Phase transition in the economically modeled growth of a cellular nervous system

Vincenzo Nicosia[a,1], Petra E. Vértes[b,1], William R. Schafer[c], Vito Latora[a,d], and Edward T. Bullmore[b,e,f,2]

[a]School of Mathematical Sciences, Queen Mary University of London, London E1 4NS, United Kingdom; [b]Department of Psychiatry, Behavioural and Clinical Neuroscience Institute, University of Cambridge, Cambridge CB2 0SZ, United Kingdom; [c]Medical Research Council Laboratory of Molecular Biology, Cambridge CB2 0QH, United Kingdom; [d]Dipartimento di Fisica e Astronomia & INFN & Laboratorio sui Sistemi Complessi, Universita' di Catania, Via S. Sofia 61, Catania, Italy; [e]Cambridgeshire and Peterborough National Health Service Foundation Trust, Cambridge CB21 5EF, United Kingdom; and [f]GlaxoSmithKline Clinical Unit Cambridge, Addenbrookes Hospital, Cambridge CB2 0QQ, United Kingdom

Edited by Charles F. Stevens, Salk Institute for Biological Studies, La Jolla, CA, and approved April 1, 2013 (received for review January 12, 2013)

Spatially embedded complex networks, such as nervous systems, the Internet, and transportation networks, generally have nontrivial topological patterns of connections combined with nearly minimal wiring costs. However, the growth rules shaping these economical tradeoffs between cost and topology are not well understood. Here, we study the cellular nervous system of the nematode worm *Caenorhabditis elegans*, together with information on the birth times of neurons and on their spatial locations. We find that the growth of this network undergoes a transition from an accelerated to a constant increase in the number of links (synaptic connections) as a function of the number of nodes (neurons). The time of this phase transition coincides closely with the observed moment of hatching, when development switches metamorphically from oval to larval stages. We use graph analysis and generative modeling to show that the transition between different growth regimes, as well as its coincidence with the moment of hatching, may be explained by a dynamic economical model that incorporates a tradeoff between topology and cost that is continuously negotiated and renegotiated over developmental time. As the body of the animal progressively elongates, the cost of longer-distance connections is increasingly penalized. This growth process regenerates many aspects of the adult nervous system's organization, including the neuronal membership of anatomically predefined ganglia. We expect that similar economical principles may be found in the development of other biological or manmade spatially embedded complex systems.

C. elegans | connectome | generative model | neurodevelopment | spatial network

In the past decade or so, an abundance of studies have demonstrated that superficially diverse systems share important statistical properties (1–4). Movie costar networks, transport and communication systems, gene–gene interactomes, and many other natural and manmade systems have similarly complex topological features: they generally are efficient, small-world, modular systems with a greater-than-random probability of highly connected nodes or hubs. Many but not all of these topologically complex systems also are spatially embedded (5). For example, both the Internet and the World Wide Web have nontrivial topologies, but only the Internet is physically instantiated as a network in a metric space. Spatially embedded networks generally increase in cost with increasing distance of connections between nodes; and this cost constraint must be traded off against the functional advantages of topological features such as hub nodes, robustness, and high global efficiency, that may add value but at greater than minimal cost (6). Nervous systems share these general economical properties (7): at all scales of space and time and in all species, brain networks likely are both parsimoniously wired (8) and topologically complex (3).

This was first demonstrated in the case of the network of neurons comprising the nervous system of the nematode worm, *Caenorhabditis elegans* (9, 10). The brain of the hermaphrodite worm consists of 279 neurons (excluding the pharyngeal neurons) and is a sparse network (4% of maximum connection density), with most connections being between cells separated by short distances (<10% of the overall body length of the adult worm). Both sparse connection density and low connection distance are as expected by the operation of a parsimonious drive to minimize wiring cost. However, the wiring cost of the *C. elegans* connectome is not strictly minimized (11–13): further reductions of connection distance may be achieved by rewiring the biological network *in silico*, but only at the expense of increasing the shortest topological path between neurons (14), thus reducing the overall system efficiency. To put it another way, it seems there is a tradeoff between connection distance and topological efficiency in the organization of the adult nematode worm's nervous system. Topological efficiency is theoretically advantageous for globally integrated information processing and coordinated behaviors, but it is disproportionately expensive to engineer (7, 15). It is arguable that such economical tradeoffs between topological value and physical cost likely are a general selection pressure on formation of spatially embedded and topologically complex networks. More specifically, we predicted that economical principles applied dynamically over the course of developmental time (hundreds of minutes after fertilization) might provide a reasonable account of the emergence of multiple observed features of the growth and adult configuration of the nematode's nervous system.

Results

Here, we investigate the growth of the *C. elegans* connectome, from the moment of fertilization through hatching of the egg and larval elongation to adulthood (16, 17). Importantly, we note that the physical distances between neurons increase as a function of the increasing overall length of the worm's body as it matures (Fig. 1*A*). The cells of the adult nervous system are concentrated in the head and the tail of the worm, with a series of neurons running along the length of the body to innervate local muscle groups (the ventral cord). This system may be decomposed into 10 ganglia (or neuronal groups) based on anatomical properties (18, 19) (Fig. 1*B*). The birth times of each neuron tend to cluster in two time windows, separated by a "quiet" period that includes the time of hatching (800 min after fertilization) (Fig. 1*C*). The developmental changes in the number of nodes (N) and edges (K) in the network occur in the context of progressive elongation of the

Author contributions: V.N., P.E.V., V.L., and E.T.B. designed research; V.N., P.E.V., V.L., and E.T.B. performed research; P.E.V., W.R.S., and E.T.B. contributed new reagents/analytic tools; V.N., P.E.V., W.R.S., V.L., and E.T.B. analyzed data; and V.N., P.E.V., W.R.S., V.L., and E.T.B. wrote the paper.

Conflict of interest statement: E.T.B. is employed part time by GlaxoSmithKline and part time by the University of Cambridge.

This article is a PNAS Direct Submission.

[1]V.N. and P.E.V. contributed equally to this work.

[2]To whom correspondence should be addressed. E-mail: etb23@cam.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1300753110/-/DCSupplemental.

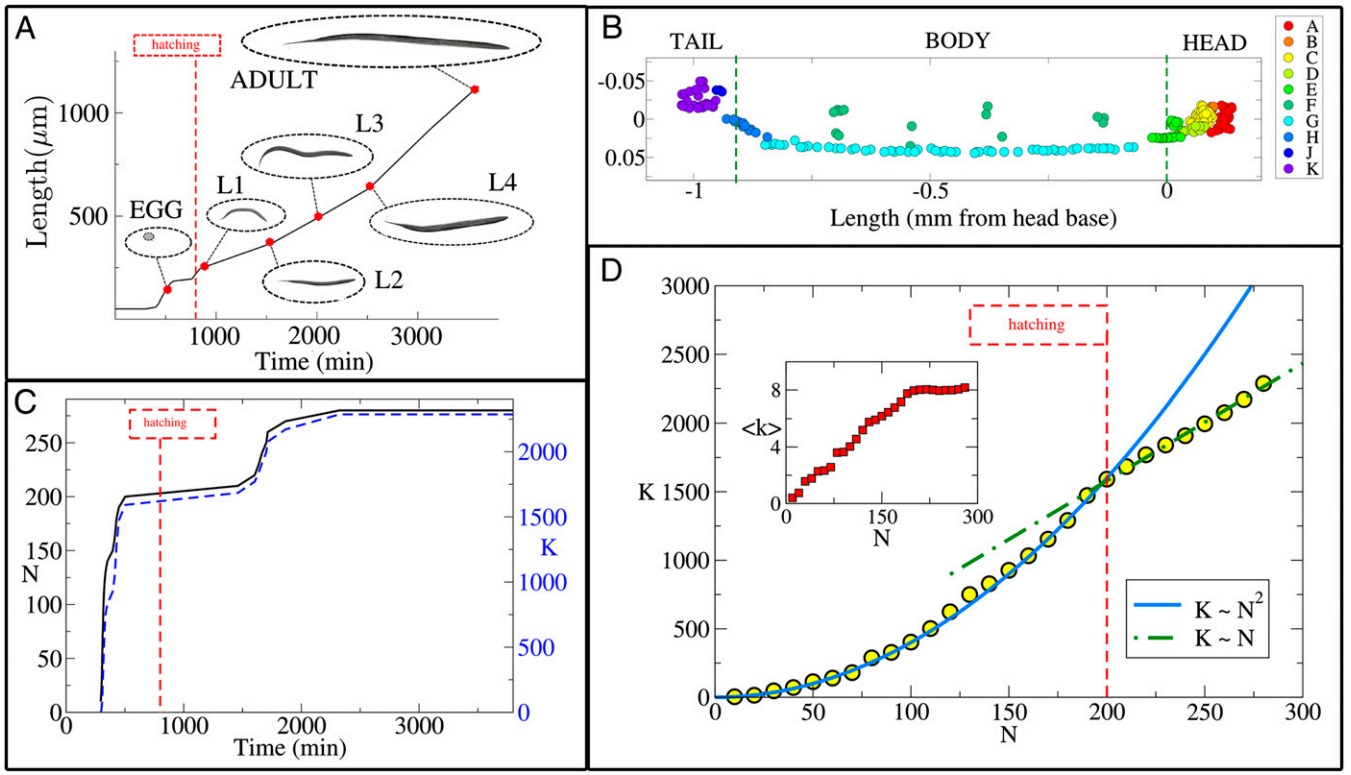

Fig. 1. Development of the *C. elegans* nervous system. (*A*) *C. elegans* reaches maturity roughly 63 h after fertilization. During this time, its body length increases from 50 μm to 1,130 μm (22–24). (*B*) In the adult hermaphrodite worm, more than 60% of the neurons are located in the head and about 15% are found in the tip of the tail (based on data modified according to ref. 17, axis arbitrarily centered such that the origin is at the base of the head). Neurons are colored by ganglion membership (16): anterior [A], dorsal [B], lateral [C], ventral [D], retrovesicular [E], ventral cord [G], posterior lateral [F], preanal [H], dorsorectal [J], and lumbar [K]. (*C*) The total number of neurons (*N*, solid black), and connections (*K*, dashed blue), grows rapidly between 250 and 500 min after fertilization. Another burst of neurogenesis is observed at the end of the L1 larval stage (using data from ref. 17). (*D*) Plotting the number of synapses as a function of the number of neurons (yellow ●) reveals the presence of a phase transition. Before hatching, *K* grows as N^2 (solid blue line), whereas after hatching, *K* grows linearly with *N* (dashed green line). (*Inset*) Plot of the average nodal degree vs. *N*.

worm's body, from less than 50 μm before hatching to more than 1 mm in the adult.

The two growth spurts in neuronal number, before and after hatching, are paralleled by a roughly synchronous increase in the total number of synaptic connections between neurons (Fig. 1*C*). However, the form of the relationship between *N* and *K* evidently is different before and after hatching, as shown in Fig. 1*D*. The initial increase in *K* is well described by a quadratic function of *N*, implying that the average node degree increases linearly as the network grows (Fig. 1*D*, *Inset*). Then, at $N \simeq 200$, hatching takes place, marking the metamorphic change of the worm from egg to larva. This event coincides with a discontinuous change in growth rules: after hatching, *K* increases linearly with *N*, so the average node degree remains constant. This experimental evidence suggests that sharp qualitative changes can indeed affect the growth rules governing the development and the formation of complex networks (1, 2, 20). In this case, the transition from one growth regime to another coincides with a metamorphic change of the worm, from egg to larva.

Although it is tempting to assume that it is a biological "trigger" or discontinuity associated with hatching that underlies the emergence of this biphasic growth curve, here we have assessed the ability

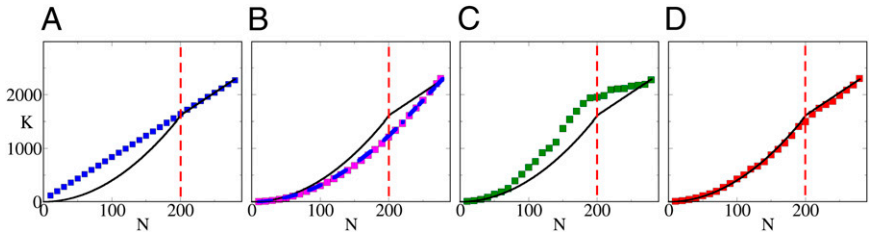

Fig. 2. Modeling network growth. (*A*) The linear preferential attachment model (BA, blue ■) fails to reproduce the biphasic growth observed (solid line). (*B*) In the BAG model (magenta ■) and the HAG model (dashed blue line), the average node degree increases linearly with the size of the network. (*C*) The ESG model (green ■) exhibits a biphasic behavior, yielding a transition from quadratic to nearly linear growth at $N \lesssim 180$, but fails to capture the details of the observed growth. (*D*) The ESTG model (red ■) accurately reproduces the details of the biphasic growth trajectory; for example, the inflection point of the modeled developmental curve corresponds closely to the moment of metamorphosis (hatching). The red dashed line in each panel indicates the number of nodes at the time of hatching ($N \simeq 200$). The SE of each growth curve is smaller than the size of the symbols used to plot it and is not reported.

of several simple and continuous models of network formation to reproduce this observed behavior without incorporating further biological detail (Fig. 2 and *Materials and Methods*). We deliberately decided to restrict ourselves to stochastic one-parameter models: first, because our aim was to isolate the fundamental ingredients that might be responsible for the observed discontinuous growth, and second, because as we show in the following, a one-parameter model indeed was enough to reproduce both the biphasic growth and many of the structural properties of the adult *C. elegans* neuronal network.

The first model we considered was the linear preferential attachment model, introduced by Barabási and Albert (BA) (20), which has been used successfully to describe the development of many different complex networks, from the World Wide Web to the Internet and citation networks. The BA model assumes that the growth of a network is driven only by its topological properties: specifically, newborn neurons are more likely to form connections to neurons that are already well connected. This model predicts a linear relationship between N and K, which closely matches the posthatching phase of worm brain development but does not provide a satisfactory fit to the prehatching phase. Conversely, the binomial accelerated growth (BAG) model, which assumes that the probability of a connection between a new neuron and any preexisting neuron is constant, predicts that K increases as a quadratic function of N (21). Similarly, we observe a quadratic dependence of K on N also in a modified version of accelerated growth [hidden-variable accelerated growth (HAG)], which additionally reproduces the node degree distribution of the adult worm. Accelerated growth models thus can reproduce the prehatching phase of the worm brain's growth but fail to accommodate the transition to linear scaling of K with N in the posthatching phase.

We found that economical tradeoff models that account for the spatial location of neurons while allowing some long-distance connections to high-degree nodes could reproduce biphasic growth more accurately. As a first approximation, we defined the economical spatial growth (ESG) model, which assumes that the probability of a connection forming between newborn neuron i and preexisting neuron j is a product of the degree of the jth node in the adult nervous system, and a decreasing exponential function of the Euclidean distance $d_{i,j}^{(\mathrm{ad})}$ between nodes i and j in the adult worm. Although the modeled growth exhibits two phases, the transition between quadratic and linear phases occurs before hatching. Therefore, we considered a more refined economical spatiotemporal growth (ESTG) model where $d_{i,j}$ is estimated by the Euclidean distance between neurons i and j at the time of birth of the newborn neuron, thereby adjusting for the fact that the connection distance between any pair of neurons will be shorter at earlier stages of development, before the worm becomes elongated. We extrapolated the position of each neuron during growth from its position in the adult worm, assuming that each neuron's position was shifted along the longitudinal axis in proportion to the overall changes in body length (Fig. 1*A*), which we collated from the literature (ref. 22 for the prehatching stage and ref. 23 after hatching) using a linear interpolation between larval stages (24). Although the penalty on connection distance remains fixed in this model, its effect on connectivity as a function of the overall scaling of the system is evolving dynamically. Indeed, the tradeoff between distance and topological degree is increasingly biased in favor of minimizing connection distance as development proceeds and the worm becomes longer overall. The model provides an excellent fit to the two observed scalings of K as function of N in the biological data, including a good approximation of the moment of hatching to the transition point from one growth regime to the other.

This suggests that the discontinuity in the growth curve is not explained by biological triggers related to hatching but instead is a consequence of the spatial properties of the system. In particular, the average distance of newly born neurons relative to all other neurons is much greater after hatching, so the distance penalty term begins to dominate the tradeoff embodied in the spatial growth rules. This is especially obvious in the ESTG model, in which the worm's elongation causes distances to increase in the interim between the two bursts of neurogenesis. Note, however, that a transition already is visible in the ESG model. This may be explained by noting that most neurons born after hatching are located along the body of the worm rather than in the head (*SI Appendix*, section S1 and Fig. S1), so the average distance between these newly born neurons and all others is again increased after hatching. We also tested the ability of other one-parameter models to reproduce the observed growth curve (*SI Appendix*, section S2); in particular, we tried to encode the cost of long connections through a power-law decay instead of an exponential one, but none of the alternative models could accommodate the abrupt change in the functional relation between K and N with the same accuracy obtained by ESTG (*SI Appendix*, section S4 , Table S-II, and Fig. S2).

The ESTG model also provides a good account of several other features of the adult nervous system's organization, including the statistical distributions of node degree, node efficiency, and edge length in the adult worm brain (Fig. 3). According to the results obtained through the computation of the symmetrized Kullback–Leibler divergence, ESTG is the model that most closely reproduces the distributions of node degree, edge length, and node efficiency (*SI Appendix*, section S5, Table S-III, and Figs. S3–S5).

Moreover, the model can provide a reasonable account of finer-grained details of the adult system, such as the anatomical variation in the average node degree and nodal efficiency along the length of the worm. Networks simulated by the model also had a mesoscopic structure closely resembling the pattern of clustered connectivity between neurons belonging to one of 10 ganglia previously defined on biological grounds. Neurons belonging to the same ganglion in the worm brain tend to have high connectivity with one another and relatively sparse connectivity to neurons in other ganglia (18, 19). This biological pattern and the neurons belonging to each specific ganglion were reproduced quite accurately by the ESTG model (Fig. 3).

Discussion

We have shown that a fairly simple economical model was adequate to account for many aspects of the spatial and topological development of the nervous system of the nematode worm, *C. elegans*. We describe this generative model as economical because it represents the formation of synaptic connections probabilistically as a tradeoff between topological value and wiring cost. More specifically, the model accommodates the potentially competitive tendencies of each new neuron to connect to topologically important hub neurons, which may be a long distance away (~1 mm), versus connecting only to neurons that are spatially adjacent (<0.1 mm), which will conserve wiring costs. Crucially, in estimating the connection cost between pairs of neurons, we have used prior data on the birth time of each neuron and the progressive elongation of the worm's body to estimate the distance between each pair of neurons at the time of synapse formation. This measure of connection cost was traded off against a topological bias (preferential attachment) for new neurons to connect to high-degree hub neurons of the adult nervous system. As the worm's body progressively elongates, the cost penalty predominates and long-distance connections, even to hub nodes, become less likely. This simple but unique model of a dynamically evolving economical tradeoff between cost and topology has allowed us to reproduce a phase transition in the growth of the *C. elegans* cellular connectome coinciding closely with the moment of hatching, or metamorphic transition from egg to larval stages of development. Dynamical economical growth processes also simulated several aspects of the configuration of the adult nervous system.

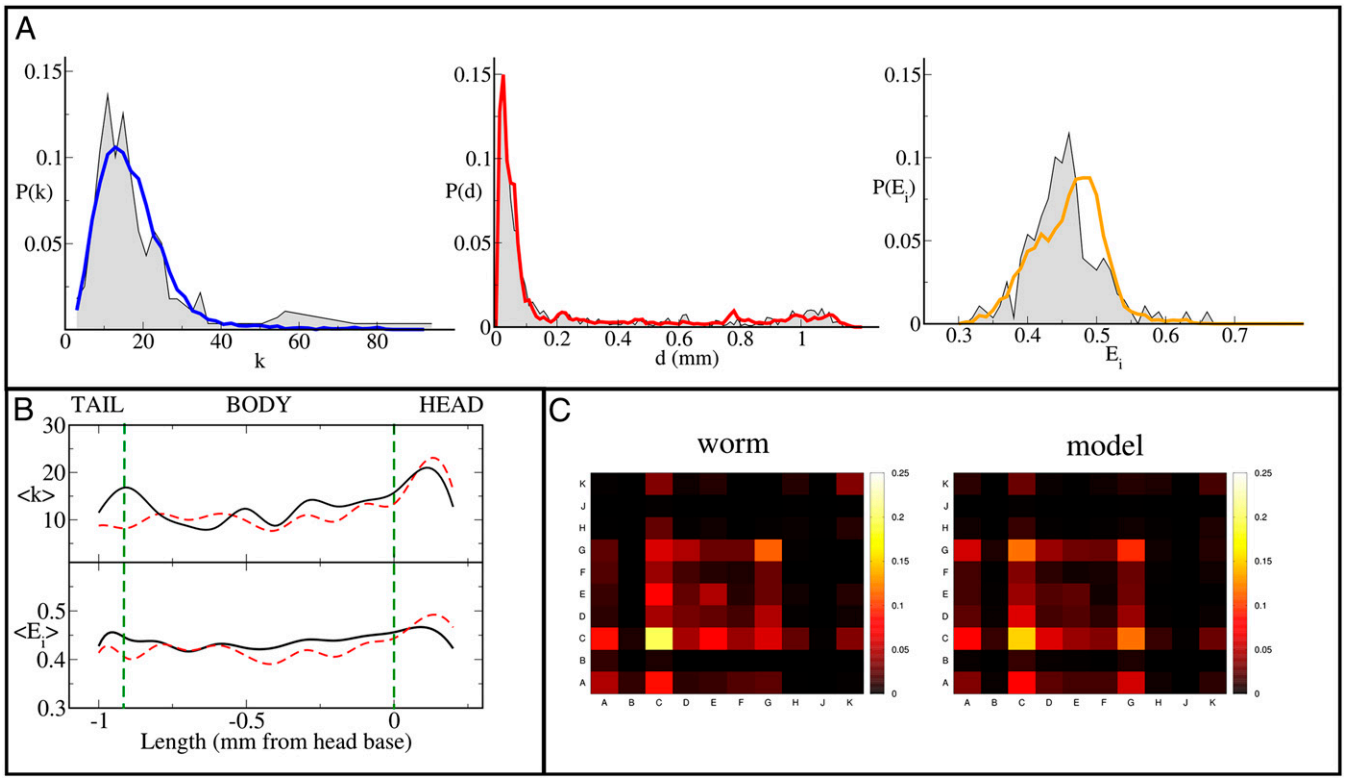

Fig. 3. Local and mesoscopic network structures. (*A*) The distributions of node degree (*Left*, blue), connection distance (*Center*, red), and node efficiency (*Right*, orange) of model-generated networks closely match those observed in the *C. elegans* neuronal network (shown in gray). (*B*) This panel shows how the average node degree (*Upper*) and the average node efficiency (*Lower*) vary along the length of the *C. elegans* body (solid black lines) and in networks generated using the ESTG model (red dashed lines). (*C*) Networks created using the ESTG model (*Right*) also reproduce the pattern of intra- and interganglia connections observed in *C. elegans* (*Left*). Brighter colors indicate higher connection density; letters A–K denote neuronal ganglia as defined in legend to Figure 1.

The principle that nervous systems conserve wiring costs dates back to the seminal work of Ramón y Cajal in the 19th century, and it has been experimentally validated and theoretically developed extensively since then. Many aspects of brain organization, ranging from the placement of neurons in the adult *C. elegans* nervous system (8) to the shape of dendritic trees (25) and the modular architecture of large-scale human brain networks (26), have been attributed plausibly to a parsimonious drive to minimize wiring costs. However, a strictly cost-minimal network would have a regular, lattice-like topology. Synaptic connections would be clustered between spatially and topologically neighboring neuronal nodes, with none of the long-distance axonal projections needed to mediate topologically efficient communication between widely separated neurons. However, this is not a recognizable description of nervous system topology. In many species, and at many scales of space and time, it has been found that brain structural and functional networks have a shorter average path length or greater efficiency than a regular lattice. Brain networks also consistently have nonregular properties, such as high-degree hubs in a fat-tailed degree distribution, and a modular community structure entailing long-distance intermodular connections between neurons in anatomically distributed modules. Many of these topological features are more than minimally expensive or incur a premium in wiring cost, but they may add value to the overall performance of the system. For example, high-degree hub nodes of the *C. elegans* nervous system include many of the so-called command interneurons that play a key role in the adaptive function of coordinated forward and backward movement of the worm (15, 16). Topological efficiency of human brain networks has been positively correlated with normal variation in IQ (more intelligent people tend to have more efficient structural and functional networks) (7, 27). Tradeoffs between cost and efficiency have been shown to be heritable properties of human brain networks derived from functional MRI (fMRI) data (28), and economical models of network formation can reproduce the (somewhat different) statistical properties of fMRI networks in both healthy adults and patients with schizophrenia (29). These and other observations support the general idea that nervous systems are selected to negotiate an economical tradeoff between wiring cost (usually measured by connection distance) and topological value (which might be measured by degree, efficiency, or several other network properties related to adaptive brain function).

Therefore, the basic principles of the economical model investigated here are not new to the neuroscience literature (7). However, there are several distinctive aspects of our results. Firstly, this work is an innovative demonstration that economical models can account for the growth of a nervous system described quite concretely and exactly at the cellular scale of synaptic connections between neurons. Many of the previous studies of economical tradeoffs in brain networks were based on analysis of statistical associations (so-called functional connectivity) between fMRI time series recorded at different spatial locations (28), or on analysis of large-scale axonal projections rendered by tractography algorithms applied to diffusion imaging data (30). Such human neuroimaging results indicate that economical principles may apply to network formation at macroscopic scales, but the neuronal substrate of networks based on imaging statistics remains unresolved. The demonstration here of economical principles applying to a connectome described with much greater precision at

a cellular scale somewhat validates the prior neuroimaging results. Moreover, it suggests that the same competitive selection criteria may inform nervous system formation over multiple spatial scales. Brain networks may have a scale-invariant or fractal economy of organization.

More broadly, these results are innovative in demonstrating directly how simple economical growth models can provide a reasonable account of complex growth curves, such as the nonlinear processes of nervous system maturation and metamorphosis, from egg to adult worm. Nematodes, like all members of the superphylum Ecdysozoa, develop through discrete stages (egg, several juvenile stages, adult) separated by molting events. The situation in *C. elegans* is most closely analogous to hemimetabolous insects (with "incomplete metamorphosis") because the juvenile stages resemble the adults apart from the absence of mating/reproductive structures. However, each molt may be considered metamorphic, with the L1–L2 and L4–adult molts, in particular, known to involve both the addition of new cells and formation of new synaptic links. The special significance of the egg–L1 transition perhaps has been less appreciated up to now, and as such represents a unique finding of this work.

Our more realistic modeling of connection cost, taking into account the changing spatial constraints during the growth of the system, also shines a different light on the many previous studies of connection cost (8, 11, 16) in this paradigmatic complex system. Further work is needed to test the hypothesis that the specific parameters of this model correspond to discrete molecular or genetic signals. It is imaginable, for example, that a penalty on long-distance connections might be biologically coded by the spatial gradient of an axonally attractive molecule diffusing from neurons, or that neurons destined to have high degree in the adult system express distinctive cell surface markers from birth that favor synaptic formation.

We have compared the performance of the dynamically evolving economical model with that of several other models and, as expected theoretically, found that simpler models based on preferential attachment rules could reproduce one of the two phases of network development (quadratic or linear) before or after hatching. However, only economical models that traded off connection distance with preferential attachment bias could reproduce both phases, and the timing of phase transition was reproduced accurately only by the dynamic linkage between interneuronal connection distance and progressive developmental elongation of the whole organism. For this reason, we consider that the modeling results affirm our hypothetical prediction that development of this cellular connectome can be accounted for by continual renegotiation of an economical tradeoff between connection cost and the formation of high-degree hubs. This affirmation is conditional on the caveat that not all possible models have been evaluated comparatively. It is possible that a better model, perhaps incorporating a few more relevant biological details (such as type of synapse, electrical or chemical), might be developed in the future.

We note that economical principles of network formation demonstrated here for the growth of the nervous system of the nematode worm are not necessarily limited to this system. Many other systems, besides brains, are both spatially embedded and topologically complex. We anticipate that economical growth models of the potentially changing tradeoffs between physical connection cost and topological value also may contribute to future understanding of the development and evolution of transport, computational, and infrastructural systems.

Materials and Methods

Data. We have used the most up-to-date map of the *C. elegans* connectome (16), consisting of 279 somatic neurons interconnected through 6,393 chemical synapses, 890 gap junctions, and 1,410 neuromuscular junctions. Because gap junctions often overlap with synapses and synaptic connections often are reciprocated, we have considered only the backbone network, in which all the synapses and gap junctions between each pair of neurons are represented by a single undirected edge, obtaining a graph with $N = 279$ nodes and $K = 2,287$ edges in total (neuromuscular connections were excluded). Information about the growth of the neuronal network, particularly the exact time of birth of each neuron, has been reconstructed from recent literature (17).

Linear Preferential Attachment. The BA model assumes that the growth of a network is driven solely by its topological structure, and produces random graphs with a power-law degree distribution $p_k \sim k^{-\gamma}$, where $\gamma \simeq 3$ (20). In the model, a new node is added at each time and is connected to m existing nodes. The probability for the new node i to be connected to an existing node j is a linear function of the degree k_j, namely

$$\Pi_{i\to j}^{BA} = k_j / 2K, \quad [1]$$

where K denotes the total number of links when the new node arrives. Because each node chooses m neighbors with which to connect, the total number of links increases linearly with the size of the network, and the average node degree remains constant.

Accelerated Topological Growth. Traditionally, network growth is said to be accelerated if the average node degree increases with the size of the network. Acceleration has been observed in many complex networks, and different models of scale-free networks with acceleration have been proposed so far (21). We have considered two different accelerated growth models. In the first model, called binomial accelerated growth (BAG), a new node i tries to establish a connection with each of the existing nodes, and a link to node j is created with probability p, namely

$$\Pi_{i\to j}^{BAG} = p. \quad [2]$$

The BAG model produces networks in which the number of links increases as the square of N. In fact, the expected number of links established when the network has N nodes is

$$K(N) = p\sum_{i=1}^{N}(i-1) = p\frac{N(N-1)}{2}. \quad [3]$$

The BAG model produces networks with a binomial degree distribution, because it is equivalent to an Erdös–Rényi random graph model, in which each of the $N(N-1)/2$ potential links appears with probability p (31).

We also introduced a second model of accelerated growth, called hidden-variable accelerated growth (HAG). In general, hidden-variable models produce networks with a prescribed degree distribution: the HAG model grows random networks having, on average, the same degree distribution observed in the adult *C. elegans* neural network. The model works as follows. We assign to each node j of the network, once and for all, a hidden variable h_j. In particular, we set $h_j = k_j^{\mathrm{ad}}$, where k_j^{ad} is the degree of node j in the adult worm. When a new node i arrives, it tries to establish a link with each of the nodes in the network, and a link to node j is created with probability

$$\Pi_{i\to j}^{HAG} = p\frac{h_j}{h_{max}}, \quad [4]$$

where h_{max} is the maximum of h_j over j and p is appropriately selected to reproduce the final number of links. It is possible to prove that the final degree of a node i over different network realizations is Poisson distributed around an average value equal to k_i^{ad}. Consequently, networks produced by HAG show an accelerated growth similar to that generated by the BAG model while also preserving the actual degree distribution of the *C. elegans* neural network.

ESG. To create networks embedded in Euclidean space (5, 6), we considered the economical spatial growth model, which is based on a tradeoff between the tendency to create topologically important connections to hubs and the physical distance between neurons. When a new node i arrives, it is placed in the position it occupies in the adult worm, and a link to each of the existing nodes is created with probability

$$\Pi_{i\to j}^{ESG} = \frac{h_j}{h_{max}} e^{-\frac{d_{ij}^{\mathrm{ad}}}{\delta}}, \quad [5]$$

where the values h_j are assigned as in the HAG model and δ is a parameter tuning the typical connection distance. Namely, the probability of creating a link exponentially decreases with the Euclidean distance d_{ij}^{ad} that separates

i and j in the adult worm, and is weighted by the hidden variable $h_j = k_j^{ad}$ (to preserve the actual degree distribution of the *C. elegans* neural network).

ESTG. The economical spatio-temporal growth model, using information about the length of the worm at different stages, takes into account the actual spatial position of each neuron while the worm grows over time. When a new node i arrives, it is placed in the position it occupies in the *C. elegans* neural network at time t, and a link to each of the existing nodes is created with probability

$$\Pi_{i\to j}^{ESTG} = \frac{h_j}{h_{max}} e^{-\frac{d_{ij}(t)}{\delta}}, \quad [6]$$

where the values h_j are assigned as in the HAG model and δ is a parameter tuning the typical edge length. Notice that the probability to establish a link depends on the time at which node i appears, because the distance $d_{ij}(t)$ depends on the relative positions of i and j, which change over time because of elongation of the worm's body. We considered the real length of the worm at each time, and we estimated the position of each node at that time using linear interpolation and assuming a uniform expansion of the worm along the longitudinal axis.

Parameter Tuning. The first requirement of any suitable model for the *C. elegans* neuronal network growth is to produce networks having $N = 279$ nodes and, on average, $K = 2,287$ edges, as observed in the adult worm. We used Monte Carlo simulations and iterative bisection to identify the interval in the parameter space for which the expected total number of edges $\tilde{K}$ of the generated networks was equal to $2,287 \pm 1\%$; see *SI Appendix*, section S3 for methodological details and the optimal parameter values for each of the eight models in *SI Appendix*, Table S-I.

Degree Distribution. Given an undirected graph $G(V, E)$ associated with the symmetric adjacency matrix $A = \{a_{ij}\}$, the degree of a node i is defined as the number of edges incident on i, and is denoted by $k_i = \sum_j a_{ij}$. The degree distribution $P(k)$ of the graph indicates, for each value of k, the probability of finding a node whose degree is equal to k.

Connection Distance Distribution. Given two directly connected nodes i and j of a spatially embedded network, we define the distance of the edge (i, j) as the Euclidean distance d_{ij} separating node i and node j. The distance distribution $P(d)$ is the probability of finding an edge whose distance is exactly equal to d.

Node and Graph Efficiency. Given an undirected and unweighted graph G, the efficiency of a node is defined as

$$E_i = \frac{1}{N-1} \sum_{\substack{j=1 \\ j\neq i}}^{N} \frac{1}{\lambda_{ij}}, \quad [7]$$

where λ_{ij} is the path length between node i and node j, measured as the number of edges in the shortest path connecting i to j. The smaller the λ_{ij}, the larger the contribution of node j to the efficiency of i. The efficiency of a graph is defined as the average efficiency of its nodes.

ACKNOWLEDGMENTS. V.N. and V.L. acknowledge the support of the EU Project LASAGNE, Contract no. 318132 (STREP). The Behavioural and Clinical Neuroscience Institute is supported by the Medical Research Council (United Kingdom) and the Wellcome Trust.

1. Albert R, Barabási A-L (2002) Statistical mechanics of complex networks. *Rev Mod Phys* 74(1):47–97.
2. Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang D-U (2006) Complex networks: Structure and dynamics. *Phys Rep* 424:175–308.
3. Bullmore E, Sporns O (2009) Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci* 10(3):186–198.
4. Barabási A-L (2012) The network takeover. *Nat Phys* 8:14–16.
5. Barthélemy M (2011) Spatial networks. *Phys Rep* 499:1–101.
6. Barthélemy M (2003) Crossover from scale-free to spatial networks. *EPL* 63(6): 915–921.
7. Bullmore E, Sporns O (2012) The economy of brain network organization. *Nat Rev Neurosci* 13(5):336–349.
8. Chen BL, Hall DH, Chklovskii DB (2006) Wiring optimization can relate neuronal structure and function. *Proc Natl Acad Sci USA* 103(12):4723–4728.
9. Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. *Nature* 393(6684):440–442.
10. Latora V, Marchiori M (2003) Economic small world behaviour in weighted networks. *Eur Phys J B* 32(2):249–263.
11. Pérez-Escudero A, de Polavieja GG (2007) Optimally wired subnetwork determines neuroanatomy of Caenorhabditis elegans. *Proc Natl Acad Sci USA* 104(43):17180–17185.
12. Pérez-Escudero A, Rivera-Alba M, de Polavieja GG (2009) Structure of deviations from optimality in biological systems. *Proc Natl Acad Sci USA* 106(48):20544–20549.
13. Sporns O (2011) *Networks of the Brain* (MIT Press, Cambridge, MA).
14. Kaiser M, Hilgetag CC (2006) Nonoptimal component placement, but short processing paths, due to long-distance projections in neural systems. *PLOS Comput Biol* 2(7):e95.
15. Towlson EK, Vertes PE, Ahnert SE, Schafer WR, Bullmore ET (2013) The rich club of the C. elegans neuronal connectome. *J Neurosci* 33(15):6380–6387.
16. Varshney LR, Chen BL, Paniagua E, Hall DH, Chklovskii DB (2011) Structural properties of the Caenorhabditis elegans neuronal network. *PLOS Comput Biol* 7(2):e1001066.
17. Varier S, Kaiser M (2011) Neural development features: Spatio-temporal development of the Caenorhabditis elegans neuronal network. *PLOS Comput Biol* 7(1):e1001044.
18. Arenas A, Fernández A, Gómez S (2008) A complex network approach to the determination of functional groups in the neural system of C. elegans. *Bio-Inspired Computing and Communication*, eds Liò P, Yoneki E, Crowcroft J, Verma DC (Springer, Berlin) pp 9–18.
19. Arenas A, Fernández A, Gómez S (2009) An optimization approach to the structure of the neuronal layout of C. elegans. *Handbook on Biological Networks*, eds Boccaletti S, Latora V, Moreno Y (World Scientific, London), Vol 10, pp 243–257.
20. Barabási A-L, Albert R (1999) Emergence of scaling in random networks. *Science* 286(5439):509–512.
21. Dorogovtsev SN, Mendes JFF (2001) Effect of the accelerating growth of communications networks on their structure. *Phys Rev E Stat Nonlin Soft Matter Phys* 63(2 Pt 2):025101.
22. McKeown C, Praitis V, Austin J (1998) sma-1 encodes a betaH-spectrin homolog required for Caenorhabditis elegans morphogenesis. *Development* 125(11):2087–2098.
23. Altun ZF, Hall DH (2012) Handbook of C. elegans anatomy. *WormAtlas*. Figure 6. Available at www.wormatlas.org/hermaphrodite/hermaphroditehomepage.htm.
24. Altun ZF, Hall DH (2012) Handbook of C. elegans anatomy, Figure 6. *WormAtlas*. Available at www.wormatlas.org/hermaphrodite/hermaphroditehomepage.htm.
25. Cuntz H, Forstner F, Borst A, Häusser M (2010) One rule to grow them all: A general theory of neuronal branching and its practical application. *PLOS Comput Biol* 6(8):e1000877.
26. Bassett DS, et al. (2010) Efficient physical embedding of topologically complex information processing networks in brains and computer circuits. *PLOS Comput Biol* 6(4):e1000748.
27. van den Heuvel MP, Stam CJ, Kahn RS, Hulshoff Pol HE (2009) Efficiency of functional brain networks and intellectual performance. *J Neurosci* 29(23):7619–7624.
28. Fornito A, et al. (2011) Genetic influences on cost-efficient organization of human cortical functional networks. *J Neurosci* 31(9):3261–3270.
29. Vértes PE, et al. (2012) Simple models of human brain functional networks. *Proc Natl Acad Sci USA* 109(15):5868–5873.
30. van den Heuvel MP, Kahn RS, Goñi J, Sporns O (2012) High-cost, high-capacity backbone for global brain communication. *Proc Natl Acad Sci USA* 109(28): 11372–11377.
31. Erdös P, Rényi A (1960) On the evolution of random graphs. *Publ Math Inst Hung Acad Sci* 5:17–61.

Supporting Information to
"Phase transition in the economically modeled growth of a cellular nervous system"

Vincenzo Nicosia,[1,*] Petra E. Vertes,[2,*] William R. Schafer,[3] Vito Latora,[1,4] and Edward T. Bullmore[2,5,6,†]

[1] *School of Mathematical Sciences, Queen Mary University of London, London E1 4NS, United Kingdom*
[2] *Department of Psychiatry, Behavioural and Clinical Neuroscience Institute, University of Cambridge, Cambridge CB2 0SZ, United Kingdom*
[3] *Medical Research Council Laboratory of Molecular Biology, Cambridge CB2 0QH, United Kingdom*
[4] *Dipartimento di Fisica e Astronomia & INFN & Laboratorio sui Sistemi Complessi, Università di Catania, Via S. Sofia 61, Catania, Italy*
[5] *Cambridgeshire and Peterborough National Health Service Foundation Trust, Cambridge CB21 5EF, United Kingdom*
[6] *GlaxoSmithKline Clinical Unit Cambridge, Addenbrookes Hospital, Cambridge CB2 0QQ, United Kingdom*

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Section S1. Location of neurons born after hatching

In Fig. S-1 we show the spatial configuration of neurons before and after hatching. Notice that the majority of the neurons born before hatching are concentrated in the head and in the tail region, while most of the neurons appearing after hatching are instead placed in the body to form the ventral cord. This explains the relative higher distance from newly added neurons to existing ones observed after hatching.

Section S2. Additional one-parameter models

We present here three additional growth models which have been tested during this study, namely the Simple Spatial Growth (SSG), Spatial Growth with Elongation (SGE) and Power-law Economical Growth (PEG). We also discuss their ability to reproduce the developmental growth of the *C. elegans* neuronal network, and we will compare them with the other five models described in the main text, i.e. Barabási-Albert (BA), Binomial Accelerated Growth (BAG), Hidden-variable Accelerated Growth (HAG), Economical Spatial Growth (ESG) and Economical Spatio-Temporal Growth (ESTG). Notice that all the models considered in this study have only one free parameter. Nevertheless some of these models, and in particular the ESTG, are exceptionally accurate at reproducing the structure and development of the *C. elegans* neuronal network.

Simple Spatial Growth (SSG). This model makes the assumption that upon arrival a new node i is placed in the same position at which it appears in the adult worm. Then, node i creates an edge to each of the already existing nodes j with probability:

$$\Pi_{i\to j}^{SP} = e^{-\frac{d_{ij}^{\mathrm{ad}}}{\delta}} \tag{S-1}$$

where d_{ij}^{ad} is the distance between node i and node j in the adult worm and δ is a parameter tuning the typical edge length. Since the connection probability decreases exponentially with the distance between nodes in the adult worm, the resulting networks exhibit very few medium- and high-distance links, which are instead relatively frequent in the real *C. elegans* neuronal networks.

* These authors have equally contributed to this work
† To whom correspondence should be addressed. Email: etb23@cam.ac.uk

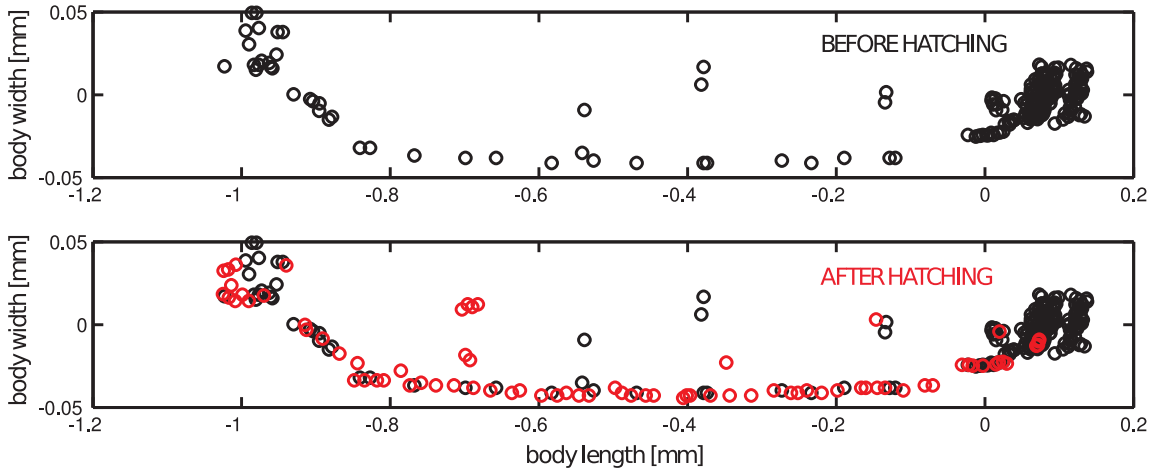

Figure S-1. **Position of neurons born after hatching.** The large majority of neurons born after hatching are located throughout the worm's body, while most of the neurons born before hatching are concentrated in the head and in the tail. The x-axis represents the distance in millimeters from the base of the head. Positive values indicate points in the worm's head, while negative values correspond to the body and the tail.

Spatial Growth with Elongation (SGE). This model uses information about the length of the worm at different stages. The node i arriving in the network at time t is placed in the position it occupies in the neural network at that time, and the probability for i to connect to an existing node j is defined as:

$$\Pi_{i\to j}^{SPE} = e^{-\frac{d_{ij}(t)}{\delta}} \tag{S-2}$$

where $d_{ij}(t)$ is the distance between node i and node j at time t and δ is a parameter. Notice that $d_{ij}(t)$ is a function of time, so that the probability to create an edge between a newly arrived node i and an existing node j depends on the time at which node i arrives in the network and on the relative positions of i and j at that time. This makes possible the creation of edges between nodes which are actually separated by a relatively large distance in the adult worm but have been closer in space in earlier developmental stages.

Power-law Economical Growth (PEG). This model implements a trade-off between the tendency to create edges to hubs and the relative distance of the nodes, and takes into account the elongation of the worm during development. Differently from the Economical Spatio-Temporal Growth model presented in the main text, in which the connection probability is a decreasing exponential function of distance, in this model the probability to connect to a distant node decreases as a power-law:

$$\Pi_{i\to j}^{PEG} = \frac{h_j}{h_{max}}\left[1-\left(\frac{d_{ij}(t)}{L_t}\right)^{\alpha}\right] \tag{S-3}$$

Here, h_j is the hidden degree of node j, which is set equal to the degree of node j observed in the adult worm, while h_{max} is maximum node degree in the adult neural network. As for the ESTG model, $d_{ij}(t)$ is the distance between node i and node j in the worm at time t. L_t is the total worm length at time t and α is the exponent of the power-law. Notice that the attachment probability Π^{PEG} approaches 0 when the distance $d_{ij}(t)$ is comparable with the length of the worm, while the hidden degree of the destination node plays a more important role if the two nodes are closer in space. Thanks to the preferential attachment term, based on the hidden degree of the nodes, this model tends to preserve the degree distribution of the original network.

Section S3. Parameter tuning

In this study we considered only one-parameter randomized growth models. In general, a randomized model generates an ensemble of graphs having certain characteristics. If the model has a tunable parameter, each value of the parameter generates a family of graphs sharing similar structural properties. For instance, the Binomial Accelerated Growth (BAG) model produces networks in which the number of edges grows quadratically with the number of nodes, but the expected number of edges in the final network, i.e. when the number of nodes is equal to $N = 279$, depends on the actual value of the attachment probability p.

Since a randomized one-parameter model generates a family of graphs for each value of the parameter, its ability to reproduce the structure of a given network cannot be assessed through a direct comparison of the original graph with a single realization of the model. Instead, the comparison should be performed by taking into account the expected structural properties of the ensemble of networks generated by the model, for each value of the parameter, averaging over a sufficiently large number of realizations. The first requirement of any suitable model for the *C. elegans* neural network growth is to pro-

duce networks having $N = 279$ nodes and, on average, $K = 2287$ edges. This constraint has been used to find the optimal parameter of each considered model.

We employed a two-step parameter optimization process. In the first step we used a Monte-Carlo approach to identify the interval in the parameter space for which the expected total number of edges $\tilde{K}$ of the generated networks was equal to $2287 \pm 5\%$. In this step, we considered 20 networks for each value of the parameter. In the second step we iteratively shrunk the parameter interval using the bisection method, in order to identify the value for which the difference between $\tilde{K}$ and $K = 2287$ was smaller than 1%. In this step we generated 500 networks for each value of the parameter. The optimal parameter values for each of the eight models are reported in Table S-I.

Model	**Optimal parameter**
BAG	$p = 0.0575$
HAG	$p = 0.302$
BA	$m_0 = 8, m = 8$
SSG	$\delta = 0.01365$
SGE	$\delta = 0.00235$
PEG	$\alpha = 0.0232$
ESG	$\delta = 0.0858$
ESTG	$\delta = 0.0126$

Table S-I. **Optimal model parameters.** The optimal parameter of a model guarantees the generation of networks having the same number of edges as the *C. elegans* adult neural network, with an error smaller than 1%.

Section S4. Model comparison

Since our aim was to reproduce as closely as possible the developmental growth of the *C. elegans* neuronal network, and in particular the abrupt transition in the number of edges in the graph as a function of the number of nodes, we defined a measure to quantify how closely each model matches the curve $\mathcal{K}(N)$, which indicates the number of edges in the *C. elegans* neuronal network when N nodes have been born.

We denote by $\mathcal{K}_M(N)$ the family of curves of K over N obtained using a certain model M and setting the value of the model parameter according to Table S-I. We computed, for each value of N, the expected number $\mu(\mathcal{K}_M(N))$ of edges in the network generated by model M when N nodes have been added to the graph, averaging over 500 realizations. Using this notation, $\mu(\mathcal{K}_M(100))$ is the expected number of edges in the graphs generated by model M when the first $N = 100$ nodes have been added to the graph.

In Fig. S-2 we report the average curve $\mu(\mathcal{K}_M(N))$ for each of the eight considered models, together with the original curve $\mathcal{K}(N)$ corresponding to the growth of the *C. elegans* neural network. By visual inspection, we conclude that the model which best fits the developmental growth of the original network and the phase transition at hatching is the Economical Spatio-Temporal Growth.

In order to quantify the discrepancy between $\mathcal{K}(N)$ and $\mathcal{K}_M(N)$ we computed, for each model and for each value of N, the difference $\xi(N)$:

$$\xi(N) = |\mathcal{K}(N) - \mu(\mathcal{K}_M(N))| \tag{S-4}$$

and we considered the expected value $\mu[\xi(N)]$ and the standard deviation $\sigma[\xi(N)]$ of $\xi(N)$. In Table S-II we report the values of $\mu[\xi(N)]$ and $\sigma[\xi(N)]$ for the eight models considered. In general, smaller values of $\mu[\xi(N)]$ and $\sigma[\xi(N)]$ indicate a closer match of the original growth curve. In agreement with the conclusions drawn after visual inspection of Fig S-2, which suggested that ESTG was the model which most closely reproduced the growth curve, the smallest values of $\mu[\xi(N)]$ and $\sigma[\xi(N)]$ are indeed obtained by the Economical Spatio-Temporal Growth model. The networks generated by all the other models fail to follow the original growth curve by a large extent, and they consequently exhibit larger values of $\mu[\xi(N)]$ and $\sigma[\xi(N)]$.

Model	$\mu[\xi(N)]$	$\sigma[\xi(N)]$
BAG	154.2	123.7
HAG	154.2	123.7
BA	216.7	150.7
SSG	205.2	167.1
SGE	89.5	73.7
PEG	209.4	168.4
ESG	215.6	172.9
ESTG	**37.3**	**31.6**

Table S-II. **Quality of growth fit.**Average and standard deviation of the point-to-point difference between the observed growth curve $\mathcal{K}(N)$ and the average curve corresponding to each of the eight considered models. The model parameters are set according to Table S-I. The smaller the value of $\mu[\xi(N)]$, the more closely a model can reproduce the growth of the *C. elegans* neural network. The Barabasi-Albert model (BA) exhibits the highest average point-to-point distance, while the Economical Spatio-Temporal Growth model (ESTG) largely outperforms all the other models.

Section S5. Node degree, edge length and node efficiency

Here we compare the structure of the networks produced by each of the eight models described in this study with that observed in the adult *C. elegans* neural network, by using three classical network metrics. The first metric is the degree distribution. Given an undirected graph $G(V, E)$ associated with the symmetric adjacency matrix $A = \{a_{ij}\}$, the degree of a node i is defined as the number of edges incident on i, and is denoted by $k_i = \sum_j a_{ij}$. The degree distribution $P(k)$ of the graph indicates, for each value of k, the probability of finding a node whose degree is equal to k. The second metric is the distribution of connection distances. Given two directly connected nodes i and j of a spatially-embedded network, we define the distance of the edge (i, j) as the Euclidean distance d_{ij} separating node i and node j. The

Model	$D_{KL}\left(P(k), P_M(k)\right)$	$D_{KL}\left(P(d), P_M(d)\right)$	$D_{KL}\left(P(E_i), P_M(E_i)\right)$
BAG	0.875	0.346	0.966
HAG	0.301	0.290	0.545
BA	0.309	0.176	0.226
SSG	0.710	0.884	0.611
SGE	0.428	0.269	1.447
PEG	0.149	0.322	0.214
ESG	0.708	0.685	0.361
ESTG	0.143	0.099	0.223

Table S-III. **Kullback-Leibler divergence.** The symmetrized Kullback-Leibler divergence between the degree, edge length and node efficiency distributions of the adult *C. elegans* neural network and the corresponding average distributions of the networks generated through each of the eight models. Smaller values of symmetrized divergence indicate higher similarity between the two distributions. The best and second-best values are highlighted in green and yellow, respectively, while the worst and second-worst are marked in red and orange, respectively. BAG and SSG exhibit the worst values of divergence. Interestingly, besides being the best model at fitting the developmental growth of the *C. elegans* neural network (as shown in Fig. S-2 and in Table S-II) ESTG performs more consistently than any of the other models in reproducing the structural properties of the adult worm's nervous system.

associated distance distribution $P(d)$ is the probability of finding an edge whose distance is exactly equal to d. The third metric is node efficiency. Given an undirected and unweighted graph G, the efficiency of a node is defined as:

$$E_i = \frac{1}{N-1} \sum_{\substack{j=1 \\ j \neq i}}^{N} \frac{1}{\lambda_{ij}} \tag{S-5}$$

where λ_{ij} is the distance between node i and node j, measured as the number of edges in the shortest path connecting i to j. The node efficiency of i measures how easy it is to reach any other node in the graph by starting from i and traveling across shortest paths. In general, the smaller the distance between i and j, the higher the contribution of j to the efficiency of node i. If the graph is not connected and node i and j belong to two different connected components then there exists no path connecting them. In this case, the distance λ_{ij} is conventionally set to ∞, and the contribution of node j to the efficiency of i is equal to $1/\infty \equiv 0$.

In Fig. S-3, S-4 and S-5 we show, respectively, the average degree distribution, length distribution and node efficiency distribution of the networks generated by each of the eight models, together with those observed in the adult *C. elegans* neural network (reported in each panel in shaded grey). By visual inspection, we notice that ESTG seems to be the model which most closely reproduces all these distributions.

In order to quantify the difference between the distributions of degree, length and node efficiency of synthetic graphs with those of the *C. elegans* neural network we used the Kullback-Leibler divergence. Given two probability distributions $P = \{p_i\}$ and $Q = \{q_i\}$, the Kullback-Leibler divergence of Q from P is defined as:

$$D_{KL}(P||Q) = \sum_i p_i \log \frac{p_i}{q_i} \tag{S-6}$$

The Kullback-Leibler divergence measures the information lost when Q is used as an approximation of P, and is non-symmetric, i.e. $D_{KL}(P||Q) \neq D_{KL}(Q||P)$. Since we are interested in measuring the similarity between two distributions, and not the relative information lost when using one of them as a predictor of the other, we opted for the symmetrized Kullback-Leibler divergence, which is defined as follows:

$$D_{KL}(P,Q) = \frac{D_{KL}(P||Q) + D_{KL}(Q||P)}{2} \tag{S-7}$$

In general, the smaller the value of $D_{KL}(P,Q)$, the more similar the two distributions P and Q. If we denote by $P(c)$ the distribution of the generic quantity c in the *C. elegans* neural network and by $P_M(c)$ the distribution of the same quantity c in networks generated through model M, the symmetrized Kullback-Leibler divergence between $P(c)$ and $P_M(c)$ is denoted as $D_{KL}\left(P(c), P_M(c)\right)$. In Table S-III we report, for each model, the values of the symmetrized Kullback-Leibler divergence between the degree, edge length and node efficiency distributions of the adult *C. elegans* neural network and the networks generated by each of the eight models, which are respectively denoted by $D_{KL}\left(P(k), P_M(k)\right)$, $D_{KL}\left(P(d), P_M(d)\right)$ and $D_{KL}\left(P(E_i), P_M(E_i)\right)$. The best and the second-best value of $D_{KL}(P,Q)$ for each metric are highlighted in green and yellow, respectively. Notice that the smallest values of the symmetrized Kullback-Leibler divergence are consistently obtained by the ESTG model, with the only exception being node efficiency for which PEG outperforms ESTG by a small amount.

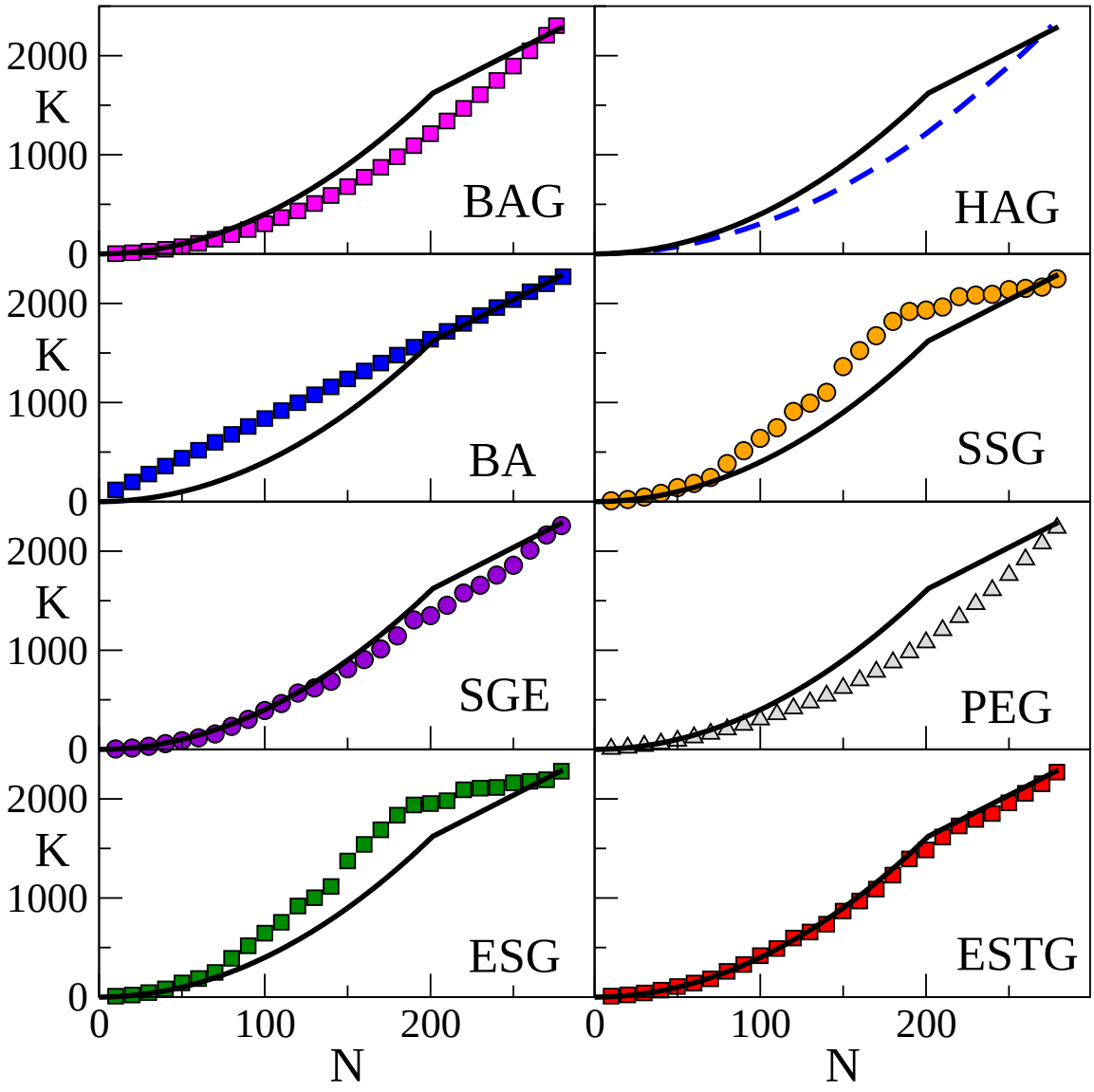

Figure S-2. **Growth curves.** The average total number of edges $\mathcal{K}_M(N)$ as a function of N for each of the eight models. The original growth curve of the *C. elegans* neural network is reported for reference in each panel, as a solid black line. The SSG, SGE, ESG and ESTG models exhibit a transition from a quadratic to a linear increasing regime, but only ETSG is able to closely match the growth curve observed in the original graph.

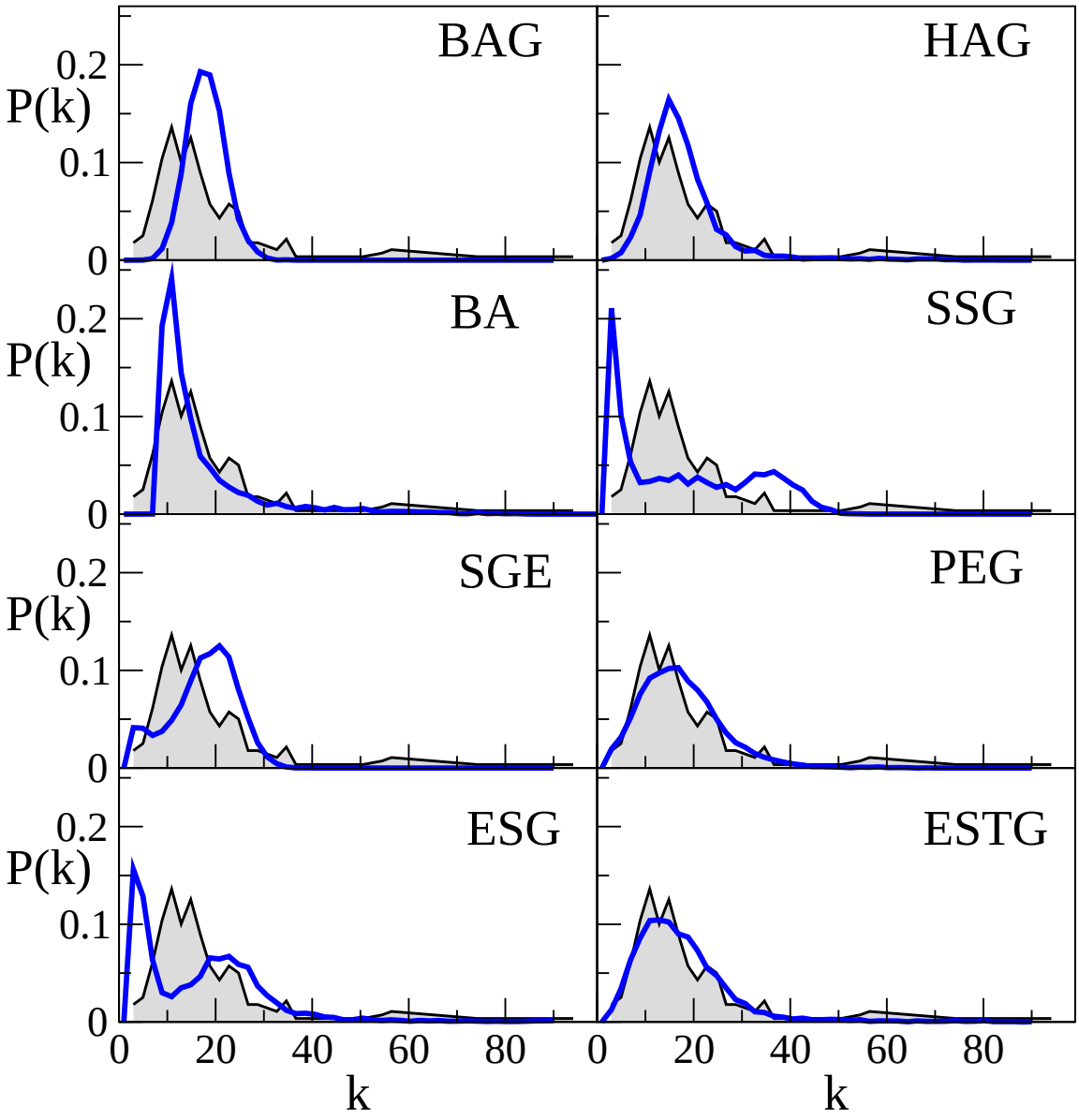

Figure S-3. **Degree distributions.** The average degree distribution of the networks generated by each of the eight models. The degree distribution of the adult *C. elegans* neural networks is reported in each panel in shaded gray, for comparison. Only the models based on hidden-variables, i.e. HAG, PEG and ETSG, are able to reproduce the degree distribution of the worm more closely. In the BA, SSG and ESG models low-degree nodes are over-represented, while in the BAG and SGE models low-degree nodes are substantially under-represented.

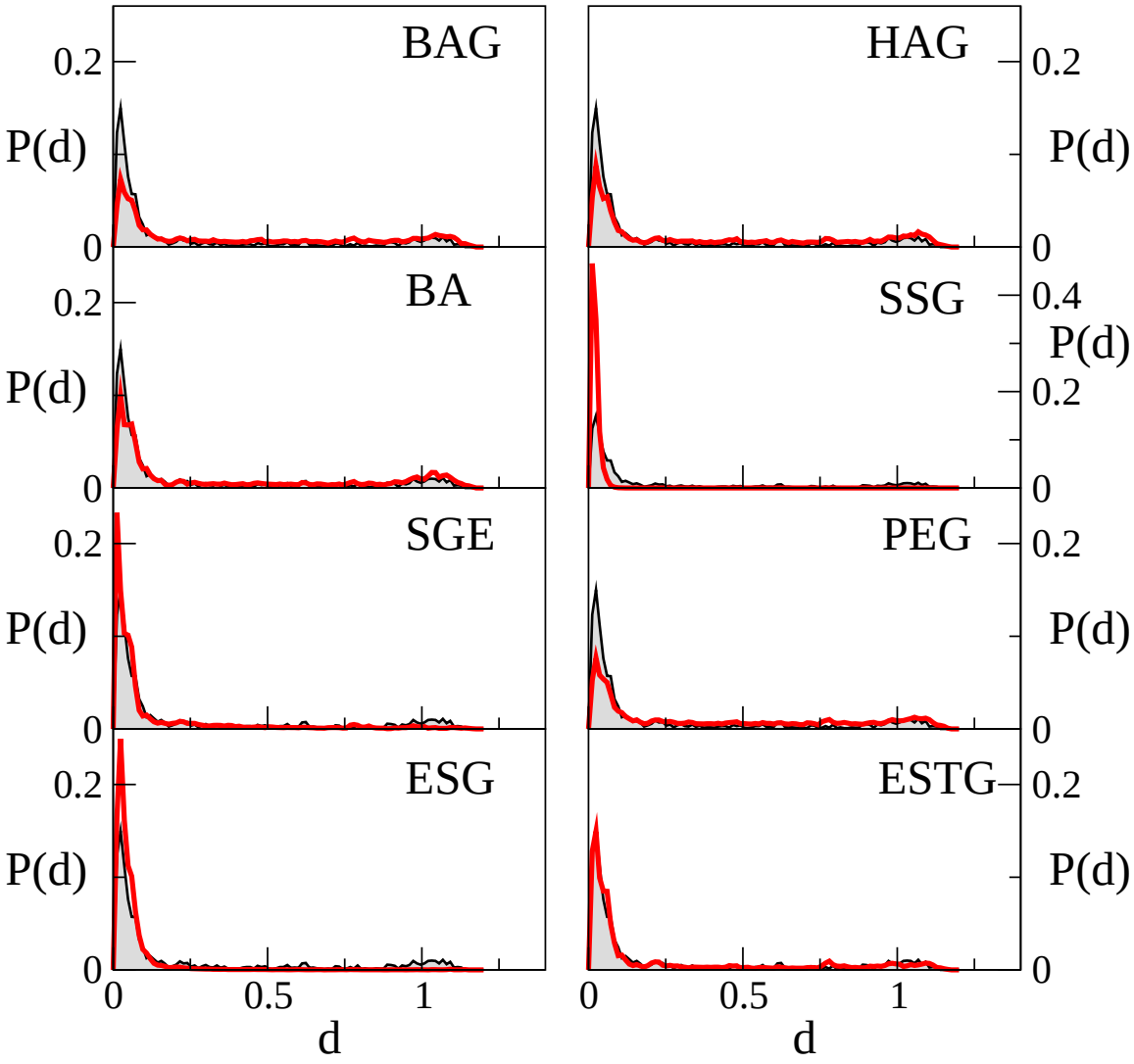

Figure S-4. **Edge length distributions.** The average distribution of edge length in the networks generated by each of the eight models, compared with the distribution of edge length observed in the adult *C. elegans* network (reported in shaded gray). BAG, HAG, BA and PEG produce networks with substantially longer links, while SSG, SGE and ESG exhibit a substantially larger percentage of short links (notice the different scale of the y-axis in the SSG panel). The only model which closely matches the distribution of edge length is ESTG.

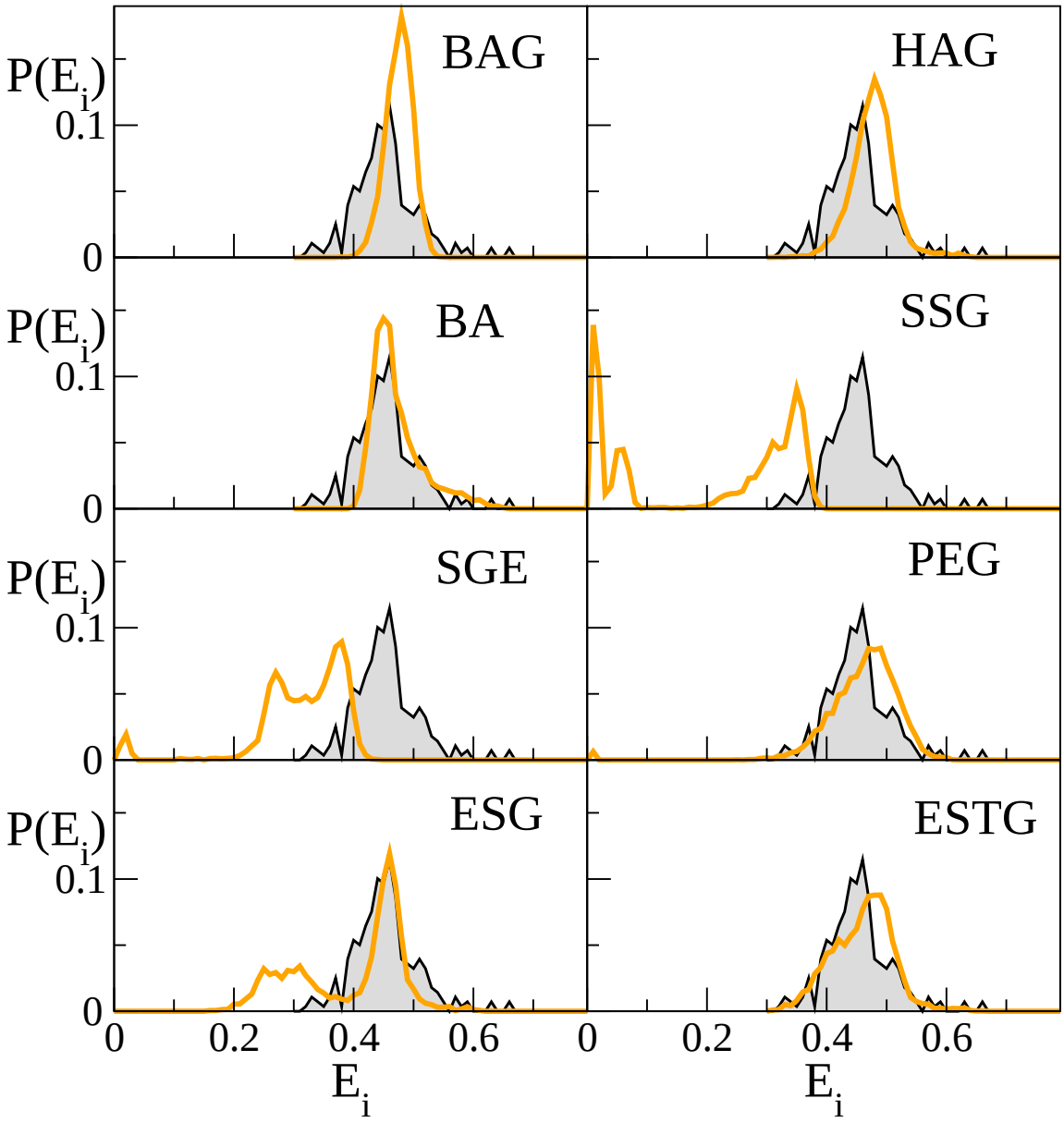

Figure S-5. **Node efficiency distribution.** The distribution of node efficiency of networks generated with each of the eight models, compared with that observed in the adult *C. elegans* (shaded gray). Both BAG and HAG produce binomial distributions of edge efficiency; for SSG, SGE and ESG models the distribution of efficiency is skewed towards smaller values while BA is able to capture the peak around $E_i = 0.47$. PEG and ESTG reproduce the original distribution in a more balanced way, even if nodes with efficiency around 0.5 are substantially over-represented while the peak around 0.47 is missing.