# A principled approach to community detection in interareal cortical networks

Jorge Martinez Armas[a,b], Kenneth Knoblauch[c,d], Henry Kennedy[c,e,f], Zoltan Toroczkai[a,1]

[a]Department of Physics, University of Notre Dame, Notre Dame, IN 46556, USA;
[b]Departments of Developmental Biology and Neuroscience, Washington University in St. Louis School of Medicine, St. Louis, MO 63110, USA
[c]Univ Lyon, Université Claude Bernard Lyon 1, INSERM, Stem Cell and Brain Research Institute U1208, 69500 Bron, France;
[d]National Centre for Optics, Vision and Eye Care, Faculty of Health and Social Sciences, University of South-Eastern Norway, Kongsberg, Norway
[e]Institute of Neuroscience, Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai 200031, China;
[f]Shanghai Center for Brain Science and Brain-Inspired Intelligence Technology, Shanghai 200031, China;

[1]To whom correspondence may be addressed. Email: toro@nd.edu

ORCIDs:
J.M.A.: https://orcid.org/0000-0002-9758-967X ;
K.K.: https://orcid.org/0000-0002-4681-4638 ;
H.K.: https://orcid.org/0000-0002-2796-7663 ;
Z.T.: https://orcid.org/0000-0002-6602-2849 ;

**This PDF file includes:**

Main Text (5611 words)
Figures 1 to 7
Methods (3911)
References (75)

**Extended Data** contains Figures 1 to 9, Tables 1

## Abstract

Structural connectivity between cortical areas, as revealed by tract-tracing is in the form of highly dense, weighted, directed, and spatially embedded complex networks. Extracting the community structure of these networks and aligning them with brain function is challenging, as most methods use local density measures, best suited for sparse graphs. Here we introduce a principled approach, based on distinguishability of connectivity profiles using the Hellinger distance, which is relatable to function. Applying it to tract-tracing data in the macaque, we show that the cortex at the interareal level is organized into a nested hierarchy of link-communities alongside with a node-community hierarchy. We find that the ½-Rényi divergence of connection profiles, a non-linear transform of the Hellinger metric, follows a Weibull-like distribution and scales linearly with the interareal distances, a quantitative expression between functional organization and cortical geometry. We discuss the relationship with the extensively studied SLN-based hierarchy.

Understanding how the brain works is an inverse problem in which, using observational data, we are trying to back-engineer functionality and principles of information processing. As the brain is a complex networked system, one approach is to study this network and how its structure deviates from that of other, random networks, serving as null models. The deviations and their heterogeneity across scales reflect structural and functional specificities of the brain (Sporns, Tononi and Kötter, 2005; Bullmore and Sporns, 2009; He and Evans, 2010; Bassett and Sporns, 2017; Váša and Mišić, 2022).

A heavily used method in studying a network's structural organization is identifying its communities (Fortunato, 2010). In their original definition, communities (or clusters) are subsets of nodes that are more interconnected amongst themselves than with the rest of the network. A more general definition, which we also adopt, considers communities to be subsets of network objects (nodes, edges, subgraphs, etc.) that are “closer” to one another by some pre-defined measure, called *similarity* measure, than to the rest of the network. When the “network objects” are themselves communities, we are dealing with communities of communities, that is, a hierarchical organization of the network (Ravasz *et al.*, 2002; Clauset, Moore and Newman, 2008; Ravasz, 2009; Nikola T. Markov *et al.*, 2014; Harris *et al.*, 2019).

Communities in networked systems have been associated with components that perform specific functions (Strong, Lawton and Southwood, 1984; Mesulam, 1990; Fortunato, 2010; Ruan, Dean and Zhang, 2010; Onnela *et al.*, 2012; Li and Pang, 2013; Henry *et al.*, 2016), simply because heterogeneous structures have a much larger capacity to encode information than homogeneous ones (having no community structure). This property also holds for neuronal networks, a well-documented observation (Wu *et al.*, 2011; Sporns, 2013; Pessoa, 2014; Sporns and Betzel, 2016). Since in the brain information is encoded by the activities of neuronal populations (Buzsáki, 2010), the structure of the network is crucial in the specificity and tunability of those activities (Michalareas *et al.*, 2016; Vezoli *et al.*, 2021). As the principal manner in which a function performing network differs from a homogeneous random network is via the existence of communities, we posit that they form a key organizational feature in the brain. Moreover, we hypothesize that this property holds across all scales in the brain, including at the mesoscale, where nodes represent functional areas and links represent physical connectivity in form of axonal bundles, forming the so-called interareal network.

Here, we present a novel method for analyzing the community structure of neuronal networks to help better understand the role of the connections and their weights (axon bundle densities) in shaping the network's functions. We focus on the interareal network, given the datasets available to us, but our methodology is applicable at any scale. Due to the nature of the mammalian interareal networks, which are dense, directed and weighted graphs, such as in rodents and non-human primates (N. T. Markov *et al.*, 2014; Horvát *et al.*, 2016; Gămănuţ *et al.*, 2018), most state-of-the-art community detection algorithms are unsuitable, since they rely on local network density as a similarity measure (Newman, 2004; Palla *et al.*, 2005; Rosvall, Axelsson and Bergstrom, 2009; Ahn, Bagrow and Lehmann, 2010; Bodlaj and Batagelj, 2015). Such measure, however, does not give a sufficiently contrasting signal in networks that are already dense. Moreover, we want to develop a principled approach, where we choose our criteria for community definition related to the roles performed by these structures; thus, expectedly, the communities and the hierarchies obtained should be a direct reflection of information processing in the brain. It is important to mention that the detection of communities and their hierarchical organization is not unique, as it strongly depends on the definition of the similarity measure and on the community merging criteria. For example, in a social network, communities obtained by physical co-location (families) are not the same as those obtained by professional memberships.

Our approach is to first find the communities formed by the links/edges (Ahn, Bagrow and Lehmann, 2010b), then from that, to generate the node-communities and capture their hierarchical organization (if any). This is because links connecting similarly

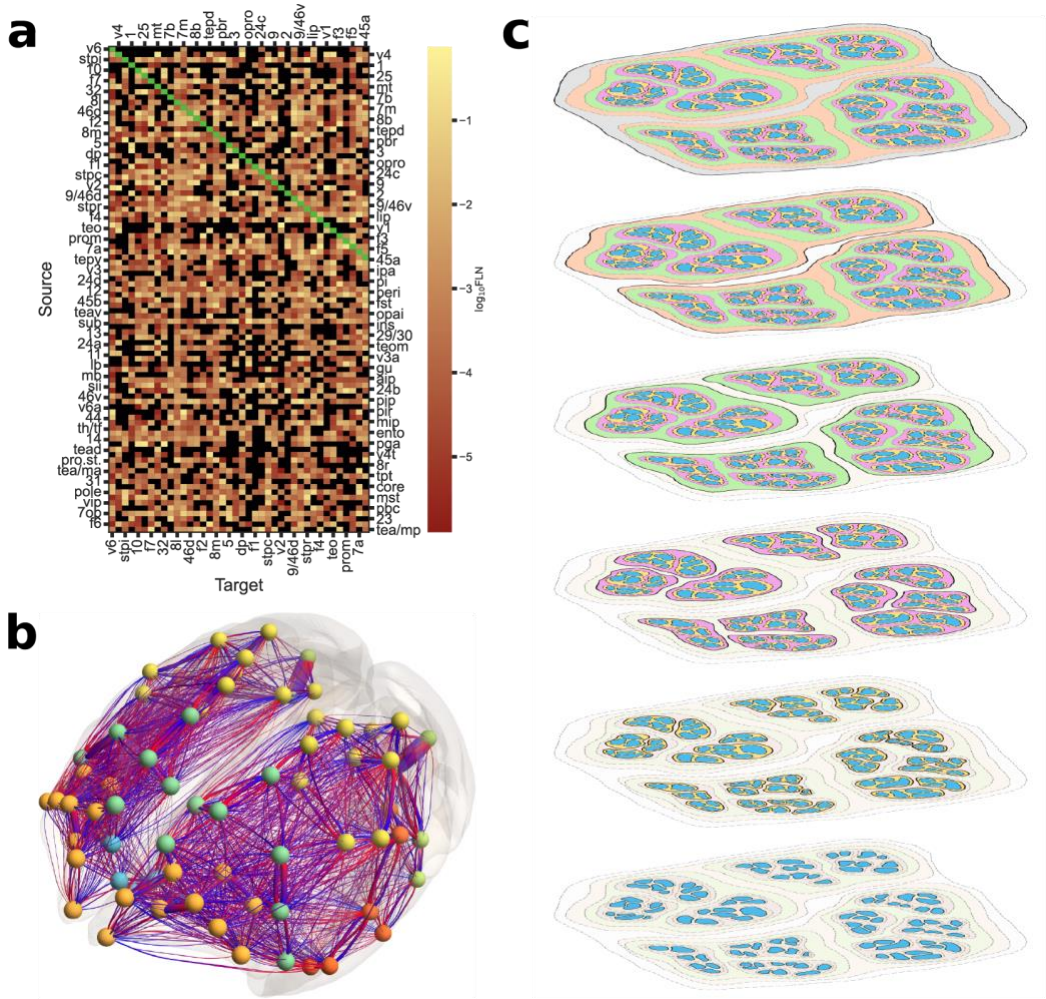



**Fig. 1 | The network community hierarchy.** **a**, The matrix of connection weights $G^{FLN}_{91\times 40}$ (in random order) obtained retrogradely from 40 target area injections, on an atlas of 91 areas in the macaque monkey. Weights are represented by colors, as shown in the color bar, with black and green denoting absent and intrinsic connections, respectively. **b**, The edge-complete network part $G^{FLN}_{40\times 40}$ of the data from **a** represented as a network with approximate 3D layout of the area's barycenters (network nodes) over both hemispheres. Node colors represent different brain regions. Cylindrical tubes are connections, whose blue-red color gradient and widths depict the direction from source area to target area and weight, respectively. **c**, A schematic of the nested network community hierarchy structure (edges are not shown) studied here.

structured network neighborhoods will have similar roles, given that in the brain, the network structure is a major part of the algorithm itself (Horvát *et al.*, 2016; Laydevant *et al.*, 2024; Molnár *et al.*, 2024). Using link communities is also advantageous, as there are many more links than nodes (hence potentially more structural information), especially in denser graphs. In the macaque dataset studied here, shown in **Fig. 1a,** the edge-complete part of the network (see Glossary in Methods), has only 40 nodes but 999 directed edges; also shown as a graph **Fig. 1b**, with an approximate 3D spatial layout.

After generating the link communities and their hierarchical organization, we show how to obtain the node communities and their hierarchy, in a principled way. Naturally, node communities interact through network links and thus the link-community structure will induce a partitioning of the nodes into groups. Interestingly, this approach, however, has not yet been proposed to our best knowledge. Note, the type of hierarchy discussed here is different from the standard view of areal hierarchy in neuroscience, introduced by Felleman and Van Essen (Felleman and Van Essen, 1991) for the visual system, later expanded and called the "SLN-hierarchy" by Barone et al. (Barone *et al.*, 2000), using the feedforward (FF) and feedback (FB) classification of connections. Our hierarchy is defined as that of "nested" communities, through the relationship of inclusion, as illustrated in **Fig. 1c**. Previous link community papers either use a human binary (unweighted) connectome obtained dMRI (de Reus *et al.*, 2014) without identifying node communities or hierarchies (of neither type, link- or node- based), or use human fmri data (Meunier, Lambiotte and Bullmore, 2010), finding a hierarchical organization in human brain functional networks. However, the latter does not directly imply anything about the physical or anatomical network itself, instead, it provides an indirect support to the observations presented here.

The article is organized as follows. We first discuss the principles behind our approach and introduce a novel dissimilarity measure based on the so-called Hellinger distance (Hellinger, 1909; Cam and Yang, 2000), a measure of distinguishability of two probability distributions. This allows us to use an information theoretic measure, the ½-Rényi divergence $D_{1/2}$, related to the Hellinger distance via a non-linear transformation, to provide an explicit relationship between the geometrical organization of the neuronal network and its functional properties. Note, our approach is generally applicable to all spatial networks in which link weights have a probabilistic interpretation. Next, we generate the link-community hierarchy using the Ahn-Bagrow-Lehmann (ABL) algorithm (Ahn, Bagrow and Lehmann, 2010) with the novel similarity measure and compare it to a null-model showing that there is significant amount of structural heterogeneity and specificity in the original brain data, at the level of link communities. We also compare it to the case where we use physical distances between areas' barycenters (Ercsey-Ravasz *et al.*, 2013) as a (dis)similarity measure and discuss the differences. We then extract the communities at the nodal level, resulting in the node-level community hierarchy of the macaque, as one of our main results. Next, we show how to select an optimal level of the node-community hierarchy and discuss its properties. Finally, we discuss the relationship of our node-community hierarchy with the SLN-hierarchy.

## Results

### A network interpretation of function

Before developing our community detection method, we first discuss the notion of function, a notion not easily quantifiable in biology. Function within a biological (or sociological) system implies some kind of systematic organization, so that the system's dynamics has usually a well-defined and repeatable outcome and thus function is connected to the notion of order. While we can quantify disorder with various notions of entropy, we do not have a systematic quantification of order, as it can be of many different types; some networks may look rather random, while in fact they are ordered structures performing a multitude of functions. Here we propose a more practical treatment of the notion of function. The observation we are making is that function is not solely an attribute of an object or a system. It is instead, a *relationship* attribute, that is of the interactions between two or more objects. Consider, for example, the myoglobin molecule. It binds, stores and releases an oxygen molecule, which by itself is a useless property. However, it becomes highly beneficial in oxygen storage and transport in the muscle tissue, thus attaining its

function via its interactions with the muscle tissue. Note, that when specifying this function, we use the sequence: myoglobin → regulation of $O_2$ and $NO$ → muscle tissue contraction (Kamga, Krishnamurthy and Shiva, 2012). Using network language, we propose that function is a property of an ordered pair of nodes $(i, j)$, or equivalently, that node $i$ performs a function towards node $j$. Regarding nodes as input/output units, the activity of node $i$ towards node $j$ serves as the input for the activity of $j$. In the brain, this is information or influence transfer (modulation). Here, we quantify the functional role of a node $i$ in the network by the list of its outgoing connections and their properties, i.e., by its out-link profile. The list of incoming links to node $i$, or its in-link profile, represents the functional feedback of the network to node $i$. In our analysis we use both types of link profiles. In the brain, we assume that all connections have a sustained role/function (otherwise they would not exist), and thus, the network of interareal connections is a direct reflection of the functional organization of the brain at the areal (mesoscopic) level. The links transfer information or modulation, which changes the state of the receiving (downstream) nodes,

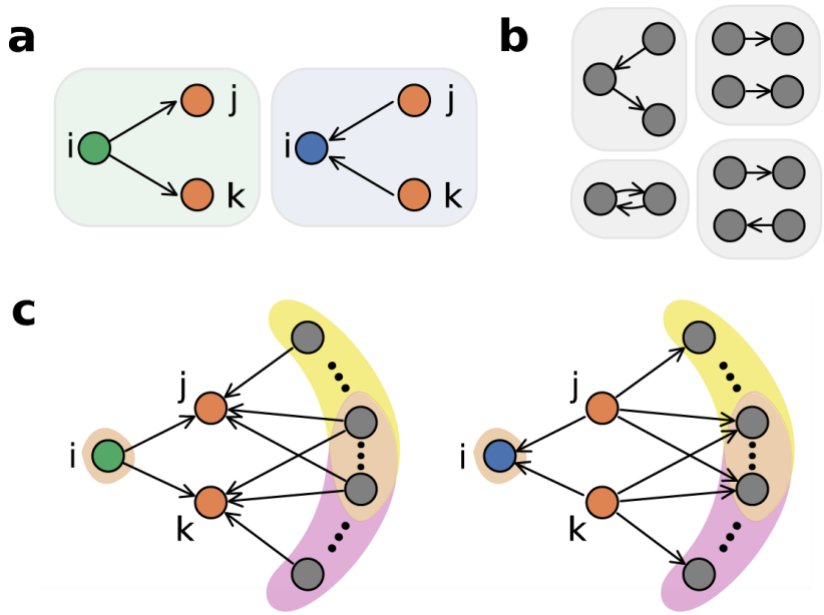



**Fig. 2 | Defining the link similarity measure. a**, Neighbor links have either a common start-node for out-links or a common end-node for in-links. **b**, All other pairs of links are disjoint. **c**, Similarity between two neighbor links is determined by the overlap between their weighted connectivity profiles of their non-common nodes $j$ and $k$, using the Hellinger distance (see Methods section).

which in turn modulates/changes their output downstream towards other nodes, etc.

Functions are composable in the sense that the union of activities of groups of nodes towards the rest of the nodes represents another function, possibly more complex/simple, depending on the higher/lower degree of connectivity of that group with the rest of the network. This property allows us to track the function-based hierarchy of the cortical network. While this network interpretation of function does not provide a translation between connectivity and behavior (which would require characterization of the nodal states and their dynamics), it is self-consistent, and it can be analyzed/quantified and even compared across species if a mapping can be made between the sets of nodes of the neuronal networks of the two species.

## Link similarity

The clusters found in community detection reflect the network's organization from the point of view of the clustering criteria used. The criteria are formulated via a similarity measure, allowing us to decide whether two network elements (nodes, links, subgraphs) belong to the same cluster (i.e., they are more similar) or not. Thus, if we want to understand how function is organized in the interareal network, our criteria and our similarity measure have to be brain function based. Since links represent function, one needs to define a (dis)similarity or distance measure between the links. We say that two directed links are *neighbor links* if they share either a common starting node but different end-nodes (called out-neighbor links) or share a common end-node but different starting nodes (in-neighbor links). **Fig. 2a** shows two neighbor out-links $(i, j)$ and $(i, k)$, and two neighbor in-links $(j, i)$ and $(k, i)$. All other pairs of links are disjoint, see **Fig. 2b**. Only neighbor links can have non-zero similarity; disjointed links have zero similarity. We define the amount of similarity between two neighbor links by the *node-similarity* of their un-common nodes ($j$ and $k$). In turn, we define the similarity of two nodes ($j$ and $k$) via the overlap between their connectivity profiles, i.e., as given by all their weighted connections (including connections from/to the common node $i$), see **Fig. 2c**. In other words, the two out-going (in-coming) neighbor links' similarity is determined by the node-similarity of their end (start) nodes.

## The Hellinger distance as a connection distinguishability measure

We define the (dis)similarity of two nodes with the help of the Hellinger distance (Hellinger, 1909). The Hellinger distance $H(P, Q)$, used for example in nonparametric density estimation, is a measure of the dissimilarity of two probability distributions $P$ and $Q$, see Methods. For our purposes it has several advantages: *i)* it is a true metric, i.e., it is non-negative, symmetric, obeys the triangle inequality, i.e., $H(P, Q) + H(Q, R) \geq H(P, R)$, for any $P, Q, R$ probability distributions, *ii)* it is a bounded metric with $0 \leq H(P, Q) \leq 1$, for any $P, Q$ and *iii)* it relates to a practical statistical measure, called *sample complexity* $\mathcal{N}^{\delta}(P, Q)$. A key consequence of *i)* and *ii)* is mathematical consistency for comparisons. For example, if distributions $P$ and $Q$ are far but $Q$ and $R$ are close, then it follows that $P$ and $R$ are also far - a property not guaranteed in general for similarity measures used in the literature. In our case, we use the Fraction of Labeled Neurons (FLN) as probability distributions, which are the link weights in the interareal network datasets. These are obtained via retrograde tract-tracing injections into the cortex, see the Methods section. The FLN weights admit a probabilistic interpretation (Markov, Mária Ercsey-Ravasz, *et al.*, 2013), so that the Hellinger distance is a good dissimilarity measure to use in our context. As similarity, we employ $S(P, Q) = 1 - H^2(P, Q)$, called *affinity* or the Battacharyya coefficient (Bhattacharyya, 1946) ($0 \leq S(P, Q) \leq 1$) in statistics. By definition, disjoint links have the largest possible distance ($H = 1$) and thus the lowest similarity value ($S = 0$).

The Hellinger distance is directly connected to the notion of sample complexity (Canonne *et al.*, 2019), a measure of how many times we need to sample from one probability distribution to tell the two apart (Methods). The smaller the Hellinger distance, the more samples are needed.

**Fig. 3a** shows $H_{-}^2$ and $H_{+}^2$ for all node pairs as a function of white-matter interareal distances between the barycenters of areas (see Methods) for the nodes within the edge-complete network (ECN). One can see another advantage of using the Hellinger distance

as dissimilarity measure: it is sensitive at all scales, especially at short distances, providing a good resolution in discernibility, hence also locally.

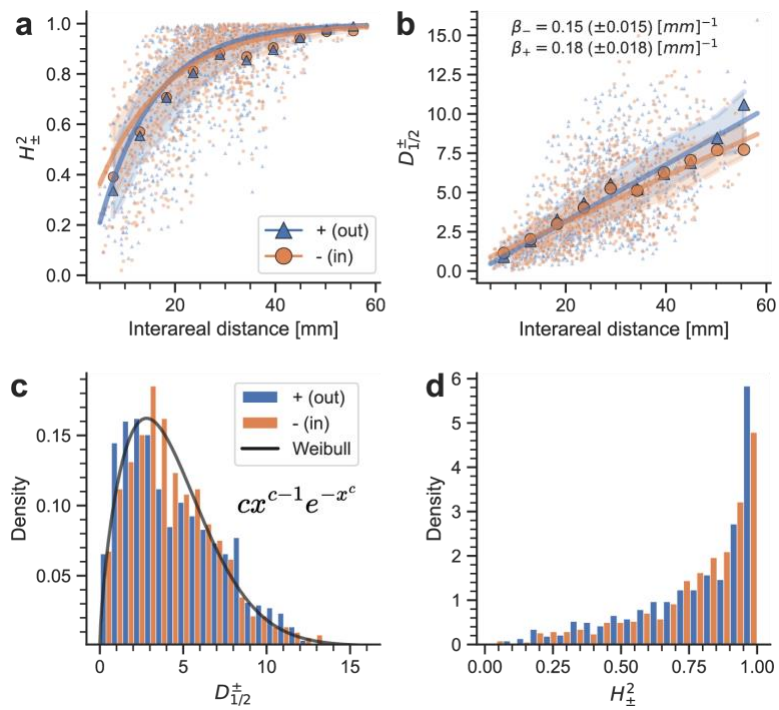



**Fig. 3 | Measures of connectivity profile differences. a**, The Hellinger distance $H^2_\pm$ as function of interareal distance for the macaque dataset. Lines, 95% confidence level (CL) uncertainty bands, and bin averages (large markers) are derived from the linear fits in panel **b,** using equation (1). **b**, ½-Rényi divergence, an information geometric measure as function of distance, showing a linear relationship. Lines and 95% CL uncertainty bands are obtained from a weighted linear regression from ten binarized averages and inverse variance weights. Slope regression coefficients $\beta_\mp$ ($\pm$ 95% CL) are shown at the top. **c**, The distributions of $D^\pm_{1/2}$ (bars) are well approximated with a Weibull distribution of shape parameter $c = 1.67$. **d**, histograms of $H^2_\pm$.

## Rényi divergence of node-connectivity profiles

There is an information-theoretic connection between the Hellinger distance and the $\alpha$-Rényi divergence (Renyi, 1961), which is a generalization of the Kullback-Leibler divergence. Namely, for $\alpha = 1/2$, the ½-Rényi divergence $D_{1/2}$ (see Methods for mathematical definition in terms of distributions $P$ and $Q$) obeys the nonlinear relationship:

$$D_{1/2} = -2\log(1 - H^2) = -2\log S\,, \tag{1}$$

(note, $\alpha = 1$ is the classic Kullback-Leibler divergence) (Nielsen and Boltz, 2011). For our purposes, the ½-Rényi divergence has two main advantages: firstly, it is symmetric (unlike the Kullback-Leibler divergence), and secondly, it can handle zero values for any of the probability distributions without going to infinity. It has many applications in information theory, hypothesis testing, multiple source adaptation, coding and machine learning through approximate inference (Csiszar, 1995; Morales, Pardo and Vajda, 2000; Mansour, Mohri and Rostamizadeh, 2009; Shayevitz, 2011; van Erven and Harremos, 2014; Li and Turner, 2016), see Methods.

Plotting the $D^{\pm}_{1/2}$ values for all ordered area pairs as a function of their distance, we obtain the plot in **Fig. 3b**. Interestingly, this shows an almost linear relationship between the Rényi divergence and distances, with slightly different slopes. However, the scatter plot is heteroskedastic with significant variability for $D^{\pm}_{1/2}$ around the average, containing additional information. Histogram plots of $D^{\pm}_{1/2}$ exhibit unimodal distribution with fast decaying tails (see **Fig. 3c**). The continuous line is a fit of the Weibull distribution $f(y) = c\, y^{c-1} e^{-y^c}$ to the standardized data $y = (x - x_{\min})/\sigma$ where $x_{\min}$ is the minimum value of $x$ (in our case $D^{\pm}_{1/2}$). $c$ and $\sigma$ are called the shape and scale parameters and fitting gives $c = 1.67$ and $\sigma = 4.75$ for our macaque dataset. In the Discussions we provide arguments for $D^{\pm}_{1/2}$ being a Weibull distribution (used in extreme value statistics) (Montgomery, 2020). **Fig. 3d** shows the histograms of $H^2_{\pm}$ for the macaque, a monotonically increasing function, indicating that the number of area pairs with a certain amount of dissimilarity between their connectivity profiles increases with their dissimilarity, resulting in improved statistics.

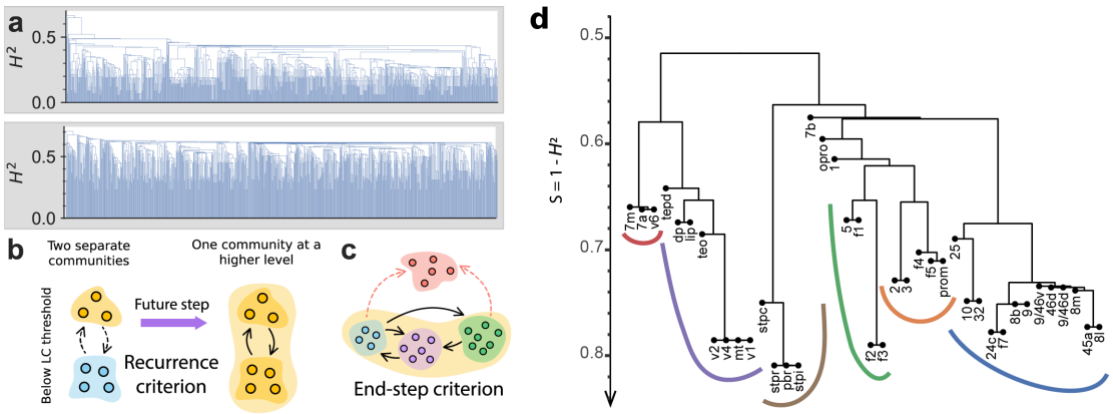



**Fig. 4 | Link- and node-community hierarchies in the macaque brain. a**, The link-community hierarchy shown as a dendrogram for the data (top) and configuration model (bottom). The vertical lines at the lowest level represent individual links (999 links). Note that in the former it is more straightforward to appreciate the link-modular structure than in the latter. **b**, Node-community merging criterion based on reciprocal connectivity. **c**, Criterion for including into the hierarchy those node communities that were not included by the previous merging criterion. **d**, The resulting node-community hierarchy. The y-axis is the link-community similarity between node communities at their merging point, increasing from top to bottom. Areas that merged towards the bottom (for example V1 and V2) have more similar in- and out-neighborhoods than the ones that merged close to the top (7b, OPRO, etc.). We note that areas with strong topographic organization tend to merge towards the bottom of the hierarchy (V1, V2, MT, 3, 2, etc.). The colored arcs correspond to functional brain regions associated with object recognition (purple), movement detection (red), hearing (brown), motor (green), somatomotor (orange), and high-level cognitive functions (blue). The hierarchy's areal horizontal arrangement is computed by a software (ggtree) (Yu et al., 2017) and do not intent to display a special order.

## The link-community hierarchy of the macaque cortex

Having defined a proper distance (or similarity) measure between the links, we can now use the ABL link-community detection algorithm (Ahn, Bagrow and Lehmann, 2010) on the ECN of the macaque dataset, see Methods for algorithm description and **Extended Data Fig. 1**.

**Fig. 4a** shows the link-community hierarchy obtained with this method for the edge-complete component of our dataset. It has many intricate structures at all levels, indicating the complexity of the functional hierarchy. However, it is difficult to interpret because it essentially shows the "cabling structure between the elements of a device," not the elements and groups of elements of the device that are working together (considered to be the natural description of the inner workings of a system). We can, however, make a comparison of the link-community hierarchy in **Fig. 4a** (top) with the same generated from a null model, called the configurational model (bottom). The configuration model is obtained from the original network by randomly rewiring its connections (maintaining edge weights) (Orsini *et al.*, 2015). The link-community hierarchy for the configuration model (typical instance), in **Fig. 4a** bottom, shows that the link-community hierarchy for the original data is more structured, indicating higher specificity and heterogeneity.

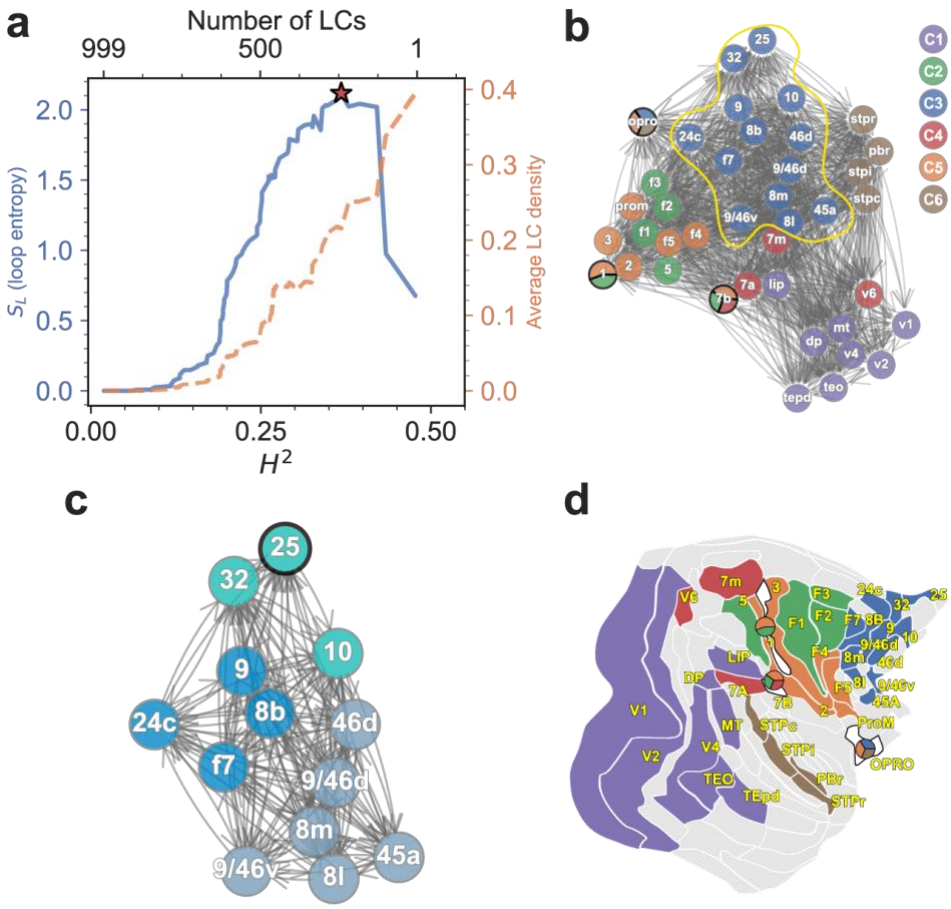



**Fig. 5 | Community structure at an optimal level of the NC hierarchy**. **a**, Distribution of the loop entropy $S_L$ (solid line) and the average LC density (Ahn, Bagrow and Lehmann, 2010) (dashed line) as function of dissimilarity $H^2$. **b**, The optimal node-community structure corresponding to the maximum loop entropy value (the star in **a**). Nodes with pie-charts in them represent nodes with overlapping communities (NOCs; see **Table 1** for more information about them). **c**, The sub-community structure (a few levels below) of the blue community from **b** within the yellow loop. Area 25 community membership was found later using the cover assignment algorithm (see Methods section for more information). **d**, The optimal NC structure on the cortical flat map. Grey areas do not belong to the 40x40 edge-complete graph.

**Extracting the node-community hierarchy from the link-community hierarchy**

As the link-community hierarchy is difficult to interpret, we need to generate a nodal community hierarchy using the link-community structure. The main criterion that we use for generating the node communities is simple: at a given level two communities of nodes are merged if they are bidirectionally connected with edges from *the same link community*. We call this the "recurrence criterion," i.e., the two groups must have recurrent connectivity within the same community of links (**Fig. 4b**). However, there are node groups that cannot be connected by this criterion because, for example, they have only in-going, or out-going connections with the rest of the network ( **Fig. 4**); they are added into the node-community hierarchy at the end sequentially, in increasing order of their index. **Fig. 4d** shows the result, with the height of the branch between two newly merged groups being proportional to the difference in the $H^2$ distance between the link-communities at the merging event. The thick-colored arcs correspond to functional regions in the cortex. The height difference between groups on this dendrogram (see Glossary of terms) corresponds to the dissimilarity change between their weighted connectivity profiles.

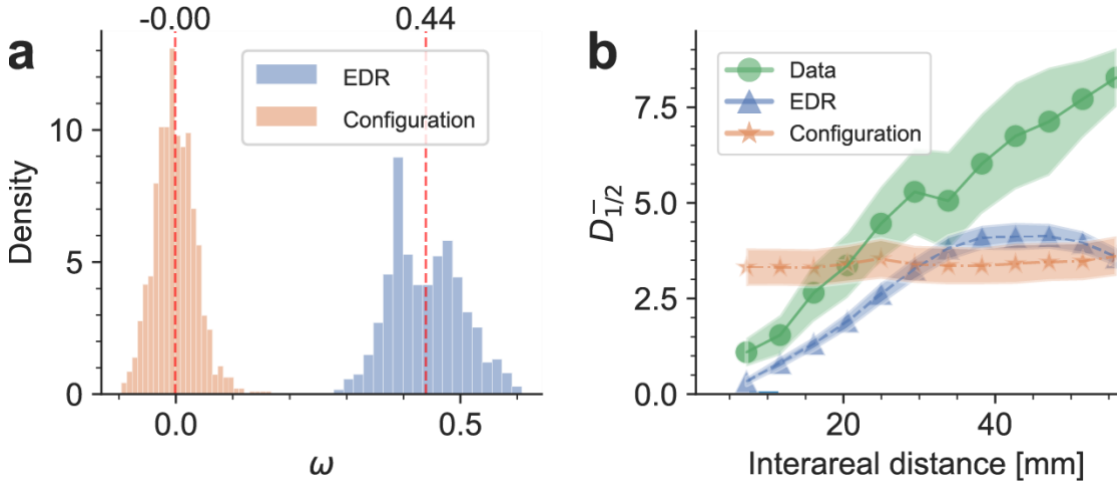



**Fig. 6 | Comparison with models**. **a**, Distribution of the omega index (a measure of the overall amount of overlap between two community structures; see Methods) for two network models, namely, the configuration model and the EDR model. The vertical dashed lines correspond to the average values after 1000 network samples. **b**, Similar plot as in **Fig. 3b** for $D^-_{1/2}$ of the data and the same for the two models. Lines and markers denote the average trends. Uncertainty band widths are equal to one standard deviation.

**An optimal level of the node-community hierarchy**

The node-community hierarchy shows the network of clusters from the most detailed level (one node in every cluster) to the least (the whole network is one cluster). The first is too detailed, but the latter is structurally uninformative. While the hierarchy as a whole provides an overall picture of the network's structural organization, we may ask if there is an "optimal" or a most informative level in the hierarchy. (Ahn, Bagrow and Lehmann, 2010) uses the average link density over all link communities to identify this level (see Methods), which, in the sparse network examples used in that study, has a peak in the middle of the hierarchy. However, for dense networks such as ours, this does not work as link densities increase all the way to the top of the hierarchy (**Fig. 5a**). Here, we modify the density measure in a way that it has a robust and guaranteed peak at a certain level, inside the hierarchy. The process used is similar to the Ahn et al. measure, considering the number of links in excess of the tree structure within each link community, but calculates an entropy measure and chooses the hierarchy level that maximizes this "loop

entropy" (see Methods). The optimal level in the link-community hierarchy is thus the one in which the distribution of the "excess" links (i.e., links generating loops and thus recurrence of information flow) is the most uniform across all link communities. The level defined in this manner is the most homogeneous from the point of view of recurrence. It is easy to show that at either end of the link-community hierarchy $S_L = 0$, and thus, it must always exhibit a single global maximum. Given that the node-community hierarchy is derived from the link-community hierarchy, we simply use this correspondence to read off the level of the optimal node-community hierarchy. This shows that at the optimal level, the macaque 40x40 dataset has a total of 6 communities (**Fig. 5b,c**). This will change as the dataset grows, but the algorithm will identify the optimal level in any version of the dataset. In the **Extended Data Fig. 2**, we show that our analysis in a 29x29 data (N. T. Markov *et al.*, 2014) results in a similar node-community hierarchical structure.

**Nodes with overlapping communities**

The algorithm generates a small number of nodes that have multiple community memberships, i.e., nodes with overlapping communities or NOCs (**Fig. 5b**). Such nodes are likely multifunctional, as shown with a pie-chart structure in **Fig. 5b**. While this set will change with the growth of the data, some may likely be preserved. See Methods for how areas are assigned to one or more communities.

**Model comparison**

In order to quantify how much overlap there is between the community structure of a model and that of the data, we use the notion of omega index (Gregory, 2011), see Methods. The value of $\omega = 1$ corresponds to perfect overlap between the model and data. In **Fig. 6a**, we show the distribution of the omega index for two network models, namely the configuration model and the EDR model. The EDR model (Ercsey-Ravasz *et al.*, 2013) is a generative interareal network model that uses the observed exponential decay of axonal bundle lengths with exponent $\lambda$ (for the 40x91 dataset, this is $\lambda = 0.16\ mm^{-1}$) and the same matrix of distances between the areas. In **Fig. 6a**, the red vertical dashed lines represent the averages of the omega index over 1000 network samples from the respective models. The fact that $\langle\omega\rangle = 0$ for the configuration model means that the connectivity randomization completely removes the community structure found in the dataset. In the EDR model $\langle\omega\rangle = 0.44$, showing that the EDR not only captures many network properties of the data (Ercsey-Ravasz *et al.*, 2013; Markov, Mária Ercsey-Ravasz, *et al.*, 2013) but also a portion of its community structure. In **Fig. 6b**, we plot the $D^{-}_{1/2}$ divergence for all area pairs for the two models and the data, as a function of the distance between the pairs, showing that randomization indeed completely removes the heterogeneity of the structural organization for the configuration model, whereas the EDR captures fairly well the same trend, up to long distances.

**Relationship with the SLN-based hierarchy**.

The fact that the interareal network of the macaque cortex is hierarchically organized has been known since the seminal work of Felleman and van Essen (Felleman and Van Essen, 1991), which was based on the Feedforward/Feedback (FF/FB) nature of the visual pathways. This has later been better specified quantitatively (Barone *et al.*, 2000; Falchier *et al.*, 2002; Nikola T. Markov *et al.*, 2014; D'Souza *et al.*, 2022), using the fraction of labeled supra-granular layer neurons (SLN), obtained from tract-tracing experiments. Links with $SLN > 1/2$ are FF connections, with $SLN < 1/2$ are FB, and with $SLN \cong 1/2$ lateral connections. We will call this hierarchy the SLN-hierarchy (SLNh). In SLNh, FF compared to FB *i)* crossed fewer levels, *ii)* are topologically more organized, *iii)* are

stronger, and *iv)* are significantly fewer. SLNh ranks but does not define communities of nodes in the traditional network science sense.

Our node-community hierarchy (NCh) uses only connection strengths (in the form of FLN values, but no SLN is used) and the connectivity matrix. It also uses an entirely different criterion for defining the levels/depth of the hierarchy, namely the Hellinger distance. Moreover, our method focuses on finding network communities based on the recurrence of connections within a strength class, not merely ranking the nodes. The fact that these communities are organized into a hierarchy is a direct outcome of the network structure. Finally, the NCh describes the organization of the whole network (its edge-complete part), not just the system. Despite these differences, however, there are similar features in the two hierarchies, as we describe below.

Since NCh does not use SLN, the FF/FB nature of connections is not resolved. However, the Hellinger distance and the absolute deviation of the SLN from 0.5 for all the links in the dataset are somewhat correlated for both FF and FB connections (**Fig. 7a**). This shows that, on average, there is a linear correlation between the depth/rankings of the two hierarchies, although with wide fluctuations (two-sided Pearson's correlation $r_{FF} = 0.43$ and $r_{FB} = 0.45$ with degrees-of-freedom, $dof_{FF} = 431$ and $dof_{FB} = 537$ for feedforward and feedback connections, respectively; $P < 10^{-5}$). The horizontal line at $H_{\underline{}}^2 = 0.37$ corresponds to the optimal level in the NCh.

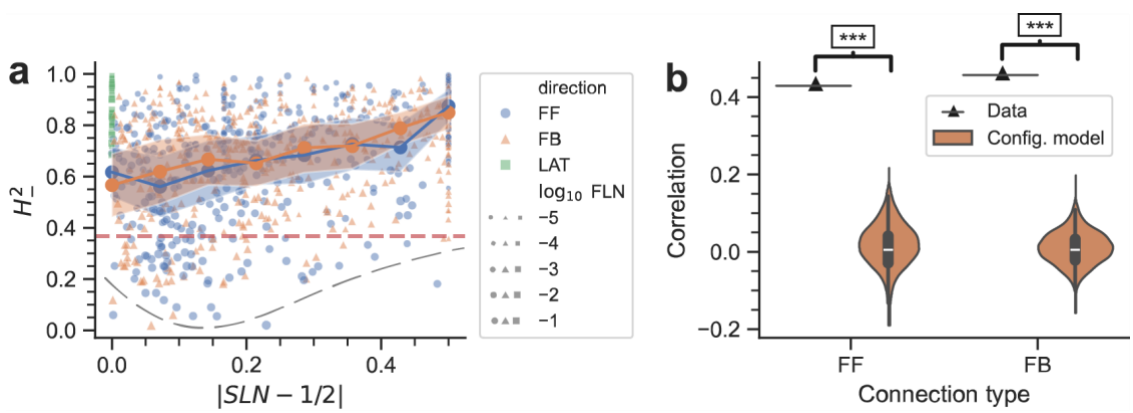



**Fig. 7| Connection with SLN measures**. **a**,The Hellinger distance between the endpoints of all the links as function of the deviation of their SLN values from ½. The size of the markers is proportional to the $log_{10} FLN$ values, i.e., the strength of the connections. Larger markers correspond to stronger connections. Blue markers denote feedforward (FF, $SLN > 0.5$) connections, whereas orange markers denote feedback (FB, $SLN < 0.5$) links. The green squares describe links corresponding to strongly lateral connections ($SLN \cong 0.5$). The red horizontal line is at $H_{\underline{}}^2 = 0.37$ corresponding to the optimal level in the NCh. The dashed curvy line is guide to the eye signaling that links within the lower levels of the NCh up to the optimal level have an $SLN$ closer to 1/2., i.e., they are not strongly FF/FB and are also of large weight. Lines show $H_{\underline{}}^2$ average trend. Uncertainty bands are the same as in **Fig. 6b**. **b**, Correlation between $H_{\underline{}}^2$ and $SLN - 1/2$ is statistically significant ($*** = P < 10^{-5}$), but it is non-existent in the configuration model.

The gray dashed line in **Fig. 7a** is a guide to the eye showing that most links at the lower levels of the hierarchy (up to the optimal one) are more balanced, that is, they are not strongly FF/FB and that they have large weight (strong links). Interestingly, while there are overall 24% more FB links than FF in the whole edge-complete dataset, under the red straight dashed line (that is, within the lower levels of the NCh), the roles are reversed: there are 31% more FF connections than FB. Furthermore, the FB connections are more scattered across the Hellinger hierarchy scale than the FF connections, which are more

organized. Thus, the Hellinger distance-based hierarchy generates similar observations to *i-iv)* above. **Fig. 7b** shows that the correlation between $H_{\underline{2}}^{2}$ and $SLN - 1/2$ is statistically significant, while for the null model (configuration model) it is non-existent.

## Discussion

In this paper we presented a principled approach to extracting the community structure of the macaque interareal cortical network and its hierarchical organization. A chief novelty of this method is to first extract the community structure at the level of links and then use the corresponding nodal community structure and hierarchy. Since in dense networks there are many more links than nodes, link communities provide richer information than node communities. As the interareal network of the mammalian cortex is extremely dense, directed, and weighted, many standard community detection methods perform poorly or fail due to using similarity measures related to local connection density, not a well-contrasting feature in ultra-dense networks.

The method introduced here is also principled, in that it employs measures and criteria that are based on notions related to the functionality of these networks, as information processing structures. Our approach embeds the nodes of the network into a metric space, allowing for consistent comparisons and a well-defined hierarchical depth scale. This is done by exploiting the fact that the link weights (axonal bundle strengths) can be interpreted as probabilities and thus allowing the use of the Hellinger metric (from statistics and information theory) to generate the distances between nodes in this space. Note, this can be used in any network in which link weights can be interpreted as probabilities. In our case, they are the fraction of labeled neurons (FLNs) obtained from retrograde tract-tracing experiments and are proxies for communication bandwidth between the functional areas (Markov, Mária Ercsey-Ravasz, *et al.*, 2013). The set of connection/communication strengths (or probabilities) between a node and the rest of the network forms its connectivity profile. Accordingly, the Hellinger distance between two nodes is a measure of dis-similarity between their connection profiles, and it is well-defined even for two non-connected nodes. Its connection to ½-Rényi divergence serves as a measure for the statistical distinguishability of areal neighborhoods and, interestingly, has a linear scaling with physical distances. This further corroborates previous observations that connection profiles diverge with spatial separation, on average (Markov, Maria Ercsey-Ravasz, *et al.*, 2013). Our contribution here quantifies this with an information-theoretic measure. Moreover, it has a practical interpretation as a sample complexity, i.e., as the number of samples we need to draw from one of the distributions in order to tell it apart from the other one. Finally, we observed that the ½-Rényi divergence is distributed according to a Weibull distribution, from extreme value statistics (goodness of fit to the Weibull has an AIC index of 3554.0, whereas for the Gamma and the Lognormal distributions are 3569.4 and 3590.0, respectively). Note that the same distribution was found to describe shortest path distributions in random networks (Bauckhage, Kersting, Rastergahpanah, 2013), a possible connection to our case here. Considering the physical distance between the areas as an alternative to the Hellinger measure does not work well, as demonstrated in the Methods section.

Function in the brain is associated with links between areas that process information. When extracting the link communities (Ahn, Bagrow and Lehmann, 2010), the sequential link merging process first merges those links that perform very similar functions, and then gradually those link communities that are increasingly different (but can be merged at that level). At this stage the loop entropy $S_L$ (**Fig. 5a**) starts to increase, meaning that recurrence between the nodes of the link community also increases and the information flow structures become more complex, likely playing the role of information

integration across computational modalities. Recurrence allows fast and reciprocal coordination between different functional modalities, for example, generating a fast motility response to a stimulus involving both visual and auditory streams. Thus, when creating the node communities from the LCh we chose a merging criterion that demands recurrence between two node communities with links that belong to a given functional modality (i.e., to a given link community at that level). As for the optimal level of the NCh, we choose the one in which recurrence is the most uniformly distributed, being a good representative of the operational community structure. To observe the importance of having a meaningful community partition, see the **Extended Data Fig. 3**. It compares a randomly ordered ECN matrix with the one obtained from the NCh. For macaque, the optimal hierarchy level is at $H^2 \cong 0.37$, corresponding to 56 link communities (see **Extended Data Fig. 4**). The associated node-community network at this level has 6 communities. As validation, when we applied our algorithm to benchmark networks (sparse) with predefined community structures in it, it correctly identified those communities (see **Extended Data Fig. 5**).

The resulting NCh shown in **Fig. 4d**, has interesting properties. The thick colored arcs depict branches in the hierarchy as generated by the algorithm and agree with known functional systems in the brain. Note, this is algorithm output, not "added in by hand". For example, the branches above the purple arc correspond to the ventral or "what" pathways, while those above the red arc to the "where" pathways. The others have similar interpretations from the auditory, somatosensory, premotor, motor and even prefrontal regions. This is remarkable, because the standard hierarchical view of processing places the primary (e.g., visual) areas as lowest level of the hierarchy, and the rest above. Our findings suggest that even those high-level areas that are not clearly associated with senses – they likely have other complicated connections with other subcortical structures – form their own pathways (sub-hierarchies), linked to attention, decision making or emotions. Our algorithm helps disentangle in a principled way across the whole network, the granular nature of functional processing in the brain. Another interesting observation is that the low-level areas in **Fig. 4d** have strong topographic properties (V1, V6, 3, 2), i.e., with highly organized inputs. In these, local changes in stimulus correspond to spatially local changes in the neurons activated, which is plausible, since the brain needs to select specific information incoming from the external world to generate the correct responses (for example, for survival). The neighbors of these areas are also topographic (**Fig. 4d**). Our algorithm also identifies nodes with overlapping communities, which we believe play a crucial role in coordinating information flow between multiple systems. Comparisons of the NCh with null models show that randomization eliminates the hierarchical organization i.e., there is strong and specific functional organization in the cortex. The EDR model captures about 44% of the organization, meaning that the rest is due to other, possibly more biological factors, which remain to be elucidated.

The SLNh tracks information flow forward and back through the links categorized by their SLN value into FF or FB. The SLN is independent from the link weights (FLN) and it is strictly a two-point descriptor. While the FLN is also associated to two points, its value is relative to the other link strengths in the target node's neighborhood (through the normalization factor). Reference (Markov et al, 2014) extracts a hierarchical scale from the SLNs using a beta-binomial model with adjustable dispersion parameter. There are two crucial differences between the SLNh and NCh. For one, the Hellinger distance is solely based on connection strengths (less informative) but provides a principled and natural metric with no adjustable parameters trying to fit to a hierarchy. Second, the SLNh places individual nodes on the levels of the hierarchy, whereas the NCh places activity-specific *communities*. The hierarchical subordination relationship in the NCh is a subset (inclusion) relationship, which is very different from the SLNh, as the latter does not identify

communities. However, there is some connection between the two, in form of an average linear correlation, illustrated in **Fig. 7**. We also discovered that connected node pairs, whose Hellinger distance is below the optimal community hierarchy threshold are 31% more FF than FB, meaning that the backbone level network is more of a driving (FF) structure than modulatory (FB), despite the fact that there are 24% more FB than FF in the *whole* network. We believe that the two hierarchies complement each other and taken together they offer a more holistic interpretation of information processing in the brain.

## References


Ahn, Y.-Y., Bagrow, J.P. and Lehmann, S. (2010) 'Link communities reveal multiscale complexity in networks', *Nature*, 466(7307), pp. 761–764. Available at: https://doi.org/10.1038/nature09182.

Ahn, Y.-Y., Jeong, H. and Kim, B.J. (2006) 'Wiring cost in the organization of a biological neuronal network', *Physica A: Statistical Mechanics and its Applications*, 367, pp. 531–537. Available at: https://doi.org/10.1016/j.physa.2005.12.013.

Barone, P. *et al.* (2000) 'Laminar Distribution of Neurons in Extrastriate Areas Projecting to Visual Areas V1 and V4 Correlates with the Hierarchical Rank and Indicates the Operation of a Distance Rule', *Journal of Neuroscience*, 20(9), pp. 3263–3281. Available at: https://doi.org/10.1523/JNEUROSCI.20-09-03263.2000.

Bassett, D.S. and Sporns, O. (2017) 'Network neuroscience', *Nature neuroscience*, 20(3), pp. 353–364. Available at: https://doi.org/10.1038/nn.4502.

Bastos, A.M. *et al.* (2012) 'Canonical Microcircuits for Predictive Coding', *Neuron*, 76(4), pp. 695–711. Available at: https://doi.org/10.1016/j.neuron.2012.10.038.

Batardiere, A. *et al.* (1998) 'Area-specific laminar distribution of cortical feedback neurons projecting to cat area 17: Quantitative analysis in the adult and during ontogeny', *Journal of Comparative Neurology*, 396(4), pp. 493–510. Available at: https://doi.org/10.1002/(SICI)1096-9861(19980713)396:4<493::AID-CNE6>3.0.CO;2-X.

Bhattacharyya, A. (1946) 'On a Measure of Divergence between Two Multinomial Populations', *Sankhyā: The Indian Journal of Statistics (1933-1960)*, 7(4), pp. 401–406.

Bodlaj, J. and Batagelj, V. (2015) 'Hierarchical link clustering algorithm in networks', *Physical Review E*, 91(6), p. 062814. Available at: https://doi.org/10.1103/PhysRevE.91.062814.

Bullmore, E. and Sporns, O. (2009) 'Complex brain networks: graph theoretical analysis of structural and functional systems', *Nature Reviews. Neuroscience*, 10(3), pp. 186–198. Available at: https://doi.org/10.1038/nrn2575.

Buzsáki, G. (2010) 'Neural Syntax: Cell Assemblies, Synapsembles, and Readers', *Neuron*, 68(3), pp. 362–385. Available at: https://doi.org/10.1016/j.neuron.2010.09.023.

Cam, L.L. and Yang, G.L. (2000) *Asymptotics in Statistics: Some Basic Concepts*. 2nd edition. New York Berlin Heidelberg: Springer.

Canonne, C.L. *et al.* (2019) 'The structure of optimal private tests for simple hypotheses', in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*. New York, NY, USA: Association for Computing Machinery (STOC 2019), pp. 310–321. Available at: https://doi.org/10.1145/3313276.3316336.

Clauset, A., Moore, C. and Newman, M.E.J. (2008) 'Hierarchical structure and the prediction of missing links in networks', *Nature*, 453(7191), pp. 98–101. Available at: https://doi.org/10.1038/nature06830.

Cormen, T.H. (ed.) (2007) *Introduction to algorithms*. 2nd. ed., 10th pr. Cambridge, Mass.: MIT Press [u.a.].

Csiszar, I. (1995) 'Generalized cutoff rates and Renyi's information measures', *IEEE Transactions on Information Theory*, 41(1), pp. 26–34. Available at: https://doi.org/10.1109/18.370121.

D'Souza, R.D. *et al.* (2022) 'Hierarchical and nonhierarchical features of the mouse visual cortical network', *Nature Communications*, 13(1), p. 503. Available at: https://doi.org/10.1038/s41467-022-28035-y.

Ercsey-Ravasz, M. *et al.* (2013) 'A Predictive Network Model of Cerebral Cortical Connectivity Based on a Distance Rule', *Neuron*, 80(1), pp. 184–197. Available at: https://doi.org/10.1016/j.neuron.2013.07.036.

van Erven, T. and Harremos, P. (2014) 'Rényi Divergence and Kullback-Leibler Divergence', *IEEE Transactions on Information Theory*, 60(7), pp. 3797–3820. Available at: https://doi.org/10.1109/TIT.2014.2320500.

Falchier, A. *et al.* (2002) 'Anatomical Evidence of Multimodal Integration in Primate Striate Cortex', *Journal of Neuroscience*, 22(13), pp. 5749–5759. Available at: https://doi.org/10.1523/JNEUROSCI.22-13-05749.2002.

Felleman, D.J. and Van Essen, D.C. (1991) 'Distributed hierarchical processing in the primate cerebral cortex', *Cerebral Cortex (New York, N.Y.: 1991)*, 1(1), pp. 1–47. Available at: https://doi.org/10.1093/cercor/1.1.1.

Fortunato, S. (2010) 'Community detection in graphs', *Physics Reports*, 486(3), pp. 75–174. Available at: https://doi.org/10.1016/j.physrep.2009.11.002.

Froudist-Walsh, S. *et al.* (2021) 'A dopamine gradient controls access to distributed working memory in the large-scale monkey cortex', *Neuron*, 109(21), pp. 3500-3520.e13. Available at: https://doi.org/10.1016/j.neuron.2021.08.024.

Gămănuţ, R. *et al.* (2018) 'The Mouse Cortical Connectome, Characterized by an Ultra-Dense Cortical Graph, Maintains Specificity by Distinct Connectivity Profiles', *Neuron*, 97(3), pp. 698-715.e10. Available at: https://doi.org/10.1016/j.neuron.2017.12.037.

Gregory, S. (2011) 'Fuzzy overlapping communities in networks', *Journal of Statistical Mechanics: Theory and Experiment*, 2011(02), p. P02017. Available at: https://doi.org/10.1088/1742-5468/2011/02/P02017.

Harris, J.A. *et al.* (2019) ‘Hierarchical organization of cortical and thalamic connectivity’, *Nature*, 575(7781), pp. 195–202. Available at: https://doi.org/10.1038/s41586-019-1716-z.

He, Y. and Evans, A. (2010) ‘Graph theoretical modeling of brain connectivity’, *Current Opinion in Neurology*, 23(4), pp. 341–350. Available at: https://doi.org/10.1097/WCO.0b013e32833aa567.

Hellinger, E. (1909) ‘Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen.’, *Journal für die reine und angewandte Mathematik*, 1909(136), pp. 210–271. Available at: https://doi.org/10.1515/crll.1909.136.210.

Henry, C.S. *et al.* (2016) ‘Microbial Community Metabolic Modeling: A Community Data-Driven Network Reconstruction’, *Journal of Cellular Physiology*, 231(11), pp. 2339–2345. Available at: https://doi.org/10.1002/jcp.25428.

Horvát, S. *et al.* (2016) ‘Spatial Embedding and Wiring Cost Constrain the Functional Layout of the Cortical Network of Rodents and Primates’, *PLOS Biology*. Edited by G. Konopka, 14(7), p. e1002512. Available at: https://doi.org/10.1371/journal.pbio.1002512.

*HS Niederrhein - Dalitz* (2020). Available at: https://lionel.kr.hs-niederrhein.de/~dalitz/data/hclust/#authors (Accessed: 28 July 2024).

Kamga, C., Krishnamurthy, S. and Shiva, S. (2012) ‘Myoglobin and Mitochondria: A relationship bound by Oxygen and Nitric Oxide’, *Nitric Oxide*, 26(4), pp. 251–258. Available at: https://doi.org/10.1016/j.niox.2012.03.005.

Lancichinetti, A. and Fortunato, S. (2009) ‘Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities’, *Physical Review E*, 80(1), p. 016118. Available at: https://doi.org/10.1103/PhysRevE.80.016118.

Li, K. and Pang, Y. (2013) ‘Communities analysis in protein-protein interaction networks’, in *2013 IEEE International Conference on Bioinformatics and Biomedicine*. *2013 IEEE International Conference on Bioinformatics and Biomedicine*, pp. 46–46. Available at: https://doi.org/10.1109/BIBM.2013.6732757.

Li, Y. and Turner, R.E. (2016) ‘Rényi Divergence Variational Inference’, in *Advances in Neural Information Processing Systems*. Curran Associates, Inc. Available at: https://papers.nips.cc/paper_files/paper/2016/hash/7750ca3559e5b8e1f44210283368fc16-Abstract.html (Accessed: 31 July 2024).

Mansour, Y., Mohri, M. and Rostamizadeh, A. (2009) ‘Multiple source adaptation and the Rényi divergence’, in *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*. Arlington, Virginia, USA: AUAI Press (UAI ’09), pp. 367–374.

Markov, N.T. *et al.* (2011) ‘Weight Consistency Specifies Regularities of Macaque Cortical Networks’, *Cerebral Cortex*, 21(6), pp. 1254–1272. Available at: https://doi.org/10.1093/cercor/bhq201.

Markov, N.T., Ercsey-Ravasz, Mária, *et al.* (2013) 'Cortical high-density counterstream architectures', *Science (New York, N.Y.)*, 342(6158), p. 1238406. Available at: https://doi.org/10.1126/science.1238406.

Markov, N.T., Ercsey-Ravasz, Maria, *et al.* (2013) 'The role of long-range connections on the specificity of the macaque interareal cortical network', *Proceedings of the National Academy of Sciences*, 110(13), pp. 5187–5192. Available at: https://doi.org/10.1073/pnas.1218972110.

Markov, N. T. *et al.* (2014) 'A Weighted and Directed Interareal Connectivity Matrix for Macaque Cerebral Cortex', *Cerebral Cortex*, 24(1), pp. 17–36. Available at: https://doi.org/10.1093/cercor/bhs270.

Markov, Nikola T. *et al.* (2014) 'Anatomy of hierarchy: Feedforward and feedback pathways in macaque visual cortex', *Journal of Comparative Neurology*, 522(1), pp. 225–259. Available at: https://doi.org/10.1002/cne.23458.

Mesulam, M.-M. (1990) 'Large-scale neurocognitive networks and distributed processing for attention, language, and memory', *Annals of Neurology*, 28(5), pp. 597–613. Available at: https://doi.org/10.1002/ana.410280502.

Meunier, D., Lambiotte, R. and Bullmore, E.T. (2010) 'Modular and Hierarchically Modular Organization of Brain Networks', *Frontiers in Neuroscience*, 4, p. 200. Available at: https://doi.org/10.3389/fnins.2010.00200.

Michalareas, G. *et al.* (2016) 'Alpha-Beta and Gamma Rhythms Subserve Feedback and Feedforward Influences among Human Visual Cortical Areas', *Neuron*, 89(2), pp. 384–397. Available at: https://doi.org/10.1016/j.neuron.2015.12.018.

Montgomery, D.C. (2020) *Introduction to statistical quality control*. Eighth edition, EMEA edition. Hoboken, NJ: Wiley.

Morales, D., Pardo, L. and Vajda, I. (2000) 'Rényi Statistics in Directed Families of Exponential Experiments*', *Statistics*, 34(2), pp. 151–174. Available at: https://doi.org/10.1080/02331880008802324.

Müllner, D. (2013) 'fastcluster: Fast Hierarchical, Agglomerative Clustering Routines for R and Python', *Journal of Statistical Software*, 53, pp. 1–18. Available at: https://doi.org/10.18637/jss.v053.i09.

Murray, G., Carenini, G. and Ng, R. (2012) 'Using the omega index for evaluating abstractive community detection', in *Proceedings of Workshop on Evaluation Metrics and System Comparison for Automatic Summarization*. USA: Association for Computational Linguistics (NAACL HLT '12), pp. 10–18.

Newman, M.E.J. (2004) 'Fast algorithm for detecting community structure in networks', *Physical Review E*, 69(6), p. 066133. Available at: https://doi.org/10.1103/PhysRevE.69.066133.

Newman, M.E.J. (2006) ‘Modularity and community structure in networks’, *Proceedings of the National Academy of Sciences*, 103(23), pp. 8577–8582. Available at: https://doi.org/10.1073/pnas.0601602103.

Nielsen, F. and Boltz, S. (2011) ‘The Burbea-Rao and Bhattacharyya Centroids’, *IEEE Transactions on Information Theory*, 57(8), pp. 5455–5466. Available at: https://doi.org/10.1109/TIT.2011.2159046.

Onnela, J.-P. *et al.* (2012) ‘Taxonomies of networks from community structure’, *Physical Review E*, 86(3), p. 036104. Available at: https://doi.org/10.1103/PhysRevE.86.036104.

Orsini, C. *et al.* (2015) ‘Quantifying randomness in real networks’, *Nature Communications*, 6(1), p. 8627. Available at: https://doi.org/10.1038/ncomms9627.

Palla, G. *et al.* (2005) ‘Uncovering the overlapping community structure of complex networks in nature and society’, *Nature*, 435(7043), pp. 814–818. Available at: https://doi.org/10.1038/nature03607.

Pensia, A., Jog, V. and Loh, P.-L. (2024) ‘Communication-Constrained Hypothesis Testing: Optimality, Robustness, and Reverse Data Processing Inequalities’, *IEEE Transactions on Information Theory*, 70(1), pp. 389–414. Available at: https://doi.org/10.1109/TIT.2023.3334024.

Pensia, A., Loh, P.-L. and Jog, V. (2022) ‘Simple Binary Hypothesis Testing under Communication Constraints’, in *2022 IEEE International Symposium on Information Theory (ISIT)*. *2022 IEEE International Symposium on Information Theory (ISIT)*, pp. 3297–3302. Available at: https://doi.org/10.1109/ISIT50566.2022.9834363.

Pessoa, L. (2014) ‘Understanding brain networks and brain organization’, *Physics of life reviews*, 11(3), pp. 400–435. Available at: https://doi.org/10.1016/j.plrev.2014.03.005.

Rao, R.P.N. and Ballard, D.H. (1999) ‘Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects’, *Nature Neuroscience*, 2(1), pp. 79–87. Available at: https://doi.org/10.1038/4580.

Ravasz, E. *et al.* (2002) ‘Hierarchical Organization of Modularity in Metabolic Networks’, *Science*, 297(5586), pp. 1551–1555. Available at: https://doi.org/10.1126/science.1073374.

Ravasz, E. (2009) ‘Detecting Hierarchical Modularity in Biological Networks’, in R. Ireton et al. (eds) *Computational Systems Biology*. Totowa, NJ: Humana Press, pp. 145–160. Available at: https://doi.org/10.1007/978-1-59745-243-4_7.

Renyi, A. (1961) ‘On measures of entropy and information’, *1961* [Preprint].

de Reus, M.A. *et al.* (2014) ‘An edge-centric perspective on the human connectome: link communities in the brain’, *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1653), p. 20130527. Available at: https://doi.org/10.1098/rstb.2013.0527.

Rosvall, M., Axelsson, D. and Bergstrom, C.T. (2009) ‘The map equation’, *The European Physical Journal Special Topics*, 178(1), pp. 13–23. Available at: https://doi.org/10.1140/epjst/e2010-01179-1.

Ruan, J., Dean, A.K. and Zhang, W. (2010) ‘A general co-expression network-based approach to gene expression analysis: comparison and applications’, *BMC Systems Biology*, 4(1), p. 8. Available at: https://doi.org/10.1186/1752-0509-4-8.

Saper, C.B., Scammell, T.E. and Lu, J. (2005) ‘Hypothalamic regulation of sleep and circadian rhythms’, *Nature*, 437(7063), pp. 1257–1263. Available at: https://doi.org/10.1038/nature04284.

Scheffé, H. (1956) ‘A “Mixed Model” for the Analysis of Variance’, *The Annals of Mathematical Statistics*, 27(1), pp. 23–36.

Shayevitz, O. (2011) ‘On Rényi measures and hypothesis testing’, in *2011 IEEE International Symposium on Information Theory Proceedings*. *2011 IEEE International Symposium on Information Theory Proceedings*, pp. 894–898. Available at: https://doi.org/10.1109/ISIT.2011.6034266.

Sporns, O. (2013) ‘Structure and function of complex brain networks’, *Dialogues in Clinical Neuroscience*, 15(3), pp. 247–262. Available at: https://doi.org/10.31887/DCNS.2013.15.3/osporns.

Sporns, O. and Betzel, R.F. (2016) ‘Modular Brain Networks’, *Annual Review of Psychology*, 67(Volume 67, 2016), pp. 613–640. Available at: https://doi.org/10.1146/annurev-psych-122414-033634.

Sporns, O., Tononi, G. and Kötter, R. (2005) ‘The Human Connectome: A Structural Description of the Human Brain’, *PLOS Computational Biology*, 1(4), p. e42. Available at: https://doi.org/10.1371/journal.pcbi.0010042.

Strong, D.R., Lawton, J.H. and Southwood, R. (1984) *Insects on plants: community patterns and mechanisms*. Cambridge, Mass: Harvard University Press.

Suresh, A.T. (2020) ‘Robust hypothesis testing and distribution estimation in Hellinger distance’. arXiv. Available at: http://arxiv.org/abs/2011.01848 (Accessed: 19 February 2024).

Váša, F. and Mišić, B. (2022) ‘Null models in network neuroscience’, *Nature Reviews Neuroscience*, 23(8), pp. 493–504. Available at: https://doi.org/10.1038/s41583-022-00601-9.

Vezoli, J. *et al.* (2021) ‘Cortical hierarchy, dual counterstream architecture and the importance of top-down generative networks’, *NeuroImage*, 225, p. 117479. Available at: https://doi.org/10.1016/j.neuroimage.2020.117479.

Wu, K. *et al.* (2011) ‘The Overlapping Community Structure of Structural Brain Network in Young Healthy Individuals’, *PLOS ONE*, 6(5), p. e19608. Available at: https://doi.org/10.1371/journal.pone.0019608.

Yatracos, Y.G. (1985) 'Rates of Convergence of Minimum Distance Estimators and Kolmogorov's Entropy', *The Annals of Statistics*, 13(2), pp. 768–774. Available at: https://doi.org/10.1214/aos/1176349553.

Yu, G. *et al.* (2017) 'ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data', *Methods in Ecology and Evolution*, 8(1), pp. 28–36. Available at: https://doi.org/10.1111/2041-210X.12628.

## Methods

### Data

We rely on two datasets from retrograde tract-tracing experiments in the macaque monkey, obtained with consistent methods. The datasets are the matrices $G^{\mathrm{FLN}}_{91\times 40}$ and $G^{\mathrm{SLN}}_{91\times 40}$ of FLN and SLN values $FLN_{ij}$ and $SLN_{ij}$ for 40 injected target areas ($i$ is a source area, projecting into target/injected area $j$), in a 91-area parcellation. The second one is the dataset that contains all pairwise distances along estimated shortest physical paths avoiding anatomical obstacles, between area barycenters, recorded in the matrix $D_{91\times 91}$. All data is available from (Froudist-Walsh *et al.*, 2021).

Community detection must be done in systems or subsystems with complete connectivity information. For that reason, from $G^{i}_{91\times 40}$ where $i \in \{\mathrm{FLN}, \mathrm{SLN}\}$, we extracted the subnetwork with 40 injected areas $G^{i}_{40\times 40}$ - also referred to as the edge-complete network (ECN) - whose connectivity and connection strength information have been measured between every pair of its nodes. Therefore, the communities and hierarchies are representative of the ECN, and that is not going to change when future injections are added to the dataset.

### Probability mass functions of area connections

Let us define the set of areas from the whole network as $U$ and from the ECN as $V \subseteq U$ with sizes $|U| = 91$ and $|V| = 40$, respectively. Each area $i \in V$ has two neighborhoods, one for its in- and another for its out-links. The in-neighborhood of (the target) area $i$ is defined as

$$P_i^- \equiv \frac{\{FLN_{1i}, \ldots, FLN_{|U|i}\}}{\sum_{j=1}^{|U|} FLN_{ji}} = \{FLN_{1i}, \ldots, FLN_{|U|i}\}.$$

Note that this probability mass function has data from the 91 areas, with $P_i^-(j)$ representing the probability that a neuron is projecting from source area $j$ into target area $i$, with normalization $\sum_{j=1}^{|U|} P_i^-(j) = 1$, due to the definition of the FLN.

For the out-neighborhoods, the probability mass functions will cover only areas in the ECN, because an injection reveals all the inputs from the rest of the network into the target area, but not the outputs. The outputs from a given source area need to be collected across all the injections. Since there are 40/91 injections, full connectivity information (and their FLN weights) exists only for the ECN. Thus, for (the source) area $i$, it is defined as

$$P_i^+ \equiv \frac{\{FLN_{i1}, \ldots, FLN_{i|V|}\}}{\sum_{j=1}^{|V|} FLN_{ij}}.$$

Thus, $P_i^+(j)$ represents the probability that a neuron sampled from source area $i$ projects into target area $j$, given that it projects into some area of the ECN.

**Hellinger distance**
is a measure between two probability distributions (Hellinger, 1909; Cam and Yang, 2000). As explained in the main text, the Hellinger distance $H$ is a metric, i.e., it satisfies the triangle inequality, and it is also bounded $0 \leq H \leq 1$. For simplicity, let us first define the squared-difference function $h(x,y) \equiv \left(\sqrt{x} - \sqrt{y}\right)^2$. In general, the squared Hellinger distance between the probability mass functions $P$ and $Q$ with common discrete support $V$ is defined as:

$$H^2(P,Q) = \frac{1}{2}\sum_{i \in V} h(p(i), q(i)).$$

In our case the probability distributions are associated with the elements of the set of nodes $V$ of the ECN, namely we are computing the squared Hellinger distance of the probability mass functions associated with nodes $A \in V$ (that is $P \equiv P_A$) and $B \in V$ ($Q \equiv P_B$) respectively, for all pairs of nodes $(A, B)$. Thus, $p_A(i) \in P_A$ represents the probability that node $i \in V$ projects into (or from) node $A$. One can think of $i$'s link to (from) $A$ as an (projection) "event". The definition for $p_B(i) \in P_B$ is similar. Because the probability mass functions are associated to the elements of the support itself, the sum above needs to be computed with care when $i = A$ or $i = B$, because in our dataset self-connections (intrinsic connections) are zero: $p_A(A) = p_B(B) = 0$. To illustrate that, assume that nodes $A$ and $B$ have identical neighborhoods (outside of $A$ and $B$), and identical probability mass functions (including $p_A(B) = p_B(A)$). If we were to compute $H^2$ naïvely, element-wise, the terms $h(p_A(B), p_B(B))$ and $h(p_A(A), p_B(A))$ would both be non-zero, and we would obtain a non-zero squared Hellinger distance, $H^2(P_A, P_B) \neq 0$, which would be incorrect, given that the two distributions are identical. For this reason, the correct expression of the squared Hellinger distance that we need to use is:

$$H^2(P_A, P_B) = \frac{1}{2}\sum_{i \in V/\{A,B\}} h\big(p_A(i), p_B(i)\big) + \frac{1}{2} h\big(p_A(B), p_B(A)\big) \equiv \frac{1}{2}\sum\nolimits'_{i \in V} h\big(p_A(i), p_B(i)\big),$$

where we use a primed sum notation to indicate that the summation is to be taken as explained above. When applying the above in our calculations, we have to do that separately for the out-neighborhoods and the in-neighborhoods of the nodes, i.e., working with $H^2\left(P_A^{\pm}, P_B^{\pm}\right) \equiv H_{\pm}^2$.

**Sample complexity**
Is defined as the sufficient number of samples needed to correctly differentiate two distributions with a classification error rate $\delta \leq 0.1$ (see next), in a simple binary hypothesis test (Pensia, Loh and Jog, 2022). Suppose $n$ samples are drawn from one of two probability distributions $P$ and $Q$ over support $V$. The test (function) $\phi$: $\cup_{i=1}^{\infty} V^i \to \{P, Q\}$ that classifies samples $X = (x^1, \ldots, x^n) \in V^n$ to $P$ or $Q$, solves the simple binary hypothesis test problem with sample complexity $\mathcal{N}^{\delta}(P, Q)$, if it is the smallest number that satisfies

$$\delta \geq \mathbb{P}_P[\phi(X) = Q] + \mathbb{P}_Q[\phi(X) = P],$$

where $\mathbb{P}_P[\phi(X) = Q]$ represents the (failure) probability of classifying the samples as coming from $Q$ when drawn from $P$, and vice versa, for the second term (Pensia, Loh and Jog, 2022). Moreover, it is known (Canonne *et al.*, 2019) that it scales as

$$\mathcal{N}^{\delta}(P,Q) = \Theta\left(\frac{-\log\delta}{H^2(P,Q)}\right),$$

where $\Theta(\cdot)$ is the Big-Theta notation for asymptotic tight (upper and lower) bound (Cormen, 2007).

Finally, if $\phi$ is the optimal likelihood test and $n \gtrsim \frac{-\log\delta}{H^2(P,Q)}$ for $\delta \leq 0.1$, the failure probability of the test $\phi$ with $n$ samples is less than $\delta$ (Pensia, Jog and Loh, 2024). In the case of the Scheffé's test $\mathbb{I}(x) = 1$ if $P(x) \geq Q(x)$ and $0$ otherwise (Scheffé, 1956), it is at most $O\left(\frac{1}{H^4(P,Q)}\right)$ (Pensia, Loh and Jog, 2022; Pensia, Jog and Loh, 2024) for $\delta = 0.1$, where $O(\cdot)$ stands for the Big-O notation for the asymptotic upper bound (Cormen, 2007); but it has been found that it can be a tight bound for certain distributions (Suresh, 2020).

**$\alpha$-Rényi divergence**

For two probability distributions $P_A$ and $P_B$, the $\alpha$-Rényi divergence (Renyi, 1961) is

$$D_{\alpha}(P_A||P_B) = \frac{1}{\alpha-1}\log\left(\sum_i' \frac{(p_{A(i)})^{\alpha}}{(p_{B(i)})^{\alpha-1}}\right),$$

where we already used the primed sum as defined for the Hellinger distance. This function has several properties. In particular, for any $P$ and $Q$, $D_{\alpha}(P||Q)$ increases as a function of $\alpha$. For $\alpha = 1$, $D_1(P||Q)$ is the Kullback-Leibler divergence, $D_{KL}(P||Q)$. For our purposes, the Rényi divergence of interest is the one with order parameter $\alpha = 1/2$, i.e.,

$$D_{1/2}(P_A||P_B) = -2\log\left(\sum_i' \sqrt{p_A(i)p_B(i)}\right).$$

We have chosen $\alpha = 1/2$ because it is the only Rényi divergence that is symmetric. This property is important since in our case, probability mass functions are associated with node neighborhoods. By comparing them, we do not aim to measure their information compression properties – a classical application of information divergence – but to measure how different they are. Moreover, since the probabilities of the same event from both probability mass functions, $p_A(i)$ and $p_B(i)$, are multiplied and not divided as in the Kullback-Leibler divergence (or any $\alpha \geq 1$), we can still measure their dissimilarity even if there are mutually exclusive events, i.e., events that have zero probability in one and non-zero in the other distribution. These two properties make the ½-Rényi divergence useful, since it makes it possible to relate network related quantities such as the similarity between node neighborhoods to information theory and statistical inference. It is associated with the squared Hellinger distance (Nielsen and Boltz, 2011) via equation (1). One simple interpretation of the ½-Rényi divergence $D_{1/2}(P \parallel Q)$ is that is quantifies the error incurred when replacing the distribution $P$ with distribution $Q$, assuming that the true distribution is $P$ and has many applications. For example, consider two sensors with the same sampling rate, monitoring the occurrence of some phenomenon in a noisy environment. $D_{1/2}$ in this case provides a lower bound on the worst-case misdetection rate of the phenomenon (Shayevitz, 2011). These functions help us to decode the functional organization of the network from an information theoretic perspective (Shayevitz, 2011; Canonne *et al.*, 2019).

**Similarity between directed links**

The similarity of out-links $e_{ij}$ and $e_{ik}$ from common a node $i$ is defined as $S(e_{ij}, e_{ik}) = 1 - H^2(P_j^-, P_k^-) = 1 - H_-^2$. For in-links $e_{ji}$ and $e_{ki}$, the similarity measure is $S(e_{ji}, e_{ki}) = 1 - H^2(P_j^+, P_k^+) = 1 - H_+^2$. The reason for using the in-neighborhoods of nodes $j$ and $k$ when measuring the similarity between the out-links $e_{ij}$ and $e_{ik}$ from node $i$, is because $e_{ij}$ and $e_{ik}$ are in-links to nodes $j$ and $k$, and in this way we are comparing in-links with in-links (one belonging to $j$, the other to $k$). In other words, the dissimilarity between directed links is the squared Hellinger distance of the uncommon nodes.

**Link community algorithm example in a toy model**

The link-community algorithm works as follows. At start, each link forms its own community, and we set the link community merging threshold value $th$ at $th = 0$. We define the distance between two LCs $C_i$ and $C_j$, $d(C_i C_j)$, as the minimum distance (in terms of either $H_+^2$ and $H_-^2$) over all pairs of links, one from $C_i$ the other from $C_j$, called *single-linkage criterion* in the literature (Müllner, 2013). Next, we increase $th$ until it reaches the smallest value among all the link-community distances at that step, i.e., until $th = min_{i,j}\{d(C_i C_j)\}$. We then merge all pairs of communities separated by this distance value $th$ into a new community at the next level, and we update the community list. This process is repeated until there is only one community left.

We next illustrate the algorithm on a simple toy model using a directed binary network with three nodes and five links which is shown in the **Extended Data Fig. 1**a. First, each link is associated with a link community. Then, we create a complete graph whose nodes are network's links and whose link weights are the dissimilarities between them (**Extended Data Fig. 1**b). The algorithm starts by selecting a dissimilarity threshold $th$ which is the lowest dissimilarity of the complete network (i.e., maximum similarity, $th = 0$). Then, link communities whose dissimilarities are equal to $th$ are merged pairwise (link merging process), creating new link communities (**Extended Data Fig. 1**b-1)-3)). Next, the dissimilarity threshold is increased to the next minimum link community dissimilarity (in the example, $th$ increases to $th = 0.29$), and the link merging process starts again (**Extended Data Fig. 1**b-4)). Note that as long as the dissimilarity between two members, one for each link community, equals to $th$, those link communities are merged. This is known as the single-linkage criterion (Müllner, 2013). The link merging process and single-linkage criterion is repeated until there is only one link community left (**Extended Data Fig. 1**b-5)). The information as to which link communities merged and at which $th$, is then used to create a dendrogram, also referred as link-community hierarchy (**Extended Data Fig. 1**c).

**Extracting the node-community hierarchy from the link-community hierarchy**

As stated in the Results section, to derive the node-community hierarchy using the link community algorithm requires specifying how node communities merge through the link merging process. We use two criteria. $i$) The *recurrence criterion* where two node communities merge if they are connected by a link community with recurrent connections, i.e., links (from the same link community) directed both ways between the node communities; and $ii$) the *end-step criterion*, in which node communities that did not merge by the end of the link merging process are then merged pairwise and sequentially. An illustration of the node merging process can be seen in **Fig. 4b-c**. Note, the node-community hierarchy depends on these criteria. We choose the recurrence criterion because this is a key property in several multi-agent systems as it allows coordination of information/balance; an example is the mutual inhibition of the neurons in the ventrolateral

preoptic and monoaminergic subcortical structures during sleep-wake transitions (Saper, Scammell and Lu, 2005). Moreover, it provides the advantage of stability (once two communities are merged, they will not split at any higher threshold value). Therefore, the node-community hierarchy is well-defined.

**An organizational analogy**

To illustrate the relationship between the link and node level community hierarchies we consider the university structure in terms of interactions between professors at a university. At the smallest research group level, the professors (nodes) are "performing research in the field of physics and subfield of condensed matter". In another group, other professors are "performing research in the field of physics and subfield of particle physics," etc. In the chemistry department, a similar structure holds, with different research groups. Within each group there are numerus back-and-forth interactions (frequent similar interactions). There are fewer interactions between the research groups, and those are typically at faculty meetings where decisions are made affecting all groups. In this case, from the point of view of the department chair, these are groups "performing research in physics," only – on a higher level of the functional hierarchy. The nature of discussions during faculty meetings will be specific to physics-based research needs. In the chemistry department, the same process happens and at their faculty meetings the discussions (interactions, functional links) will be specific to chemistry-based research needs. The two types of discussions in the two departments will clearly be different, although they will have commonalities. The two departments form two different nodal communities. The next level is at the College of Science, which houses (inclusion) both departments (and others). The dean is a mediator between these departments and when the dean has a meeting of the chairs (each representing their departments), the discussions are at the level of departmental research needs and the potential return on those research activities (and less on research group specifics within the departments). When an issue arises that is beyond a particular research area (such as research space availability, graduate student stipends, etc.) then these functional activities (links) are running between all the departments, because decisions here globally impact the College of Science. At the next level, that of the provost, similar interactions occur, but now between the deans of the different colleges (Science, Arts and Letters, Engineering, etc.) through links that affect all at that level. If there was only links going in one direction between two groups than those groups do not belong to the same level of the hierarchy, one is subordinate to the other.

**Loop entropy**

We introduce here the notion of loop entropy, that will provide us with a criterion that allows us to select an "optimal" level in the link-community hierarchy. We start with the definition that trees (which are connected networks whose number of links $M$ are related to its number of nodes $N$ by the equation $M = N - 1$), have no (i.e., zero) community structure. To further simplify matters we will ignore the directionality of the links in what follows. We define the number of excess links $M_{exc}$ of a network $G$ with $M$ links and $N$ nodes as $M_{exc}(G) = M - N + 1$ (which is zero for a tree).

Suppose a network $G$ is partitioned in $L$ mutually exclusive subgraphs $g^L = \{g_i\}_{i=1}^{L}$ i.e., $G = \bigcup_{i=1}^{L} g_i$ and $\emptyset = g_i \cap g_j \; \forall i, j \in \{1, \dots, L\}$. Then, the loop entropy, $S_L$, of that partition is defined as

$$S_L(g^L) = -\sum_{i=1}^{L} p_{exc}(i) \log p_{exc}(i) - p_{tree}(g^L) \log p_{tree}(g^L),$$

where $p_{exc}(i) = \frac{M_{exc}(g_i)}{M_{exc}(G)}$ is the probability of sampling excess links in the subgraph $g_i$ from the total number of excess links in $G$. Finally, $p_{tree}(g^L) = 1 - \sum_{i=1}^{L} p_{exc}(i)$ is a normalization factor.

The loop entropy is a measure of the distribution of excess links among the subnetworks, i.e., how much the partition differs from the tree (null) model which does not encode community information (recall, communities must have recurrent links). Since it is a Shannon-Gibbs type entropy, it is maximized when the distribution of excess links of the partition $P_{exc}(g^L) = \{p_{exc}(i)\}_{i=1}^{L} \cup p_{tree}(g^L)$ is uniform of the community structure.

Note, we have not proved that $S_L$ can always be computed since it might occur that there are partitions where $M(g) > M_{exc}(G)$ for a particular subgraph $g$, or that $\sum_{i=1}^{L} M_{exc}(g_i) > M_{exc}(G)$ (in this case we no longer work with probabilities). However, we found that the score did not have such issues in the networks analyzed by us. Here we use the loop entropy as a community quality score for link community partitions in various network models and benchmarks (Lancichinetti and Fortunato, 2009) and in all cases the $S_L$ was well defined. Until there is a mathematical proof that $S_L$ can be computed for any arbitrary mutually exclusive partition, we recommend monitoring the $S_L$ values to find possible anomalies, when using it with the link community algorithm. See the **Extended Data Fig. 6** for an example.

**Identifying the optimal node-community hierarchy level from the optimal link-community hierarchy**

We can use the loop entropy to find an optimal LCh level. However, finding the correspondent NCh level is not straightforward. In the **Extended Data Fig. 7**, we explain our method with an example.

**Cover assignment algorithm**

At a selected level (other than the first) from the NCh, there might be single-node communities. However, it is still desired to assign them to one or several non-single node communities, i.e., with more than one member. In the **Extended Data Fig. 8** and its caption we explain how the algorithm works, with an illustration.

**Omega index**

It is a measure of the agreement between community partitions of a network when nodes can belong to more than one group or covers. To our best knowledge, it was published in reference (Gregory, 2011a). Assume there is a network with $N$ nodes partitioned into two cover structures $C = \{C_i\}_{i=1}^{k}$ and $C' = \{C'_{i'}\}_{i'=1}^{k'}$ such that $C_i$ ($C'_{i'}$) is the set of nodes in cover with index $i$ ($i'$) (not necessarily mutually exclusive). Then, their omega index is

$$\omega(C, C') = \frac{\omega_u(C,C') - \omega_e(C,C')}{1 - \omega_e(C,C')},$$

where

$$\omega_u(C, C') = \binom{N}{2}^{-2} \sum_{j=0}^{\max(k,k')} \left| t_j(C) \cap t_j(C') \right|,$$

and

$$\omega_e(C,C') = \binom{N}{2}^{-2} \sum_{j=0}^{\max(k,k')} |t_j(C)||t_j(C')| \ .$$

In the above equations, $t_j(\cdot)$ represents the set of node pairs that appear exactly $j$ times in a cover. For example, suppose there are three nodes $a$, $b$, and $c$ with two cover structures $C = \{(a,b,c),(c,d)\}$ and $C' = \{(a,b),(c,d),(a,d)\}$, with two and three covers, respectively. Then, $j \in \{0,1,2,3\}$. For $j = 0$, $t_0(C) = \{(a,d),(b,d)\}$ and $t_0(C') = \{(a,c),(b,d),(bc)\}$. So, $|t_0(C) \cap t_0(C')| = 1$, which means that only the pair $(b,d)$ appears zero times in both cover structures. Note, that if a node pair belongs to $t_0$, it cannot belong to other $t_j$ for any $j > 0$, i.e., $t_j(C)$ are disjoint sets (mutually exclusive).

Then, $\omega_u$ represents the probability of selecting randomly a node pair with equal frequency which is observed in both cover structures for some $t_j(\cdot)$; $\omega_e$ denotes the probability of sampling a node pair with equal frequency but which might not exist in both cover structures for some $t_j(\cdot)$. Finally, the term $1 - \omega_e$ is the normalization factor for the case $\omega_u = 1$. Notice that $\omega = 1$ if and only if the node pair appearance frequency sets $t_j(\cdot)$ have the same elements for both cover structures and for all $j$s, denoting a strong similarity/overlap.

**Validation**

To validate the performance of the algorithm, we generated weighted and directed Lancichinetti-Fortunato (LF) benchmarks with ground-truth node community memberships, allowing some nodes to participate in more than one cover (Lancichinetti and Fortunato, 2009) (see **Extended Data Fig. 5**). The benchmark generates networks with power-law degree and power-law community-size distributions. Here, we choose the degree and community size distribution exponents to be $t_1 = 2$ and $t_2 = 1$, respectively.

Additionally, the benchmark has topological mixing parameter $\mu_t$ and weight mixing parameter $\mu_w$ that control the complexity of the generated networks (explained next). Suppose $k_i$ is the total degree of node $i$. Also, define the total strength of the node to be $s_i \equiv k_i^{\beta}$. Here, $\beta = 2$. Then, the internal degree $k_i^{(in)}$ and internal strength $s_i^{(in)}$ are the number of neighbors of node $i$ and the sum of the weights of the links between node $i$ and all the nodes of the network within the same cover as node $i$. They are defined as $k_i^{(in)} \equiv (1-\mu_t)k_i$ and $s_i^{(in)} \equiv (1-\mu_w)s_i$ with $\mu_t, \mu_w \in [0,1]$. By definition, the external degree $k_i^{(ext)}$ and external strength $s_i^{(ext)}$ are $k_i^{(ext)} = k_i - k_i^{(in)}$ and $s_i^{(ext)} = s_i - s_i^{(in)}$, respectively. Clearly, low values of $\mu_t$ or/and $\mu_w$ imply that most of the links or link strengths of a node are distributed within the communities belonging to the same cover as the node itself, signaling a sharp community structure. On the other hand, large values of $\mu_t$ or $\mu_w$ represent a more diffuse community structure.

Now, let $w_{ij}$ denote the link weight of the link from node $i$ to node $j$. The set of link weights $\{w_{ij}\}$ are generated randomly and then modified to minimize the function $\mathrm{Var}(\{w_{ij}\}) = \sum_i (s_i - \rho_i)^2 + \left(s_i^{(in)} - \rho_i^{(in)}\right)^2 + \left(s_i^{(ext)} - \rho_i^{(ext)}\right)^2$, where $\rho_i = \sum_j w_{ij}$, $\rho_i^{(in)} = \sum_j w_{ij}\kappa(i,j)$, $\rho_i^{(ext)} = \sum_j w_{ij}(1-\kappa(i,j))$, and $\kappa(i,j) = 1$ if nodes $i$ and $j$ have at least one common node-community membership, and 0 otherwise.

Here, $\mu_t$ was varied between 0.1 (compact communities; low complexity) to 0.8 (diffuse communities; high complexity) in steps of 7/90 (10 variations) while $\mu_w = 0.01$ is fixed. We choose a low $\mu_w$ because these networks are sparse, otherwise the external strength $s^{(ext)}$ would be distributed among only a small number of inter-community links, making them much stronger than the within-community links, a biologically non-plausible scenario.

Other parameters are the average and maximum degree $\langle k \rangle = \{5, 7, 10\}$ and $k_{\max} = \{30, 50\}$ (**Extended Data Fig. 5**), and the minimum and maximum community size parameters. Here they are 5 and 25, respectively.

We choose the generated benchmarks to have 100 nodes, where 5% of them participate in 2, 3 or 4 overlapping community memberships (Om), shown as columns in **Extended Data Fig. 5** and the rest of the nodes participate in only one. For each point in the parameter space, 100 instances were generated.

After generating the networks, we used our algorithm to find their cover structure, including multi-community memberships. We did that by computing the NCh of the networks and finding the optimal LCh level with the maximum average link community density $\langle D \rangle$ (Ahn, Bagrow and Lehmann, 2010) (as explained below) and loop entropy, $S_L$.

The average link-community density $\langle D \rangle$ on level $l$ of the LCh in a network with $M$ links is

$$\langle D \rangle_l = \sum_{c \in C(l)} D_c \frac{m_c}{M},$$

where $C(l)$ is the set of link communities at that level, $D_c$ is the "effective" density of the link community $c \in C(l)$. We call it "effective" because it is the density relative to the link community tree null version (Ahn, Bagrow and Lehmann, 2010), i.e.,

$$D_c = \frac{m_c - (n_c - 1)}{n_c(n_c - 1) - (n_c - 1)} = \frac{m_c - (n_c - 1)}{(n_c - 1)^2},$$

where $m_c$ and $n_c$ are the number of links and nodes of the link community $c \in C(l)$. As mentioned earlier, the number of links of the tree null version of a link community is $n_c - 1$, and the total possible number of links between the nodes of the link community is $n_c(n_c - 1)$ because the graph is directed in this case.

Once the optimal level is found in the LCh, we find the corresponding optimal level of the NCh. Then, we use the cover assignment algorithm to assign covers to possible single-community nodes and compute the omega index $\omega$ between the predicted and the ground-truth community structures. In the **Extended Data Fig. 5**, we show that $S_L$ performs better on average than $\langle D \rangle$, in most of the cases, except the ones where communities are very dense such as the last row of **Extended Data Fig. 5**, where the large $\langle k \rangle = 10$ and the small $k_{\max} = 30$ results in compact node communities. In all other cases, $S_L$ is a better score to identify the optimal node community partition.

Finally, we note that the weighted and directed networks with nodes that participate in multiple communities generated from the LF benchmark are very challenging for most state-of-the-art community detection algorithms; and none of those algorithms, as far as we know, work in dense networks. Unlike the algorithm presented here, they are not able to produce a NCh, encoding community information at multiple levels.

**Using interareal distance as a dissimilarity measure**

Here we briefly consider an alternative to the Hellinger distance (which ultimately is based on connectivity strengths), the physical interareal separation distance as "dissimilarity" measure. At first sight this seems acceptable, as according to the general observation areas that are more dissimilar (in function) are also further apart in space. Moreover, according to **Fig. 3**b, $D_{1/2}$ scales linearly, on average, with the interareal distance. However, also note that there is a significant amount of variation around the average, which likely holds valuable information, that the average does not capture. And this indeed is the case, as shown in **Extended Data Fig. 9** and described in its caption, in more detail.

## Glossary of terms

Retrograde tract-tracing: It is an experimental procedure (Batardiere *et al.*, 1998; Barone *et al.*, 2000; Falchier *et al.*, 2002; Markov *et al.*, 2011) in which chemical markers injected into a region/area of the brain, called target, label the cell-bodies of the cells whose projections end in the injected area. The labeling happens through axonal transport from the terminal arbors to the soma of the cell. There is no secondary labeling.

Fraction of labeled neurons (FLN): Given a connection directed from node $i$ to node $j$, its FLN value $f_{ij}$ is the ratio between the number of neuronal labels found in source area $i$ given an injection in target area $j$ and the total number of labels extrinsic to $j$. It can be interpreted as the probability that neural activity in area $i$ will induce (suppress) neural activity in area $j$ (Falchier *et al.*, 2002; Markov *et al.*, 2011).

Edge-complete network (ECN): In this type of network, the connection and connection attributes, e.g., weight, between each network's node is known. Usually important when working with networks that are edge-incomplete, i.e., with missing connections.

EDR network model: The exponential distance rule (EDR) is the empirical observation (Ahn, Jeong and Kim, 2006; Ercsey-Ravasz *et al.*, 2013; Horvát *et al.*, 2016) that the distribution of the axonal lengths in the cortex decays exponentially, with a rate $\lambda$: $p(l) = ce^{-\lambda l}$, where $l$ denotes the axonal (bundle) length, from the soma of the presynaptic neuron to the synapse with the postsynaptic neuron. The EDR network model is a generative interareal cortical random network model based on the maximum entropy principle that incorporates the EDR and the matrix of distances/lengths between the barycenters of the areas, see References (Ercsey-Ravasz *et al.*, 2013; Markov, Mária Ercsey-Ravasz, *et al.*, 2013; Horvát *et al.*, 2016; Gămănuţ *et al.*, 2018) for details on the model.

Dendrogram: Tree-like data structure that stores the relationships between datapoints.

Cover: It is a listing of the identities of all the node communities of a node. When the node belongs to just one community, its cover is formed by the id of that community; in case it belongs to multiple communities, its cover is the union of all identities.

Supragranular labeled neurons (SLN): Given a connection directed from node $i$ to node $j$, its SLN value $s_{ij}$ is the number of supra-granular neurons (above layer IV) divided by the total number across the cortical column found in source area $i$ given an injection in target area $j$ (Nikola T. Markov *et al.*, 2014). Its value is associated with two connection types with different physiological functions in the context of predictive coding. Values below 0.5 are called feedback (FB) and they typically propagate prediction probabilities from high- to low-level cortical areas; and above 0.5 are feedforward (FF) and typically propagate errors from the predictions made in low-level areas to high-level areas and have the role to update their internal state (Rao and Ballard, 1999; Barone *et al.*, 2000; Bastos *et al.*, 2012).

## Software packages

We used Python 3.9.18 and libraries SciKit-Learn version 1.1.1, SciPy version 1.13.1, NetworkX version 2.8.8, and Pybind11 2.10.0. For plotting, we used Matplotlib version 3.5.2 and Seaborn version 0.13.2, and R-library ggtree version 3.8.2 (Yu *et al.*, 2017). Benchmarks were generated using the Lancichinetti-Fortunato model for directed,

weighted, networks with nodes with multiple communities (Lancichinetti and Fortunato, 2009). We also used the hclust-cpp library that implements the hierarchical clustering algorithm (Müllner, 2013; *HS Niederrhein - Dalitz*, 2020). To compute the omega index, we used a publicly available code (Murray, Carenini and Ng, 2012).

**Table 1. Cover neighborhood direction of the nodes with overlapping community membership in Fig. 5b.** At the optimal loop entropy level, three areas do not merge with any non-single node community. As seen in **Fig. 5b**, the non-single node community ids are C1, C2, C3, C4, C5, and C6. Using the cover assignment algorithm, those areas are assigned to some non-single node communities by maximizing their average in/out-neighborhood similarity, or both (see Methods section). The neighborhood direction information is shown here.

| Area | In | Out | Both |
|---|---|---|---|
| 7b | C2 and C4 | - | C5 |
| 1 | C2 | - | C5 |
| OPRO | - | C3 and C6 | C5 |

**Data availability**

The network connectivity data is available from https://core-nets.org/.

**Code availability**

The codes used in this study can be obtained via request to JMA.

**Acknowledgments**

This work was supported by the National Science Foundation (NSF) grant IIS-1724297 (J.M.A., Z.T.), French National Research Agency (ANR) grants A2P2MC ANR-17-NEUC-0004 (K.K., H.K.), ANR-17-FLAG-ERA-HBP-CORTICITY (to H.K. ANR-19-CE37–0025-DUAL_STREAMS (to K.K.).

**Author contributions**

The authors would like to express their thanks for discussions with B. Molnár, L. Magrou, Y. Hou. JMA and ZT designed the research, JMA wrote all the algorithms and ran the simulations, JMA performed the computational and data analysis with guidance from ZT, KK and HK provided the experimental datasets, JMA and ZT wrote the paper with assistance from KK and HK, and all authors contributed to editing the paper.

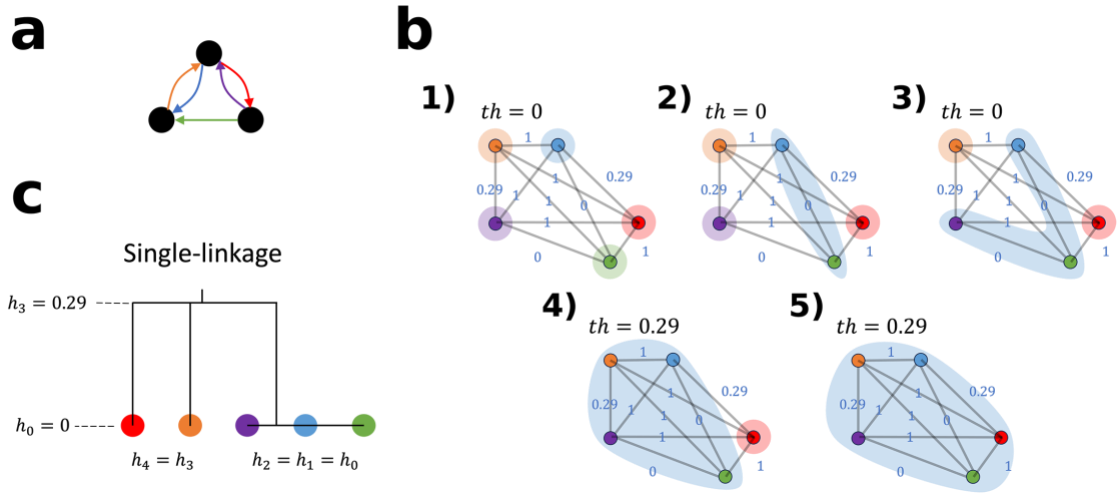
**Competing interests**

The authors declare no competing interests.

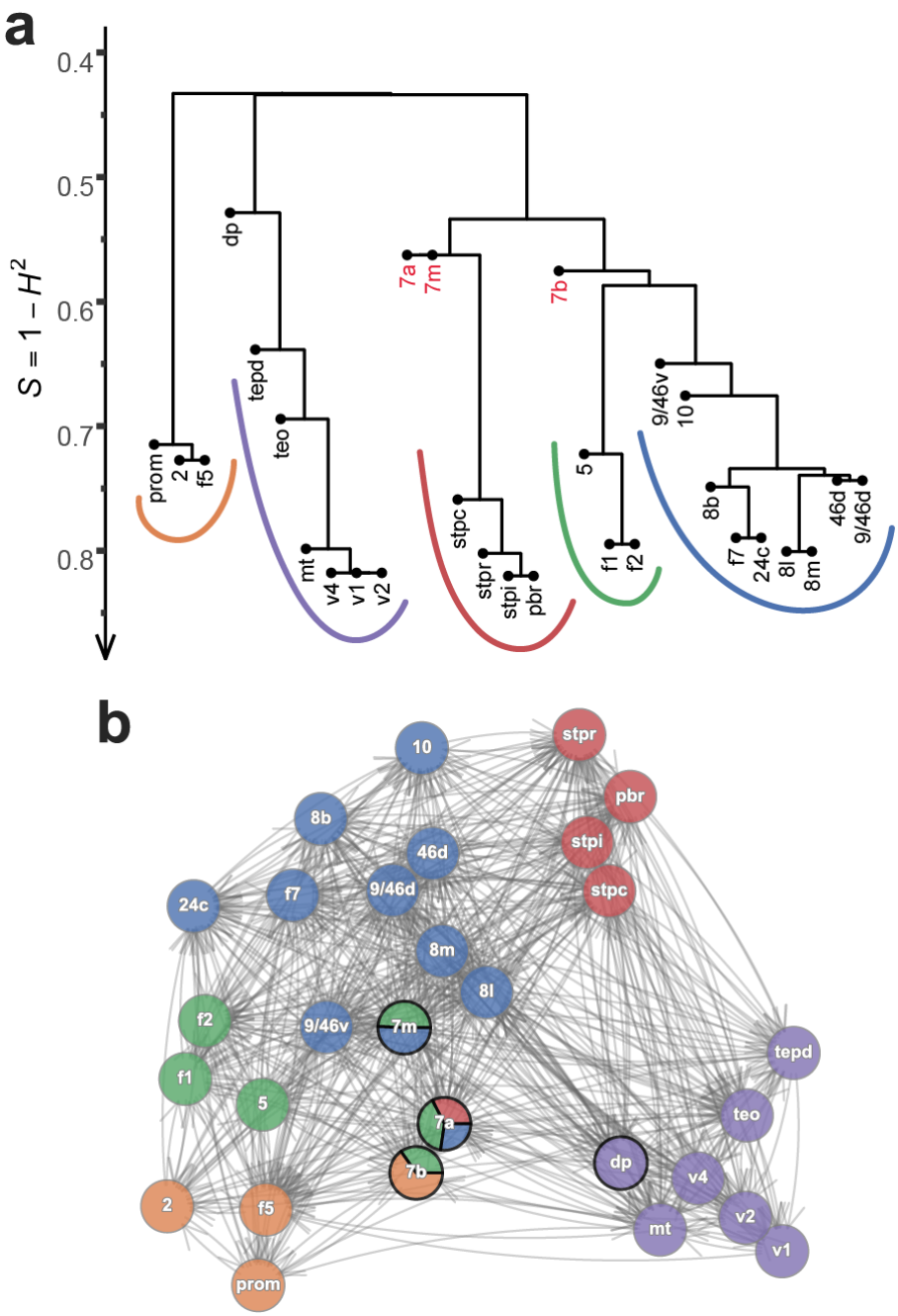
**Additional information**

**Extended data** are found below.

**Correspondence and request for materials** should be addressed to Z.T.

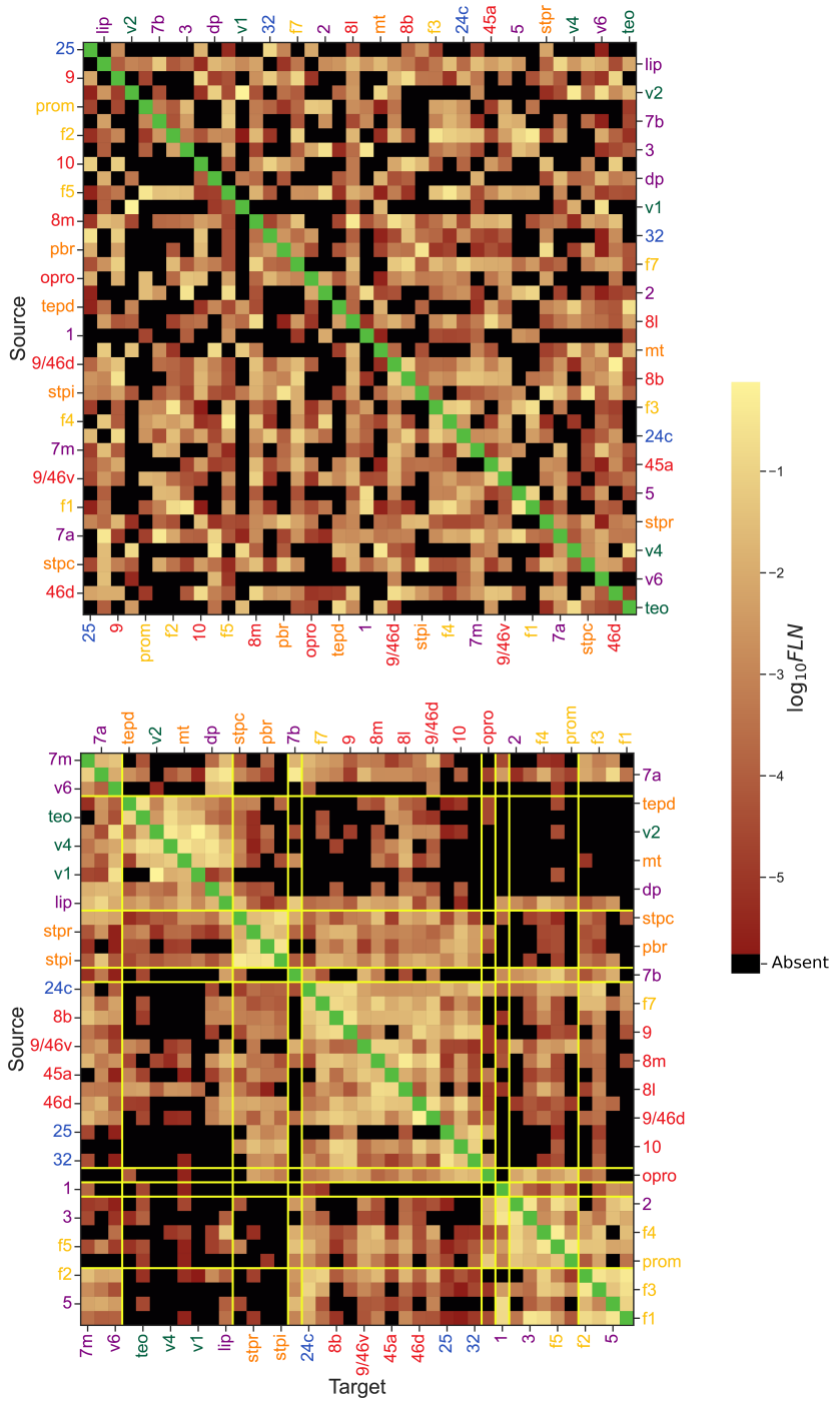



**Extended Data Fig. 1 | Link community algorithm example. a,** To illustrate how the link community algorithm works, we use a toy binary directed network with three nodes and five links. Each link is color differently. In **b**, the network's nodes are associated with the links in **a**. The $H^2$ dissimilarity between links is written with a blue number. **b1)-b3),** The algorithm starts out by setting the lowest link dissimilarity threshold $th = 0$ then merging pairwise those links whose dissimilarity is equal to the threshold. **b4),** If no more link community pairs can be merged, the threshold is increased to the next lowest link dissimilarity value. Note that as soon as the dissimilarity between two members (one for each link community) equals to $th$, the link communities are merged. This rule is known as the single-linkage criterion. **b5),** The process of merging link community pairs and increasing the threshold continues until there is only one link community left. **c,** The information about which link community pairs merged at which $th$ value is stored in the LCh.

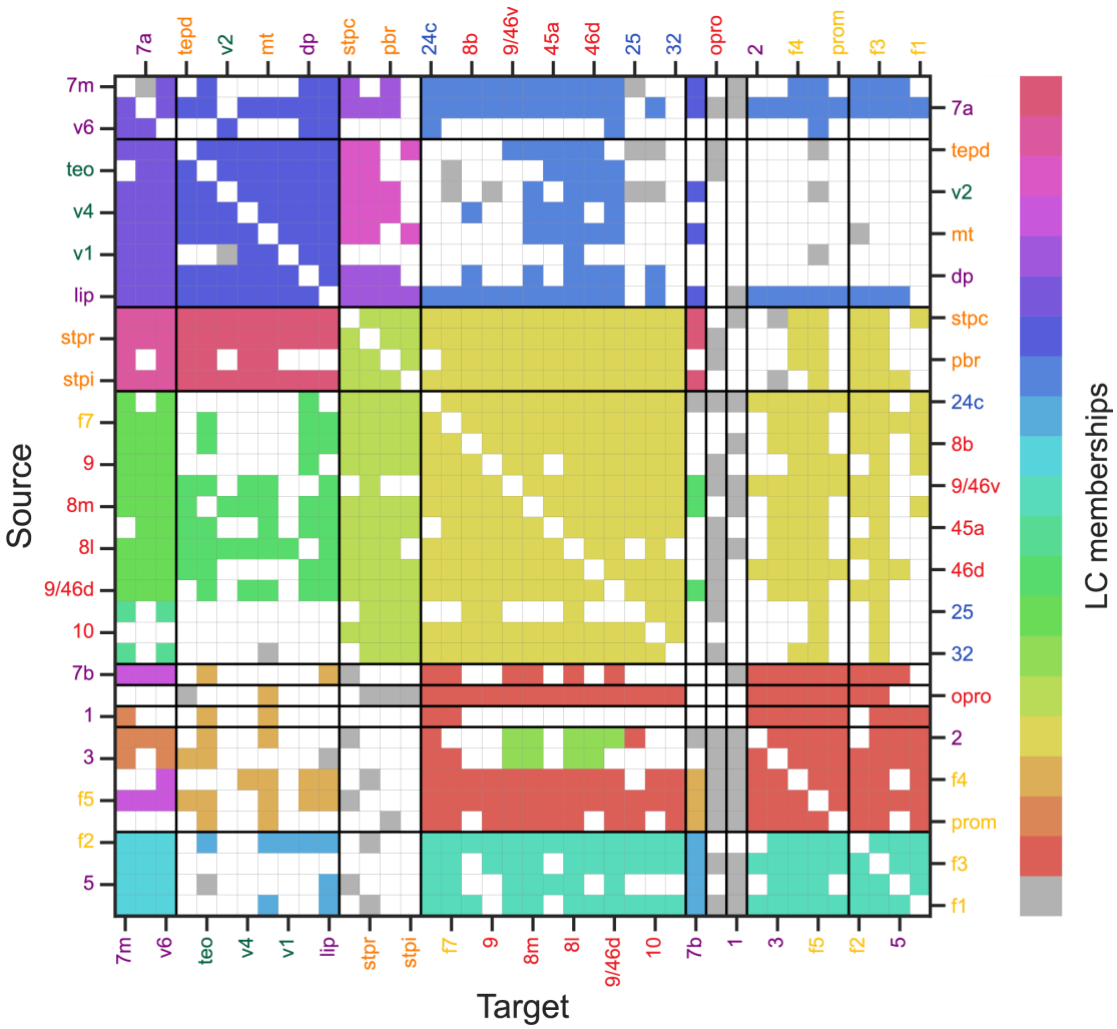



1182

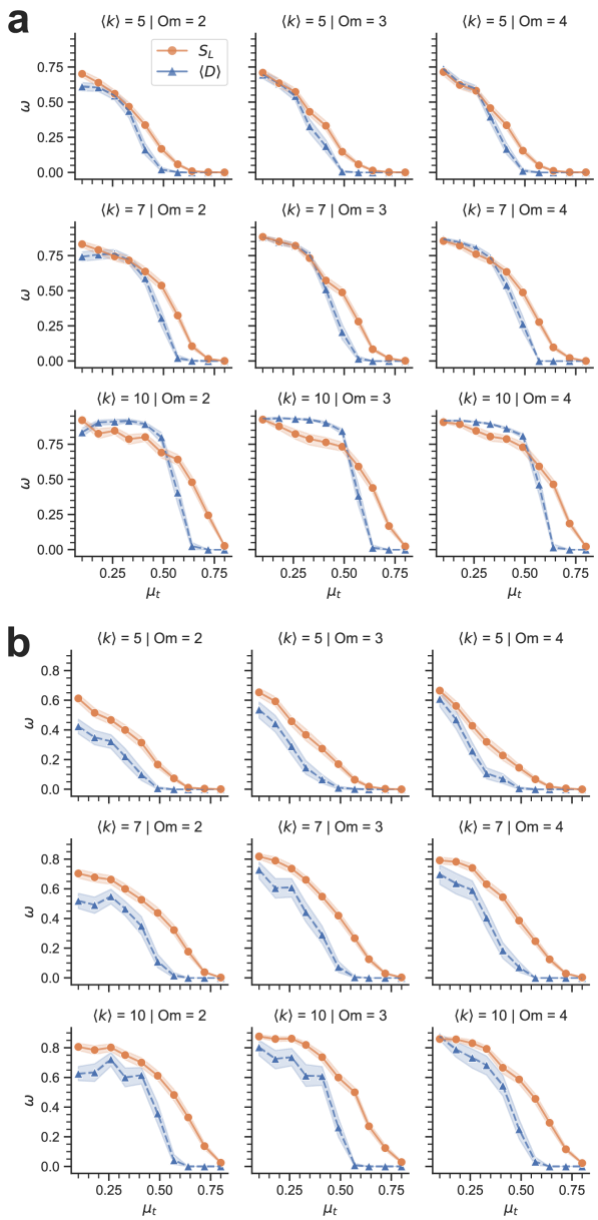
**Extended Data Fig. 2 | Community analysis of the edge-complete network of a previous FLN macaque dataset with 29 injected areas (N. T. Markov *et al.*, 2014a). a**, All the branches of this NCh correspond to one branch of **Fig. 4d**. This shows that the hierarchical community structure of the interareal cortical network is robust since 11 extra injections does not change substantially the structure already observed with 29. Red areas are the NOCs shown in panel **b**. **b**, At the optimal loop entropy partition, there are five node communities and three NOCs. Only area 7b is also classified as a NOC in the ECN with 40 injected areas. As expected, NOCs are more sensible to network changes. However, some are conserved and will be present even in the final dataset, after all the areas have been injected.



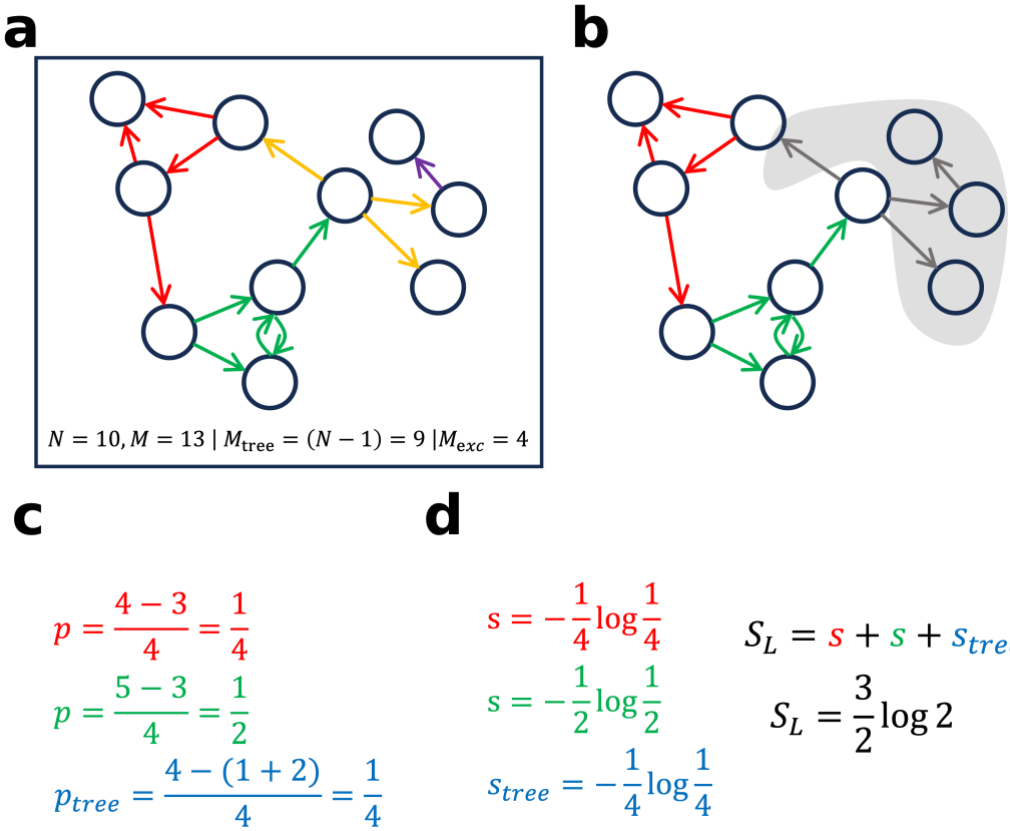

**Extended Data Fig. 3 | Connection strength heatmaps of the ECN.** Rows and columns are arranged randomly at the top while organized following the order of the NCh at the bottom. Yellow vertical and horizontal lines show the borders between node communities at the optimal loop entropy level. Absent and intrinsic connections are colored in black and green, respectively. Without information about the network's community structure, it is challenging to distinguish it from a structureless random graph.



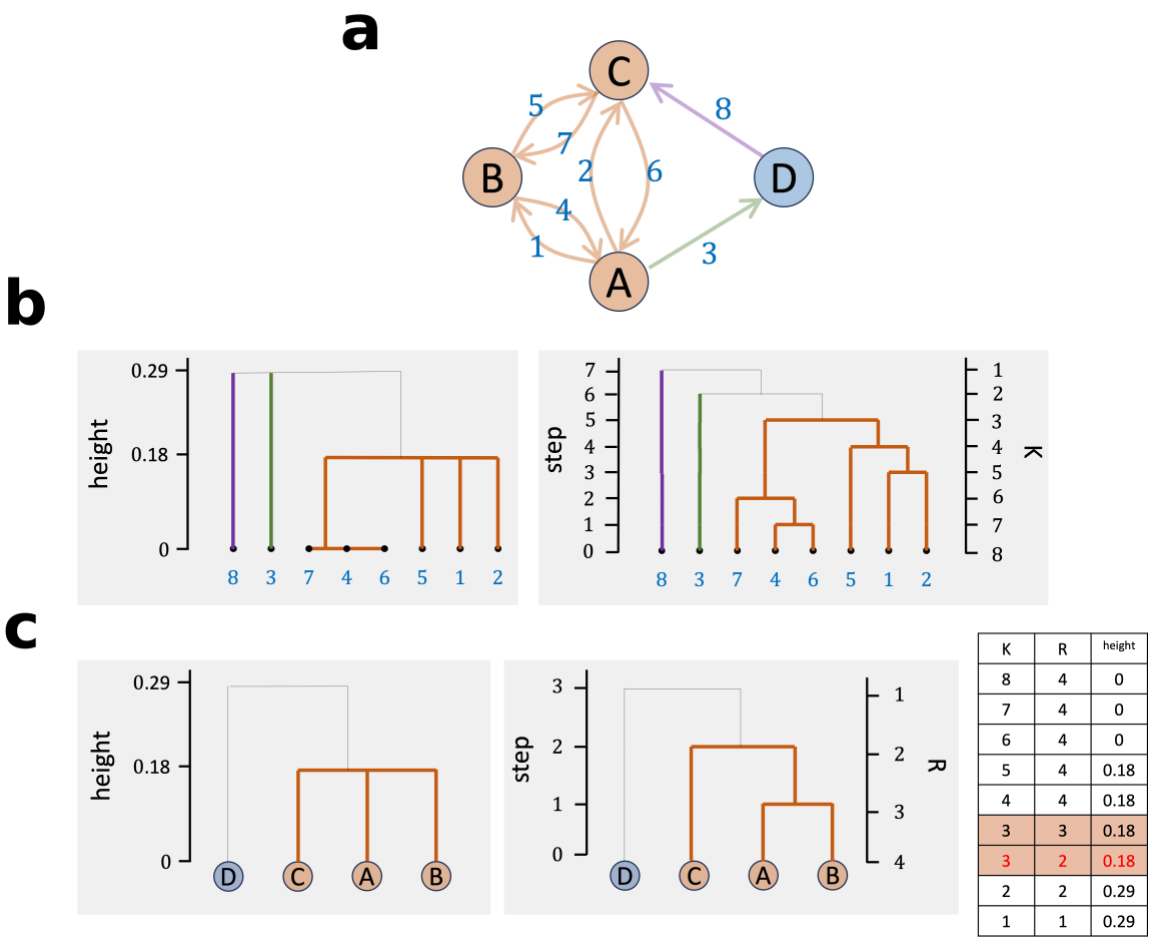

**Extended Data Fig. 4 | Link communities at the optimal level.** Links are colored by their LC membership except those in gray, corresponding to links that belong to a tree link community. At the maximum loop entropy level, there are 56 link communities, 20 of which belong to non-tree LCs (non-grey colored). Rows and columns are the same as in the **Extended Data Fig. 3** . Note that, using our methodology, LCs tend to connect nodes from the same NC or between NCs – a key observation for interpretability of the LCs, which otherwise might be challenging to interpret. Additionally, links of single-node communities tend to be trees, showing high dissimilarity with other network links and which might be following other organizational principles.




**Extended Data Fig. 5 | Algorithm validation. a-b,** We used the LF benchmark (Lancichinetti and Fortunato, 2009) to generate weighted and directed random networks with ground-truth community structure, allowing some nodes to belong to more than one community. The networks have 100 nodes, 5% of them have 2, 3, or 4 overlapping community memberships (Om; columns). The average degree $\langle k \rangle$ of the networks is 5, 7, or 10 (rows). We varied the topology mixing parameter $\mu_t$ (see Methods section) from 0.1 to 0.8 in steps of 7/90 to assess the performance in networks from low to high complexity. Panel **a** and **b** are for networks where the maximum degree of a node is 30 and 50, respectively. For each point in the parameter space, we generated 100 instances, computed the node community hierarchy, found the optimal level using the average link community density $\langle D \rangle$ (see Methods section) and loop entropy $S_L$, assigned covers to any single-community node, and compared the predicted and ground-truth cover structures using the omega index $\omega$. The analysis shows that $S_L$ tends to perform better than the $\langle D \rangle$ in most cases, except for the last row of panel **a**, where due to the effect of creating networks with average degree close to the maximum degree, which itself is of low value (sparser network), there are few external links (between nodes participating in different communities), and in this case local density can be a more suitable community score. Markers and shades represent averages and one standard deviation region, respectively.

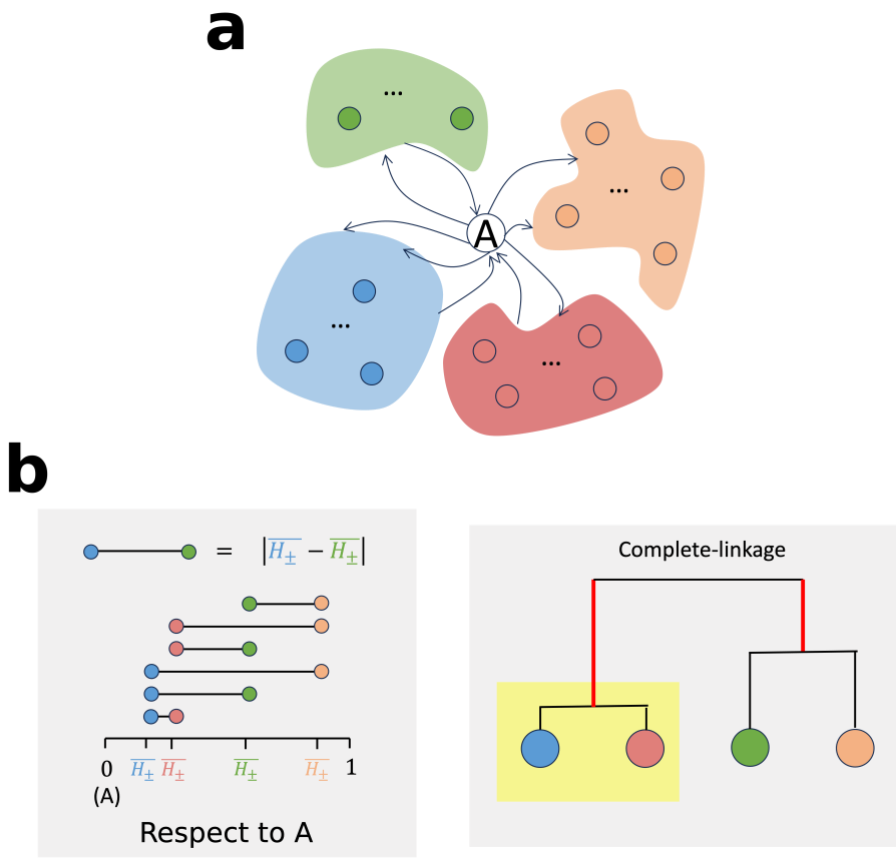



**Extended Data Fig. 6 | Loop entropy example. a,** Toy network with 10 nodes and 13 directed links. Links are distributed in five link communities. The number of excess $M_{exc}$ links of this network is four. **b,** The red and green link communities are the only ones that are not trees because their $M_{exc} > 0$. **c,** We compute the probability of sampling excess links in the red and blue green link communities and the normalization factor $p_{tree}$. **d,** Finally, the loop entropy $S_L$ is simply the Shannon-Gibbs entropy of the excess-link probability distribution.

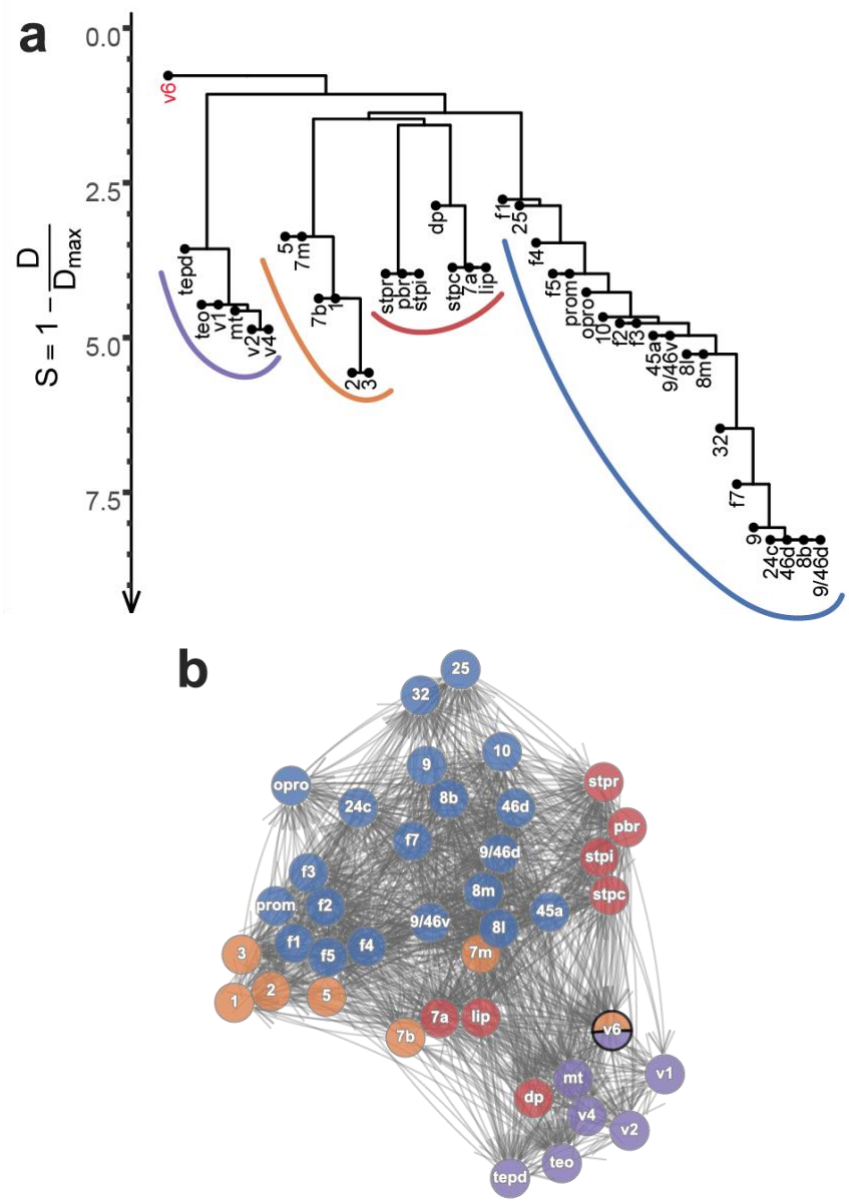



| K | R | height |
|---|---|---|
| 8 | 4 | 0 |
| 7 | 4 | 0 |
| 6 | 4 | 0 |
| 5 | 4 | 0.18 |
| 4 | 4 | 0.18 |
| 3 | 3 | 0.18 |
| 3 | 2 | 0.18 |
| 2 | 2 | 0.29 |
| 1 | 1 | 0.29 |

**Extended Data Fig. 7 | Finding a node-community hierarchy level from a link-community hierarchy level. a,** Toy network with four nodes and eight directed links. Links are numbered and their colors represent link communities found at the optimal level. Node communities associated with that level are similarly colored. Node "D" is a single-node community. **b**, (left) LCh of the network where levels are written in terms of the threshold (height). **b**, (right) Same LCh but its levels have been separated by link merging step and number of link communities, K, as shown by the left and right vertical axis, respectively. Each node at the bottom is one link from the toy network. Branches show the progression of the link merging process, and thick colored ones represent different link communities. **c**, Same as **b**, but for the NCh. Hierarchy levels separated by (left) height; **c**, (center) node merging steps, and number of node communities, denoted by R. Thick colored branches are associated with the non-single node community. **c**, (right) By storing in a table the information of K, R, and height across the link merging process, we can find how many node communities R existed at any LCh level K. If there are more than one node merging event at a particular K, we select the one with lowest R.




**Extended Data Fig. 8 | Cover assignment algorithm. a**, In this illustration, node "A" is a single-node community, i.e., no other node in the network participates with "A" in the same node community. We want to assign to some of the existing non-single node communities, shown by colored shapes. **b**, (left) To that aim, for each directed neighborhood, the mean Hellinger distance ($\overline{H_{\pm}}$) between the single-node community and the nodes in each non-single node communities is computed. The distance represents the dissimilarity centroid (DC) between the single-node community and a non-single node community. Given that $H$ is a metric, the relative difference between each $\overline{H_{\pm}}$ is computed and will satisfy the triangle inequality. (right) For each direction, the DCs are clustered using a hierarchical agglomeration algorithm with complete-linkage criterion, i.e., the dissimilarity between groups is the maximum dissimilarity between group's elements (Müllner, 2013), creating a DC hierarchy. The criterion is chosen since DCs are expected to associate with highly distinguishable clusters. To finish, the DC hierarchy is partitioned at the level before the largest dissimilarity threshold increment (thick red branches), and the single-node community receives the membership(s) of the DCs in the branch with the most similar DC.




**Extended Data Fig. 9 | Community analysis using interareal distance as dissimilarity measure.** **a**, NCh derived by replacing areal similarities by one minus the interareal distance divided by the maximum whole brain interareal distance. There are some commonalities and differences with respect to the NCh using the Hellinger dissimilarity measure (H-NCh). For example, the visual ventral system, above the purple arc, is almost the same as in the H-NCh. However, the rest of the branches tend to differ. A marked discrepancy is seen in the branch above the blue arc, which consists of prefrontal areas, because it does not show any sub-branches, i.e., a lack of functional specialization. The red area V6 is a NOC in this model, in contrast with the H-NCh. **b**, At the optimal loop entropy level, four communities are found, which is less than the six shown in **Fig. 4d**. Additionally, the omega index of this partition and the one from data is $\omega = 0.48$, which is close to the value of the one for the EDR model ($0.44$), see **Fig. 6a**. This shows that besides the ventral visual system, the model does not reproduce well the expected functional components of the macaque ECN. Therefore, the interareal distance-based hierarchy encodes around 50% of the functional structure found with Hellinger distances.