Spatial embedding of structural similarity in the cerebral cortex

H. Francis Song[a], Henry Kennedy[b,c], and Xiao-Jing Wang[a,d,1]

[a]Center for Neural Science, New York University, New York, NY 10003; [b]Stem Cell and Brain Research Institute, INSERM U846, 69500 Bron, France; [c]Université de Lyon, Université Lyon I, 69003 Lyon, France; and [d]NYU-ECNU Institute of Brain and Cognitive Science, NYU Shanghai, Shanghai, China

Recent anatomical tracing studies have yielded substantial amounts of data on the areal connectivity underlying distributed processing in cortex, yet the fundamental principles that govern the large-scale organization of cortex remain unknown. Here we show that functional similarity between areas as defined by the pattern of shared inputs or outputs is a key to understanding the areal network of cortex. In particular, we report a systematic relation in the monkey, human, and mouse cortex between the occurrence of connections from one area to another and their similarity distance. This characteristic relation is rooted in the wiring distance dependence of connections in the brain. We introduce a weighted, spatially embedded random network model that robustly gives rise to this structure, as well as many other spatial and topological properties observed in cortex. These include features that were not accounted for in any previous model, such as the wide range of interareal connection weights. Connections in the model emerge from an underlying distribution of spatially embedded axons, thereby integrating the two scales of cortical connectivity—individual axons and interareal pathways—into a common geometric framework. These results provide insights into the origin of large-scale connectivity in cortex and have important implications for theories of cortical organization.

cortex | large scale | complex network | monkey | model

The cerebral cortex can be divided into a number of distinct areas according to anatomy and function. Understanding the complex pattern of connections among these areas is necessary for elucidating the structural basis of distributed processing in the brain, yet the principles that govern this large-scale organization are not fully understood. In particular, although cortical organization has been characterized extensively within the framework of "complex networks" (1, 2), there are few generative models that explain how the observed features of areal connectivity may arise (3–5).

In contrast to the large and sparsely connected architecture of many networks (including the neuronal network of the brain), cortical areal networks are relatively small and densely connected: the mouse neocortex consists of roughly 40 areas per hemisphere with ~50% of the possible connections present (6, 7), and the macaque neocortex consists of roughly 100 areas per hemisphere (8) with ~60% of the connections present (9). In such networks, the properties that define conventional complex networks—the degree distribution (number of areas connected to an area), average path length (smallest number of connected steps between a pair of areas), and clustering (density of connections among areas connected to the same area)—are not very informative. For instance, in a network with overall connection density p_0 greater than 0.5, almost all pairs of areas are connected by a path of length one or two, and the average path length is close to $1 \cdot p_0 + 2 \cdot (1 - p_0) = 2 - p_0$ whether the structure is random, regular, or in between (10). As a result, canonical complex networks such as the Barabási–Albert model of "scale-free" networks (11) and the Watts–Strogatz model of "small-world" networks (12), which were developed specifically to account for the behavior of these quantities, have limited applicability to the areal network of cortex (5, 7, 10).

Fundamentally, existing complex network models provide an incomplete description of the cortical areal network because they neglect the underlying spatial structure that shapes connectional topology (4, 5, 13, 14). Recent data indicate that interareal pathways consist of spatially heterogeneous axonal projections whose distribution exhibits two striking properties: first, the number of axons that project from one area to another varies over several orders of magnitude across cortex (7, 15), and second, connections between different cortical areas correspond to only a small fraction of all axons, most of which project to within the same area and contribute to the local circuit (9, 16). Thus, a proper description of the cortical network requires integrating interareal pathways and the axons that compose them into a common framework encompassing both global and local structure.

To address these challenges, we present an analysis of cortical connectivity that reveals a key organizing principle common to three mammalian species with widely different brain sizes, namely monkeys, humans, and mice. Our starting point was that connectional similarity of pairs of areas, defined by their shared inputs or outputs ("connectional fingerprints"), reflects the functional organization of cortex (17–19). A major finding of this work is a systematic relation between the occurrence of connections from one area to another and their similarity distance as defined by the amount of shared connections. This relation, which is observed in all datasets we examined, has its basis in the dependence of connections on wiring distance (4). We developed a simple, spatially embedded random network model that gives rise to the structures revealed by our analysis. Notably, to our knowledge, it is the first generative model that relates binary features to the underlying weighted structure through a well-defined geometric coarse-graining.

The model we propose reproduces numerous spatial and topological properties of the macaque cortex. These include graph-theoretic measures such as degree sequence, path length, clustering coefficient, and motif distribution, as well as features that were not accounted for in any previous model, such as the wide

Significance

The cerebral cortex can be divided into a number of distinct areas on the basis of anatomy and function. Understanding the complex pattern of connections among these areas is essential to uncovering how the brain performs its distributed computations. We report a systematic relation between the connectivity and functional similarity of cortical areas in the monkey, human, and mouse cortex. Motivated by observations that the cortical areal network is densely connected and that connections have a strong dependence on wiring length, we introduce a spatially embedded, generative model of the areal network that accounts for many observed features of cortical connectivity.

range of interareal connection weights. Because our model cortex is embedded in a spatial continuum, moreover, it can be partitioned into regions with an arbitrary spatial resolution. Such an approach makes it possible to study the same axonal network using alternative methods of parcellation, an important issue for modeling data obtained from diffusion-based tractography at varying spatial scales (20). Taken together, our results provide a promising direction for the investigation of the structure and origin of areal connectivity in cortex.

Results

Similarity Distance. We first investigated the binary connection structure of the macaque cortex, the proper understanding of which is critical to developing a theory of its weighted connectivity. The binarized intrahemispheric connection matrix C, where $C_{ts} = 1$ if there is a directed connection from source area s to target area t and 0 otherwise (Fig. 1A), was obtained from recent retrograde tracer injections in 29 representative target areas and labeling in all 91 source areas (9). The connectivity is dense: 62% of the possible interareal connections are present, which is a much higher proportion than in previously available data [e.g., from CoCoMac (17, 21)] as a result of the consistency of the cortical parcellation, comprehensive hemispheric examination, and optimized tract-tracing methods (22).

We used "cosine similarity distance" to quantify the amount of shared outputs between areas s and s':

$$D_{ss'} = 1 - \frac{\sum_t C_{ts} C_{ts'}}{\sqrt{\sum_t C_{ts}^2} \sqrt{\sum_t C_{ts'}^2}}. \qquad [1]$$

Here the sum over t runs through the target areas excluding s and s', i.e., $D_{ss'}$ does not depend on $C_{ss'}$ and $C_{s's}$. Thus, the similarity distance between two areas is not an explicit function of whether connections between them are present.

Because the connection matrix in Fig. 1A reflects injections into a subset of all cortical areas, it is a priori unclear how reliable the resulting similarity distances are. To ascertain this, we computed the analogously defined input similarity distances, whose "true" values can be determined from Fig. 1A, but only between pairs of injected areas. A comparison of these values with those obtained by sampling (*SI Appendix*, Fig. S1) suggests that output similarity

distances computed from the sampled target areas are representative of the true distances we would obtain if the full connection matrix were available. Moreover, cosine similarity distance is only one of many possible measures of shared connections, or "structural equivalence" (23, 24), and our results are qualitatively the same for other normalized measures (*SI Appendix*).

Structural Similarity and the Functional Organization of Cortex. As a measure of the amount of shared outputs, the similarity distance between two areas indicates their functional similarity. This is revealed in a classical multidimensional scaling (MDS) analysis of "similarity space," which places areas in a Euclidean space most compatible with the interareal distance relations given by Eq. 1 (Fig. 1B). Although we must be cautious in drawing detailed conclusions from such visualizations, they capture the broad functional layout of cortex. For instance, in this map, many sensory and motor areas are grouped by function and located along the periphery, with primary sensorimotor areas at the edge and areas involved in "higher-level" processing closer to the center. The most highly connected—and interconnected—areas are placed at the origin of the MDS map due to the requirement that they be close to many areas. Interestingly, the center of the map also comprises areas from physically distant parts of cortex, showing that the relation between similarity distance and anatomical distance is not trivial [their correlation is significant ($P < 0.01$) but moderate, Spearman's $r = 0.42$ (see Fig. 3A, Inset)].

The definition in Eq. 1 can also be applied to the weighted connection matrix from which the binarized matrix was derived. The result (*SI Appendix*, Fig. S2) is a more clustered map with fewer areas in the center, which can be understood from the fact that the connection weights vary over five orders of magnitude across cortex (15). Thus, the similarity distances and resulting MDS map are dominated by the strongest connections, many of them between nearby areas (19).

We applied the same analysis to a recently published connection matrix of the mouse cortex (6). The known functional modules of the mouse cortex are again grouped in the MDS map of similarity space, and suggest certain homologies with the macaque cortex (Fig. 1C; see *SI Appendix*, Fig. S3 for all modules). *SI Appendix*, Fig. S4 presents the corresponding map of human cortical connectivity obtained from diffusion spectrum
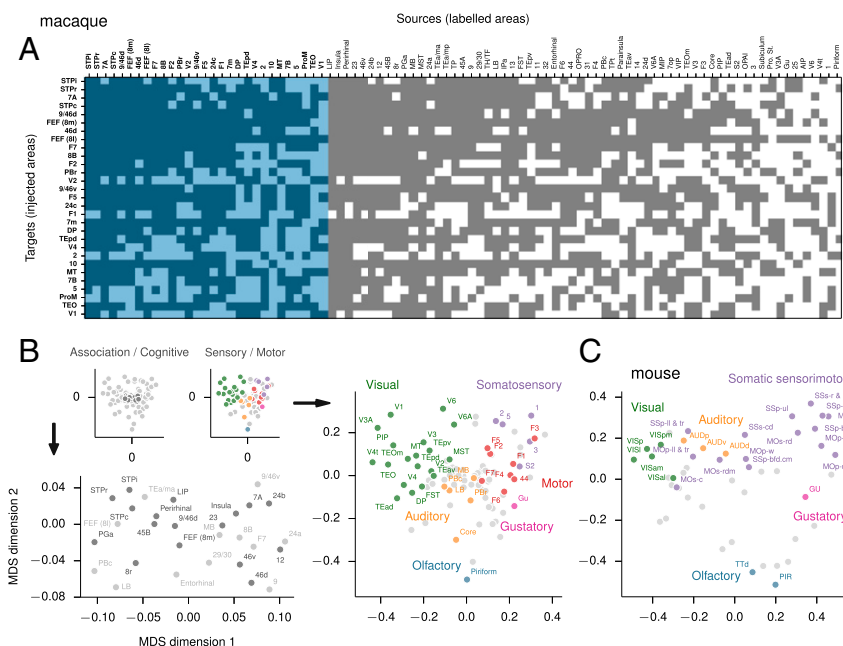
Fig. 1. Similarity distances reflect the functional organization of cortex. (A) Binarized intrahemispheric connection matrix of the macaque cortex. The 29 × 29 edge-complete subnetwork is highlighted in blue. (B) First two dimensions of the classical MDS analysis of similarity distances (23), accounting for ~46% of the variance. Areas closest to the origin ("association and cognitive" areas) and a subset of those on the periphery ("sensory and motor" areas) are highlighted to emphasize the overall organization of the map, but these distinctions are not meant to be definitive. For the former, lighter gray indicates areas that are closest to the origin only in the 2D projection and not in the full MDS coordinate system. Note the difference in scale. (C) MDS map (47% of variance) of mouse cortex; see the original reference (6) for a list of areas and details of subnetwork and module assignments. Colors were selected to highlight homologies with the macaque cortex in B.

imaging (DSI) (25), for which we can additionally incorporate interhemispheric connections.

Similarity Distance and Connectivity. It was previously observed that similarity distances are correlated with the presence of connections (26), which is intuitively understood from the fact that areas separated by small similarity distances tend to be functionally similar, and functionally similar areas are likely to be connected. Indeed, we found that a systematic relation exists between the occurrence of a connection from one cortical area to another and the similarity distance between them.

We established this in two different ways (Fig. 2A, *Left*): first, by binning the similarity distances and computing for each bin the fraction of all possible connections between pairs of areas that are present in the data, and second, by applying binomial regression with a probability of connection—a modified logistic function—that depends on the similarity distance (*SI Appendix*). This relation is a natural generalization of clustering, which quantifies the connectedness of a set of areas having one shared connection (12) but is uninformative in a highly connected network. The results are in strong contrast to an uncorrelated model such as the Erdős–Rényi (ER) random network (red), in which every connection is present with an independent probability that is trivially a constant function of the similarity distance. Furthermore, similarity distances in the macaque cortex have a smooth and broad distribution (*Inset*). This is again in contrast to an ER network, for which the distribution can be calculated theoretically (*SI Appendix*) and is narrowly peaked around the mean because every area looks equally similar to every other area (*Inset*, red).

An interesting feature of the connection matrix of Fig. 1A, specifically the edge-complete subnetwork for which both inputs and outputs are known, is that many areal pairs are not reciprocally connected (9). If we assume that connections between two areas with similarity distance d occur with independent probability $\hat{p}(d)$ in the two directions, then the relation in Fig. 2A (*Left*) can be used to predict the occurrence of unidirectional $[2\hat{p}(1-\hat{p})]$ and reciprocal $(\hat{p}^2)$ connections (*Right*). Moreover, although we have focused on output similarity distances to include all 91 areas in the analysis, Fig. 2A also holds if input similarity distances are used instead (Fig. 2B).

These results are not unique to the macaque cortex. In data from human DSI (25), the relation between connectivity and similarity distance is qualitatively similar (Fig. 2C). Although the distributions of similarity distances differ slightly (*Inset*), the discrepancy can be largely accounted for by first, symmetrizing the edge-complete subnetwork of the macaque connection matrix to account for the absence of directional information in DSI, and second, removing its weakest connections to match the overall density of connections in the human data. Here the weakest connections were defined as those with the smallest number of neurons labeled in the retrograde tracer study of the macaque cortex. In contrast, removing randomly chosen connections leads to a profile similar to that of the ER network in Fig. 2A (red curves).

The corresponding analysis for a connection matrix derived from the CoCoMac database of macaque connectivity (17, 21) likewise shows that discrepancies are attributable to the absence of weak, long-distance connections (Fig. 2D) (19, 22). In this case, the distribution of similarity distances is narrowly peaked near $d = 1$, in contrast to the broad distribution of similarity distances observed in the recent macaque (Fig. 2A, *Inset*), human (Fig. 2C, *Inset*), and mouse (Fig. 2E, *Inset*) data. Comparable relations between the proportion of connections and similarity distance also are found in the mouse cortex (Fig. 2E); here the availability of the full, directed connection matrix confirms that inputs and outputs have similar statistics (see also *SI Appendix,* Fig. S5), which is somewhat surprising given the presence of many nonreciprocally connected pairs of areas.

The physical origin of the relation between connectivity and similarity distance may be partly inferred by noting that the occurrence of connections has, qualitatively, the same dependence on interareal, center-to-center wiring distance (Fig. 3A) as it does on similarity distance (Fig. 3C). This is somewhat obscured in the data in part due to the difficulty of estimating wiring distances through the white matter (as opposed to Euclidean distances between the centers), particularly for spatially distant areas. However, the simultaneous presence of both relations—which is also the case in other types of spatial networks (*SI Appendix*, Fig. S6)—suggests that the dependence on similarity distance is intuitively related to the dependence on wiring distance (Fig. 3A, *Inset*): cortical areas tend to form connections with spatially

Fig. 2. Proportion of connections and similarity distance. (A, *Left*) Symbols show the number of present connections divided by the number of possible connections in each distance bin. The ML fit (solid blue) was obtained from binomial regression of unbinned data (*SI Appendix*). The same relation for the equivalent ER network is shown in red. (*Inset*) Distribution of similarity distances fit to the distribution of distances in an ellipsoid with radii $(1, \lambda_1, \lambda_2)/2$. The theoretical result for the ER network (*SI Appendix*) is shown in red. (*Right*) Symbols show binned estimates of the proportion of unidirectionally and reciprocally connected pairs in the edge-complete subnetwork as a function of similarity distance. Solid lines are predictions from the ML curve in A. Self-connections are considered reciprocal. (B) Same as A but for input similarity distances, as indicated by the empty symbols. (C) Same as A, *Left*, but for the left and right hemispheres of the human cortex (25). "Matched macaque" is the result of symmetrizing and removing the weakest connections to match the overall proportion of connections in the human. The solid line is the ML fit to the matched macaque. (D) Same as A, *Left*, but for a 71-area connection matrix derived from CoCoMac (17, 21). Recent macaque data were matched to the same connection density as the CoCoMac network by removing the weakest connections. (E) Same as A but for the mouse cortex (6), showing that the proportion of connections as functions of output (filled symbols) and input (empty symbols) similarity distances are similar. The solid blue line is the ML fit to the former. (*Inset*) The distributions of output (shaded) and input (solid line) similarity distances.
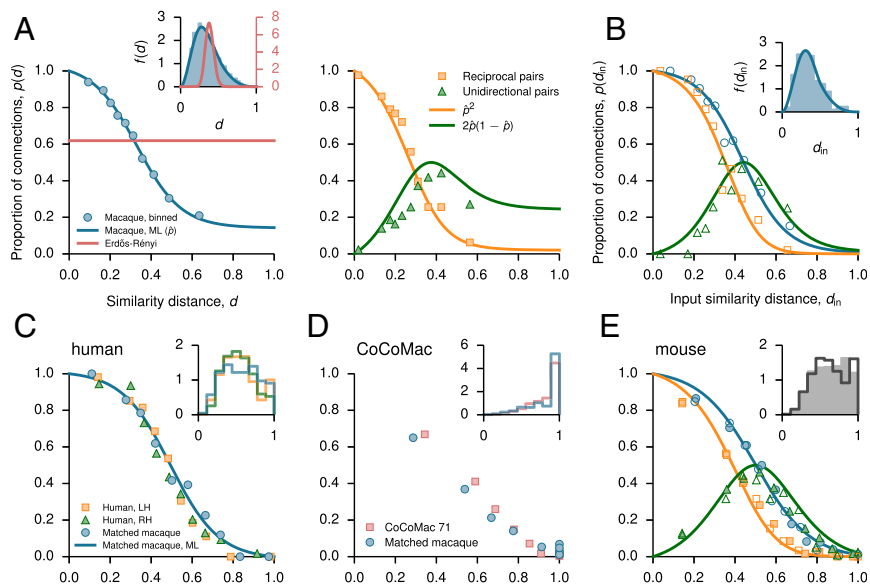
Song et al.

Fig. 3. Weighted, spatially embedded random network model of the macaque cortex. (*A*) Proportion of directed connections and occurrence of reciprocal and unidirectional pairs as a function of interareal (center-to-center) wiring distance, data. (*Inset*) Relationship between similarity (d_s) and wiring (d_w) distances in the edge-complete subnetwork. (*B*) Distribution of wiring distances fit by the distribution of distances in a spheroid (*C* and *D*; from Fig. 2 *A* and *B*) Same as *A* but as a function of (*C*) output and (*D*) input similarity distances. (*Insets*) Distribution of similarity distances. (*E*) Same as *A*, for the model. (*F*) Schematic of the model, displayed for two dimensions. A cortical area is defined by a "center" (plus symbols) and all points in the spheroid closer to it than to any other center. (*Left Inset*) Distribution of center-to-center distances in the model network, with the same fit (solid line) from *B*. Axonal connections are established in the direction determined by the sum of attractive forces originating from all areal centers, with lengths sampled from an exponential distribution (*Right Inset*) (4). (*G* and *H*) Same as *C* and *D* but for the model. In all relevant cases, the occurrence of reciprocal (squares) and unidirectional (triangles) connections are compared with $\hat{p}^2$ (orange line) and $2\hat{p}(1-\hat{p})$ (green line), respectively, where $\hat{p}$ is the ML estimate of the proportion of directed connections (blue line). See *SI Appendix*, Figs. S7 and S8 for the same figure but using alternative measures of similarity distance.
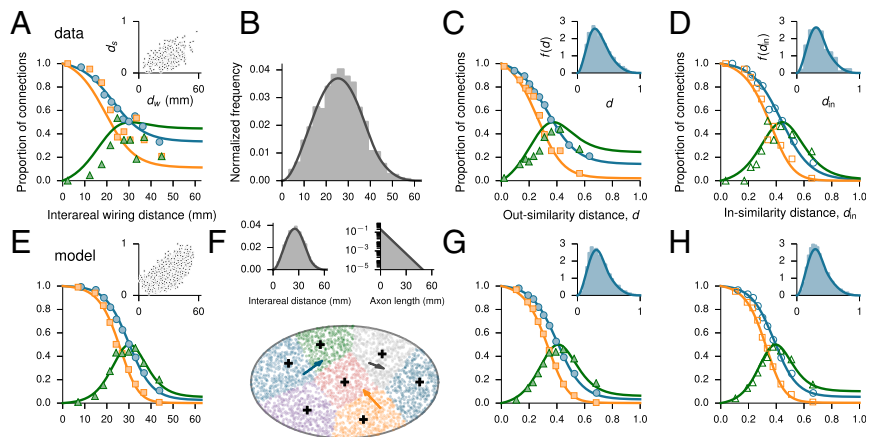
nearby areas, so two spatially nearby areas are likely to form connections with the same areas including each other.

Generative Model. Motivated by the results of our data analysis, we developed a simple, spatially embedded model of weighted cortical networks. Here we describe the minimal version of this model, whose specification requires only six parameters.

Our starting point was the observation that the overall distribution of interareal wiring distances is well described by the distribution of distances in a spheroid (i.e., a rugby ball) with radii $a(1, \lambda, \lambda)$ (Fig. 3*B*). We therefore approximated this distribution in the model by randomly placing the "centers" of the N areas at positions $\vec{R}_i$, $i = 1, \ldots, N$, in a spheroid (Fig. 3*F*, *Left Inset*). Cortical area i is then defined as all points in the spheroid closer to $\vec{R}_i$ than to any other center, i.e., the model parcellation is mathematically a Voronoi partition (27) of the spheroid. Although the true cortical sheet is nearly 2D, this 3D Euclidean manifold provides a useful approximation to its convoluted geometry with respect to axonal pathways. More formally, this space may be considered an abstract embedding space with an unknown dimensionality that can, in principle, be inferred from the data by using a model comparison procedure (28).

For each axon generated as described below, its source and target areas are determined by the axon's initial and terminal coordinates (Fig. 3*F*). The initial positions of axons are sampled uniformly in the spheroid. Inspired by the growth of real axons that are guided by a variety of attractive and repulsive molecular cues (29, 30), we assume that each area exerts a distance-dependent attractive "force," and the trajectory of each axon is determined by the sum of these forces. The effect on an axon originating at $\vec{s}$ is summarized by the vector

$$\vec{L}(\vec{s}) = -\sum_{i=1}^{N} \frac{\vec{s} - \vec{R}_i}{\left|\vec{s} - \vec{R}_i\right|^{\beta+1}}, \qquad [2]$$

where $|\cdot|$ indicates the vector's magnitude and β determines the distance dependence of the cue, respectively. The direction of the axon's growth is given by the unit vector $\hat{\ell} = \vec{L}/|\vec{L}|$ computed from the sum over forces in Eq. **2**, while the length of the axon, ℓ, is drawn from an exponential distribution with scale parameter Λ (Fig. 3*F*, *Right Inset*). The axon terminates at position $\vec{t} = \vec{s} + \ell\hat{\ell}$, with the constraint that axons must remain within the bounds of the spheroid. A simple but important consequence of the combination of Eq. **2** (specifically, the exponent β) and the exponential distribution of axon lengths is that most axons terminate in the

same area from which they originate, capturing the predominantly local character of axonal projections observed in the macaque data (15, 16). The process is not strongly affected by small, "noisy" rotations of $\hat{\ell}$ in a random direction. Fig. 4 presents a step-by-step illustration of the model.

We emphasize that the generation of axons as described here is not dynamic: both the direction and length of an axon are calculated once, and in particular, the forces that determine the direction of axon growth are evaluated only at the initial position of the axon. For the simple cortical manifold used here, this effective mechanism was sufficient to reproduce many features of the data while retaining the advantage of being computationally fast. We expect that a dynamical growth model will be necessary in future studies that incorporate a more realistic cortical geometry.

Comparison of Model and Data. Fig. 3 shows that a realization of the model closely reproduces the macaque data. The number of
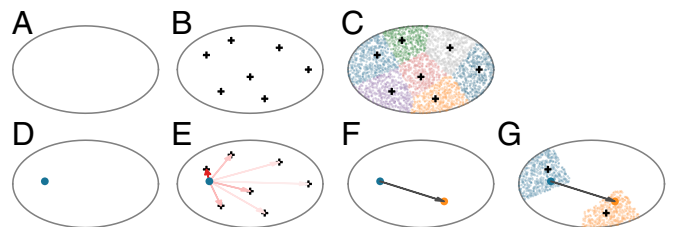

Fig. 4. Model illustration. (*A*) The model cortex is a continuous volume in a 3D Euclidean space in the shape of a spheroid with major radius a and aspect ratio λ, here represented in 2D as an ellipse for illustration purposes. (*B*) N areal centers are chosen randomly from the spheroid (plus symbols). (*C*) The configuration of the areal centers defines the parcellation of the model cortex into N areas (various colors) through a Voronoi partition of the spheroid, i.e., each area is the set of points closer to a given center than to any other center. (*D–G*) How a single axon is generated; this is repeated $n_a = 2 \times 10^6$ times in the model shown in Fig. 3. (*D*) The source of an axon (blue dot) is sampled uniformly from within the spheroid. (*E*) The direction of the axon is determined by the sum of the forces which attract the axon to the areal centers (see Eq. **2**). The individual forces decay with the distance to the areal centers (arrows) according to an inverse power law with exponent β, with the strengths represented here by the red intensity from light (weak) to dark (strong) red. (*F*) With the direction fixed, the axon extends from the source (blue dot) to the target (orange dot); the length is determined by sampling from an exponential distribution with spatial scale Λ. (*G*) The areas corresponding to the source and target of the axon are assigned according to the parcellation from *C*.
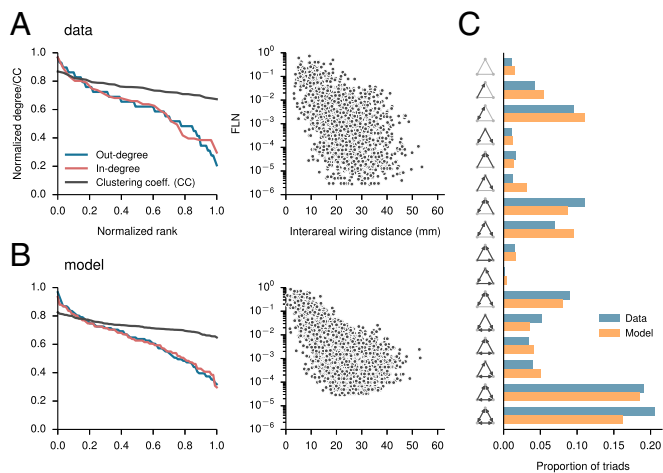
Fig. 5. Comparison of graph-theoretic measures and connection weights between model and data. (*A*) Data; (*B*) model. (*Left*) Normalized in- and out-degree sequences and clustering coefficients (for areas in the edge-complete subnetwork). In-degree represents input from all 91 source areas, for the 29 injected areas; out-degree represents output to the 29 injected areas, from all 91 source areas. (*Right*) Dispersion of FLN as a function of interareal wiring distance. (*C*) Comparison of the triad distribution in model and data.

areas in the model network is the same as the parcellation of Fig. 1*A*, $N = 91$. The spheroid dimensions, $a = 31.4$ mm and $\lambda = 0.69$, were estimated by maximum likelihood (ML) from the empirical distribution of interareal wiring distances (*SI Appendix*). The remaining three parameters—$\Lambda = 5$ mm, $\beta = 2.5$, and total number of axons $n_a = 2 \times 10^6$—were chosen to approximate the overall proportion of connections (62% in both model and data), the fraction of within-area connections [68–89% in the eight areas in which this was measured (15), $69 \pm 19\%$ in the model], and the shape of the similarity distance distribution observed in data. Thus, three of the six parameters were fixed quantitatively by the data, whereas the other three—corresponding to spatial properties that could not be inferred easily from the data—were not. In future studies, it will be of interest to use a more formal procedure for fitting all model parameters to best reproduce a selected set of summary statistics computed from the data (28).

The variability in the gross statistical properties of the resulting networks is generally small: the overall proportion of connections, for instance, was $p_0 = 62.0 \pm 3.2\%$ over 1,000 realizations of the model with the same parameters. This binary connection density depends on the number of axons; moreover, the larger the number of axons, the more likely "weak" connections will exist. Because the number of axons n_a is large, the main source of variability across different realizations is the random configuration of areal centers; for the same cortical parcellation, and hence "growth environment," simulated networks are quite similar at the areal level despite the inherent randomness of the model at the axonal level. These results shed light on how the brain may be wired consistently through a combination of specific patterning and general connection rules (31).

The model also reproduces many common measures of network structure (1, 2), such as the in- and out-degree sequences (degree in descending order, normalized by 91 sources and 29 targets, respectively) and clustering coefficients (Fig. 5 *A* and *B*, *Left*; see *SI Appendix* for the precise definition). The model captures the wide range in the number of axons that connect different areas, which can be compared with data using the fraction of labeled neurons (FLN) (Fig. 5 *A* and *B*, *Right*). FLNs, which are derived from the number of connections and represent normalized information "bandwidth" rather than actual (synaptic) strengths, are defined relative to a target area as the fraction of axons originating from each external area (15); they span

approximately five orders of magnitude in both model and data. Finally, the distribution of three-area motifs (32) is well described by the model (Fig. 5*C*).

One indication that the proposed model applies to the mammalian cortex in general rather than being specific to the macaque cortex is that the macaque connectivity data can be "matched" to the mouse connectivity data systematically. To illustrate this, we observe that the small discrepancy between data and model for the triad distribution in Fig. 5*C* can be largely eliminated (Pearson's $r = 0.98$, $P < 0.01$; Fig. 6*A*) by matching the density of connections in the edge-complete subnetwork of the macaque connection matrix to the model density. As before, this is done by removing the weakest connections. Likewise, matching the overall connection density of the macaque to the mouse results in triad distributions that are strikingly similar (Pearson's $r = 0.99$, $P < 0.01$; Fig. 6*B*). Matching the model network of Fig. 3 to the mouse connection matrix thus gives similar results (Pearson's $r = 0.99$, $P < 0.01$; Fig. 6*C*), suggesting that weak, mostly long-distance connections present in the primate cortex but absent in the rodent cortex may play a role in shaping functional differences between the two species.

Discussion

Our analysis of the macaque, human, and mouse cortical areal networks suggests that there is a common logic to the large-scale organization of cortex. Many aspects of this structure are captured by a simple, generative random network model in which interareal connectivity emerges as a coarse-grained description of heterogeneously distributed, spatially embedded axons. The model highlights the need for caution in interpreting certain features of the cortical network: many such features are surprising only if our expectations are based on inappropriate null models (33). For instance, motif distributions observed in the macaque and mouse cortex are reproduced easily with our random network model, indicating that they do not necessarily reflect specialized computational roles.

Previous modeling studies have used connection rules with a distance dependence at the level of areas (3, 5). However, the distance traveled by an axon between two cortical areas may be substantially different depending on where in the respective areas the axon originates and terminates. Thus, it may be more realistic to assume an effective distance distribution at the level
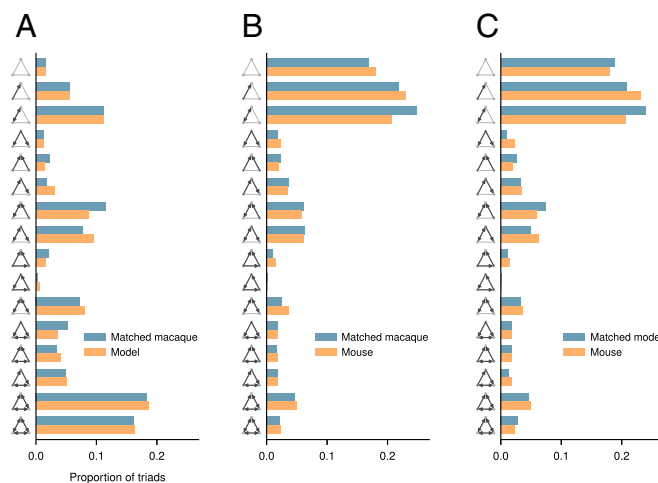

Fig. 6. Similarity of triad distributions in different networks. (*A*) Edge-complete subnetwork of macaque (overall density of connections $p_0 = 66\%$) matched to model ($p_0 = 62\%$). (*B*) Edge-complete subnetwork of macaque ($p_0 = 66\%$) matched to mouse ($p_0 = 33\%$). (*C*) Model network from Fig. 3 ($p_0 = 62\%$) matched to mouse ($p_0 = 33\%$). In all cases, the weakest connections in the more densely connected network were removed to match the overall proportion of connections of the less densely connected network.

of axons. Our model uses an "exponential distance rule" (EDR) for the length of spatially embedded axons, which has strong empirical support and was already shown to describe many properties of the macaque cortex (4). We have demonstrated that a model incorporating EDR into a fully geometric framework generates, in a natural way, the quantitative relation between interareal connectivity and similarity distance (Fig. 2*A*), and between interareal connectivity and center-to-center wiring distance (Fig. 3*A*). Furthermore, this geometric approach has the advantage that once the axons are generated, the cortex can be divided into any number of areas with arbitrary spatial resolution, independent of the parcellation that generated the connection matrix.

The fundamental advantage of using a generative model of connectivity is that one can evaluate the evidence for competing models and use Bayesian model comparison to ask which aspects of the model are important for explaining empirical observations (28). For example, it might be that one, two, or four dimensions for the embedding manifold provide a better explanation for the observed distributions, which could be evaluated by using the probability of empirical distributions (say, of similarity distances) under each model. After optimizing the model parameters using their ML values, the resulting likelihood may be used as a proxy for model evidence. Future work will investigate the use of Bayesian model comparison to optimize formal aspects of the model and address related questions.

Elaborations of the minimal model will also allow future work to incorporate additional properties of the cortical network. For instance, the uniform sampling of the initial positions of axons might be modulated to better reflect the heterogeneity of connection patterns within areas (9). The model cortex might expand during the axon generation process to mimic simplified notions of growth and development. The distribution of axon lengths can depend on the total cue strength, and there can be multiple cues with both repulsive and attractive interactions (34), allowing for greater specificity of connections. The distinction between superficial- and deep-layer axons, which are essentially geometric properties and are strongly related to whether the axons are feedforward or feedback in nature (35), can also be included in the model. Realistic cortical geometries are clearly important. Together with dynamical axon growth and the inclusion of subcortical structures (particularly thalamus), such extensions would allow the model to more faithfully capture the process of cortical wiring as it occurs in the brain.

Conclusion

The present work highlights functional similarity represented by shared connections as an important principle for understanding the large-scale structural organization of the mammalian cortex, and demonstrates that simple generative principles respecting the spatially embedded nature of cortex can account for numerous features of cortical connectivity. The dynamical and functional implications of this structure remain important questions for future investigation.

Materials and Methods

All data are available from the references given in the main text; see *SI Appendix* for further details. *SI Appendix* additionally describes the ML estimation procedure used to establish the relation between the proportion of connections and similarity (wiring) distance, and the derivation of the distribution of similarity distances in ER networks. Alternative definitions of similarity distance, formulas for the distribution of distances in the spheroid and ellipsoid, and the definition of clustering coefficient are also summarized in *SI Appendix*.

1. Bullmore E, Sporns O (2009) Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci* 10(3):186–198.
2. Bressler SL, Menon V (2010) Large-scale brain networks in cognition: Emerging methods and principles. *Trends Cogn Sci* 14(6):277–290.
3. Kaiser M, Hilgetag CC (2004) Spatial growth of real-world networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 69(3 Pt 2):036103.
4. Ercsey-Ravasz M, et al. (2013) A predictive network model of cerebral cortical connectivity based on a distance rule. *Neuron* 80(1):184–197.
5. Klimm F, Bassett DS, Carlson JM, Mucha PJ (2014) Resolving structural variability in network models and the brain. *PLOS Comput Biol* 10(3):e1003491.
6. Zingg B, et al. (2014) Neural networks of the mouse neocortex. *Cell* 156(5):1096–1111.
7. Oh SW, et al. (2014) A mesoscale connectome of the mouse brain. *Nature* 508(7495):207–214.
8. Van Essen DC, Glasser MF, Dierker DL, Harwell J (2012) Cortical parcellations of the macaque monkey analyzed on surface-based atlases. *Cereb Cortex* 22(10):2227–2240.
9. Markov NT, et al. (2014) A weighted and directed interareal connectivity matrix for macaque cerebral cortex. *Cereb Cortex* 24(1):17–36.
10. Markov NT, et al. (2013) Cortical high-density counterstream architectures. *Science* 342(6158):1238406.
11. Barabási AL, Albert R (1999) Emergence of scaling in random networks. *Science* 286(5439):509–512.
12. Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. *Nature* 393(6684):440–442.
13. Chklovskii DB, Schikorski T, Stevens CF (2002) Wiring optimization in cortical circuits. *Neuron* 34(3):341–347.
14. Raj A, Chen YH (2011) The wiring economy principle: connectivity determines anatomy in the human brain. *PLoS One* 6(9):e14832.
15. Markov NT, et al. (2011) Weight consistency specifies regularities of macaque cortical networks. *Cereb Cortex* 21(6):1254–1272.
16. Douglas RJ, Martin KAC, Whitteridge D (1989) A canonical microcircuit for neocortex. *Neural Comput* 1:480–488.
17. Young MP (1993) The organization of neural systems in the primate cerebral cortex. *Proc Biol Sci* 252(1333):13–18.
18. Passingham RE, Stephan KE, Kötter R (2002) The anatomical basis of functional localization in the cortex. *Nat Rev Neurosci* 3(8):606–616.
19. Markov NT, et al. (2013) The role of long-range connections on the specificity of the macaque interareal cortical network. *Proc Natl Acad Sci USA* 110(13):5187–5192.
20. Mars RB, et al. (2011) Diffusion-weighted imaging tractography-based parcellation of the human parietal cortex and comparison with human and macaque resting-state functional connectivity. *J Neurosci* 31(11):4087–4100.
21. Sporns O, Zwi JD (2004) The small world of the cerebral cortex. *Neuroinformatics* 2(2):145–162.
22. Kennedy H, Knoblauch K, Toroczkai Z (2013) Why data coherence and quality is critical for understanding interareal cortical networks. *Neuroimage* 80:37–45.
23. Cox T, Cox MAA (2001) *Multidimensional Scaling* (CRC, Boca Raton, FL), 2nd Ed.
24. Leicht EA, Holme P, Newman ME (2006) Vertex similarity in networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 73(2 Pt 2):026120.
25. Hagmann P, et al. (2008) Mapping the structural core of human cerebral cortex. *PLoS Biol* 6(7):e159.
26. Costa LdaF, Kaiser M, Hilgetag CC (2007) Predicting the connectivity of primate cortical networks from topological and spatial node properties. *BMC Syst Biol* 1:16.
27. Okabe A, Boots B, Sugihara K, Chiu SN (2000) *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams* (Wiley, Chichester, UK), 2nd Ed.
28. Friston KJ, Kahan J, Razi A, Stephan KE, Sporns O (2014) On nodes and modes in resting state fMRI. *Neuroimage* 99:533–547.
29. Price DJ, et al. (2006) The development of cortical connections. *Eur J Neurosci* 23(4):910–920.
30. Stiles J, Jernigan TL (2010) The basics of brain development. *Neuropsychol Rev* 20(4):327–348.
31. Braitenberg V, Schüz A (1998) *Cortex: Statistics and Geometry of Neuronal Connectivity* (Springer, New York), 2nd Ed.
32. Milo R, et al. (2004) Superfamilies of evolved and designed networks. *Science* 303(5663):1538–1542.
33. 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(5687):1107, author reply 1107.
34. Wei Y, Tsigankov D, Koulakov A (2013) The molecular basis for the development of neural maps. *Ann N Y Acad Sci* 1305:44–60.
35. Markov NT, et al. (2014) Anatomy of hierarchy: Feedforward and feedback pathways in macaque visual cortex. *J Comp Neurol* 522(1):225–259.

NEUROSCIENCE