels (i.e., through how many synaptic contacts) the conversation is conducted. On the other hand, the individual strengths of the many synaptic interactions remain largely unknown. Here we are limiting our attention merely to the list of postsynaptic neurons contacted by any given presynaptic neuron, in other words, to the connectivity (or adjacency) matrix (V_{ij}).

We shall not be concerned with the multiplicity of synaptic contacts of one cell upon another; rather, we only care whether such contacts are – or are not – made. Nevertheless, some comments on the issue of multiplicity may be useful for clarification. Physically, an *individual* synaptic contact involves the conjunction of an

axonal digit of the presynaptic cell with the dendritic tree or soma of the postsynaptic cell. In the *C. Elegans* system, multiplicity is found to range between 1 and 19, with an average of five [16]. Multiplicity might serve as an appropriate measure of the strength or weight of the connection from the presynaptic to the postsynaptic neuron. It can be difficult to determine, since postsynaptic partners of a given presynaptic neuron may be far away. Moreover, the presynaptic cell may well deliver different neurotransmitter chemicals at its different axon terminals, which may contain neurotransmitter vesicles of different sizes.

7. NEURAL CONNECTIVITY OF C. ELEGANS

Our study of *C. Elegans* as a potential small-world network is based on connectivity data taken from the Wormbase web site [14], and more specifically from the data on the pages [15]. This source provides us with complete connectivity information for a subset of 180 nerve cells considered to form the “main” nervous system. The data takes account of directed as well as reciprocal connections and therefore allows us to construct a connectivity matrix (V_{ij}) for this primary collection of neurons. Of the remaining 122 cells of the full nervous system, some 85 are motor neurons, which by definition have terminals outside the nervous system proper and thus might sensibly be omitted in an analysis that acknowledges the possible directionality of connections. Also omitted is the essentially independent subsystem of 20 pharyngeal neurons. The network selected for study involves 2406 outgoing connections from its neurons and 2403 incoming connections to its neurons. There are 340 gap junctions and 222 reciprocal synapses (counting gap junctions separately), for a total number of 562 bi-directional connections. If the directionality of connections is not distinguished, the number of edges is 1836.

Figure 5 shows the distance distribution $\rho(d)$ derived from repeated multiplication of the empirically determined matrix (V_{ij}) by itself. This matrix has a total of $180 \times 180 = 32,400$ elements and allows for the same number of connections. However, only 2406 of these elements are nonzero, corresponding to the existing interneuronal connections; this density of connections corresponds in turn to a connection probability m of 0.0742. Also shown in Figure 5 is the distance distribution for a random network with the same connection probability. We observe that the actual distribution is somewhat broader, although evaluation of the average distance $\langle d \rangle$ based on this distribution gives a value 2.50, not much larger than the result 2.26 obtained for the random case from Eqs. (6)–(7). The standard deviation associated with the former mean value is 0.732, i.e., the spread is significant.

The process of repeated self-multiplication of the connection matrix does not converge to a multiple that is free of zeros. This means that for some neuron pairs ij it is not possible to find a continuous directed pathway from one member of the pair to the other; in other words, our assumption that all interneuron distances are finite is not strictly correct. However, some violation must be expected, since any functioning nervous system must have an input interface and an output interface where its neurons interact with their environment (whether internal or external to the organism). For example, all inputs to a sensory system could originate from the

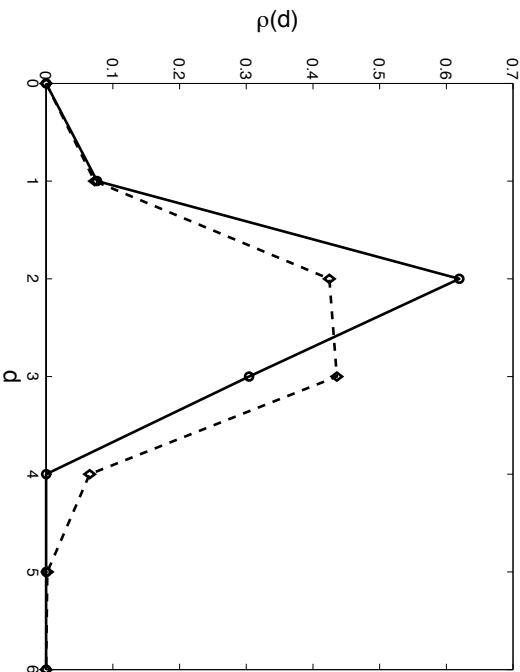


Figure 5. Distance distribution $\rho(d)$ versus distance d for the 180-neuron subnet of the nervous system of *C. Elegans* (diamonds and dashed line). Plotted for comparison is the average distance distribution obtained from Eqs. (6) and (7) for an ensemble of random networks constructed using the same connectivity parameter m (circles and solid line).

visual environment. At any rate, the violation observed in the present case is not very serious. Recall that in the 180-neuron system, there are 2406 nonzero elements in the connection matrix (V_{ij}). Raising (V_{ij}) to successive integral powers, saturation is reached at the fifth power, in the sense that no further matrix elements change from zero to nonzero values upon forming any higher power. A residue of less than 400 nonzero elements remains, implying that more than 98% of the 32,400 ordered (i, j) pairs are actually connected by some admissible, directed pathway.

Figures 6 and 7 provide data on two degree distributions associated with the chosen 180-neuron network of *C. Elegans*. Figure 6 refers to the distribution in the number of efferent (outgoing) connections of neurons of the system; and Figure 7, to the distribution in the number of afferent (incoming) connections. The average number of input connections per neuron is 13.37, and the number of output connections per neuron is 13.35. These numbers need not coincide, since some connections may terminate outside (or originate from outside) the 180-neuron network. At any rate, we may infer an empirical value $m = 0.0742$ for the connection probability introduced in Section 2.

For each of the degree distributions derived from the *C. Elegans* data, comparison is made with the corresponding Poisson distribution predicted from Eq. (8) for a random network at large N but finite Nm . Substantial deviation from the Poisson form is observed in both cases; in particular, the fall-off at large degree appears to be slower than exponential.

The clustering coefficient C determined from the connectivity data on the 180-neuron system is found to be 0.299, a value considerably larger than the result 0.0742 for the corresponding random case. Of the three broad types of neurons alluded to above, the *sensory neurons* have, on average, the highest clustering

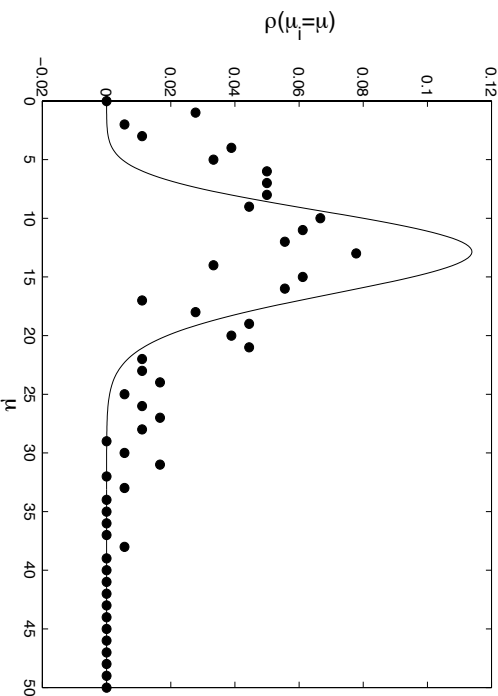


Figure 6. Degree distribution $\rho(\mu_i = \mu)$ of the *outgoing* connections of an arbitrary neuron i , versus the number μ of efferent connections emanating from the neuron. The corresponding Poisson distribution, expected for a random network in the limit of large N but finite M , is plotted for reference (solid curve).

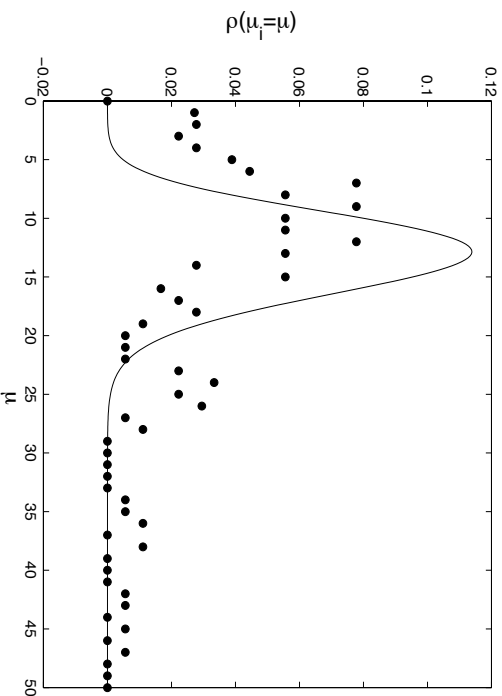


Figure 7. Degree distribution $\rho(\mu_i = \mu)$ of the *incoming* connections to an arbitrary neuron i , versus the number μ of afferent connections impinging on the neuron. The corresponding Poisson distribution, expected for a random network in the limit of large N but finite M , is plotted for reference (solid curve).

coefficients, the average over all cells of this type being 0.332. The largest clustering coefficient found was 0.6, for the sensory neuron named ASSJR [15]. The strong clustering of sensory neurons is not unexpected, in view of the rapid diffusion of sensory information coming from the environment, which requires a high density of connections with the affected ganglionic cells.

The clustering coefficients for *interneurons* and *motor neurons* present in the 180-neuron system are somewhat smaller, with the respective average values 0.278 and 0.275. It is understandable that these type of neurons should participate in

a smaller fraction of transitive triples. Roughly speaking, the set of interneurons serves collectively as the switchboard for communication within the worm’s nervous system, providing for both positive and negative feedback. Although these cells have many neighbors, they connect preferentially with remote parts of a nerve bundle, or with cells in an entirely different neighborhoods. In fact, the interneuron RMGR has the lowest coefficient (0.154) of any cell studied. Turning to the motor neurons, we need merely observe that their primary connections are made outside the nervous system. (It was for this very reason that not many motor neurons were included in the 180-neuron system chosen for “small-world” analysis.) Their function is to transmit signals to the muscles, usually those on the opposite side of the body, by means of axon fibers that pass through tiny holes in a ridge of skin-type cells that line the ventral side of the creature. We must note, of course, that since the bulk of motor neurons were disregarded in our study, the clustering coefficients we have derived for them may not be representative.

Setting aside the relatively minor differences in local connectivity found for the three categories of neurons, it is well to emphasize the high *average* clustering coefficient determined for the 180-neuron network chosen for investigation. This finding is in complete accord with the following observation quoted from “The Mind of the Worm” [13]:

One of the striking features of the connectivity diagrams is the high incidence of triangular connections linking three classes. These structures may occur frequently as a consequence of the organization of the neuropile. A typical neuron in C. elegans is accessible (i.e. adjacent) to a fairly limited subset of the total complement of neurons but is fairly highly locally connected within this subset. Thus, if a neuron has synaptic contacts with two partners, these two partners must be neighbours to the neuron and therefore are likely to be neighbours themselves. It is therefore quite probable, given the high level of local connectivity, that there will be a synaptic contact between them, which will close the triangle. The abundance of triangular connections in the nervous system of C. elegans may thus simply be a consequence of the high levels of connectivity that are present within neighbourhoods.

8. CONCLUSIONS

Utilizing connectivity data available on the world wide web, we have carried out an analysis of the complex network represented by the nervous system of the nematode *C. Elegans* based on the concepts and techniques developed and reviewed in Refs. [1,4] and presented in Sections 2–4. The results of a similar analysis of the worm’s nervous system were reported in the path-breaking *Nature* article [1] of Watts and Strogatz (WS). Both treatments offer clear and convincing evidence for the small-world character of this fundamental biological nerve network. Nevertheless, there are some key differences in the two treatments that deserve comment.

One significant difference is that the present analysis takes into account the intrinsic directionality of chemical synapses, whereas WS suppress directionality and regard the system as an undirected graph. Clearly, our analysis is more realistic in

this sense. On the other hand, WS consider a larger portion of the full neural net of *C. Elegans*, including 282 neurons compared to our 180. Different data sources are chosen for the two studies; in particular, our data has been taken from the updated Wormbase on-line resource [15], whereas WS use the connectivity matrix provided in the invaluable book by Achacoso and Yamamoto [16]. The network analyzed by WS omits only the 20 neurons of the pharyngeal nervous system, whereas ours deletes those and most of the motor neurons of the central nervous system. On this score, the WS choice might be viewed as the more realistic.

Remarkably, the two analysis are in rather close agreement in terms of the essential measures that signal small-world behavior. For the average distance we find 2.50, compared to the WS value of 2.62 (and the random-network value 2.26). For the average clustering coefficient, our result of 0.299 is to be compared with their 0.28 (and our random-network value 0.05). One would naturally expect an increase in average distance when the directionality of connections is acknowledged, so the very small difference between the $\langle d \rangle$ values found in the two studies – with our result actually being the *larger* – is perhaps surprising. On the other hand, the expected increase may well be largely compensated by the omission from our study of most of the motor neurons, which are by nature “less connected” than the bulk of the other neurons in the central nervous system.

The results of a repetition of our analysis for the larger data set considered in Ref. [2] will be reported in a future article.

ACKNOWLEDGMENTS

This research was supported in part by the U. S. National Science Foundation under Grant No. PHY-0140316. JWC is grateful for travel support that enabled him to participate in CMT26, jointly provided to him by the U.S. Army Research Office–Research Triangle Park (through a grant to Southern Illinois University–Carbondale) and by Fundação Luso-Americana para o Desenvolvimento (FLAD). He is also indebted to FLAD and to the Fundação para a Ciência e a Tecnologia (FCT) for financial support of his participation in Madeira Math Encounters XXIII at the University of Madeira, where part of this work was done. He thanks both Professor João da Providência and Professor Ludwig Streit for exceptionally generous hospitality.

REFERENCES

- [1] J. W. Clark, G. C. Littlewort, and J. Rafelski, in *Computer Simulation in Brain Science*, R. M. J. Cotterill, ed. (Cambridge University Press, Cambridge, 1988), pp. 104-118.
- [2] D. J. Watts and S. H. Strogatz, *Nature* **393**, 440 (1998).
- [3] D. J. Watts, *Small Worlds: The Dynamics of Networks Between Order and Randomness* (Princeton University Press, Princeton, NJ, 1999).
- [4] R. Albert and A.-L. Barabási, *Reviews of Modern Physics* **74**, 47 (2002).
- [5] P. Erdős and A. Rényi, *Publ. Math. Inst. Hung. Acad. Sci.* **5**, 17 (1960).

- [6] B. Bollobás, *Random Graphs* (Academic Press, NY, 1985).
- [7] S. Milgram, *Psychol. Today* **1**, 60 (1967).
- [8] <http://elegans.swmed.edu/>.
- [9] <http://www.biotech.missouri.edu/Dauer-World/Wormintro.html>;
<http://www.loci.wisc.edu/outreach/text/celegans.html>.
- [10] <http://www.nobel.se/medicine/laureates/2002/press.html>;
<http://almaz.com/nobel/medicine/2002a.html>.
- [11] <http://elegans.swmed.edu/Sydney.html>.
- [12] <http://www.wormatlas.org/>.
- [13] J. G. White, E. Southgate, J. N. Thomson, and S. Brenner, “The structure of the nervous system of the nematode *Caenorhabditis elegans*, *Phil. Trans. Royal Soc. London. Series B, Biol. Sciences* **314**, 1 (1986); this paper is easily accessed at <http://www.wormatlas.org/>.
- [14] <http://www.wormbase.org/>.
- [15] <http://www.wormbase.org/db/searches/neuron302.cgi>.
- [16] T. B. Achaoso and W. S. Yamamoto, *Neuroanatomy of C. elegans for Computation* (CRC Press, Boca, Raton, FL, 1992).