Biol Cybern (2014) 108:381–396
DOI 10.1007/s00422-014-0606-6

ORIGINAL PAPER

Biological Cybernetics

Aspects of randomness in neural graph structures

Michelle Rudolph-Lilith · Lyle E. Muller

Received: 17 May 2013 / Accepted: 21 April 2014 / Published online: 14 May 2014

Abstract In the past two decades, significant advances have been made in understanding the structural and functional properties of biological networks, via graph-theoretic analysis. In general, most graph-theoretic studies are conducted in the presence of serious uncertainties, such as major undersampling of the experimental data. In the specific case of neural systems, however, a few moderately robust experimental reconstructions have been reported, and these have long served as fundamental prototypes for studying connectivity patterns in the nervous system. In this paper, we provide a comparative analysis of these "historical" graphs, both in their directed (original) and symmetrized (a common preprocessing step) forms, and provide a set of measures that can be consistently applied across graphs (directed or undirected, with or without self-loops). We focus on simple structural characterizations of network connectivity and find that in many measures, the networks studied are captured by simple random graph models. In a few key measures, however, we observe a marked departure from the random graph prediction. Our results suggest that the mechanism of graph formation in the networks studied is not well captured by existing abstract graph models in their first- and second-order connectivity.

Keywords Graph theory · Network structure · Random graphs · Scale-free graphs · Mammalian brain · *C. elegans* · Network models

Electronic supplementary material The online version of this article (doi:10.1007/s00422-014-0606-6) contains supplementary material, which is available to authorized users.

M. Rudolph-Lilith (✉) · L. E. Muller
CNRS, Unité de Neurosciences, Information et Complexité (UNIC), 1 Ave de la Terrasse, 91198 Gif-sur-Yvette, France
e-mail: which.lilith@gmail.com; rudolph@unic.cnrs-gif.fr

1 Introduction

Since Stanley Milgram's six degrees of separation (Milgram 1967), the characterization of topological structure has become a major focus of graph-theoretic investigations in complex networks (Costa et al. 2007). In recent years, studies of this kind have begun to play an important role in a wide variety of disciplines, ranging from communications and power systems engineering to molecular and population biology (Albert et al. 1999; Albert and Barabási 2002; Dorogovtsev and Mendes 2002; Alm and Arkin 2003; Alon 2003; Bray 2003; Newman 2003; Barabási and Oltvai 2004). Often, by applying simple graph-theoretic measures, it is possible to find similarities in real-world graphs describing systems in many different domains, and also to separate these graphs into a number of representative classes, by highlighting their differences. Several studies have moved forward to connect such shared similarities to abstract, theoretical models of graph generation, which in turn can then be used to further investigate real-world graphs beyond the limitations imposed by the technologies currently available for the large-scale reconstruction of connectivity.

Two of the most successful of these models are the small-world and scale-free graphs (Watts and Strogatz 1998; Albert et al. 1999; for a general review, see Boccaletti et al. 2006; Newman 2010). In particular, scale-free graphs are viewed as a crucial prerequisite for complex dynamical behaviors and have been identified as a unifying feature of many real-world graphs (Barabási and Bonabeau 2003; Amaral and Ottino 2004). In recent years, however, several studies have challenged the empirical support for scale-free properties in many real-world graphs and their mechanistic backing (Clauset et al. 2009; Lima-Mendez and Helden 2009; Stumpf and Porter 2012). There is a growing consensus that the evidence for scale-free properties needs to be carefully reconsidered. The

insights gained from this may in turn lead to a deeper understanding of the underlying mechanisms producing the large-scale structure seen in real-world systems.

Among the most challenging real-world systems for graph-theoretic characterization are the strongly interconnected networks of the nervous system. Here, the analysis of the structural makeup of these graphs has shown no consistent evidence for scale-free properties. One of the reasons for this lack could be the severe undersampling due to technical limitations in the experimental reconstructions. A few moderately robust experimental reconstructions do exist, however, both of neural connectivity graphs (Varshney et al. 2011) and areal connectivity maps (Modhaa and Singh 2010), and these data suggest that scale-free organization is rather unlikely, as the number of connections per graph node generally does not span multiple scales. Furthermore, it is known in the mammalian cortex that the typical number of synaptic connections for a single neuron varies over one or, at most, two orders of magnitude (Braitenberg and Shüz 1998), ruling out the possibility of power-law organization over multiple scales in this structural network. While this argumentation does not necessarily apply to functional brain networks, and there has been evidence recently presented for their scale-free organization (e.g., see Eguíluz et al. 2005), other studies have report conflicting observations (Lima-Mendez and Helden 2009; Stumpf and Porter 2012) and this remains an open question (for review, see Bullmore and Sporns 2009).

In this study, we provide a comparative analysis of several "historical" reconstructions of structural neural graphs, including areal connectivity maps of the cat and macaque monkey cortex, as well as the neural connectivity graph of the nematode *Caenorhabditis elegans*. The general subject of interest here is the assessment of the structural connectivity pattern in these graphs. To this end, we introduce a mathematically consistent set of measures for relational graphs (directed or undirected, with or without self-loops), for characterizing the basic features of connectivity in these graphs. Specifically, we consider the node degree distributions, the structural equivalence of graph nodes, as well as nearest-neighbor degree and assortativity. Throughout this work, all measures are defined in their most general fashion, for directed graphs, but yield their forms known from the literature when applied to undirected graphs. By then applying the same measures to both the original directed and symmetrized undirected versions of each considered graph, we demonstrate that the process of symmetrization not only places limits on the characterization of graphs, but also introduces a systematic bias in measurement.

We find that the investigated networks share a strong component of randomness in their structural makeup, suggesting a mechanism of their formation that is much less constrained than that required for scale-free graphs. The graph structures observed in these data, furthermore, differ from that of the Erdős-Rényi random graphs. In particular, they differ from the latter by their specific node degree distribution and strong correlations of incoming and outgoing connections for individual nodes.

2 Methods

2.1 Graph theory preliminaries

A graph or network is comprised of a set of nodes that are linked by a set of edges. Two general types of graphs can be distinguished: undirected graphs, for which all edges act as bidirectional links between two nodes, and directed graphs (digraphs), in which case each edge is endowed with a direction pointing from a source node to a target node. In both cases, the spatial position of nodes can be considered (spatial graphs) and edges can exhibit properties such as a delay (delayed graphs) or weight (weighted graphs). Relational graphs are those excluding these additional properties, taking only the relations, or adjacencies, between nodes into account.

In this work, only relational graphs will be considered. In this case, the relationship between nodes can be mathematically formulated using an adjacency matrix a_{ij}, $i, j \in [1, N_N]$, where N_N denotes the number of nodes in the given graph. If node i makes a connection to node j, then $a_{ij} = 1$, otherwise $a_{ij} = 0$. Undirected graphs are a special case of digraph with symmetric adjacency matrix, i.e., $a_{ij} = a_{ji}$.

Special attention is required for the diagonal elements a_{ii} of the adjacency matrix, which describe self-loops. In the case of digraphs, a self-looped node acts both as target and source, so that nonzero diagonal elements contribute always two edges to a graph. For undirected graphs, this also leads to the contribution of two edges per self-looped node. This definition, however, deviates from that commonly used, because historically, undirected graphs were considered first, and most notions in modern graph theory will use only one edge per undirected link. In this paper, we will follow the latter notion with the exception of self-loops, which contribute two edges, and present all measures utilized in both their undirected and directed forms.

For a general introduction to graph theory, its measures, and applications, we refer to Diestel [2000], Boccaletti et al. [2006], Newman [2010].

2.2 Analysis methods

In this study, we used graph data that are publicly available. No modifications of the original data were performed. However, some of the graphs experienced a certain level of modifications since their first investigation. Therefore, numerical results reported in this study might deviate slightly from results reported in earlier studies.

It is a common practice in applying graph-theoretic measures to empirical data to symmetrize the adjacency matrix prior to analysis. Throughout this paper, both the original, directed graphs and their undirected, symmetrized versions are considered. Undirected graphs were symmetrized by setting $a_{ij} = 1$ for each pair (i, j) with $a_{ji} = 1$. The reduction to the giant component and subsequent analysis was performed on the symmetrized versions of the original graphs.

In some cases, we constructed corresponding Erdős-Rényi graph models for the given graphs with the same number of nodes, edges, and node degree distributions (node in/out-degree distributions for digraphs). These exact degree-matched Erdős-Rényi graph models (EDM) were obtained using the *cygraph* implementation of a sophisticated model introduced in Genio et al. [2010] and Kim et al. [2012]. When considering EDM graphs, 1,000 random realizations were used for each parameter set to ensure statistical stability.

Numerical analysis was performed using the custom software *cygraph* and *Mathematica*. A *cygraph* binary (Mac OSX), all graph data, and analysis protocols are available for download.[1]

2.3 "Historical" biological neural graphs

We study the structural aspects of a number of publicly available biological neural graphs used in the literature in the past two decades. These include areal connectivity graphs of the cat and macaque monkey cortex, as well as the neuronal connectivity graph of the nematode *C. elegans*. Below, we briefly describe these graphs and the notation used throughout this paper. More information on the data sources can be found in the references.

Cat neural graphs

The first set contains the areal connectivity graph including all cortical and thalamic areas of the cat brain (CC1), and a graph containing only the 52 cortical areas (CC2). Structural connection data for both CC1 and CC2 were first reported in Scannell et al. [1999] and obtained by analyzing a large collection of individual connection tracing studies done in the cortical and thalamic nuclei of the cat cerebral hemisphere. Available connection matrices[2] describe graphs containing 95 nodes and 2126 directed edges (CC1), and 52 nodes and 818 directed edges (CC2). Both graphs were studied in detail in Sporns and Zwi [2004] and Sporns and Ktter [2004].

C. elegans neural graphs

Three variants of the neuronal connectivity graph of the nematode worm *C. elegans* most often used throughout the literature were studied. Data for the first two (CE1 and CE2) are based on the experimental data from White et al. [1986] and were modified and made public in Watts and Strogatz [1998]. Available connection matrices[3] describe graphs containing 306 nodes and 2,345 directed edges (CE1), and 297 nodes with 2,345 directed edges (CE2).

The third dataset (CE3) constitutes the most recent and complete connectivity graph of *C. elegans* and was first discussed in Chen et al. [2006] (for a comprehensive review, see Varshney et al. 2011). The available connection matrix[4] describes a graph containing 279 nodes and 2,996 directed edges.

All graph data describe the synaptic connections between neurons of the *C. elegans* brain, with distinction of directed chemical synapses and undirected electrical junctions. In this paper, we will not consider this distinction, but view both chemical synapses and electrical junctions as part of the same connectivity structure (see Varshney et al. 2011; Rudolph-Lilith et al. 2012 for an analysis of both subgraphs).

Macaque monkey neural graphs

Various graphs of the macaque brain were considered. The most complete dataset (MB1) describes the macaque brain's long-distance areal connections and was first described in Modhaa and Singh [2010]. The obtained connectivity data[2] were assembled from the *CoCoMac (Collation of Connectivity data on the Macaque brain)* database. The latter is a growing collection of annotated information about a large number of published tracing studies performed in the macaque brain (Stephan et al. 2001; Kötter 2004). The investigated graph contains 383 nodes describing the various brain regions of the macaque monkey and 6,602 directed edges.

A second graph (MC1) describes the areal connectivity pattern of the macaque cortex, based on the original data published in Young [1993], and investigated in detail in Sporns and Tononi [2002] and Sporns [2002]. The available connection matrix[2] describes a graph containing 71 nodes and 746 directed edges.

A third graph (MC2) contains the macaque cortical connectivity within one hemisphere, based on the data from Choe et al. [2004], Kötter [2004], and Kaiser and Hilgetag [2006].

[1] http://www.cydyns.com;http://www.newscienceportal.com/MLR.

[2] https://sites.google.com/site/bctnet/datasets.

[3] CE1: http://wiki.gephi.org/index.php/Datasets; CE2: http://www-personal.umich.edu/~mejn/netdata/ with modifications by M. Newman.

[4] http://wormatlas.org/neuronalwiring.html.

Table 1 Basic graph-statistical measures were applied to the original directed and symmetrized (undirected) versions of the considered neural graphs

	N_N	N_L	Directed			Undirected	
			A	$\mathscr{A}$	Co	A	Co
CC1	95	0	2,126	0.1829	0.2331	2,340	0.2566
CC2	52	0	818	0.4117	0.2968	1,030	0.3737
CE1	306	0	2,345	0.9083	0.0250	4,296	0.0457
CE2	297	0	2,345	0.9083	0.0265	4,296	0.0486
CE3	279	3	2,996	0.6917	0.0384	4,580	0.0586
MB1	383	0	6,602	0.7323	0.0449	10,416	0.0708
MC1	71	0	746	0.2968	0.1459	876	0.1714
MC2	94	0	2,390	0.4224	0.2676	3,030	0.3393
MNC1	47	0	505	0.3866	0.2238	626	0.2775
MVC1	30	0	311	0.3632	0.3344	380	0.4086
MVC2	32	0	315	0.3763	0.2983	388	0.3674

Shown are values for the number of nodes N_N, number of self-loops N_L, total adjacency A (Eq. 1; N_E given in Eq. 2), asymmetry index $\mathscr{A}$ (Eq. 3; $\mathscr{A} = 0$ for undirected graphs) and connectedness Co (Eq. 4)

Table 2 Basic node degree analysis of original directed and symmetrized (undirected) versions of the considered neural graphs

	Directed					Undirected		
	δ^{in}	Δ^{in}	δ^{out}	Δ^{out}	$\langle a_i \rangle$	δ	Δ	$\langle a_i \rangle$
CC1	2	55	2	52	22.38	2	61	24.63
CC2	7	32	3	34	15.73	7	37	19.81
CE1	0	134	0	39	7.66	0	134	14.04
CE2	0	134	0	39	7.90	1	134	14.46
CE3	0	83	0	57	10.73	2	93	16.41
MB1	0	105	0	109	17.24	0	149	27.20
MC1	0	26	1	28	10.51	1	28	12.34
MC2	0	73	1	54	25.43	1	74	32.23
MNC1	1	23	2	23	10.74	3	27	13.32
MVC1	2	19	4	20	10.37	5	22	12.67
MVC2	0	19	2	20	9.84	2	22	12.13

Listed are the minimum and maximum node degree δ and Δ, respectively (δ^α and Δ^α for digraphs; $\alpha \in \{\text{in}, \text{out}\}$), and the average node degree $\langle a_i \rangle$ (Eq. 14)

The available dataset[5] describes a graph containing 94 nodes and 2,390 directed edges.

Finally, we analyzed the visual and sensorimotor area corticocortical connectivity graph of the macaque neocortex (MNC1) and two areal connectivity graphs of the macaque visual cortex (MVC1, MVC2). The MNC1 graph was first studied and made public in Honey et al. [2007]. The available dataset[2] describes a graph containing 47 nodes and 505 directed edges. MVC1 and MVC2 are two variants of the visual cortical connectivity originally published in Felleman and Essen [1991] and investigated in detail in Sporns et al. [2000] and Sporns and Ktter [2004]. The available datasets[2] describe graphs containing 30 nodes and 311 directed edges (MVC1), and 32 nodes with 315 directed edges (MVC2).

The basic graph-theoretic properties of these graphs are listed in Tables 1 and 2, and further discussed below.

2.4 Connected components

A (strongly) connected component is defined as a subgraph consisting of a set of nodes from which all other nodes in the subgraph can be reached, and which can be reached from all other nodes, by following existing edges. Typically, the set of (strongly) connected components of a graph will be dominated by a giant (strongly) connected component of size S_{gcc}, defined as the number of nodes in this component (Boccaletti et al. 2006). We calculated the number of connected components (strongly connected components, in the case of digraphs) N_{cc} and the size of the giant connected component (giant strongly connected component for digraphs) S_{gcc}.

Table 3 summarizes the numerical results for the considered biological neural graphs and their symmetrizations, along with the asymmetry index, minimal, maximal, and average node degrees, δ, Δ, and $\langle a_i \rangle$, respectively. Naturally, the connectedness of the giant connected component is slightly larger than that of the original graphs, whereas the asymmetry index $\mathscr{A}$ is slightly smaller for graphs whose size of the giant connected component is smaller than the N_N of the original graph.

Throughout this work, we restrict our analysis to the giant connected (for undirected versions of the considered graphs) and giant strongly connected (for the original digraphs) components. Furthermore, as indicated in Table 3, the giant (strongly) connected components of CE1 and CE2 are identical, and in the following, only CE1 will be considered.

3 Adjacency, connectance, asymmetry

In a first step, we analyzed all considered graphs with respect to basic graph-statistical measures. These include the number of self-loops N_L, defined as the number of nonzero diagonal elements a_{ii} in the adjacency matrix, and the total adjacency A, defined as the sum over all entries in the adjacency matrix, with diagonal elements (self-loops) counting two:

$$A = \sum_{i,j=1}^{N_N} a_{ij} + N_L \,. \tag{1}$$

Using the total adjacency, the number of edges N_E is defined as

[5] http://www.biological-networks.org/?page_id=25.

Table 3 Connected component analysis of original directed and symmetrized (undirected) versions of various neural graphs

	Directed								Undirected					
	N_{cc}	S_{gcc}	A	$\mathscr{A}$	Co	$\delta/\Delta^{\text{in}}$	$\delta/\Delta^{\text{out}}$	$\langle a_i \rangle$	N_{cc}	S_{gcc}	A	Co	δ/Δ	$\langle a_i \rangle$
CC1	1	95	2,126	0.1829	0.2331	2/55	2/52	22.38	1	95	2,340	0.2566	2/61	24.63
CC2	1	52	818	0.4117	0.2968	7/32	3/34	15.73	1	52	1,030	0.3737	7/37	19.81
CE1	66	239	1,912	0.8864	0.0333	1/42	1/38	8.00	10	297	4,296	0.0485	1/134	14.46
CE2	57	239	1,912	0.8864	0.0333	1/42	1/38	8.00	1	297	4,296	0.0485	1/134	14.46
CE3	6	274	2,962	0.6871	0.0393	1/82	1/57	10.80	1	279	4,580	0.0586	2/93	16.41
MB1	33	351	6,491	0.7265	0.0525	1/103	1/108	18.49	24	360	10,416	0.0801	1/149	28.93
MC1	2	70	745	0.2952	0.1499	2/26	2/28	10.64	1	71	876	0.1714	1/28	12.34
MC2	10	85	2,356	0.4092	0.3223	1/65	1/54	27.72	1	94	3,030	0.3393	1/74	32.23
MNC1	1	47	505	0.3866	0.2238	1/23	2/23	10.74	1	47	626	0.2775	3/27	13.32
MVC1	1	30	311	0.3632	0.3344	2/19	4/20	10.37	1	30	380	0.4086	5/22	12.67
MVC2	3	30	311	0.3632	0.3344	2/19	4/20	10.37	1	32	388	0.3674	2/22	12.13

Shown are values for the number of connected components N_{cc}, the size of the giant connected component S_{gcc}, total adjacency A (N_E given in Eq. 2), asymmetry index $\mathscr{A}$ ($\mathscr{A} = 0$ for undirected graphs), connectedness Co, the minimum and maximum node degree δ and Δ, respectively ($\delta^{\{\text{in,out}\}}$ and $\Delta^{\{\text{in,out}\}}$ for directed graphs), and the average node degree $\langle a_i \rangle$ for the giant connected components

$$N_E = \begin{cases} A & \text{directed} \\ A/2 & \text{undirected}\,. \end{cases} \tag{2}$$

The asymmetry index $\mathscr{A}$ quantifies the ratio between the number of non-symmetrical edges N_A and symmetrical edges N_S, and is given by Wasserman and Faust [1994], Newman et al. [2002], Serrano and Boguñá [2003], but see Garlaschelli and Loffredo [2004]

$$\mathscr{A} = \frac{N_A}{A - N_S}\,, \tag{3}$$

where N_A is the number of node pairs $(i, j|j \geq i)$ for which $a_{ij} \neq a_{ji}$, and N_S is the number of node pairs $(i, j|j \geq i)$ for which $a_{ij} = a_{ji} = 1$. It can be shown that $0 \leq \mathscr{A} \leq 1$, and that Eq. 3 holds for self-looped and non-self-looped graphs.

Finally, the graph connectedness (or connectance) Co, a measure of relative graph connectivity, is defined as (Boccaletti et al. 2006; Newman 2010)

$$Co = \frac{N_E}{N_E^{\max}} = \begin{cases} \frac{A}{N_N(N_N+1)} & \text{self-looped} \\ \frac{A}{N_N(N_N-1)} & \text{not self-looped}, \end{cases} \tag{4}$$

where $N_E^{\max}$ denotes the number of possible edges in a complete, i.e., maximally connected, graph:

$$N_E^{\max} = \begin{cases} N_N(N_N+1) & \text{directed, self-looped} \\ N_N(N_N-1) & \text{directed, not self-looped} \\ \frac{1}{2}N_N(N_N+1) & \text{undirected, self-looped} \\ \frac{1}{2}N_N(N_N-1) & \text{undirected, not self-looped}. \end{cases}$$

It can be shown that $0 \leq Co \leq 1$. We note that Eq. 4 generalizes the commonly used definition of the connectedness to graphs containing self-loops (e.g., see Boccaletti et al. 2006; Newman 2010).

The basic graph-statistical properties of both the directed and symmetrized (undirected) versions of the investigated neural graphs are listed in Table 1. Of these graphs, only CE3 has self-loops, accounting for about 0.1 % of the graph's edges, which stem from electrical junctions connecting a node with itself. If not specified otherwise, these self-loops were included in the analysis.

Naturally, the total adjacency A is larger for the undirected version of the corresponding graphs, as symmetrization of the adjacency matrix only adds edges to a given digraph. By symmetrizing a graph, the total adjacency and, thus, the connectedness can increase by more than 50 %, as in the case of the *C. elegans* neural graph (CE1 and CE2), where the connectedness of the symmetrized undirected graph (Co^{ud}) is about 1.828 times greater than that of the original directed version (Co^{d}), in addition to other instances (CE3: $Co^{ud} \sim 1.526\, Co^{d}$, MB1: $Co^{ud} \sim 1.577\, Co^{d}$). Thus, the consideration of undirected versions of digraphs might already at this level introduce a significant mischaracterization of investigated graphs.

Due to the symmetrization procedure, the connectedness is intrinsically dependent on the asymmetry of the considered graph. The higher the asymmetry $\mathscr{A}$, the more edges will be added, yielding a higher connectedness in the undirected version of a given digraph (Fig. 1a). A theoretical relationship between the ratio of the connectedness for digraphs and their undirected equivalents, and the asymmetry index can be obtained by observing that the total adjacency $A^d = N_A + 2N_S$ for digraphs takes after symmetrization the form $A^{ud} = 2(N_A + N_S)$. This yields

$$A^d = A^{ud}\left(1 - \frac{1}{2}\mathscr{A}\right).$$

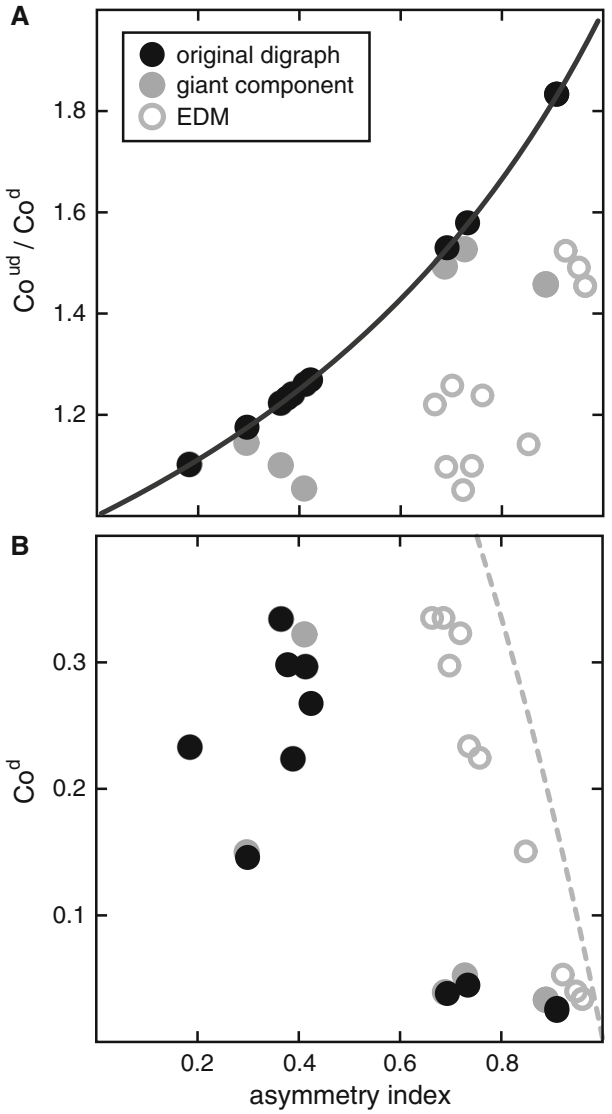

Fig. 1 Relation of connectedness and asymmetry index in various biological neural graphs and their giant connected components. **a** The ratio between connectedness of the symmetrized (undirected) version of a graph (Co^{ud}) and its directed original (Co^{d}) increases with the asymmetry index of the digraph. The *solid line* shows the theoretical relation, Eq. 5. **b** In biological neural digraphs, high connectedness appears to be linked with a lower asymmetry index. The *dashed line* shows the analytical result for a Erdős-Rényi graph model, Eq. 9. In both panels, results are shown for the original digraphs (*black dots*), their giant connected components (*gray dots*), and corresponding exact degree-matched Erdős-Rényi graph models (*open dots*)

If we assume that the number of nodes in both the directed and its symmetrized version is the same, this gives, together with Eq. 4, the desired relation

$$\frac{Co^{ud}}{Co^{d}} = \frac{2}{2-\mathscr{A}}, \tag{5}$$

shown in Fig. 1a (solid). However, the numerical results for the giant component, as well as the corresponding EDM graphs, deviate from Eq. 5 (Fig. 1a, compare gray and open dots with black solid line). The reason for this deviation is simply that, in the case of undirected graphs, the giant connected components were obtained from their original digraphs after symmetrization (see Methods). This leads to a change in the number of nodes in the constructed giant components of corresponding digraph and undirected graph, therefore Eq. 5 no longer applies.

Interestingly, a weak relationship between connectedness and asymmetry index can also be found when considering digraphs only (Fig. 1b), with graphs of higher connectedness being associated with a weaker asymmetry. Such a link is expected, however, and can be calculated analytically in the case of directed Erdős-Rényi (ER) graphs. Excluding (for simplicity) self-loops, the total adjacency is $pN_N(N_N-1)$, where p denotes the connection probability of a classical ER graph. With the total number of possible edges in a directed, not self-looped ER graph being $N_E^{\max} = N_N(N_N-1)$, the connectedness $Co = p$. Node pairs (i, j) with $a_{ij} = 1 \wedge a_{ji} = 1$ occur here with a probability of p^2. These are the only contribution to the number of symmetric edges N_S, thus yielding

$$N_S = p^2 \frac{N_E^{\max}}{2}. \tag{6}$$

In a similar fashion, node pairs (i, j) with $a_{ij} = 1 \wedge a_{ji} = 0$ or $a_{ij} = 0 \wedge a_{ji} = 1$ occur with a probability $p(1-p)$ each and contribute to the number of non-symmetrical edges N_A, yielding

$$N_A = 2p(1-p)\frac{N_E^{\max}}{2}. \tag{7}$$

With this, the asymmetry index (Eq. 3) of a ER graph is given by

$$\mathscr{A} = 2\frac{1-Co}{2-Co}, \tag{8}$$

which yields the desired relationship between connectedness and asymmetry

$$Co = 2\frac{\mathscr{A}-1}{\mathscr{A}-2}. \tag{9}$$

The theoretical result for classical ER graph models (Eq. 9) is independent of the number of nodes and is shown in Fig. 1a (dashed line). Although displaying the same qualitative behavior, namely a decrease in the connectedness for increasing asymmetry index, the quantitative results for the investigated neural graphs deviate significantly from the theoretic expectation for ER graph models. Even after the incorporation of the exact node degree distribution using the EDM graph models (Fig. 1a, gray circles), the results still deviate markedly from that observed in their biological counterparts. This suggests that a random distribution of edges

with a given degree distribution cannot account for the relationship between connectedness and asymmetry observed in biological neural digraphs. However, adding a correlation between node in and out degree will increase the probability of occurrence of node pairs (i, j) with $a_{ij} = a_{ji} = 1$, thus increase the number of symmetric edges N_S and proportionally lower the number of non-symmetric edges N_A. According to Eq. 3, this will effectively lead to a decrease in $\mathscr{A}$ for a given connectedness. As demonstrated in the Supplementary Information (Graph assortativity), such a correlation between a node's in and out degree is indeed what we observe in the investigated biological graphs.

4 Node degrees

To further characterize the structural aspects of biological neural digraphs, we calculated the node in and out degrees

$$a_i^{\text{in}} = \sum_{j=1}^{N_N} a_{ji} \tag{10}$$

$$a_i^{\text{out}} = \sum_{j=1}^{N_N} a_{ij} \tag{11}$$

as well as the node degree

$$a_i = \sum_{j=1}^{N_N} a_{ij} + a_{ii} \tag{12}$$

for the corresponding symmetrized graphs. Note that the definition of the node degree in Eq. 12 deviates from that commonly employed through the inclusion of self-loops, which are considered as contributing two edges to adjacency relations (one in-edge and one out-edge pointing to the same node, see above). This definition is more natural, as it is a direct result of the definition of a_i for digraphs in Eqs. 10 and 11. By defining terms in this way, undirected graphs become a special case of digraphs, which then assume the more fundamental role. With this, the handshaking lemma, which provides a consistency relation linking the sum over all node degrees with the total adjacency of a graph, takes a more general form valid for both directed and undirected self-looped graphs:

$$\sum_{i=1}^{N_N} a_i = 2(A - N_L)\,, \tag{13}$$

where $a_i = a_i^{\text{in}} + a_i^{\text{out}}$ for digraphs.

Given the node (in/out-) degrees of a graph, we define the minimum and maximum node (in/out-) degree, respectively, as δ (δ^α) and Δ (Δ^α), $\alpha \in \{\text{in}, \text{out}\}$. Furthermore, the average node in/out degree $\langle a_i^\alpha \rangle$ for directed and average node degree $\langle a_i \rangle$ for undirected graphs is given by

$$\langle a_i \rangle = \begin{cases} \frac{1}{N_N} \sum_{i=1}^{N_N} a_i^{\text{in}} = \langle a_i^{\text{in}} \rangle = \langle a_i^{\text{out}} \rangle = \frac{1}{N_N} \sum_{i=1}^{N_N} a_i^{\text{out}} \\ \qquad\qquad \text{Directed} \\ \frac{1}{N_N} \sum_{i=1}^{N_N} a_i \qquad \text{Undirected.} \end{cases} \tag{14}$$

Note that due to the handshaking lemma, Eq. 13, we have $\langle a_i^{\text{in}} \rangle = \langle a_i^{\text{out}} \rangle$.

Results for the investigated biological graphs are summarized in Table 2. In two of the investigated graphs (CE1 and MB1), the minimal node degrees δ^α and δ in both the directed and undirected version, respectively, are zero. Furthermore, the minimal total node degree in these graphs is also zero, indicating the existence of nodes without edges. Further analysis revealed that in the directed version of CE1, the number of weakly connected components, i.e., subgraphs whose nodes are connected by at least one directed edge to other nodes in the same subgraph, is 10, with the size of the largest weakly connected component being 297 nodes. Thus, with a total of 306 in this graph, the remaining 9 components share 9 nodes, i.e., each of the remaining weakly connected components contains only one isolated node. The same argumentation applies to the undirected version of CE1. The MB1 graph has one giant weakly connected component with 351 nodes, and the remaining 32 connected components share 32 isolated nodes. The existence of these isolated nodes in the neural graphs suggests that the mapping of these graphs is incomplete, as such nodes are very unlikely to have a functional or structural meaning. Therefore, in the remainder of this study, we will focus our analysis on the giant (strongly) connected component of each investigated graph (see Methods, Connected components).

5 Node degree distributions

A prevalent theme in the literature of the past two decades is scale-free properties of various real-world systems, typically investigated by fitting corresponding physical quantities with power-law distributions. However, recently it was pointed out that in many cases, such a fit provides only a poor description of the true behavior, or at best a faithful representation in only a narrow region of the investigated quantities' value range (Clauset et al. 2009; Lima-Mendez and Helden 2009). This is especially crucial when considering small systems, for which boundary effects cannot feasibly be neglected. Moreover, claims of scale-free properties, with little or no support from experimental data, may distract further search for mechanisms by which such networks form and develop.

In order to assess the extent to which the power law provides a valid description of structural characteristics of biological neural graphs, we studied the node degree probability density functions (PDFs: node in-degree and out-degree

PDFs for digraphs) of these graphs. The power-law model is defined by

$$\rho^{pl}(a;\alpha) = a_{\min}^{\alpha-1}(\alpha-1)a^{-\alpha}\,, \tag{15}$$

where a denotes the node degree and $a_{\min}$ the lower bound of the fitting interval. In addition, we applied other fitting models proposed in the literature (see Clauset et al. 2009). The second model considered was the "power law with cutoff," defined by

$$\rho^{plwc}(a;\alpha,\lambda) = \frac{1}{\Gamma[1-\alpha,\lambda a_{\min}]}\lambda^{1-\alpha}a^{-\alpha}e^{-\lambda a}\,, \tag{16}$$

where $\Gamma[s,x]$ is the incomplete gamma function. We note that due to the exponential term, this model carries none of the implications commonly associated with the power-law model (i.e., scale-free characteristics), as this term replaces the power-law properties at both tails of the distribution, leaving only the dominance of the power-law behavior within a certain characteristic scale. However, to avoid confusion and to remain in accordance with the literature (Clauset et al. 2009), we retain this terminology through the remainder of the paper.

Further models utilized were the stretched exponential model

$$\rho^{se}(a;\beta,\lambda) = \beta\lambda e^{\lambda a_{\min}^{\beta}}a^{\beta-1}e^{-\lambda a^{\beta}} \tag{17}$$

and gamma model

$$\rho^{g}(a;\theta,k) = \frac{1}{\Gamma[k]}a^{k-1}e^{-a/\theta}\theta^{-k}\,. \tag{18}$$

The latter is equivalent to the power law with cutoff model when considering $k \leftrightarrow (1-\alpha)$ and $\theta \leftrightarrow 1/\lambda$. Finally, the log-normal model

$$\rho^{ln}(a;\mu,\sigma) = N\frac{1}{a}\exp\left[-\frac{(\ln a-\mu)^2}{2\sigma^2}\right] \tag{19}$$

with

$$N = \sqrt{\frac{2}{\pi\sigma^2}}\left(\mathrm{erfc}\left[\frac{\ln a_{\min}-\mu}{\sqrt{2\sigma^2}}\right]\right)^{-1}$$

and the Poisson model

$$\rho^{p}(a;\mu) = \tilde{N}\frac{1}{\Gamma[a+1]}\mu^{a} \tag{20}$$

with

$$\tilde{N} = \left(e^{\mu} - \sum_{k=0}^{a_{\min}-1}\frac{\mu^k}{\Gamma[k+1]}\right)^{-1}$$

were considered. In all cases, $a_{\min} = 1$ was used, except for the power-law model, which only allowed fitting the tail of the degree distributions. Moreover, all fitted models were constrained by the normalization condition

$$\int_{a_{\min}}^{\infty}\rho^{\alpha}(a;\bullet)\,da = 1\,, \tag{21}$$

where $\alpha \in \{plwc, se, g, ln, p\}$ and $\bullet$ stands for the set of parameters of the corresponding model, with the exception of the power-law model. As the latter fits only the tail of a given PDF, the normalization constant was adjusted to the fraction of the node degree PDF $\rho(a_i)$ above the lower bound, i.e.,

$$\int_{a_{\min}}^{\infty}\rho^{pl}(a_i;\alpha) = \int_{a_{\min}}^{\infty}\rho(a_i)\,. \tag{22}$$

Representative examples of the node degree PDFs for the cat cortex graph (CC1) and the neural connectivity graph of *C. elegans* (CE3) are shown in Fig. 2a, b, respectively. Among the graphs considered, only the node degree distributions of MC2 did not allow for a reasonable fit with any of the above models, both in the directed and undirected version (Fig. 2c). This hints either at a very peculiar connectivity pattern in this graph, as it describes the cortical connectivity pattern in only one hemisphere, or its incomplete representation due to missing experimental data. Additionally, in no case, the Poisson model (Eq. 20) delivered an acceptable fit of the node degree PDFs (Fig. 2, black dashed), for which reason it was excluded from further consideration.

The obtained best fitting parameters of the node degree models, using the nonlinear least-squares method in *Mathematica*, are summarized in Tables 4, 5 and 6, and visualized in Fig. 3a. A detailed presentation of the fitting results using the various models, including the standard error, t-statistics, and P-value of the fitted parameters, as well an analysis of the decomposition of the variation in the data attributable to the fitted function and to the residual errors (ANOVA test) along with the root-mean-square difference between actual and predicted values can be found in the Supplementary Information, Fitting of node degree distributions.

Interestingly, in all fitted models, except the power law, a weak correlation between fitted parameters and the graph connectedness Co for both node in/out degree (digraphs) and node degree (undirected graphs) PDFs was found (Fig. 3b). Specifically, the α parameter in the power law with cutoff model appears to decrease with increasing connectedness (Fig. 3b, top left). This may reflect the fact that smaller α values lead to a shift of the mean degree to higher values, thus in turn reflecting the higher connectivity. Due to the relation $k = 1-\alpha$ between α of the power law with cutoff model and gamma model, a similar relation holds for gamma fits of the PDFs, with k increasing for higher Co (Fig. 3b, top right).

Both parameters of the stretched exponential model, β and λ, show a weak dependency on the connectedness

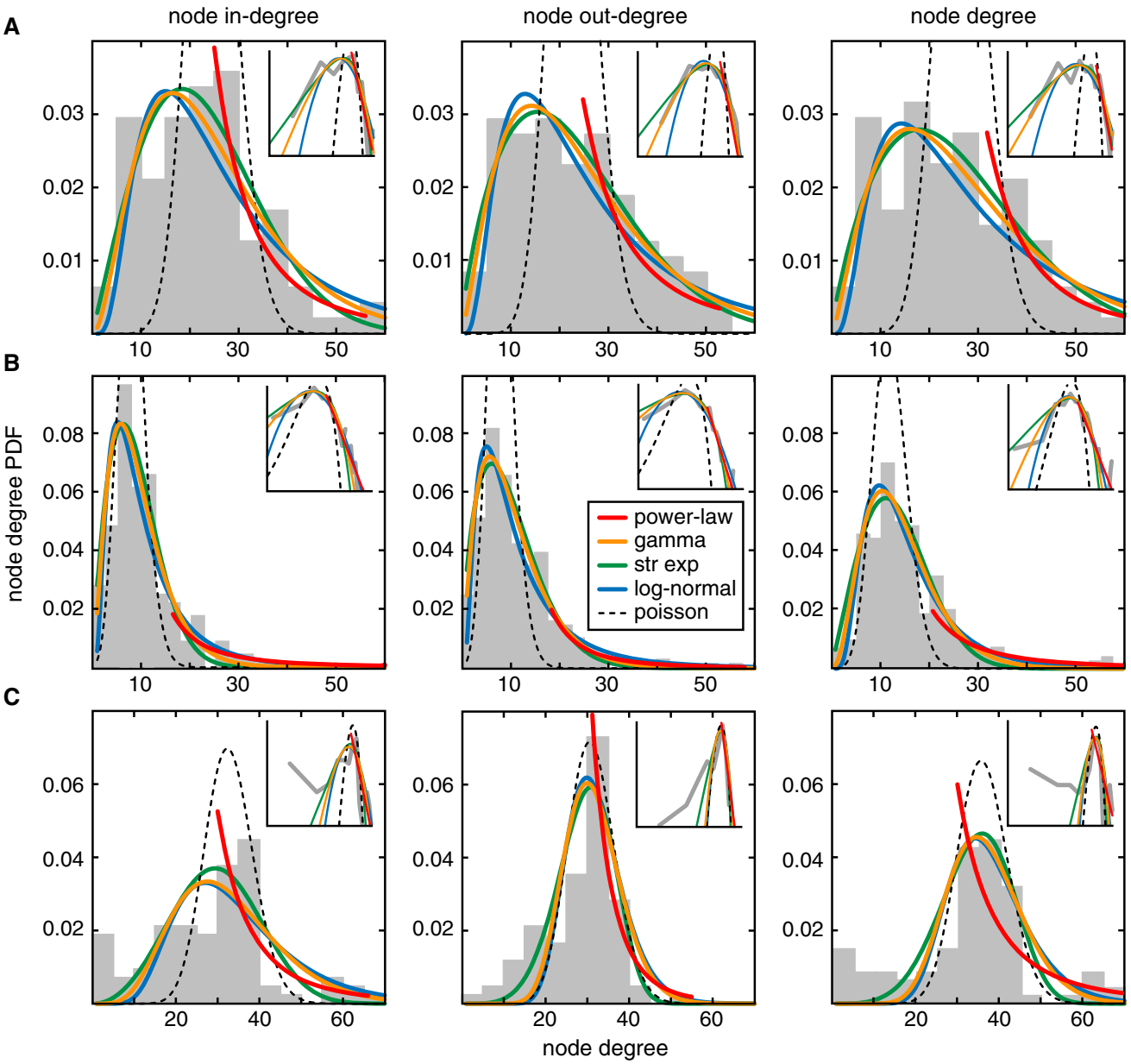

Fig. 2 Representative examples of node degree probability density functions and their fits. **a** CC1, **b** CE3, **c** MC2. For digraphs, the PDFs (*gray bar plots*) of node in degree (*left*) and node out degree (*middle*) are shown, for undirected graphs, the PDFs of the node degree (*right*). Best fits of the node degree distributions are provided by the power law with cutoff/gamma model (Eq. 16; *orange solid*), followed by the stretched exponential model (Eq. 17; *green solid*) and log-normal model (Eq. 19; *blue solid*). The power-law model (Eq. 15; *red solid*) provides an approximate fit of the tail of a given node degree PDF only. In all considered graphs, the Poisson model (Eq. 20; *black dashed*) did not deliver an acceptable fit of the data. *Insets* show the corresponding data in log–log representation

of the graphs, with β increasing and λ decreasing for increasing Co (Fig. 3b, bottom left and middle). For λ, this dependency reflects again the shift in the mean degree to higher values for more highly connected graphs. Interestingly, for the β parameter in this model, this relation appears to be inverted, with smaller β values leading to broader distributions but being associated with smaller connectedness. However, the stretched exponential model displays a strong correlation between the fitted parameters (Fig. 3b, bottom middle, inset). This is not the case in the other two-parameter models. Thus, the impact of both β and λ on the shape of the distribution cannot be considered as independent; this explains the peculiar behavior when both parameters are considered to be independent. Finally, a weak correlation between connectedness and σ of the log-normal model was found, with broader distributions (smaller σ) being associated with higher connectedness.

Table 4 Power-law fits of the tail degree distributions for various biological neural graphs

	Power-law model		
	α^{in}	α^{out}	α
CC1	3.4448	2.9610	3.8817
CC2	3.4922	3.7824	2.3411
CE1	2.8122	2.8068	3.1565
CE3	2.7072	3.4023	2.7054
MB1	2.1281	2.2697	2.3336
MC1	3.0592	2.7606	2.6497
MC2	4.0459	(6.6290)	3.6015
MNC1	1.7702	1.7275	2.0251
MVC1	1.9798	2.7790	2.7924
MVC2	1.7711	2.7790	2.7573

The values give the best fitting parameters $\alpha^{\{\text{in,out}\}}$ and α according to Eq. 15 for the node in/out degree and node degree PDFs of the directed and undirected versions of the graphs, respectively. Values in parentheses are excluded from Fig. 3

For assessing the quality of the different models, we compared the root mean squares of the fit residuals, i.e., differences between the actual and predicted node degree values. We found that the power law with cutoff and gamma model provided, on average, the best fits, closely followed by the stretched exponential and log-normal model. This evaluation is consistent with the conclusion reached in Clauset et al. [2009]. Most interestingly, the nature of these node degree distributions (gamma or power law with cutoff) could be consistent with a simple local mechanism responsible for generating neural graphs. In this way, we may conceive of a graph generation mechanism more parsimonious than those currently in the literature, such as preferential attachment (first discussed as the "Matthew effect" in Merton 1968; see also Barabási and Albert 1999). In this model, knowledge about the degree distribution of the whole network must be present at the individual nodes in order to bias new connections toward those that are already highly connected. We term such knowledge about the whole network that must be present at the level of individual nodes "nonlocal," to contrast with "local" generation mechanisms that can be implemented without additional information at the single-node level (see Discussion).

6 Structural equivalence

In order to assess the similarity of the connectivity pattern of individual nodes, various measures of structural equivalence were defined and used in the literature. Here, two nodes are defined as structurally equivalent if they share the same pattern of relationships with all other nodes in a given graph. A first coarse measure quantifying a pattern of relationships among nodes in digraphs is the Euclidean distance between rows and columns of the adjacency matrix (Boccaletti et al. 2006), defined as

$$D_{ij}^{\text{in-in}} = \left\{ \sum_{k=1}^{N_N} (a_{ki} - a_{kj})^2 \right\}^{1/2} \tag{23}$$

$$D_{ij}^{\text{in-out}} = \left\{ \sum_{k=1}^{N_N} (a_{ki} - a_{jk})^2 \right\}^{1/2} \tag{24}$$

$$D_{ij}^{\text{out-out}} = \left\{ \sum_{k=1}^{N_N} (a_{ik} - a_{jk})^2 \right\}^{1/2} \tag{25}$$

Table 5 Power law with cutoff and gamma fits of the node degree distributions for various biological neural graphs

	Power law with cutoff model						Gamma model					
	α^{in}	λ^{in}	α^{out}	λ^{out}	α	λ	θ^{in}	k^{in}	θ^{out}	k^{out}	θ	k
CC1	−1.9885	0.1212	−1.4334	0.0995	−1.4159	0.0885	8.2456	2.9891	10.0439	2.4338	11.3001	2.4163
CC2	−5.4192	0.4135	−1.8189	0.1680	−3.6622	0.2293	2.4169	6.4233	5.9539	2.8193	4.3598	4.6636
CE1	−0.7030	0.2277	−0.1054	0.1333	−2.1114	0.2281	4.1422	1.7442	6.3163	1.1787	4.3814	3.1114
CE3	−1.6139	0.2769	−1.1791	0.2084	−2.7632	0.2612	3.5851	2.6229	4.7285	2.1953	3.8276	3.7635
MB1	0.3211	0.0351	0.3322	0.0281	−0.1318	0.0361	21.8775	0.7280	28.9742	0.7063	26.8478	1.1439
MC1	−1.6364	0.2462	−1.2988	0.2077	−1.3638	0.1684	4.0623	2.6365	4.8144	2.2989	5.8766	2.3823
MC2	−5.4764	0.1988	(−20.7122)	0.6916	(−15.7716)	0.4560	5.0266	6.4798	1.4455	(21.7190)	2.1922	(16.7780)
MNC1	−0.6411	0.1180	−1.7654	0.2445	−1.7342	0.1913	−6.4050	1.6074	4.0904	2.7656	5.2265	2.7344
MVC1	−4.4855	0.5045	−3.4864	0.4100	−7.7057	0.6615	1.9820	5.4858	2.4385	4.4869	1.5100	8.7159
MVC2	−4.4855	0.5045	−3.4864	0.4100	−6.1531	0.5291	1.9820	5.4858	2.4385	4.4869	1.8898	7.1533

The values give the best fitting parameters $\alpha^{\{\text{in,out}\}}$, $\lambda^{\{\text{in,out}\}}$ and α, λ according to Eq. 16 as well as $\theta^{\{\text{in,out}\}}$, $k^{\{\text{in,out}\}}$ and θ, k according to Eq. 18 for the node in/out degree and node degree PDFs of the directed and undirected versions of the graphs, respectively. Both fitting models are qualitatively equivalent. Values in parentheses are excluded from Fig. 3

Table 6 Stretched exponential and log-normal fits of the node degree distributions for various biological neural graphs

	Stretched exponential model						Log-normal model					
	β^{in}	λ^{in}	β^{out}	λ^{out}	β	λ	μ^{in}	σ^{in}	μ^{out}	σ^{out}	μ	σ
CC1	2.0180	1.4140×10^{-3}	1.7297	3.5818×10^{-3}	1.7717	2.6172×10^{-3}	3.1299	0.6495	3.0851	0.7146	3.2057	0.7395
CC2	2.8725	3.3298×10^{-4}	1.8293	4.9790×10^{-3}	2.3767	6.7242×10^{-4}	2.7072	0.4171	2.7145	0.6526	2.9535	0.4941
CE1	1.3953	5.6656×10^{-2}	1.0540	1.0567×10^{-1}	1.9149	6.1237×10^{-3}	1.8429	0.9189	1.8158	1.1163	2.5433	0.5951
CE3	1.8187	1.5470×10^{-2}	1.6044	2.0917×10^{-2}	2.1206	3.0620×10^{-3}	2.1561	0.7096	2.2170	0.7773	2.6015	0.5483
MB1	0.7751	1.0642×10^{-1}	0.7816	8.9299×10^{-2}	1.1078	2.1859×10^{-2}	2.5100	1.5886	2.7913	1.8178	3.2205	1.3011
MC1	1.7632	1.3072×10^{-2}	1.6654	1.6043×10^{-2}	1.7744	8.4554×10^{-3}	2.2461	0.6905	2.2642	0.7655	2.5209	0.7640
MC2	3.1522	(1.6068×10^{-5})	(5.0710)	(2.2765×10^{-8})	(4.6665)	(4.3361×10^{-8})	3.4657	0.4108	3.4417	0.2109	3.5959	0.2497
MNC1	1.4838	1.9558×10^{-2}	1.8212	1.0289×10^{-2}	1.8120	7.0028×10^{-3}	2.5110	0.8737	2.3030	0.6657	2.5353	0.6771
MVC1	2.5811	1.8229×10^{-3}	2.3898	2.8284×10^{-3}	3.4395	1.1885×10^{-4}	2.3429	0.4509	2.3330	0.5116	2.5502	0.3503
MVC2	2.5811	1.8229×10^{-3}	2.3898	2.8284×10^{-3}	3.0058	3.4508×10^{-4}	2.3429	0.4509	2.3330	0.5116	2.5726	0.3887

The values give the best fitting parameters $\beta^{\{\text{in,out}\}}$, $\lambda^{\{\text{in,out}\}}$ and β, λ according to Eq. 17 as well as $\mu^{\{\text{in,out}\}}$, $\sigma^{\{\text{in,out}\}}$ and μ, σ according to Eq. 19 for the node in/out degree and node degree PDFs of the directed and undirected versions of the graphs, respectively. Values in parentheses are excluded from Fig. 3

Note that here $[D_{ij}^{\text{in}-\text{out}}]^T = D_{ji}^{\text{in}-\text{out}} = D_{ij}^{\text{out}-\text{in}}$, leaving the three independent measures of Euclidean distance in Eqs. 23–25. The above definition holds for digraphs with self-loops. If self-loops are excluded, the sum in Eqs. 23–25 runs over $k \neq \{i, j\}$. The above definitions hold for undirected graphs as well. However, due to the symmetry of the adjacency matrix in this case, we have in addition the relation $D_{ij}^{\text{in}-\text{in}} = D_{ij}^{\text{out}-\text{out}} = D_{ij}^{\text{in}-\text{out}}$, thus leaving only one independent Euclidean distance measure. Moreover, in the case of undirected graphs, $D_{ii}^{\text{in}-\text{out}} = 0$, which reflects the fact that here, for each given node, the columns and rows of the adjacency matrix are identical.

According to the notion of structural equivalence, two structurally perfectly equivalent nodes will have identical entries in their corresponding rows and columns in the adjacency matrix. With Eqs. 23 and 25, one thus expects an Euclidean distance $D_{ij}^{\text{in}-\text{in}} = D_{ij}^{\text{out}-\text{out}} = 0$. A similar conclusion can, however, not be made for $D_{ij}^{\text{in}-\text{out}}$, as the notion of perfect structural equivalence between two nodes does not require a matching pattern in the incoming connection of one node and outgoing connection of another node.

We calculated the Euclidean distance of node adjacencies, Eqs. 23–25, for the giant connected component of the investigated biological neural graphs. To statistically evaluate the distance between two nodes, we considered various subsets of the obtained $N_N \times N_N$ matrices $D_{ij}^{\alpha-\beta}$, $\alpha, \beta \in \{\text{in, out}\}$. First, $D_{ij}^{\text{in}-\text{in}}, i \neq j$ provides the Euclidean distances between the incoming edges of two different nodes. Secondly, $D_{ij}^{\text{in}-\text{out}} = D_{ji}^{\text{out}-\text{in}}$ provides the Euclidean distance between incoming and outgoing edges of two nodes, including the same node. $D_{ij}^{\text{out}-\text{out}}, i \neq j$ yields the distances between outgoing edges of two different nodes. Finally, $D_{ii}^{\text{in}-\text{out}}$ contains the Euclidean distances between incoming and outgoing edges of the same node. For each of these subsets of Euclidean distances, we calculated the mean, standard deviation, minimum and maximum value, first and third quartile, and median.

Representative examples of the Euclidean distance statistical analysis are shown in Fig. 4 (left; a complete representation of Euclidean distances can be found in the Supplementary Information, Data Tables). We found that in digraphs, the mean and median of $D_{ij,i\neq j}^{\text{in}-\text{in}}$, $D_{ij}^{\text{in}-\text{out}}$ and $D_{ij,i\neq j}^{\text{out}-\text{out}}$, and in undirected graphs, the mean and median of $D_{ij,i\neq j}^{\text{in}-\text{in}}$ and $D_{ij}^{\text{in}-\text{out}}$ are almost identical, a behavior expected from a random, i.e., independent, distribution of edges in different nodes. This behavior was shared among all investigated graphs (Fig. 4, middle and right).

In not self-looped random graphs with connectedness Co, $a_{ij} = 1$ with probability Co (see above). Thus, two adjacencies with $a_{ij} = 1$ and $a_{mn} = 0$ will occur with probability $Co(1 - Co)$. As the latter adjacency relations constitute the only contributions to the Euclidean distance,

$$D_{ij}^{\alpha-\beta} = \sqrt{2N_N Co(1-Co)}\,, \tag{26}$$

$\alpha, \beta \in \{\text{in, out}\}$, for Erdős-Rényi graphs (Fig. 4, left, red dotted). The relative deviation

$$\Delta^{ER} D^{\alpha-\beta} = \frac{\left|\overline{D_{ij}^{\alpha-\beta}} - \overline{D_{ij}^{ER\,\alpha-\beta}}\right|}{\overline{D_{ij}^{ER\,\alpha-\beta}}}\,, \tag{27}$$

where $\overline{D_{ij}^{\alpha-\beta}}$ and $\overline{D_{ij}^{ER\,\alpha-\beta}}$ denote the mean of the Euclidean distances $D_{ij}^{\alpha-\beta}$ for a given graph and its corresponding Erdős-Rényi graph was found to be almost zero (Fig. 4, right, blue end of color range). This observation provides further

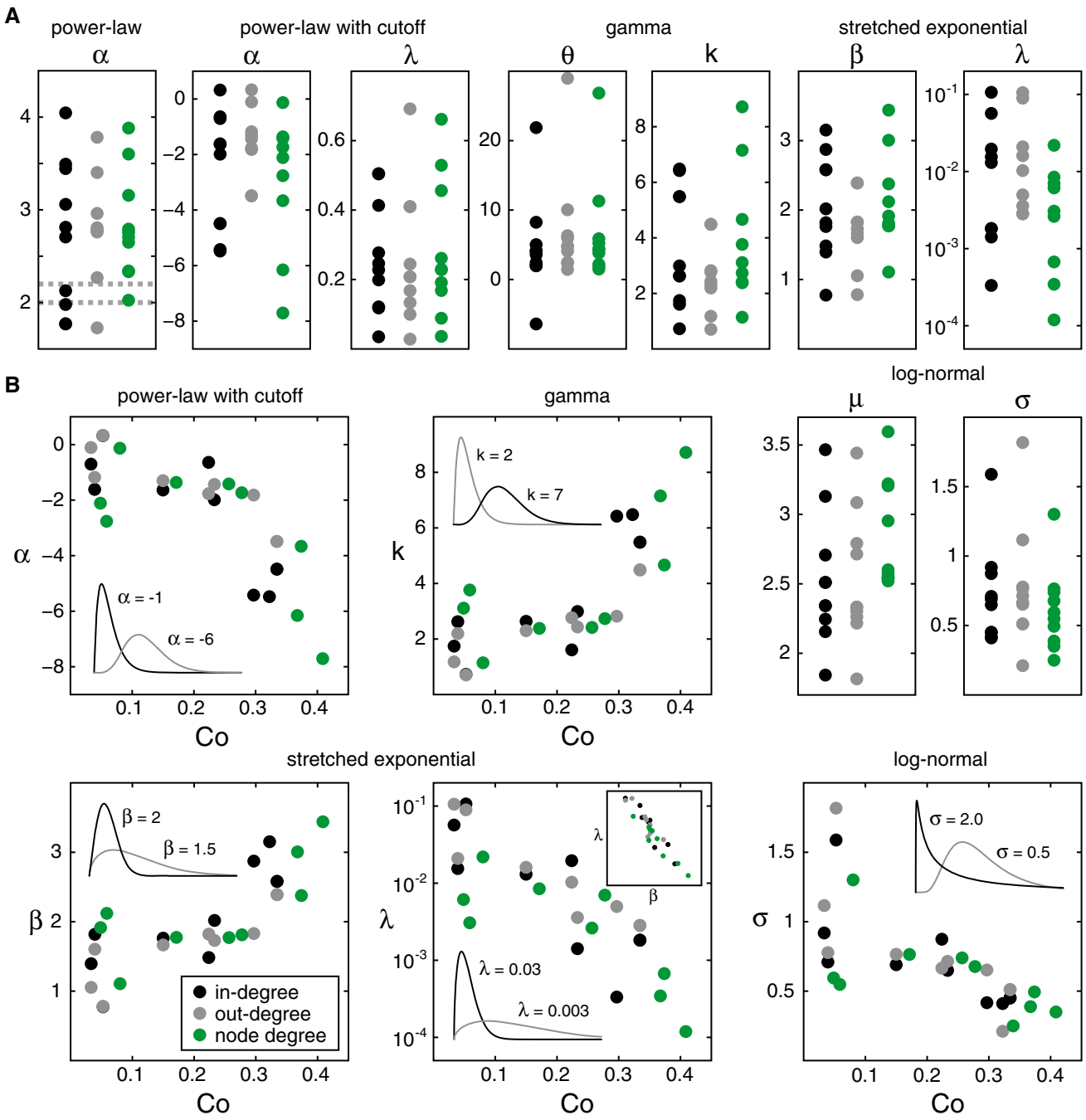

Fig. 3 Best parameter fits for node degree PDFs of various biological neural graphs. **a** Distribution of best fitting parameters for various models (Eqs. 15–19). **b** For some models, a correlation between fitted model parameters and graph connectedness Co was found. The *insets* show the qualitative role of the parameter on the model distribution

evidence suggesting that the distribution of edges in different nodes of biological neural graphs follows a simple random pattern.

When considering in edges and out edges of the same node, however, we find that this is not the case. Here, by definition, one expects $D_{ii}^{\text{in-out}} = 0$ in undirected graphs, whereas in digraphs, if assuming a random distribution of edges, $D_{ii}^{\text{in-out}}$ should take values in accordance with Eq. 26. Our analysis reveals that the mean and median of $D_{ii}^{\text{in-out}}$ is significantly lower than expected from a pure random assignment of edges, leading to $\Delta^{ER} D^{\alpha-\beta}$ significantly larger than zero (Fig. 4, right). The latter suggests that the distribution

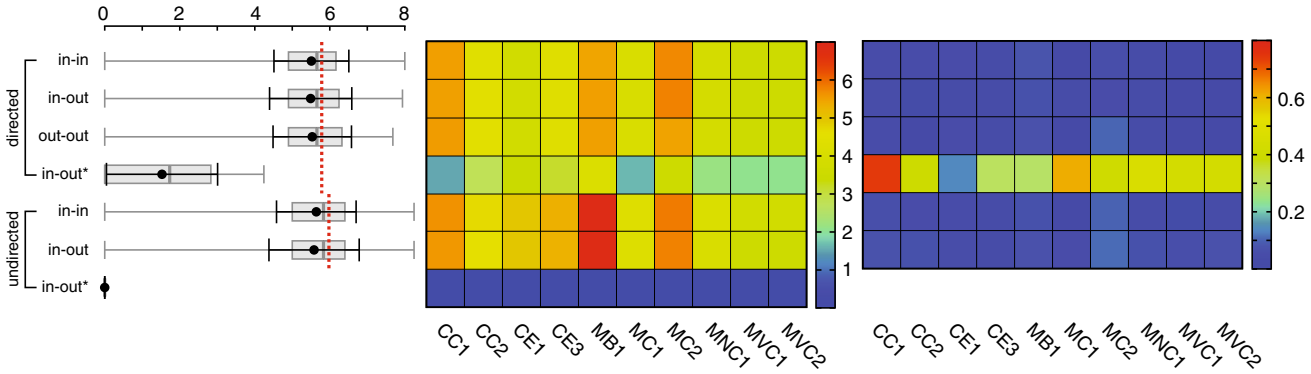

Fig. 4 Structural equivalence analysis of neural graphs. *Left* Representative box plots (*gray*), mean and standard deviation (*black*) of Euclidean distance measures, Eqs. 23–25 (in-in: $D^{\text{in-in}}_{ij,i\neq j}$, in-out: $D^{\text{in-out}}_{ij}$, out-out: $D^{\text{out-out}}_{ij,i\neq j}$, in-out*: $D^{\text{in-out}}_{ii}$), for directed and undirected versions of CC1. The *red line* indicates the value expected for corresponding Erdős-Rényi graph models. *Middle* Mean of Euclidean distance of node adjacencies for all investigated biological neural graphs. *Right* relative deviation $\Delta^{ER} D^{\alpha-\beta}$ (Eq. 27) from corresponding Erdős-Rényi graph models

of edges in biological neural graphs, although being consistent with a minimal random model, exhibits a significant correlation between in and out edges for individual nodes.

Finally, as a confirmation and extension of the structural equivalence analysis presented in this section, we replicated these results using a similar measure of structural equivalence (see Supplementary Information, Structural Correlation Analysis) and then performed a thorough analysis of assortativity in the neural graphs under consideration (see Supplementary Information, Graph Assortativity). In the latter analysis, we observe a weak dissortative tendency throughout the graphs considered, indicating that nodes make connections, on average, to others with the same, or slightly higher, degree. The magnitudes observed here, however, are smaller than those previously reported (Newman 2002), indicating that the magnitude of this effect is generally weaker than that of the other effects considered in this study.

7 Discussion

In this work, we have completed a detailed comparative analysis of several networks fundamental to the application of graph theory in neuroscience, while challenging the pertinence of several established graph-theoretic concepts in the context of neural connectivity patterns. Contrary to many results reported in the neuroscientific literature, the biological graphs studied here show in many measures a consistency with randomness, as opposed to a consistency with simple models of graph construction, such the scale-free graph (Barabási and Albert 1999).

Firstly, we found that fits of the node degree distributions are in accordance with a gamma model, supporting the idea of a simple local mechanism responsible for generating neural graphs. Notably, the degree distributions observed here differ from those of the Erdős-Rényi graph widely used in computational neuroscience, which may have an effect on the dynamics of the networks modeled (Roxin 2011). Secondly, the Euclidean distance of node adjacencies and node degree correlations was observed to be consistent with an independent random distribution of node connections for different nodes, but with strong correlations between incoming and outgoing connections for the same node. Taken together, these two observations indicate that at the first and second orders of connectivity, neural graphs can be largely described by a random process at the level of single node, coupled with a bias toward reciprocal connections.

This consistency with randomness is not fully surprising, as any conceivable mechanism that could give rise to such a structural makeup will be subject to fewer constraints compared to graph models conceived to fulfill a specific set of structural requirements. In particular, for scale-free graphs, various generating algorithms have been proposed, ranging from static models to evolving models more closely reflecting processes found in nature. Typically, static models construct scale-free graphs by imposing global constraints, such as the scale-free node degree distribution itself (Aiello et al. 2000; Chung and Lu 2002) or fitness (Goh et al. 2001; Caldarelli et al. 2002). The most prominent model of evolving scale-free graphs is the classical growth and preferential attachment model (Barabási and Albert 1999), originally studied as "Matthew effect" (Merton 1968) or "cumulative advantage" (Solla Price 1965), and its generalizations (Dorogovtsev et al. 2000; Ravasz and Barabási 2003). Here, the probability of linking two nodes is (linearly) proportional to the actual node degree, requiring the generating algorithm to keep track of all node degrees and, thus, nonlocal information about the graph at any stage of its construction. Further studies on generative mechanisms for scale-free graphs share this requirement, including models utilizing accelerated growth (Dorogovtsev and Mendes 2001), which require the knowledge of the network size at any stage of construction, finite

node memory models (Klemm and Eguíluz 2002), which require the knowledge of the activity state of each node in the graph, or duplication and divergence models (Goh et al. 2002; Vázquez et al. 2003), which require copies of arbitrarily selected graph nodes. Importantly, all these models are crucially dependent on information about the degree distribution of the network to be present at the individual node during graph construction, a requirement not necessary for generating graphs consistent with the observations presented in this study.

Perhaps the most surprising result from this work is that certain graph-theoretic quantities, such as the form of the degree distribution and the abundance of reciprocal connections, seem to be conserved across networks spanning vastly different spatial scales and functional roles. The conserved form of a unimodal, slightly skewed degree distribution may reflect a locally random process, where the elements of circuit construction interact with their immediate surroundings. At the microscopic level, such deviation from the classical random model may have important consequences for the state space of the dynamics in spiking networks, as for example addressed by interpolating between binomial and power-law degree distributions in Roxin [2011]. Reciprocal connectivity patterns have been discussed previously in studies ranging from local cortical microcircuits (Song et al. 2005; Ko et al. 2011) to areal connectivity (Felleman and Essen 1991; Essen 2005). In this work, we present the first systematic analysis to confirm the prevalence of this reciprocal connectivity in neural graphs spanning multiple spatial scales, and to exclude the generality of the other two-edge connectivity patterns in a thorough fashion.

In recent years, deviations from the occurrence of two-edge motifs in purely random connectivity have begun to receive interest in the computational literature and have been found to have consequences on macroscopic properties of the network dynamics (Hennequin et al. 2012; Pernice et al. 2013), specifically in relation to synchronization (Zhao et al. 2011; Hu et al. 2013). Furthermore, in a previous experimental study (Song et al. 2005), the implications and possible relations of two-edge motifs to known plasticity mechanisms were emphasized through the analysis of the synaptic weight distribution of bidirectionally connected neuron pairs. Emergence of bidirectional connections resulting from spike-timing-dependent plasticity has also been observed in simulations of small-scale networks (Clopath et al. 2010; Bourjaily and Miller 2011; Vasilaki and Giugliano 2012), and studied mathematically in Gilson et al. [2009] and Babadi and Abbott [2013], though this remains to be thoroughly studied in large-scale simulations of networks with biologically realistic activity states.

Finally, we note that some studies have critiqued the validity of the random graph null hypothesis (Artzy-Randrup et al. 2004), noting specifically that the spatial nature of certain networks could confound the statistical comparison to a random graph for measures taken from real-world graphs (as in Milo et al. 2002). In this work, we compare the measures for structural similarity to those of an equivalent random graph; however, we note that it is additionally possible with such an analysis to detect statistical differences between subsets of the Euclidean distance and structural correlation matrices (see Fig. 4, left panel). Such a comparison, free from a random graph null hypothesis, can be explored in future work.

In conclusion, while there has been a great interest in recent years in the possibility that structural graphs share important features in common with abstract models of graph generation, we hope that, in future work, not only greater care will be taken in the support of such claims, but also more measurement and theory will be developed toward the discovery of new, specific graph-theoretic models with explanatory power able to meet the challenges of the next generation of large-scale experimental network reconstructions.

Acknowledgments The authors wish to thank OD Little for inspiring comments and A Destexhe for continuing support. Work supported by the CNRS and the European Community (BrainScales project, FP7-269921). LM is a PhD fellow from École des Neurosciences de Paris (ENP).

References

Aiello W, Chung F, Lu L (2000) A random graph model for massive graphs. In: Proceedings of the 32nd annual ACM symposium on theory of computing, association of computing machinery. New York, pp 171–180

Albert R, Barabási A-L (2002) The statistical mechanics of complex networks. Rev Mod Phys 74:47–97

Albert R, Jeong H, Barabási A-L (1999) Diameter of the world-wide web. Nature 401:130–131

Alm E, Arkin A (2003) Biological networks. Curr Opinion Struct Biol 13:193–202

Alon U (2003) Biological networks: the tinkerer as engineer. Science 301:1866–1867

Amaral LAN, Ottino JM (2004) Complex networks. Eur Phys J B38:147–162

Artzy-Randrup Y, Fleishman SJ, Ben-Tal N, Stone L (2004) Comment on "Network motifs: simple building blocks of complex networks" and "Superfamilies of evolved and designed networks". Science 305:1107

Babadi B, Abbott LF (2013) Pairwise analysis can account for network structures arising from spike-timing dependent plasticity. PLoS Comput Biol 9:e1002906

Barabási AL, Albert R (1999) Emergence of scaling in random networks. Science 286:509–512

Barabási A-L, Bonabeau E (2003) Scale-free networks. Scientific American, pp 50–59

Barabási A-L, Oltvai Z (2004) Network biology: understanding the cell's functional organization. Nature Rev Gen 5:101–113

Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang D-U (2006) Complex networks: structure and dynamics. Phys Rep 424:175–308

Bourjaily MA, Miller P (2011) Excitatory, inhibitory, and structural plasticity produce correlated connectivity in random networks trained to solve paired-stimulus tasks. Front Comput Neurosci 5:37

Braitenberg V, Shüz (1998) Cortex: statistics and geometry of neuronal connectivity (revised, 2nd edition of anatomy of the cortex—statistics and geometry, 1998). Springer, Berlin
Bray D (2003) Molecular networks: the top–down view. Science 301:1864–1865
Bullmore E, Sporns O (2009) Complex brain networks: graph theoretical analysis of structural and functional systems. Nat Rev Neuro 10:186–198
Caldarelli G, Capocci A, De Los Rios P, Muñoz MA (2002) Scale-free networks from varying vertex intrinsic fitness. Phys Rev Lett 89:258702
Chen BL, Hall DH, Chklovskii DB (2006) Wiring optimization can relate neuronal structure and function. PNAS 103:4723–4728
Choe Y, McCormick BH, Koh W (2004) Network connectivity analysis on the temporally augmented C. elegans web: a pilot study. Soc Neurosci Abstr 30:921.9
Chung F, Lu L (2002) The average distances in random graphs with given expected degrees. Proc Natl Acad Sci USA 99:15879–15882
Clauset A, Shalizi CR, Newman MEJ (2009) Power-law distributions in empirical data. SIAM Rev 51:661703
Clopath C, Büsing L, Vasilaki E, Gerstner W (2010) Connectivity reflects coding: a model of voltage-based STDP with homeostasis. Nature Neurosci 13:344–352
Costa LF, Rodrigues FA, Travieso G, Villas Boas PR (2007) Characterization of complex networks: a survey of measurements. Adv Phys 56:167–242
de Solla Price DJ (1965) Networks of scientific papers. Science 149:510–515
Del Genio CI, Kim H, Toroczkai Z, Bassler KE (2010) Efficient and exact sampling of simple graphs with given arbitrary degree sequence. PLoS One 5:e10012
Diestel R (2000) Graph theory. Springer, New York
Dorogovtsev SN, Mendes JFF (2001) Giant strongly connected component of directed networks. Phys Rev E 63:025101
Dorogovtsev SN, Mendes J (2002) Evolution of networks. Adv Phys 51:1079–1187
Dorogovtsev SN, Mendes JFF, Samukhin AN (2000) Structure of growing networks: exact solution of the Barabási-Albert model. Phys Rev Lett 85:4633–4636
Eguíluz VM, Chialvo DR, Cecchi GA, Baliki M, Apkarian AV (2005) Scale-free brain functional networks. Phys Rev Lett 94:018102
Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex 1:1–47
Foster JG, Foster DV, Grassberger P, Paczuski M (2010) Edge direction and the structure of networks. Proc Natl Acad Sci USA 107:10815
Garlaschelli D, Loffredo MI (2004) Patterns of link reciprocity in directed networks. Phys Rev Lett 93:268701
Gilson M, Burkitt AN, Grayden DB, Thomas DA, van Hemmen JA (2009) Emergence of network structure due to spike-timing-dependent plasticity in recurrent neuronal networks IV: structuring synaptic pathways among recurrent connections. Biol Cybern 101:427–444
Goh KI, Kahng B, Kim D (2001) Universal behavior of load distribution in scale-free networks. Phys Rev Lett 87:278701
Goh KI, Kahng B, Kim D (2002) Fluctuation-driven dynamics of the internet topology. Phys Rev Lett 88:108701
Hennequin G, Vogels TP, Gerstner W (2012) Non-normal amplification in random balanced neuronal networks. Phys Rev E 86:011909
Honey CJ, Kötter R, Breakspear M, Sporns O (2007) Network structure of cerebral cortex shapes functional connectivity on multiple time scales. Proc Natl Acad Sci USA 104:10240–10245
Hu Y, Trousdale J, Krešimir J, Shea-Brown E (2013) Motif statistics and spike correlations in neuronal networks. J Stat Mech P03012
Johnson S, Torres JJ, Marro J, Muñoz MA (2010) Entropic origin of disassortativity in complex networks. Phys Rev Lett 104:108702
Kaiser M, Hilgetag CC (2006) Non-optimal component placement, but short processing paths, due to long-distance projections in neural systems. PLoS Comput Biol 2:e95
Ko H, Hofer SB, Pichler B, Buchanan KA, Sjöström PJ, Mrsic-Flogel TD (2011) Functional specificity of local synaptic connections in neocortical networks. Nature 473:87–91
Kim H, Del Genio CI, Bassler KE, Toroczkai Z (2012) Constructing and sampling directed graphs with given degree sequences. New J Phys 14:023012
Klemm K, Eguíluz VM (2002) Highly clustered scale-free networks. Phys Rev E65:36123
Kötter R (2004) Online retrieval, processing, and visualization of primate connectivity data from the CoCoMac database. Neuroinformatics 2:127–144
Lima-Mendez G, van Helden J (2009) The powerful law of the power law and other myths in network biology. Mol BioSyst 5:1482–1493
Merton RK (1968) The Matthew effect in science. Science 159:56–63
Milgram S (1967) The small World problem. Psychology today, May 1967, pp 60–67
Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U (2002) Network motifs: simple building blocks of complex networks. Science 298:824–827
Modhaa DA, Singh R (2010) Network architecture of the long-distance pathways in the macaque brain. PNAS 107:1348513490
Newman MEJ (2002) Assortative mixing in networks. Phys Rev Lett 89:208701
Newman MEJ (2003) The structure and function of complex networks. SIAM Rev 45:167–256
Newman MEJ (2010) An introduction., NetworksOxford University Press, Oxford
Newman MEJ, Forrest S, Balthrop J (2002) Email networks and the spread of computer viruses. Phys Rev E 66:035101
Pernice V, Deger M, Cardanobile S, Rotter S (2013) The relevance of network micro-structure for neural dynamics. Front Comput Neurosci 7:72
Ravasz E, Barabási AL (2003) Hierarchical organization in complex networks. Phys Rev E 67:026112
Roxin A (2011) The role of degree distribution in shaping the dynamics in networks of sparsely connected spiking neurons. Font Comput Neurosci 5:8
Rudolph-Lilith M, Destexhe A, Muller LE (2012) Structual Vulnerability of the Nematode Worm Neural Graph. arXiv:1208.3383v1 [cond-mat.dis-nn]. Available: http://arxiv.org/abs/1208.3383v1
Scannell JW, Burns GA, Hilgetag CC, O'Neil MA, Young MP (1999) The connectional organization of the cortico-thalamic system of the cat. Cereb Cortex 9:277–299
Serrano MA, Boguñá M (2003) Topology of the world trade web. Phys Rev E 68:015101
Song S, Sjöström PJ, Reigl M, Nelson A, Chklovskii DB (2005) Highly nonrandom features of synaptic connectivity in local cortical circuits. PLoS Biol 3:e68
Sporns O (2002) Graph theory methods for the analysis of neural connectivity patterns. In: Kotter R (ed) Neuroscience databases. A practical guide. Kluwer, New York, p 171186
Sporns O, Ktter R (2004) Motifs in brain networks. PLoS Biol 2:19101918
Sporns O, Tononi G (2002) Classes of network connectivity and dynamics. Complexity 7:28–38
Sporns O, Tononi G, Edelman GM (2000) Theoretical neuroanatomy: relating anatomical and functional connectivity in graphs and cortical connection matrices. Cereb Cortex 10:127–141
Sporns O, Zwi J (2004) The small world of the cerebral cortex. Neuroinformatics 2:145–162
Stephan KE, Kamper L, Bozkurt A, Burns GA, Young MP, Kötter R (2001) Advanced database methodology for the collation of connec-

tivity data on the Macaque brain (CoCoMac). Philos Trans R Soc Lond B Biol Sci 356:1159–1186
Stumpf MPH, Porter MA (2012) Critical truths about power laws. Science 335:665–666
Van Essen DC (2005) Corticocortical and thalamocortical information flow in the primate visual system. Prog Brain Res 149:173–185
Varshney LR, Chen BL, Paniaqua E, Hall DH, Chklovskii DB (2011) Structural properties of the C. elegans neuronal network. PLoS Comput Biol 7:e1001066
Vasilaki E, Giugliano M (2012) Emergence of connectivity patterns from long-term and short-term plasticities. In: Villa AEP et al (eds) Artificial neural networks and machine learning ICANN 2012. Springer, New York, pp 193–200
Vázquez A, Flammini A, Maritan A, Vespignani A (2003) Modeling of protein interaction networks. Complexus 1:38–44
Wasserman S, Faust K (1994) Social network analysis. Cambridge University Press, Cambridge
Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. Nature 393:440–442
White JG, Southgate E, Thompson JN, Brenner S (1986) The structure of the nervous system of the nematode caenorhabditis elegans. Phil Trans R Soc Lond 314:1–340
Young MP (1993) The organization of neural systems in the primate cerebral cortex. Proc R Soc Lond B 252:13–18
Zhao L, Beverlin B II, Netoff T, Nykamp DQ (2011) Synchronization from second order network connectivity statistics. Front Comput Neurosci 5:28