Multi-Scale Parcellation of Dynamic Causal Models of the Brain

Tahereh S. Zarghami[1*]

[1] Bio-Electric Department, School of Electrical and Computer Engineering, University of Tehran, Iran

* Correspondence: tzarghami@ut.ac.ir

Abstract

The hierarchical organization of the brain's distributed network has received growing interest from the neuroscientific community, largely because of its potential to enhance our understanding of the human cognition and behavior, in health and disease. While most multiscale connectivity analyses focus on structural and functional networks, characterizing the effective connectome across multiple scales has been somewhat overlooked – mostly for computational reasons. The difficulty of estimating large cyclic causal models on the one hand, and the scarcity of theoretical frameworks for systematically moving between scales on the other hand, have hindered progress in this direction. This technical note introduces a top-down multi-scale parcellation scheme for dynamic causal models, with application to neuroimaging data. The method is based on Bayesian model comparison, as a generalization to the well-known ΔBIC method. To facilitate the computations, recent developments in linear dynamic causal modeling (DCM) and Bayesian model reduction (BMR) are deployed. Specifically, a naïve version of BMR is introduced, using which the parcellation scheme can be scaled up to hundreds or thousands of regions. Notably, the derivations reveal an analytical relationship between the model comparison score and the concept of cut size in graph theory. This duality puts the tools of graph theory at the service of model evidence optimization and significance testing. The proposed method was applied to simulated and empirical causal models, for face and construct validity. Notably, the large causal network (inferred from a neuroimaging dataset) showed scale-invariance in multiple measures. Future generalizations of this technique and its potential applications in systems and clinical neuroscience have been discussed.

Keywords: dynamic causal modeling, Bayesian network, Bayesian model reduction, structure learning, partitioning, graph theory, Markov blanket, scale invariance

Introduction

Brain connectomic studies have been progressing towards multiscale analyses in recent years (Axer & Amunts, 2022; Betzel & Bassett, 2017; D'Angelo & Jirsa, 2022; Lu et al., 2022; Shine et al., 2021; Van Den Heuvel et al., 2019). Understanding the full richness of the brain's multiscale structure and function - across time, space and topology – has motivated a large body of empirical work (Hansen et al., 2023; Maksimenko et al., 2018; Morante, Frølich, Shahzaib, et al., 2024; Munn et al., 2024) and several methodological developments (Doucet et al., 2011; Farahibozorg et al., 2024; Morante, Frølich, & Rehman, 2024; Presigny & Fallani, 2022). Multiscale analyses are specifically attractive, because they hold the potential to illuminate the complex relationship between the hierarchical organization of the brain and human behavior and cognition: in health, disease, injury, over development and in aging (Bastiani & Roebroeck, 2015; Betzel & Bassett, 2017; Bonifazi et al., 2018; He et al., 2025; Latifi & Carmichael, 2024; Lu et al., 2022; Pines et al., 2022; Sala et al., 2017; Zhou et al., 2023; Zou et al., 2024).

While hierarchical decomposition of structural and functional networks are ubiquitous in the neuroimaging literature (Ashourvan et al., 2019; Ferrarini et al., 2009; Iraji et al., 2023; Meng et al., 2022; Meunier et al., 2010; Mirzaeian et al., 2025; Park et al., 2022; Rousseau et al., 2025), parcellating effective (i.e. causal) networks of the brain remains largely unexplored, with a few recent exceptions (D'Acunto et al., 2024; K. J. Friston et al., 2021; Wu et al., 2011; C. Wang & Shanechi, 2019). The challenges are mostly computational, as well as theoretical (Bhattacharya et al., 2024). While learning the structure and parameters of large (cyclic) Bayesian networks is computationally demanding, testing hypotheses about their potential modular organization calls for devoted theoretical work. But this journey is worth the effort because, neither structural nor functional network analyses can offer the mechanistic insight that a model-based causal framework can offer. For instance, (K. J. Friston et al., 2021) has recently shown that recursive coarse-graining and demarcation of neuronal states based on their effective connectivity (plus summarizing their dynamics at each scale based on the renormalization group theory) can explain the emergence of large-scale networks, self-organized criticality and dynamical instability in the brain. These are the sort of deep questions that only causal models can entertain.

Herein, we intend to present a top-down[1] parcellation approach. As we will see, the advantage of the top-down treatment is the systematic identification of subordinate components, as opposed to the more arbitrary designation of internal states in the bottom-up approach (K. J. Friston et al., 2021). This advantage comes at the cost of further computational challenges in the top-down scheme. Fortunately, recent developments in the estimation of large linearized DCMs (Frässle et al., 2017, 2021; K. J. Friston et al., 2021) and sparsification of densely connected Bayesian networks (K. Friston et al., 2019; Marković et al., 2024; Prando et al., 2020; Wright et al., 2024) have facilitated the estimation of large sparse DCMs – including hundreds or thousands of regions. Hence, the present paper tackles the remaining issue, i.e. revealing the hierarchical modular structure of a (large) causal network, starting at the top, i.e. the coarsest scale.

In short, the present study offers a top-down recursive parcellation scheme for a (potentially large) effective connectivity graph, such that at each scale the neuronal states are (bi)partitioned into statistically significant subordinate parcels. To ease the computations, the proposed scheme relies on a naive formulation of Bayesian model reduction (BMR) that eliminates costly matrix multiplications, and reveals an analytical relationship between the reduced model evidence and the minimum cut problem in graph theory.

The rest of this paper is organized as follows: In the Theory section, we go through the rationale of the proposed approach and its relation to the better-known ΔBIC method. Therein, we will see how causal model comparison crosses path with graph theory. Assessing the significance of the identified partitions is covered in the same section. Thereafter, illustrative examples of bi-partitioning simulated (modular and nonmodular) causal networks are presented to show the functionality of the proposed approach. Next, the same method is applied to the causal graph estimated from an empirical neuroimaging dataset - to reveal the hierarchical organization in a dynamic causal model of the brain. We will show the scaling behavior of multiple measures for this empirical causal network. Furthermore, we will study whether the partitions identified from the causal graph can be interpreted as

[1] Conventionally, (biological) systems are modeled either bottom-up or top-down. Bottom-up models explain how elementary causes lead to ensemble functions, whereas top-down models start from ensemble descriptions and try to infer the underlying components (D'Angelo & Jirsa, 2022).

statistically independent subsystems conditioned upon their boundary (i.e. mediating) states. This question is of interest because recent (simulation) work (Aguilera et al., 2022) has suggested that recurrent connections in cyclic causal models can induce statistical dependencies that cross causal boundaries. We will provide a hypothesis testing framework that can inspect this phenomenon for any given dataset and causal model. Finally, potential applications of the proposed parcellation scheme in systems and clinical neuroscience are discussed, and some future lines of research are outlined.

Theory

Dividing a group of datapoints into several cohesive groups is a recurrent problem in many domains, which has been tackled with a multitude of (supervised and unsupervised) methods. These methods fall under different umbrella terms[2]: clustering, partitioning, segmentation, community/module detection, change point detection, blind source separation, data decomposition, etc. The solution is usually formulated in order to either minimize the within-group[3] *dispersions* or maximize the within-group *cohesions*. The chosen measure of *dispersion* may be, e.g., *variance* in K-means clustering (Lloyd, 1982) or *entropy* in infomax-based ICA (Bell & Sejnowski, 1995). Likewise, *cohesion* may be defined in terms of *connectedness*, in density-based clustering (Ester et al., 1996) and modularity maximization (Newman, 2006), or data *likelihood* under distributional assumptions, in Gaussian mixture models (Dempster et al., 1977) and maximum likelihood-based ICA (Ans et al., 1985; Hyvärinen, 2013), among others.

Specifically, maximum likelihood (when penalized for model complexity) has commonly been used as a model selection criterion to divide some data (usually timeseries) into two or more coherent segments. The goal is to find the breakpoint(s) such that fitting a separate model to each segment would yield a more plausible explanation for the data, than fitting a single model to the entirety[4]. The plausibility of each model is determined by a criterion of

---

[2] Hence, the following terms are used generically and interchangeably in this paper, without intending the specialized terminology of any domain: segment, partition, module, subset, cluster, community, parcel.

[3] This usually turns out to be equivalent to maximizing the between-group separation, e.g. based on the law of total variance in kmeans clustering.

[4] This is conceptually similar to piecewise regression.

the form: $-2\log\text{likelihood} + \text{some penalty term for model complexity}$ (as a function of the number of model parameters and sample size). Depending on the form of the penalty, the criterion becomes: Akaike information criterion (AIC), Bayesian information criterion (BIC), or minimum description length (MDL) (Gao et al., 2024). For example, in a bisegmentation problem, the optimal breakpoint would maximally reduce $BIC_1 + BIC_2$ compared to BIC_0; where BIC_1 and BIC_2 result from fitting separate models to the two segments (before and after the breakpoint) and BIC_0 is computed for the single model fitted to the entire data. This approach is known as the ΔBIC method, since the objective is to minimize the following cost function (Chen & Gupta, 1997; Gao et al., 2024):

$$\Delta BIC = (BIC_1 + BIC_2) - BIC_0 < 0 \tag{1}$$

More statistically grounded, the same problem has been formulated in the hypothesis testing framework, where the null hypothesis H_0 corresponds to a model without any breakpoints and the alternative hypothesis H_a proposes the existence of a breakpoint (Gao et al., 2024). Hypothesis testing based on ΔBIC is a hybrid approach (Chen & Gupta, 1997), in which case a significance level α can be associated with the model selection result. This is realized by adding a critical value $C_\alpha \geq 0$ such that the null hypothesis is rejected only when:

$$BIC_1 + BIC_2 + C_\alpha < BIC_0 \tag{2}$$

Basically, the role of C_α is to ensure that the difference between $BIC_1 + BIC_2$ and BIC_0 is statistically convincing and not merely driven by noise. Approximate C_α values have been computed by (Chen & Gupta, 1997) for different significance levels α, based on the asymptotic null distribution of ΔBIC. The ΔBIC method – with or without hypothesis testing – has been used in different contexts to segment various sorts of temporal (Chung-Hsien Wu et al., 2006; Costa et al., 2016; Theodorou et al., 2014; Tritschler & Gopinath, 1999; Xue et al., 2010) and spatial data (Vatsavai et al., 2011; Zhao et al., 2008) – see (Gao et al., 2024) for a recent review and numerous references.

Here we will adopt a similar statistically supported model comparison approach to justify the (multi-scale) parcellation of a dynamic causal model fitted to regional timeseries. Our goal is to partition the regions associated with the timeseries. As we will shortly see, once we move beyond maximum likelihood to a fully Bayesian treatment, the heuristic

formulations of the ΔBIC method fall out naturally and turn out to be simplified forms of established constructs in Bayesian model comparison. Also, note that the same methods can be used to divide the data into more than two partitions (Chen & Gupta, 1997; Gao et al., 2024). Hence, we shall start with the bi-partition problem and later discuss the generalization to multi-partitions.

Problem statement:

Having fitted a (dynamic) causal model to the regional timeseries of a distributed network of interacting elements, how can we use this causal information to plausibly (bi)partition the system? How to score the plausibility of the partitioning?

As probably clear by the introduction above, this is at heart a model comparison problem; and the solution offered herein is inspired by the ΔBIC method. In short, we shall compare the *null hypothesis* (H_0) of fitting one DCM to all the regional timeseries versus the *alternative hypothesis* (H_a) of fitting two DCMs separately to the partitioned regions. Notably, each way of bi-partitioning the network serves as an alternative hypothesis, associated with some model evidence. The winning model, amongst the alternatives, reveals the most plausible way to divide the overall causal network into two disjoint causal models/modules. This is the main message of this paper. In what follows, we will elaborate on the implementation of this model comparison, and offer some solutions for the computational challenges, through six key points:

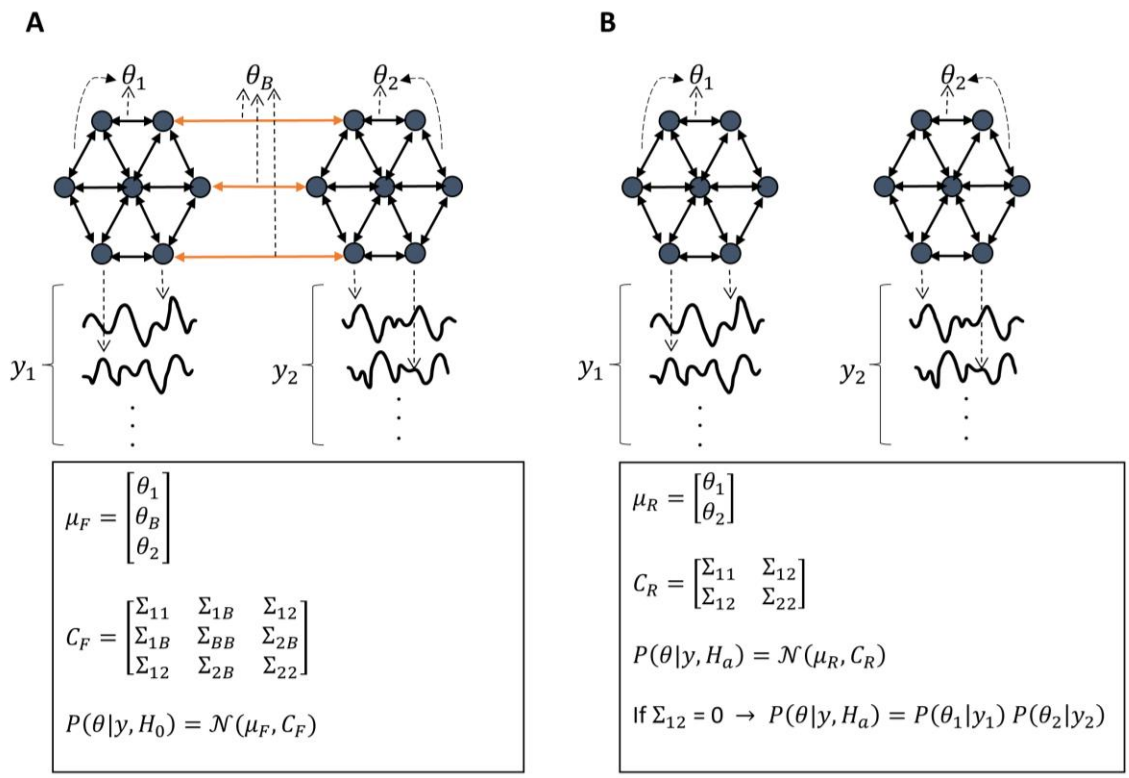

Fig. 1: Schematic of the full and reduced models. (A) A dynamic causal model (DCM) fitted to the timeseries from 14 regions/nodes. Intra-partition connections are denoted by θ_i, whereas θ_B represents between-partition connections. The causal connections are normally distributed with mean μ_F and covariance matrix C_F. The full model serves as the null hypothesis, H_0. (B) The reduced model, which serves as the alternative hypothesis H_a, is comprised of two DCMs fitted separately to the two partitions. Practically this means that between-partition connections have been removed from the full model, such that the partitioned model is *nested* inside the full model.

I) ΔF relation to ΔBIC: In general, the logarithm of the model evidence ratio (for the alternative versus the null hypothesis) can be written as:

$$\log \frac{P(y|H_a)}{P(y|H_0)} \approx \Delta F = F(H_a) - F(H_0) \tag{3}$$

where F denotes variational free energy optimized for each model, i.e., $F \approx \log P(y|model)$, and y is the observed data. Notably, ΔF has the interpretation of log Bayes factor[5], which

[5] Bayes factor is the ratio of two competing statistical models represented by their evidence, and is used to quantify the support for one model over the other (Kass & Raftery, 1995).

conventionally 'strongly' supports one hypothesis against another when it exceeds 3 nats[6], and 'very strongly' when it exceeds 5 nats (Kass & Raftery, 1995). But we also know that (negative) BIC is a special case of the free energy approximation to log model evidence - which drops all the terms that do not scale with the number of data points (Penny, 2012; Penny et al., 2004). So, under what circumstances can ΔF be decomposed as follows, like ΔBIC (Eq. 1)?

$$\Delta F = (F_1 + F_2) - F_0 \tag{4}$$

The answer lies in the implicit statistical independence assumption in the ΔBIC method. If we assume that the two local partitions can be modeled independently, then the joint distribution over their parameters can be written as the product of partition-specific distributions; i.e., $P(\theta|H_a) = P(\theta_1)P(\theta_2)$. Similarly, the generative model, as well as the true and approximate posterior distributions assume factorized forms under the alternative hypothesis:

$$P(y,\theta|H_a) = P(y|\theta,H_a)P(\theta|H_a) = P(y_1|\theta_1)P(y_2|\theta_2)P(\theta_1)P(\theta_2) = P(y_1,\theta_1)P(y_2,\theta_2)$$

$$P(\theta|y,H_a) = \frac{P(y,\theta|H_a)}{P(y|H_a)} = \frac{P(y_1,\theta_1)P(y_2,\theta_2)}{P(y_1)P(y_2)} = P(\theta_1|y_1)\,P(\theta_2|y_2)$$

$$P(\theta|y,H_a) \approx Q(\theta|H_a) = Q(\theta_1)Q(\theta_2) \tag{5}$$

where y is the observed data (i.e. timeseries), θ denotes the parameters/hidden variables[7], $Q(\theta)$ denotes the variational (i.e. approximate) posterior and subscripts denote partition-specific entities. For visualization, please see Fig. 1. As such, looking for maximally independent partitions of nodes comes down to optimizing the factorization over subsets of model parameters (i.e. causal connections) to maximize the model evidence. By simply substituting the factorized model in the free energy formula (Appendix A), the free energy of the partitioned model can be decomposed as the sum of the free energies of the two

[6] $\Delta F = 3$ nats means the alternative hypothesis is $\exp(3) \approx 20$ times more likely than the null hypothesis, reminiscent of the p = 0.05 significance level in frequentist statistics.

[7] A dynamic causal model usually includes hidden variables (ϕ) other than the causal connections (θ) as well (e.g., the hemodynamic variables), in which case the partitioned model looks like: $P(\theta,\phi|H_a) = P(\theta_1,\phi_1)\,P(\theta_2,\phi_2)$. For brevity, the ϕ variables are not shown explicitly in this paper. One can assume they have been lumped together with the θ variables.

partitions; i.e. $F(H_a) = F_1 + F_2$. Such evidence accumulation (akin to ΔBIC) is the mathematical consequence of assuming that local subsets of the data can be modeled independently. Hence, ΔBIC is a special case for the more general formulation in Eq. (3).

Additionally, note that ΔF is *negative* when we partition (i.e. break) an optimized causal model of an integrated system[8] - because none of the alternative hypotheses can (or should) win over the null hypothesis of an optimized integrated model. However, the winner model is expected to have the highest (i.e. *least negative*) ΔF among the alternatives, denoted by ΔF^{max}. So, the first step is to compute and maximize ΔF. And, the second step would be inspecting the significance of ΔF^{max}, as explained in the following.

II) Partition significance: To claim that an integrated system is partitionable (i.e. modular), we need to check whether ΔF^{max} is statistically significant, at some α significance level. This is achieved by checking whether $\Delta F^{max} > C_\alpha$, where C_α is a critical value for rejecting the null hypothesis[9], and is determined from the null distribution of ΔF^{max} such that:

$$P\ [\Delta F_{rnd}^{max} > \mathrm{C}_\alpha |\ H_0^{rnd}] = \alpha \tag{6}$$

Here H_0^{rnd} stands for the null hypothesis of optimally bi-partitioning a randomly connected model (with some properties matching the original model). We will discuss how to build the null distribution of ΔF^{max}(denoted by ΔF_{rnd}^{max} for clarity) later in this section. To summarize so far, our goal is to (1) compute and maximize ΔF and (2) make sure $\Delta F^{max} > \mathrm{C}_\alpha$, or equivalently, p-value $= P[\Delta F_{rnd}^{max} > \Delta F^{max}] \leq \alpha$. The remaining points below offer practical methods to achieve these goals.

III) Bayesian model reduction: The third point is about efficient computation of ΔF – before trying to maximize it – for any alternative hypothesis, H_a. Unlike the ΔBIC method, here we don't need to compute F_1 and F_2 by fitting separate models to subsets of the data.

---

[8] If ΔF is positive, it shows that the removed connections were *redundant* in the original causal model, which implies that the structure learning stage of the model development had been inadequate. In SPM12, a DCM can be structurally optimized post-hoc (using greedy search and Bayesian model reduction) using a routine called spm_dcm_bmr_all.m. We will talk more about Bayesian model reduction shortly.

[9] Note the resemblance with Eq. (2): $-\Delta BIC > C_\alpha$.

Since the structure of any two-partition model is *nested* inside the full (unpartitioned) model, we can use the machinery of *Bayesian model reduction* (BMR) to compute $F(H_a)$ in one shot and analytically. This is realized by treating the alternative hypothesis as a reduced (i.e. nested) *prior* in the formulation of BMR, and computing $F(H_a)$ based on the information we have from fitting the full model (i.e., its free energy, plus prior and posterior distributions), which leads to (Appendix B):

$$\Delta F = F(H_a) - F(H_0) = \log E_{Q(\theta|H_0)}\left[\frac{P(\theta|H_a)}{P(\theta|H_0)}\right] \tag{7}$$

Under Gaussian assumptions on the distributions of the causal connections (in variational Laplace), the BMR formulae admit simple forms based on the prior and posterior mean vectors and covariance matrices of the causal connections (K. Friston et al., 2007; Zeidman et al., 2023):

$$\Delta F = \frac{1}{2} Ln\,|\Pi_R P_F C_R \Sigma_F| - \frac{1}{2}(\mu_F^T P_F \mu_F + \eta_R^T \Pi_R \eta_R - \eta_F^T \Pi_F \eta_F - \mu_R^T P_R \mu_R) \tag{8}$$

For the notation and a succinct introduction to BMR, please refer to Appendix B. Now, what if the network is large and matrix multiplications hinder these computations?

IV) Naïve BMR: The fourth point is computationally crucial for large networks. When the network contains many (e.g. n = 1000) nodes, multiplying and inverting the covariance matrices (with e.g. $n^2 = 1000^2 =$ one million entries) quickly becomes prohibitive in BMR calculations. In this case, naïve mean-field approximations (MFA) (Dauwels, 2007; Geiger & Meek, 2005; Jaakkola & Jordan, 1998; Jordan et al., 1998; Marković et al., 2024; E. Xing, 2004; E. P. Xing et al., 2002) can be used to factorize the variational posterior and simplify the BMR computations. As shown in Appendix C, MFA leads to a naïve BMR formulation with an intuitive ΔF expression that avoids matrix operations:

$$\Delta F = F(H_a) - F(H_0) = -\frac{1}{2}\sum_{i\in\{C\}}\left(\sigma_i^{-2}\mu_F^2(i) + \log\left(\frac{\sigma_i^2}{\sigma_0^2}\right)\right) \tag{9}$$

Here, μ_F and σ_i^{-2} denote the posterior means and precisions of the connections, while σ_i^2 and σ_0^2 denote their posterior and prior variances, respectively. The summation is performed over the set of removed (i.e. between-partition) connections, $\{C\}$. Hence, to maximize ΔF, the causal network should be partitioned to minimize the R.H.S. of Eq. (9):

$$\max_{\{C\}} \Delta F \Leftrightarrow \min_{\{C\}} \sum_{i\in\{C\}} \left(\sigma_i^{-2}\mu_F^2(i) + \log\left(\frac{\sigma_i^2}{\sigma_0^2}\right)\right) \tag{10}$$

V) ΔF^{max}relation to minCut: This point is a bridge to the realm of graph theory, which offers an abundance of optimization heuristics. Notably, the objective function in Eq. (10) resembles the *minimum cut* (minCut) problem in graph theory (Von Luxburg, 2007), where the Cut refers to the sum of between-partition connections. In this case, the graph for which we seek the minCut is a function of the original graph, where the connections have been squared and precision-weighted ($\sigma_i^{-2}\mu_F^2$), and adjusted to reflect the change in their variances aposteriori ($\log\left(\frac{\sigma_i^2}{\sigma_0^2}\right)$). Let's unpack the implications.

Squaring the connections (μ_F^2) implies that the magnitudes (and not signs) of the connections matter, so removing weaker connections would minimize the cut. At the same time, scaling by σ_i^{-2} steers the cut towards less precise connections. On the other hand, the (log) variance ratio term advocates for removing overly precise connections (because they result in unfavorable "brittle" models by contributing to the complexity term in free energy (Penny et al., 2004). For typical DCM values, the log ratio term (which is negative[10]) is several orders of magnitude smaller than the (positive) connection squared term, which renders the former negligible in practice. As such, solving the minCut problem for the transformed graph comes down to finding and cutting through the relatively 'weak and uncertain' connections - to carve up the graph into disjoint partitions.

The well-known issue with minCut is that it becomes an NP-hard combinatorial optimization problem when the partitions are required to be balanced (Von Luxburg, 2007). Hence, a computationally efficient approximate solution, called *spectral clustering*, has been developed. In spectral clustering, the eigenvectors of the graph Laplacian matrix[11] constitute low-dimensional representations of the graph vertices, a process known as spectral or

[10] Typically, the posterior variances of the connections are several orders of magnitude smaller than the prior variance, meaning that our estimates become more *precise* after seeing the data. That is, $\sigma_i^2 \ll \sigma_0^2 \ll 1$, which leads to $\frac{\sigma_i^2}{\sigma_0^2} \ll 1$ and $\log\left(\frac{\sigma_i^2}{\sigma_0^2}\right) < 0$.

[11] The unnormalized graph Laplacian is computed as: $L = D - A$, where D is a diagonal matrix holding the node degrees, and A is the adjacency matrix.

Laplacian embedding. Then, a clustering algorithm partitions the vertices/nodes in this lower dimensional space, to minimize (a relaxed version of) the balanced cut function and yield partitions that are comparable in size. For a thorough review of spectral clustering please refer to (Von Luxburg, 2007). We intend to apply spectral clustering to the transformed graph based on Eq. (10). As traditional spectral clustering is applied to symmetric graphs[12], the transformed graph is symmetrized by adding the transpose of its adjacency matrix to itself (and dividing by two, to preserve the Cut value). In summary, applying spectral clustering[13] to the (transformed and symmetrized) causal graph provides an approximate solution to the (balanced) minCut problem, which in turn maximizes ΔF (Eq. 10). So far, we have a reasonable solution for our first goal (i.e., maximizing ΔF). What remains is comparing our winning model's ΔF^{max} against that of a null model – to check for the statistical significance of the partitions.

VI) Null distribution of ΔF^{max}: To assess the significance of the partitions, we need the null distribution of the (optimized) reduced free energy, i.e. ΔF^{max}**.** Recall that Eq. (10) had established a relationship between ΔF and the (transformed) graph's Cut value, to which we will refer as the "free energy-graph cut" duality:

$$\Delta F = -\frac{1}{2}\sum_{i\in\{C\}}\left(\sigma_i^{-2}\mu_F^2(i) + \log\left(\frac{\sigma_i^2}{\sigma_0^2}\right)\right) = -\,Cut$$

$$\Longrightarrow \Delta F^{max} = -\,minCut \tag{11}$$

As such, examining $\Delta F^{max} > C_\alpha$ is equivalent to assessing $minCut < -C_\alpha$. As we will see, the null distribution minCut is more straightforward to compute. The intuition is that, modular networks are more likely - by design - to have bottlenecks[14] in their topology, than randomly connected networks do. Since minCut solutions typically cut through these bottleneck edges, they often result in *smaller* Cut values for modular networks than for randomly connected

[12] More recently, generalized versions of spectral clustering for directed graphs have been proposed (Sevi et al., 2022).

[13] One could use, for instance, the spectral clustering MATLAB routine implemented in the Compressive Spectral Clustering toolbox (Tremblay et al., 2016): http://cscbox.gforge.inria.fr/

[14] Bottlenecks are the few links that connect different modules in the topology of a modular graph (Tavara & Schliep, 2021).

networks (with comparable degree and weight distributions). This property can be used to infer modularity - or lack thereof - in the topology of a given network, and to decide when to stop partitioning, in hierarchical schemes.

The relevant question in the modularity field is called "To cut or not to cut?" (Chang et al., 2014; Reichardt & Bornholdt, 2006), which seeks stopping criteria for graph division by evaluating the statistical significance of the modules. Methods that assess the significance of modularity draw either from random matrix theory (Alexander-Bloch et al., 2010; He et al., 2009; Meunier, 2009; Mirshahvalad et al., 2013), or network perturbation (Hu et al., 2010; Karrer et al., 2008; Lancichinetti et al., 2010; Mirshahvalad et al., 2012; Seifi et al., 2013), or analytical methods (Chang et al., 2014; Y. Li & Qi, 2020; Newman, 2006; Reichardt & Bornholdt, 2007). Although the test statistic that we are using (minCut) is not exactly the same as modularity (Newman, 2006), their goals and methods are closely related. In the following, we will elaborate on the randomization method (used to build the null distribution of minCut) and comment on the potential analytical approach in the Discussion.

In short, we shall create a large number of random graphs[15] (connected and matching the transformed graph in degree and weight distributions), bi-partition each one using spectral clustering and record the Cut value, in order to build an empirical null distribution for the minCut. Based on this null distribution, we can calculate the probability of getting minCut values as small as or smaller than our actual minCut. This probability is our p-value, and we will use the conventional $\alpha = 0.05$ significance threshold to reject the null hypothesis (of a nonmodular/indivisible graph). Based on the duality in Eq. (11), the same p-value can adjudicate the significance of ΔF^{max}.

We have now covered the main theory that we need. In the following section, we shall apply this procedure first to simulated examples and then to an empirical causal graph of the brain.

Materials and Methods

3.1. Simulated networks

[15] Using randmio_dir_connected.m routine in Brain Connectivity Toolbox: http://www.brain-connectivity-toolbox.net/

To demonstrate the functionality of the bi-partitioning scheme, first it is applied to small simulated causal networks (Fig. 2-A) - and later to a large empirical network. The first simulated network, originally used in (K. Friston et al., 2019), consists of eight interacting nodes with classical state-space dynamics of the form:

$$\frac{dx}{dt} = f(x,u) \approx Ax + Cu$$

$$y = g(x) + \epsilon$$

$$A = \frac{\partial f}{\partial x}, C = \frac{\partial f}{\partial u} \tag{12}$$

where the time dependencies of x, y, u, ϵ are implicit, for brevity. The state vector x holds some (hidden) quantity associated with each node/region. For example, if the nodes represent brain regions, x_i encodes the ensemble neuronal activity of region i - which affects and is affected by the activity in the regions connected to it. The general state equation $\dot{x} = f(x,u)$ was approximated by a Taylor series in the first line of Eq. (12); hence, the Jacobian/connectivity matrix, $A_{N\times N}$, encodes the interactions among the N nodes - such that A_{ij} is the influence from node j on node i. Moreover, matrix C parametrizes the effect of external perturbations applied to the network, encoded as time series in the columns of matrix $u(t)$. The connectivity parameters $\theta = (A, C)$ are rate constants in units of Hz. The second line of Eq. (12) describes the generation of measurements given the state vector x, through some observation function g and some measurement error ϵ (K. Friston et al., 2019).

To generate the data, the connectivity matrix A was designed to have excitatory feedforward connections (+0.5 Hz), inhibitory feedback connections (-1 Hz), and self-inhibition on each region for stability (-0.25 Hz). The network was driven by twenty 'events' of one second duration each that perturbed the first node with temporally jittered onsets ($C_{11} = 1$ Hz). The observation function g was set to identity for simplicity, and Gaussian observation noise (with variance = 1/400) was added to the measurements to get a signal to noise ratio of about 8 dB (K. Friston et al., 2019).

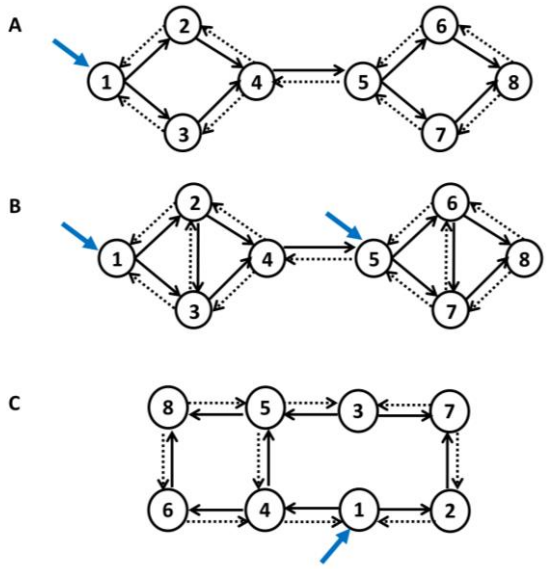

Fig. 2: Simulated causal networks. (A) An 8-node causal network, which is somewhat modular. (B) A more modular causal network. (C) A randomly connected (nonmodular) causal network. Solid black arrows represent excitatory connections, whereas dashed arrows signify inhibitory connections. Self-connections are always inhibitory, but are not shown explicitly. Blue arrows represent external perturbations. Synthetic timeseries were generated from each network. A fully-connected DCM was fitted to each network, and was pruned post-hoc (using Bayesian model reduction) to yield the optimal (i.e. most accurate and sparsest) causal structure. The estimated model structure and parameters are depicted in Fig. 3 for the first network, and in supplementary Figs. S1-S2 for the other two networks.

Having generated the synthetic data, the connectivity structure was next assumed to be unknown. To infer it, a fully-connected DCM was fitted to the data, using a variational Laplace scheme (spm_nlsi_GN.m), which estimated the posterior distribution over the parameters of this full model and a free energy approximation of its (log) evidence. The full model's posterior connectivity was then pruned using exploratory Bayesian model reduction (spm_dcm_bmr_all.m)(K. Friston & Penny, 2011; Penny et al., 2004). The results are illustrated in Fig. 3. Notably, our partitioning job starts from this point onwards; i.e. *after* having optimized the causal model's structure and parameters.

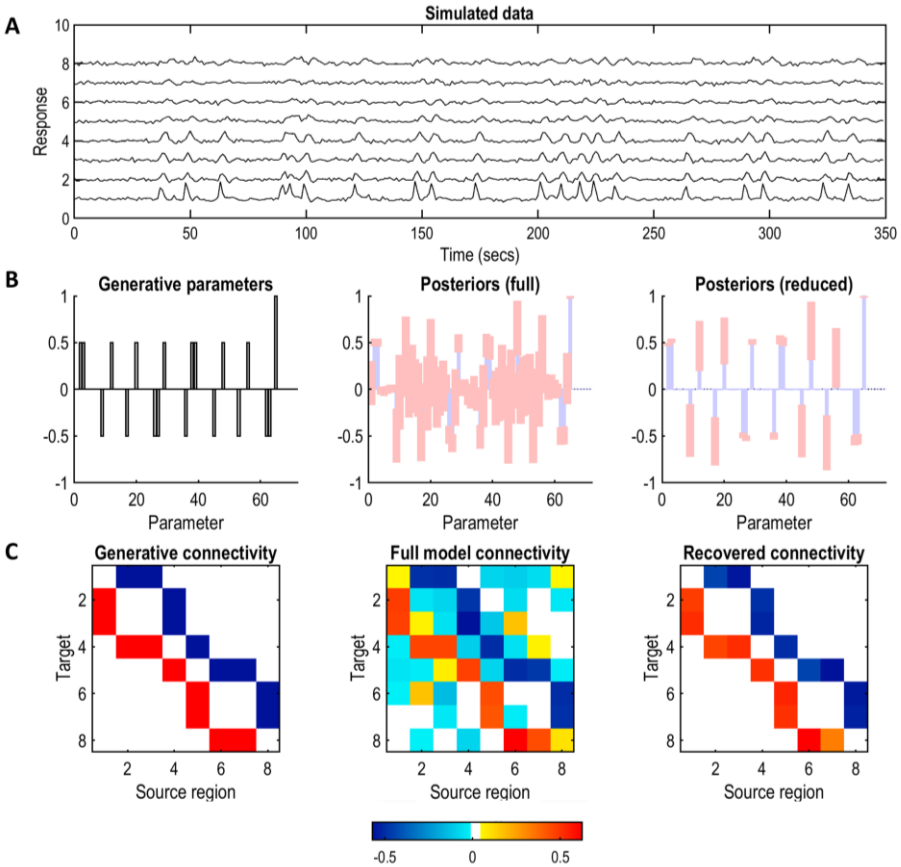

Fig. 3: Synthetic data generation and DCM model inversion. (A) Data simulated from the causal model in Fig. 2-A. Each row contains the timeseries from one of the eight nodes. (B) Left: Generative model parameters (i.e. the ground truth connections). Middle: Posterior parameter strengths, from fitting a fully-connected DCM to the synthetic timeseries. Blue bars represent expected values and pink error bars represent 95% credible intervals. Right: Posterior parameter strengths following Bayesian model reduction. Weak and uncertain parameters have been pruned automatically, to improve model evidence. (C) Left: Generative (i.e. benchmark) connectivity in matrix format. Red and blue connections represent excitatory and inhibitory effects, respectively. Middle: Connectivity matrix following the (variational) inversion of the fully-connected model. Right: The final connectivity matrix, following Bayesian model reduction. Only significant parameters (with 90% credible intervals not containing zero) have been shown. Similar results for the other two simulated networks (Fig. 2-B and Fig. 2-C) are included as supplementary Figs. S1-S2.

Since the network is small enough, we could do an exhaustive search over the space of all bi-partitions, to show that spectral clustering solves the minCut (= $max\Delta F$) problem while respecting the balance condition. To demonstrate this, each possible bisection of the graph was considered an alternative hypothesis (H_a is Eq. 3). Practically, creating random bisections is analogous to performing random cuts on the graph. Following each random cut, the corresponding Cut value was translated to a ΔF value using Eq. (11) - and called ΔF_{rnd}.

Fig. 4-B shows the distribution of ΔF_{rnd}, while ΔF^{max} (associated with the spectral Cut) is marked with a red circle.

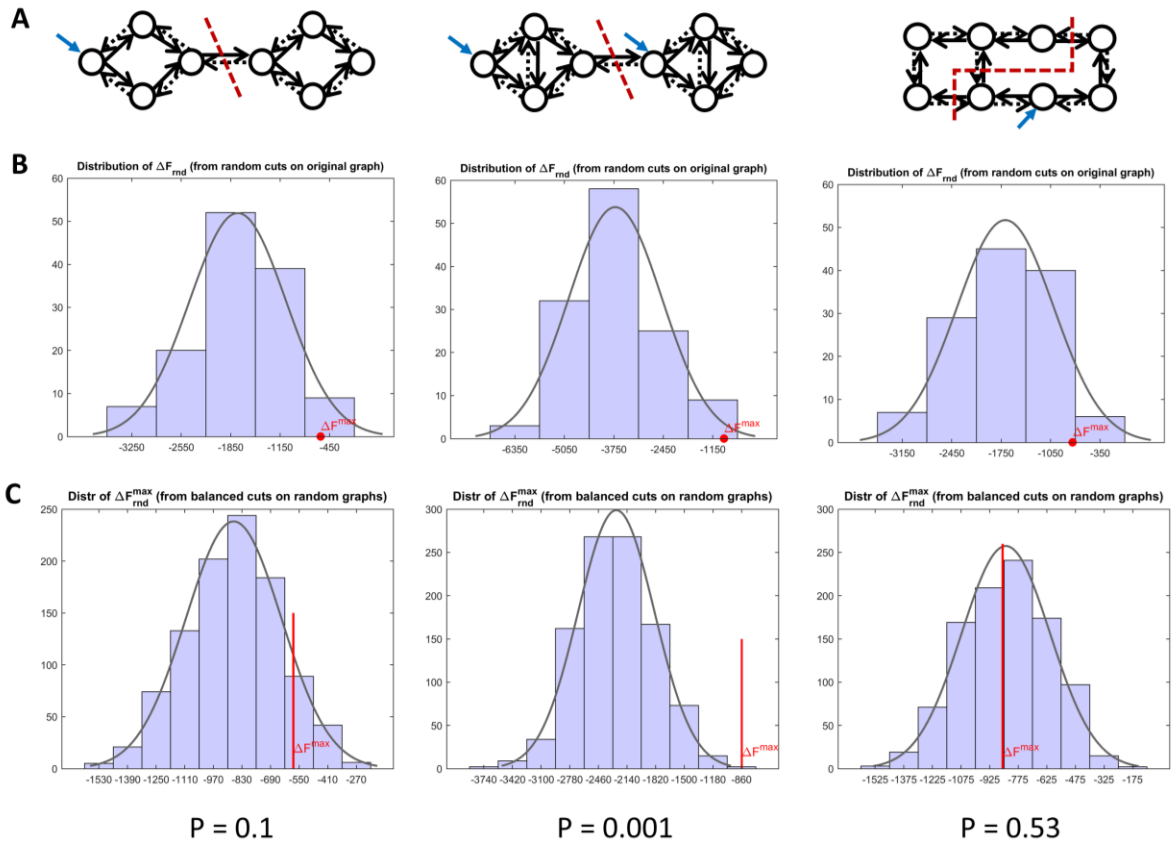

Fig. 4: Bi-partitioning results for the simulated causal networks. (A) Posterior structures of the three simulated causal networks from Fig. 2. Each network was bi-partitioned by applying spectral clustering to the transformed graph (see Theory for details). The dashed red lines represent the minCut locations. (B) Histogram of ΔF_{rnd} values, from performing random cuts on the transformed graph. Cut values were translated to ΔF values using Eq. (11). Each column corresponds to one network, shown at the top. Normal distributions have been overlaid, for comparison with theoretical predictions (Schreiber & Martin, 1999). ΔF^{max} values (following spectral clustering on the transformed graphs) are shown with red circles. Spectral clustering tries to maximize ΔF, while keeping the solution balanced (i.e., partitions of comparable size). (C) Histogram of ΔF_{rnd}^{max} values from performing balanced minCuts (using spectral clustering) on random graphs (matched in degree and weight distributions to the original transformed graphs). Cut values were translated to ΔF values using Eq. (11). Normal distributions are overlaid, for comparison with theoretical expectations (Schreiber & Martin, 1999). Red lines indicate ΔF^{max} values from balanced minCuts on the original transformed graphs. P-values, at the bottom, reflect the area lying on the right side of ΔF^{max} on each histogram.

Finally, to build the null distribution of ΔF^{max}, spectral clustering was performed on hundreds of random graphs[16] (that matched the transformed graph in degree and weight distributions, and remained connected). The resulting Cut values were mapped to ΔF values using Eq. (11) - and denoted by ΔF_{rnd}^{max}. The distribution of ΔF_{rnd}^{max} is demonstrated in Fig. 4-C. This distribution was used to assess the significance of the (spectral) bi-partition solution for the original graph, based on Eq. (6).

The same procedure was applied to a slightly more modular causal network (in Fig. 2-B) and to a randomly connected (nonmodular) network (in Fig. 2-C), to demonstrate that the hypothesis testing framework can discern between these cases and accordingly recommend when (not) to partition. The findings are reviewed in the Results section.

3.2. Empirical network

At this point, the multi-scale parcellation procedure was applied to a large DCM fitted to empirical fMRI timeseries from the infamous attention dataset[17] used in (K. J. Friston et al., 2021). To summarize, a linearized DCM (Appendix D) was fitted to 1024 regional timeseries, where each timeseries represented the first eigenvariate from the voxels in the 4-mm vicinity of the local voxel with the highest variance. The Jacobian (i.e. effective connectivity) between regions was then subjected to a Bayesian model reduction procedure, which pruned the redundant connections based on the improvement in model evidence following assumptions of reciprocal local connectivity, and yielded a sparse directed coupling matrix (Fig. 5). This sparse effective connectivity matrix was recursively bi-partitioned by applying spectral clustering to the transformed graph based on Eq. 10, and the partitions at each scale were assessed for significance based on the null distribution of minCut constructed from

[16] Using randmio_dir_connected.m routine in the Brain Connectivity Toolbox.

[17] FMRI timeseries were acquired from a normal subject, at 2 Tesla, using a gradient echo-planar sequence (TE = 40ms; TR = 3.22 seconds; matrix size = 64x64x32, voxel size 3x3x3mm). Four consecutive hundred-scan sessions were acquired, comprising a sequence of 10-scan blocks under five conditions: *Dummy, Fixation, Attention, No attention,* and *Stationary*. In the *Attention* condition, the subject was asked to detect changes in the radial velocity of moving dots. In *No attention,* the subject was asked to simply look at the moving dots. In the *Stationary* condition, the subject viewed stationary dots. The order of the conditions alternated between *Fixation* and photic stimulation. The subject fixated at the center of the screen in all conditions. No overt response was required in any condition and there were no actual speed changes in the moving dots. For more details see (Büchel & Friston, 1997; K. J. Friston et al., 2021).

(degree and weight) matched randomized graphs (Eqs. 6 & 11). The results (up to the 5th scale) have been demonstrated in Figs. 6-8. We will elaborate on the findings in the Results section.

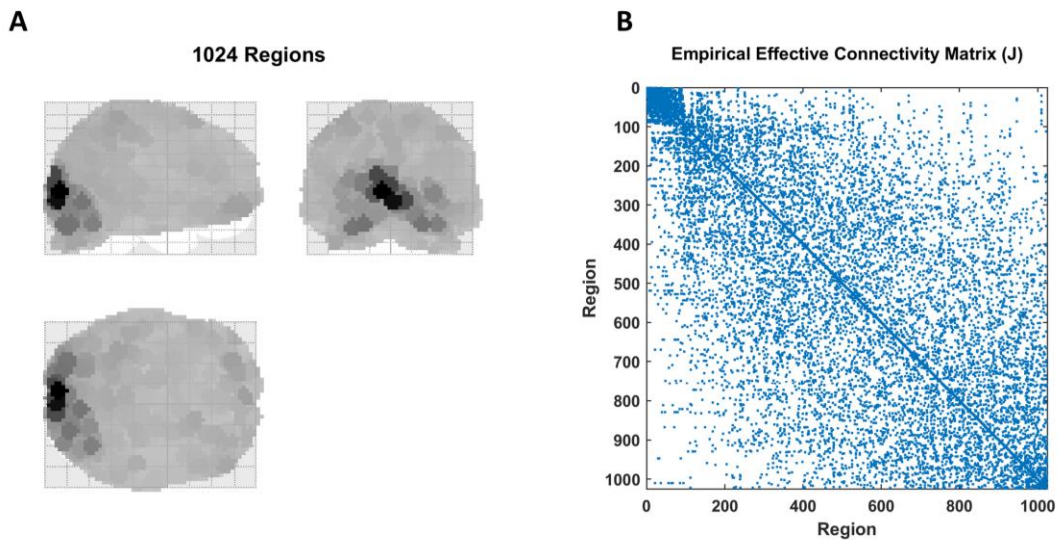

Fig. 5: Empirical causal network, inferred from the attention dataset (1st scale). (A) A maximum intensity projection of the 1024 regions that nearly cover the entire brain, where the spatial support of each region has been color coded according to the variance explained by its eigenmode. (B) Sparsity pattern of the empirical effective connectivity matrix, following DCM inversion and Bayesian model reduction. The matrix is negative definite and symmetric in structure (i.e. bidirectional connections), but nonsymmetric in coupling values. The connectivities lie in the [-1.82, +1.64] Hz range, but are shown as binary values here to improve the image contrast. This constitutes the model at the first scale of analysis, which was then recursively bi-partitioned using the procedure outlined in the Theory section. The resultant partitions at four consecutive scales have been plotted in the next few figures.

Results

4.1. Simulated networks

Figure 3 shows the model inversion results for the first causal network (in Fig. 2-A). Herein, we see the parameters of the generative model (used to simulate the data), the simulated timeseries, the parameter posteriors following DCM inversion, and the posteriors after pruning the network (using BMR). The whole procedure has faithfully recovered the original causal structure. Similar plots have been prepared for the other two networks, and provided in supplementary Figs. S1-S2.

Figure 4 contains the parcellation results for the same networks. The red dashed lines (in panel A) denote the graph cut location following spectral clustering (on the transformed graphs). The plots in Fig. 4-B show, for each causal model, the distribution of the reduced model evidence from random cuts performed on the same graph (ΔF_{rnd}) - computed using the free energy-graph cut duality in Eq. (11). The value of ΔF^{max} (= -minCut) from spectral clustering is shown as a red circle on each plot. Note that only a small number of random cuts can result in lower minCut values (equivalently higher ΔF values) than that of spectral clustering. However, these random cuts create *unbalanced* partitions (e.g. separating one node from the rest) whereas the spectral clustering solution respects the balanced partition requirement - as evinced visually by the cut locations in Fig. 4-A. Also note the Gaussian-like distribution of ΔF_{rnd} values, which is because random cut values tend to be normally distributed (in the limit, for large networks) as rigorously studied in (Schreiber & Martin, 1999).

Finally, Fig. 4-C shows the distribution of ΔF_{rnd}^{max}, pertaining to each causal model, which was computed by performing balanced minCuts (using spectral clustering) on random networks that were matched in degree and weight distributions to the transformed graph of each causal model. Again, ΔF_{rnd}^{max} values are approximately normally distributed because minCut distributions for random graphs are expected to become Gaussian, as the size of the network increases (Schreiber & Martin, 1999). The p-values reported at the bottom of each plot reflect the significance of each ΔF^{max}, based on the respective null distribution. It is apparent that the p-values reflect the level of modularity of each model. Namely, the second model is the most modular (with p = 0.001), followed by the first model which is borderline modular (p = 0.1). The third model (which was built by randomizing the connections in the first model) is clearly nonmodular, and the p-value of 0.53 confirms the insignificance of the corresponding ΔF^{max}. Accordingly, the algorithm is recommending not to partition this model.

4.2. Empirical network

The empirical effective connectivity matrix of Fig. 5 was recursively bi-partitioned, by applying spectral clustering to the transformed graphs, at multiple scales. Figs 6-8

demonstrate the resultant partitions, at four consecutive scales, projected on glass brains for anatomical visualization. Fig 6-A shows the first bi-partition (on 2nd scale), which breaks the causal model into posterior and anterior parts of the brain. The posterior partition itself (in the blue box) was further bi-partitioned, at the next (3rd) scale, to superior and inferior portions (Fig. 6-B). Likewise, the anterior partition (in orange box) was further broken into superior and inferior partitions (Fig. 6-C). At the following (4th) scale, in Fig. 7, each partition was further divided into either anterior-posterior or lateral-medial portions. The same trend continued at the subsequent (5th) scale, in Fig. 8. The bi-partitions were only accepted if ΔF^{max} was deemed significant, based on the null distribution built from graph randomization. As such, the causal model was partitioned top-down into 10 scales, before the evidence no longer supported the modularity of the remaining partitions.

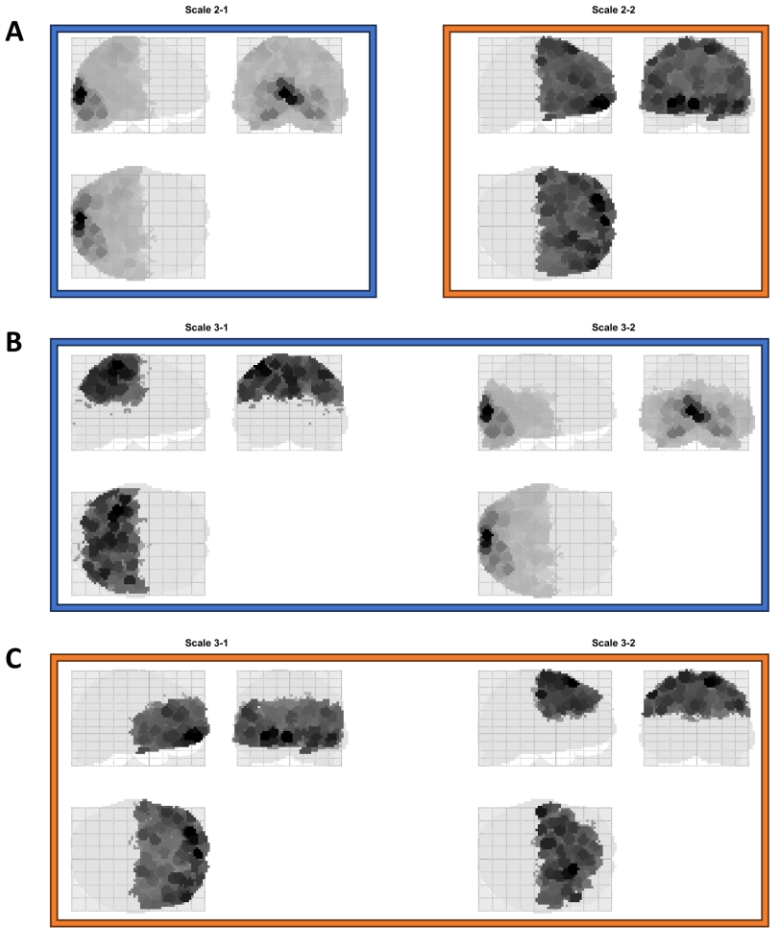

Fig. 6: Recursive bi-partitioning of the empirical causal network (2nd and 3rd scales). (A) On the 2nd scale, the brain has been partitioned into posterior and anterior parts. (B) On the 3rd scale, the posterior partition from

the 2nd scale (in blue box) has been further decomposed into superior and inferior portions. (C) Still on the 3rd scale, the anterior partition from the 2nd scale (in orange box) has also been divided into superior and inferior partitions. In each case, the glass brain shows maximum intensity projections, on the sagittal, coronal and axial planes. Each box contains the two partitions resulting from one round of bi-partitioning. Each of these partitions was further bi-partitioned, as shown in the next figure.

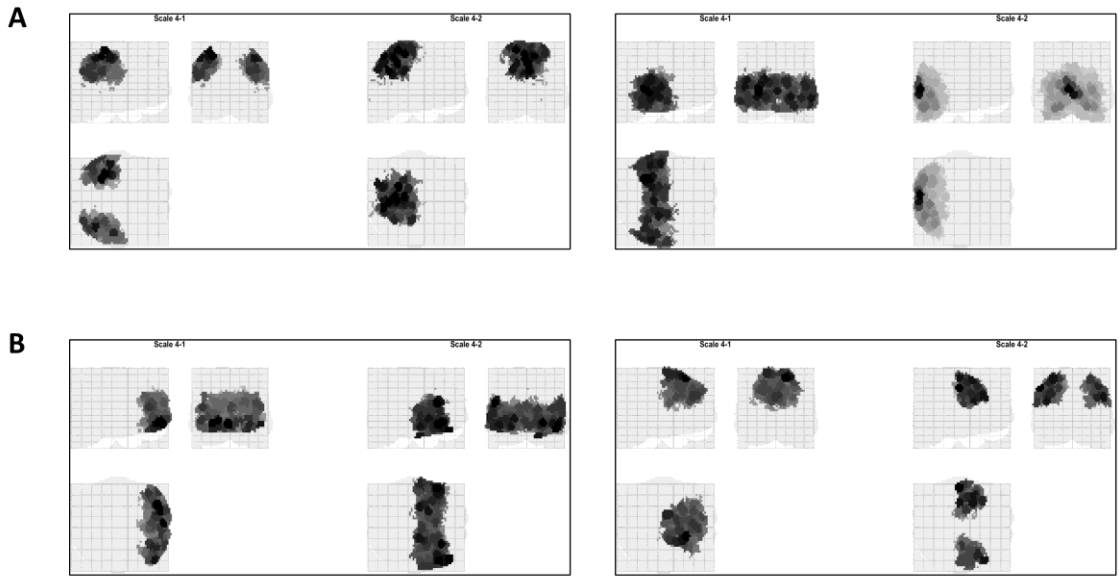

Fig. 7: Recursive bi-partitioning of the empirical causal network (4th scale). (A) On the 4th scale, the superior posterior partition from the 3rd scale has been divided into lateral and medial parts, whereas the inferior posterior partition has been divided into anterior and posterior portions. (B) Still on the 4th scale, the inferior anterior partition of the 3rd scale was further broken into anterior and posterior parts, whereas the superior anterior partition was divided into medial and lateral parts. Bi-partitioning continues in the next figure.

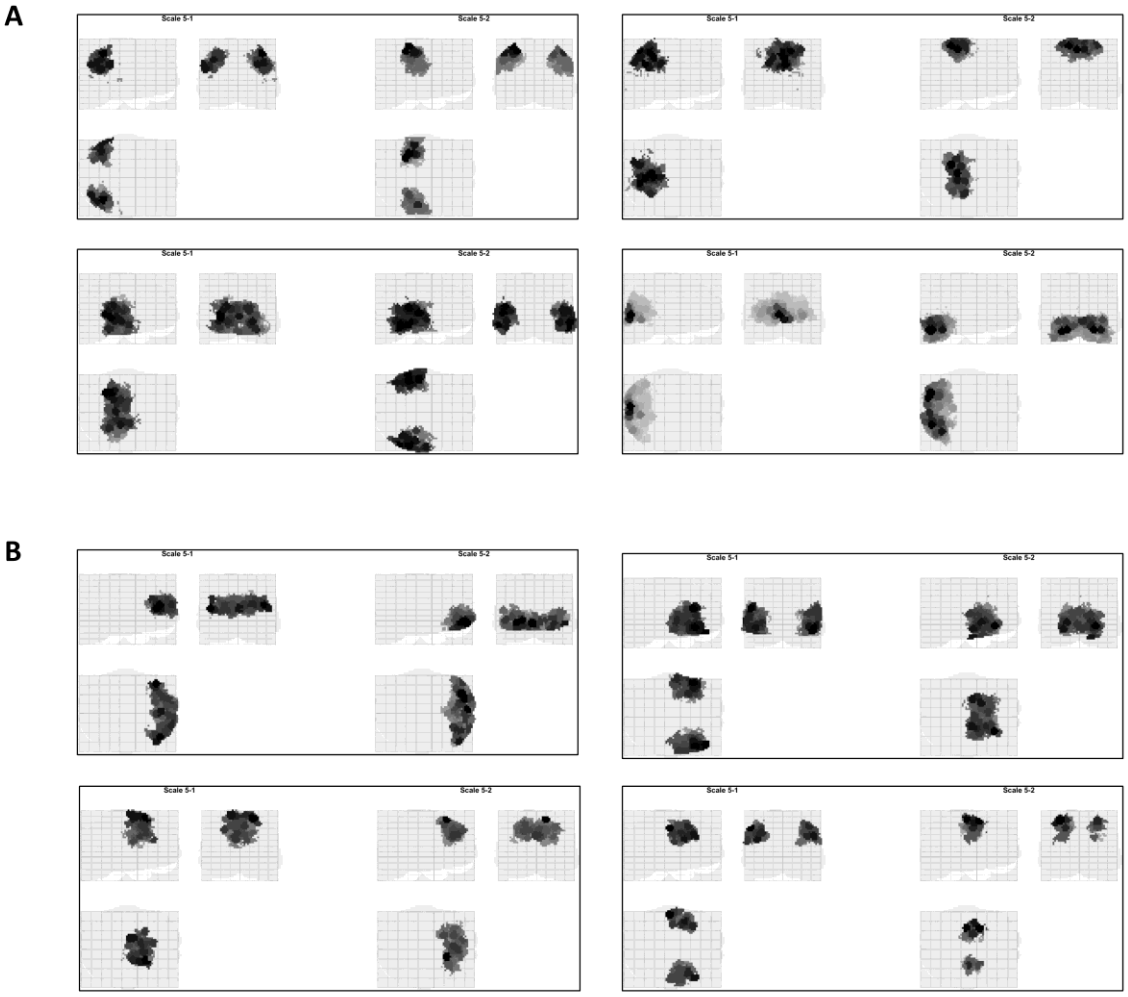

Fig. 8: Recursive bi-partitioning of the empirical causal network (5th scale). (A) Further bi-partitioning of the posterior brain. (B) Further bi-partitioning of the anterior brain. Note how in each case the divisions are either anterior-posterior, lateral-medial or superior-inferior. The number of partitions grows exponentially until finer scales, where some partitions turn out to be indivisible, as judged by the p-value of partition significance.

Scale invariance

Figure 9 shows how the average cardinality, time constant and kinetic energy of the partitions vary across 10 scales. The latter two are defined based on the eigenvalues of the Jacobian (i.e. effective connectivity matrix) in Appendix E. The linearity of the log-log plots is a signature of power-law relationships and scale invariance. The same dataset, when parcellated in a bottom-up fashion in (K. J. Friston et al., 2021) also manifested the scale invariance property. The present top-down decomposition corroborates those findings.

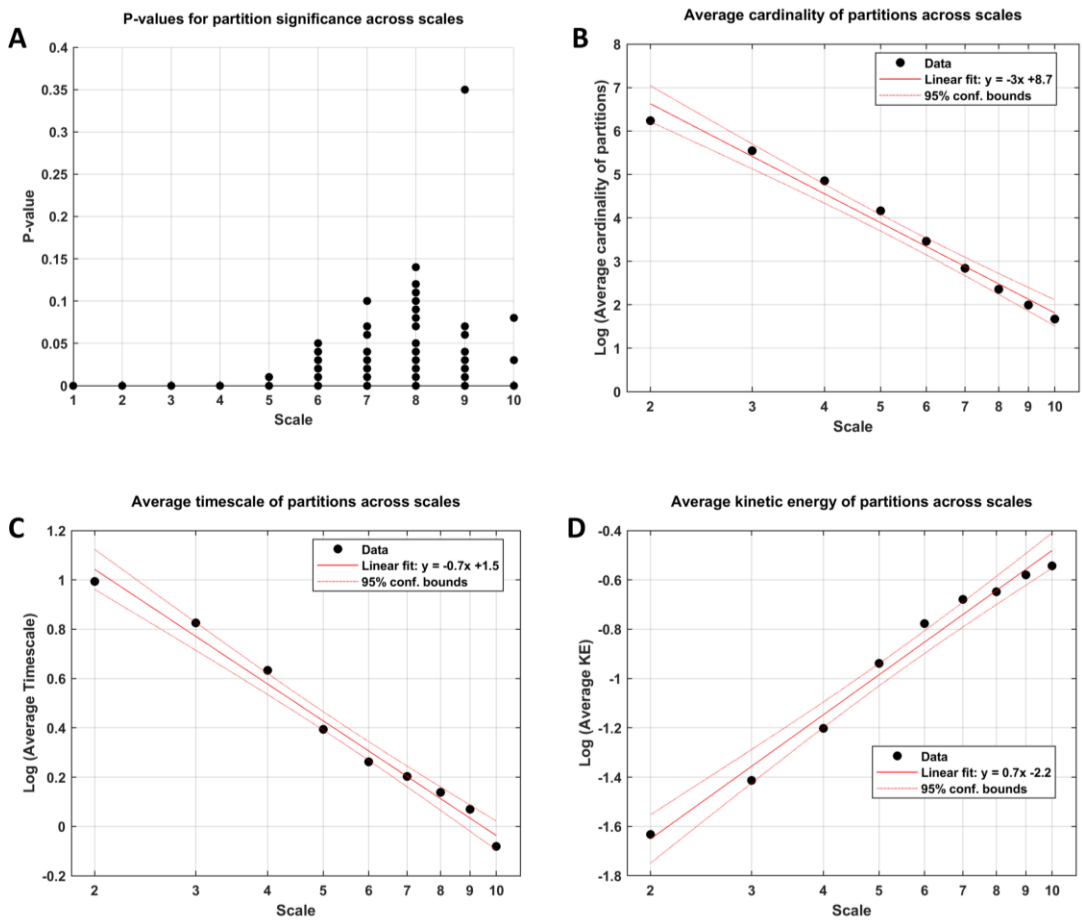

Fig. 9: Scaling behavior across 10 scales, following multi-scale parcellation. (A) P-values, from partition significance testing, at different scales. Many values overlap, so few distinct values are discernable. At finer (i.e. higher) scales, the probability of the partitions being indivisible increases, as reflected in higher p-values. (B) Average cardinality of partitions, across scales. On the log-log plot, the average cardinality drops linearly. (C) Average timescale of the partitions, across scales, decreases linearly on a log-log plot. (D) Average kinetic energy of the partitions, vs. scales, increases linearly on the log-log plot. Overall, the plots hint at scale invariant behavior.

Conditional independence and Markov blanket

Another question we can ask of the parcellation results is the conditional independence of the partitions. This question is of interest because recent (simulation) work has suggested that recurrent connections in cyclic causal models can induce statistical dependencies (i.e. functional connectivity) that cross the structural boundaries of the causal model, especially at nonequilibrium steady state (Aguilera et al., 2022). Consequently, the Markov blanket (MB) identified based on the causal (i.e. effective) connectivity may not be the same as the MB endorsed by functional couplings. Practically this means that, if certain functional couplings "cross" the borders of effective couplings, then the boundary states of the effective connectome are not *the only* MB states (see Fig. 10 for the intuition). We can ask this question of any given dataset and causal model once we cast this problem in the hypothesis testing framework:

Having identified a set of Markov blanket (i.e. boundary) states based on the *effective* couplings in a recurrent causal network, what are the odds that certain *functional* couplings cross the boundary delineated by the alleged MB states and establish themselves as additional MB states?

To answer this question, we have to compute and examine the functional network – which can be derived from the effective couplings. The details of computing the inverse covariance (i.e. precision) matrix of the neuronal states $\Pi(x) = \Sigma_{\mathrm{x}}^{-1}$ from the effective connectome $J(x)$, under Gaussian assumptions, are included in Appendix F. Having computed the precision matrix (with elements p_{ij}), the partial correlation between any two states is simply computed as:

$$\rho_{X_i X_j|Z} = -\frac{p_{ij}}{\sqrt{p_{ii}p_{jj}}} \tag{13}$$

where $\rho_{X_i X_j|Z}$ quantifies the correlation between two variables X_i, X_j, while excluding the effect of a third (set of) variable(s), Z. Under jointly normal assumption on the distribution of the variables, partial correlation coincides with conditional correlation and can therefore be used to assess conditional (in)dependence between the variables (Baba et al., 2004; Baba & Sibuya, 2005; Lawrance, 1976).

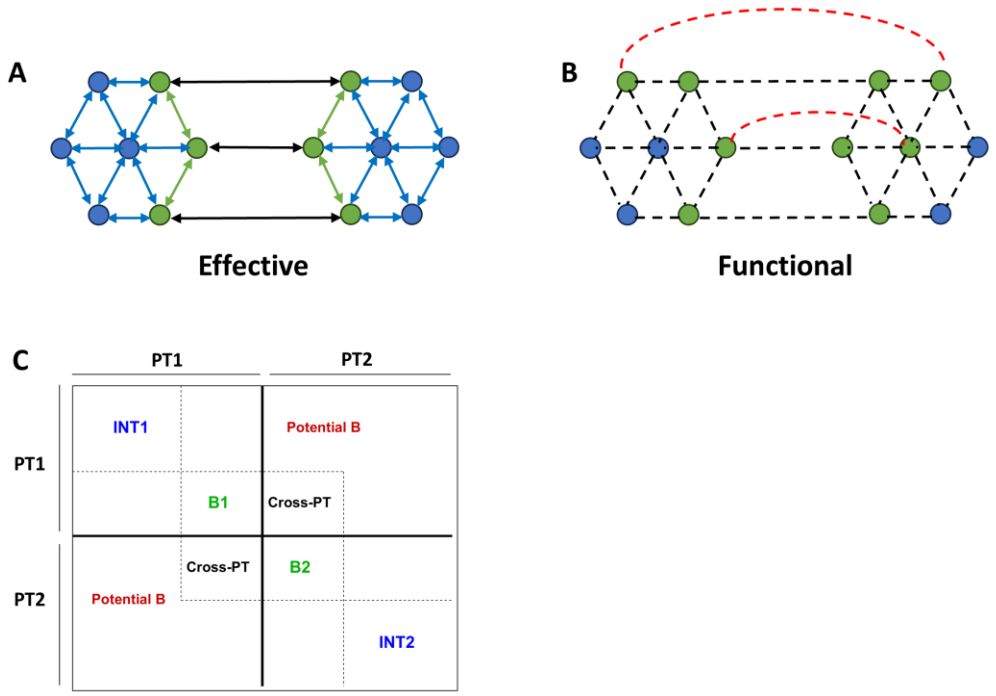

Fig. 10: Identifying Markov blanket (MB) states based on effective and functional couplings. (A) Bi-directional effective couplings within and across two partitions are shown with double-headed arrows. Blue circles denote internal (INT) states, whereas green circles represent boundary/blanket states (denoted by B). (B) Undirected functional couplings are represented by dashed lines. If functional couplings (in red) show up between INT1-INT2 states, or INT2-B1 (INT1-B2) states, then the former INT states (based on effective connectivity) turn into B states, functionally. (C) The structure of the effective coupling matrix, after sorting the states to reveal the inter- and intra- partition coupling structure. PT1 and PT2 stand for Partition 1 and Partition2, respectively. The 'Cross-PT' area contains cross-partition connections, which enable the communication between two partitions via the boundary/blanket states. The 'Potential B' area contains zeros on the effective connectivity matrix. However, some of the connections in this area might be non-zero according to the functional coupling matrix, in which case those connections should be moved from the INT set to the B (i.e. boundary/blanket) set. This format is used to visualize the findings for the empirical causal model, in the next figure.

To test whether a population partial correlation is zero, frequentist tests are available (Weatherburn, 1949, pp. 242-263). However, frequentist hypothesis testing is unable to quantify the evidence in favor of the null hypothesis. In other words, "the frequentist test does not discriminate between 'evidence of absence' and 'absence of evidence'" (Keysers et al., 2020; Kucharský et al., 2023). This limitation can be overcome in the Bayesian framework, by using Bayes factors. Here we use a recent Bayes factor test for partial correlations, which is carried out on the partial correlation coefficient itself - as opposed to testing proxy statistics such as regression coefficients. For each functional connection, we

will quantify the evidence in favor of the null hypothesis, $H_0: \rho_{X_i X_j|Z} = 0$, against the alternative hypothesis, $H_a: \rho_{X_i X_j|Z} \in (-1,1)$. For estimation, the alternative model is assumed to hold true, wherein an unknown population partial correlation is free to vary between -1 and 1. The goal is to infer this unknown parameter based on the available data. As such, the (marginal) posterior distribution of the partial correlation and an analytic Bayes factor have been derived in (Kucharský et al., 2023), for appropriate priors. Specifically, the default prior on the partial correlation is a symmetric beta distribution stretched to the interval -1 to 1. The stretched beta prior has a single hyperparameter α that governs the concentration of the distribution around zero[18]. Exemplar stretched beta priors, for different values of α, have been plotted in Fig. 12-A. The information consistency[19] desideratum for Bayes factors requires that we use $\alpha \leq \frac{1}{2}$ (Jeffreys, 1961; Kucharský et al., 2023). Note that the critical (functional) connections to be tested are situated in L-shaped regions in the lower/upper corners of J and $\rho_{X_i X_j|Z}$ (see the schematic in Fig. 10). These are the connections whose presence would reveal additional boundary/blanket states, which is why the name 'potential blanket' was chosen for the states they connect.

Figure 11 contains the results of applying this procedure to an exemplar empirical effective connectivity matrix (J) from the present dataset. In short, J was bi-partitioned using spectral clustering (on the transformed graph). Then, Σ_x^{-1} was estimated from J (Appendix F) and converted to $\rho_{X_i X_j|Z}$ (Eq. 13). At this stage, a stretched beta prior (with default $\alpha = 0.5$) was specified for ρ. Next, Bayes factor for the alternative model versus the null model (denoted by $BF10$) was computed using the formulation of (Kucharský et al., 2023) for each functional connection, and these values were converted to $\log BF10$ for the ease of interpretation. Specifically, $\log BF10 \leq -3$ is interpreted as 'strong' evidence for the null hypothesis (i.e. absence of functional coupling), whereas $\log BF10 \leq -1$ provides 'positive' evidence for the

[18] The stretched beta prior has a single hyperparameter α that governs the concentration of the distribution around zero: if $\alpha = 1$, the prior distribution is uniform; when $\alpha > 1$, the distribution is peaked and centered on the value $\rho_{X_i X_j|Z} = 0$; when $0 < \alpha < 1$ the prior is concentrated towards the extremes (-1, +1), with a lower probability of values near the center (Kucharský et al., 2023).

[19] A Bayes factor is information consistent if it diverges to infinity, thus falsifies the null, when the Bayes factor is presented with overwhelmingly informative data (Jeffreys, 1961; Kucharský et al., 2023).

null model, against the alternative. Conversely, $\log BF10 \geq 3$ and $\log BF10 \geq 1$ should be interpreted respectively as 'strong' and 'positive' evidence in favor of the alternative hypothesis (i.e. presence of functional coupling). The results are demonstrated for both thresholds (3 and 1) in Fig. 11.

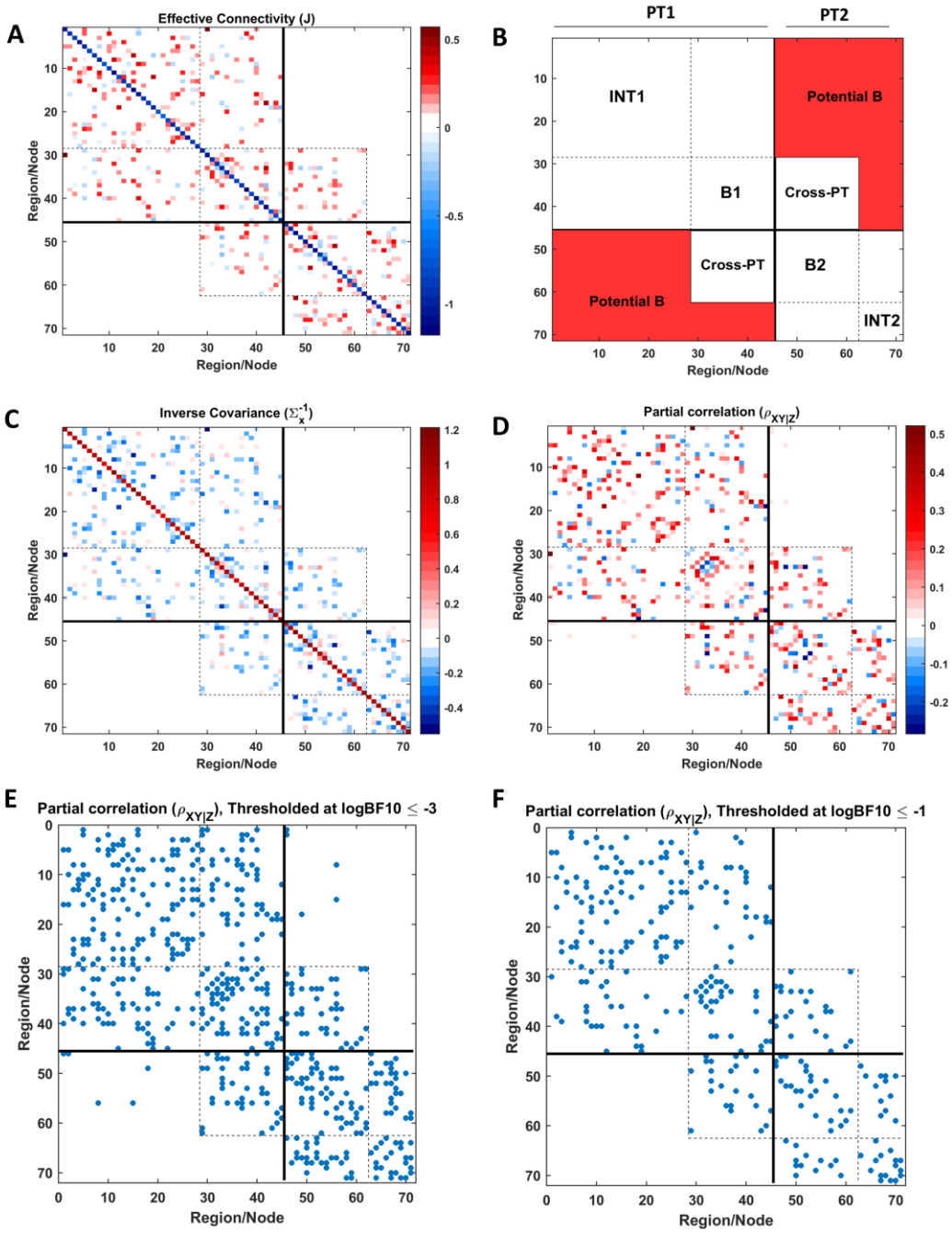

Fig. 11: Markov blanket identification procedure applied to an empirical effective connectivity matrix. (A) a 71×71 empirical effective connectivity matrix, which was bi-partitioned (by applying spectral clustering to

the transformed graph; see Theory) and sorted to reveal the modular structure of the network. (B) A visual guideline for the order and identity of the states/nodes and the connections between them. Abbreviations were defined in the caption of Fig. 10. (C) The inverse covariance matrix of the neuronal states, computed based on the relations in Appendix F. (D) Partial correlations computed using Eq. (13). Diagonal values (= -1) were removed to improve the image contrast. (E) The sparsity pattern of the partial correlation matrix, once thresholded at $\log BF10 \leq -3$, following significance testing on partial correlation values (with $\alpha = 0.5$ as hyperparameter of the stretched beta prior). Consequently, only connections for which there was 'strong' evidence in favor of the null hypothesis (of zero partial correlation) were removed. For the few remaining connections in the 'potentially boundary' region, we have investigated the evidence for their existence (i.e. non-zero partial correlation) in the next panel. (F) The sparsity pattern of the partial correlation matrix, once thresholded at $\log BF10 \leq -1$. That is, all the connections with 'positive' null evidence for partial correlation were removed. It turns out that, even though there was not 'strong' evidence for the nullity of the few 'potentially boundary' connections in panel E, there is 'positive' evidence for their absence. Hence, these connections are not eligible to be added to the boundary/blanket set. As such, the MB states identified based on the effective connectivity matrix remains unchanged and valid. This speaks to the conditional independence of the partitions identified with the proposed method, based on effective connectivity, once conditioned on the boundary states. In the next figure, we inspect the stability of these findings for different prior settings.

To show the sensitivity of the findings to the stretched beta prior specification, we also calculated $\log BF10$ for a range of $\alpha's$ less than 0.5 (to maintain information consistency). The values of $\log BF10$ versus $\rho_{X_i X_j|Z}$ have been plotted, in Fig. 12-B, for the 'potentially blanket' connections, defined earlier. The results show that: as we decrease α, $\log BF10$ decreases (i.e., evidence for the *absence* of potentially blanket connections increases), which reinforces the boundary states (identified from the effective coupling) as the only MB states of this system. These computations were repeated over a range of priors for multiple other empirical effective connectivity matrices from the current dataset, sampled from different scales of the recursive parcellation results (plots not shown for brevity). The overall conclusion was that: with the current (linear, sparse, reciprocally and locally coupled) dynamic causal model, and for the present high-dimensional neuroimaging dataset, the evidence mostly supports the *absence* of potentially blanket states, which suggests that the boundary states (based on effective connectivity) are the only MB states of the inferred causal networks. This implies that our recursive parcellation scheme is identifying conditionally independent parcels from the empirical neuroimaging dataset. The extent to which these results hold for other datasets and other DCMs remains to be explored. Although a complete characterization of this phenomenon is outside the scope of the current work, herein we have laid the necessary framework for more rigorous studies.

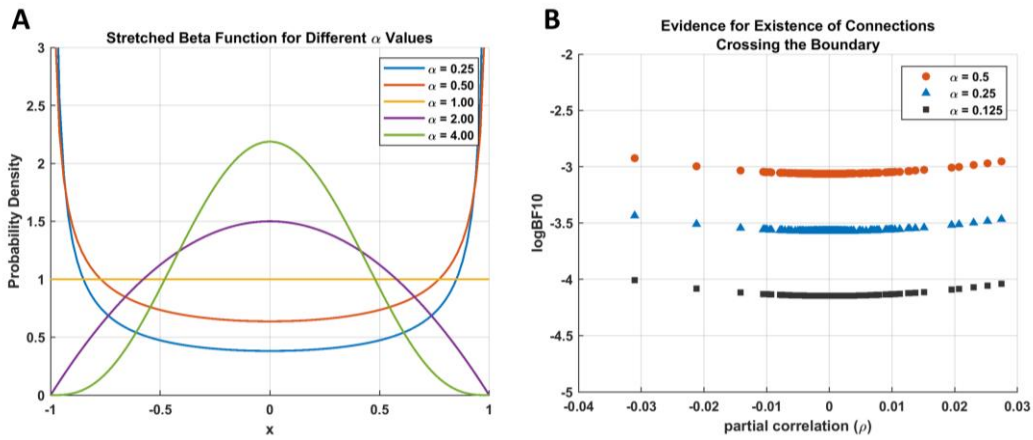

Fig. 12: Inferring the existence of 'potentially blanket' states for different prior specifications. (A) The stretched beta prior plotted for different values of the hyperparameter α. (B) $\log BF10$ versus partial correlation values, computed for the 'potentially blanket' states in Fig. 11, for several hyperparameter settings. The selected hyperparameter values ($\alpha \leq 0.5$) comply with the information consistency requirement for Bayes factor. Consequently, as α decreases, the $\log BF10$ becomes more negative (i.e., evidence for the absence of boundary-crossing connections increases).

Discussion

A top-down recursive parcellation scheme was proposed for (bi)partitioning dynamic causal models. This approach is a generalization of the ΔBIC method, framed as Bayesian model comparison. However, unlike the ΔBIC method, the proposed approach does not require fitting local models to subsets of the data, in order to compute ΔF for model comparison. Instead, model evidence for the partitioned model is computed analytically from the full (unpartitioned) model using Bayesian model reduction - because the former model is nested inside the latter. Furthermore, a naïve version of BMR was introduced, which avoids matrix multiplications and considerably reduces the computational cost for large networks. Under the naïve formulation, we derived an analytical relationship between ΔF and the graph Cut value for a transformed version of the original graph (Eq.11: $\Delta F = -Cut$). This duality facilitated both the optimization of ΔF and the assessment of its significance, using the tools from graph theory. Specifically, we chose spectral clustering to solve the (balanced) minCut problem, and indirectly maximized ΔF. Moreover, the distribution of minCut under the null hypothesis (of randomly connected graphs) served as a proxy for the null distribution of (minus) ΔF_{rnd}^{max}, which was subsequently used to inspect the significance of ΔF^{max} - i.e., to

decide whether the causal network was partitionable, at each scale. This recursive parcellation procedure was applied to simulated causal networks and to a high-dimensional causal network inferred from a neuroimaging dataset. In the following, we review the key findings and highlight some important aspects of the proposed method.

In the simulations, this approach correctly identified the location of the minCut in each case - using spectral clustering on the transformed graph. Moreover, the p-value associated with each ΔF^{max} reflected the extent of modularity of the underlying network. A relevant point to bring up is that the parcellation outcome depends on the structure (and uncertainty) of the *estimated* network, not the ground truth - to which we do not have access in real situations. This speaks to the old debate of whether a true model exists or whether some models are more useful (i.e., have higher evidence) than the others (Litvak et al., 2019). Although in the three (small) simulated networks the posterior structures were faithful to the benchmarks, in general the accuracy of the recovered network depends on the size of the network, the amount of data, the type of DCM, and the difficulty of model inversion (i.e., the ill-posed optimization problem and the local optima). In any case, our parcellation job starts *after* optimization of the causal model; so, we always assume that the estimated model at hand is the best characterization of the system that we could come up with, and that our task is to inspect its (hierarchical) modular organization.

A second point is the use of spectral clustering as an approximate solution to the (balanced) minCut problem. In general, minCut is a combinatorial optimization problem. So, finding the minCut in a bi-partition problem requires a brute-force search over all possible bi-partitions of the network, whose number grows exponentially with the network size[20]. Spectral clustering is usually considered a plausible approximate solution. Still, it is possible to improve upon this (initial) solution using incremental changes, similar to hill climbing algorithms (S. Li et al., 2014). More recent methods can even solve minCut exactly in near linear time (Henzinger et al., 2024).

The third point pertains to the free energy-graph cut duality (Eq. 11). This relationship was derived for the naïve mean field approximation (MFA) scenario (in Appendix C), and proved

[20] The number of candidate bi-partitions is defined by a Stirling number of the second kind, $S(n, 2) = 2^{n-1} - 1$.

to be useful for model optimization and statistical significance assessment. The graph theoretical link is also interesting because, previous work in the realm of integrated information theory (IIT) has provided *numerical* evidence that applying spectral clustering to the correlation matrix of neural timeseries is a good approximation for finding the minimum information bi-partition (MIB) of a network, which splits the network into two maximally independent sub-communities. The latter is a necessity for computing information integration in a given network (Toker & Sommer, 2019). The authors then proceeded to compute the "geometric" version of integrated information, which is the KL-divergence between the distributions (over the temporal evolution) of the full and disconnected models (Oizumi et al., 2016; Toker & Sommer, 2019). In comparison, we also used spectral clustering for bi-partitioning the network, but since the present approach is model-based, the (transformed) effective connectivity matrix was used instead of the correlation matrix. We further provided an *analytical* relationship between the Cut value and the change in model evidence (ΔF) from disconnecting the full model. Future work can compare the MIBs identified using the causal model (i.e. effective connectivity) against those found based on the correlation matrix (i.e. functional connectivity). A potential advantage of the former is that a model-based framework offers a standard measure (ΔF = log Bayes factor) to score and compare the plausibility of different bi-partitions against each other.

Note that without the naïve MFA assumption, the model comparison score to maximize has the general form of:

$$\Delta F = cte + \frac{1}{2}\left(\mu_R^T P_R \mu_R + \log \frac{|C_R|}{|\Sigma_R|} \right) \tag{14}$$

where the terms associated with the full model have been lumped into a constant; μ_R is the mean vector and P_R is the precision matrix of the retained/within-partition connections. So, intuitively, we have to partition the model to maximize $\sum_{i,j\in\setminus\{C\}} \mu_i \mu_j P_{ij}$ for within-partition connections, and penalize overly precise combinations through the log ratio term. This is still implicitly akin to what modularity maximization and graph partitioning algorithms try to achieve (the two are closely related, (Y. Wang et al., 2015)). So, it might still be useful to rely on graph partitioning techniques as approximate/initial bi-partition solutions, and then improve upon them iteratively – to maximize ΔF.

The fourth point is about testing the statistical significance of the identified partitions. We built the null distribution of minCut by performing spectral clustering on a large number of randomized graphs, which were connected and matched to the transformed graph in degree and weight distributions. Another useful avenue is thinking about *parametric* tests to assess the significance of the partitions - especially when the network is large and graph randomization is computationally costly. To move towards parametric testing, we need to revisit what we know about the distribution of minCut size in random graphs. For instance, at large N, the minCut distribution (for sparse random graphs) tends towards a Gaussian[21], the mean and variance of which scale linearly with N (Schreiber & Martin, 1999). Notably, as the number of vertices of the random graph $N \to \infty$, the relative fluctuations of the minCut about its mean tend towards zero, a property known as "self-averaging"[22] (Schreiber & Martin, 1999). We have shown these properties, numerically, on our randomized graphs. In supplementary Fig. S3, the average minCut values have been plotted against the number of nodes, N, for random weighted graphs that matched the transformed version of our empirical graphs in degree and weight distributions, at different scales. Note how the mean and variance of the minCut scale linearly with N, as theoretically and numerically demonstrated by (Schreiber & Martin, 1999). For the same (sparse and weighted) random graphs, Fig. S3-C demonstrates the self-averaging property and the Gaussian-like nature of the minCut distribution. While these numerical results provide useful hints, exact parametric testing requires further theoretical work to generalize the derivations in (Schreiber & Martin, 1999) to sparse *weighted* random graphs, potentially for different weight distributions. We leave this as useful future work.

Fifth, we come to the empirical findings. Recursive bi-partitioning of the empirical causal network revealed anatomically well-defined parcels, at each of the 10 scales. The spatial maps showed divisions that were either anterior-posterior, superior-inferior or lateral-

[21] Even for small random networks, minCut distribution shows Gaussian-like peakedness; see Fig. 4-C.

[22] Although the "self-averaging" property of the minCut distribution has been mainly discussed in the context of unweighted random graphs, this property is typical of processes to which many terms contribute – as discussed in (Schreiber & Martin, 1999) and shown numerically in Fig. S3. Normally, to generalize findings from unweighted to weighted graphs theoretically, it suffices to map the weighted graphs to multigraphs. A multigraph is a graph that allows multiple (parallel) edges between the same pair of vertices. More details can be found in (Newman, 2004).

medial, as depicted in Figs. 6-8. Inspecting the scaling behavior revealed that p-values generally increased for higher (i.e. smaller) scales, meaning that it was more likely for smaller partitions not to be deemed partitionable. Moreover, on the log-log plots, the average cardinality and the average timescales of the partitions decreased linearly with scale, whereas the average kinetic energy of the partitions increased linearly across scales, showing signs of the anticipated scale-invariant property. Note that similar scaling behaviors had emerged when the same dataset was partitioned in a bottom-up fashion in previous work (K. J. Friston et al., 2021). Our top-down parcellation results provides construct validity for those findings.

Sixth, putting our empirical parcels through another hypothesis testing framework showed that: the boundary states identified from the cyclic effective connectivity structure were usually the only Markov blanket states of the system – as endorsed by the nullity of partial correlations (i.e. functional connections) between the 'potentially blanket' states (Figs. 10-11). This speaks to the conditional independence of the identified partitions (based on the causal model) in the current high-dimensional neuroimaging dataset, which was modeled with a sparsely, reciprocally and locally connected linear DCM. Importantly, the general framework that we introduced here allows for further study of this phenomenon in other datasets and models. For instance, one can change the type of data, the number of regions, and the type of DCM (to e.g. non-linear); it is also possible to investigate the effect of sparsity and asymmetry of the couplings on the findings. The simulation work of (Aguilera et al., 2022) had showed that higher asymmetry in the effective couplings (which drives the system further away from equilibrium) can encourage functional couplings to cross the alleged boundaries of the effective connectome. While the authors concluded that this phenomenon may contradict the existence of a MB in the cyclic system, it might simply imply the existence of a more extended MB, than the one prescribed by the effective connectome. This is the type of question that can be entertained within the proposed hypothesis testing framework, based upon causal modeling.

Seventh, although we restricted the recursive parcellation to bi-partitions, it is straightforward to generalize to multi-partitions, in which case the model comparison would be conducted for $H_0: n = 1$ versus $H_a: n \geq 2$ partitions. Note that the derivations for ΔF using

(naïve) BMR and the $\Delta F - Cut$ duality do not depend on the *number* of cuts (which is more than one for multi-partitioning), so the same framework would still be applicable. The only difference is that we would have to conduct spectral clustering for a range of $n's \geq 2$ and compare the resulting $\Delta F(n)$'s to choose the optimal number of partitions at each scale.

Eighth, the derivations of naïve BMR are generic and can be used for purposes other than parcellation as well. We mention two such applications. For instance, if the goal is to prune an over-parametrized Bayesian network, Eq. (9) is saying that we can *sort* the causal connections based on how much free energy improvement each removal would bring about. And, to compute the free energy improvement from removing a *group* of connections, we just have to add up the (analytically computed) individual improvements. Another potential application of naïve BMR is in conducting group Bayesian analyses (using a random effects scheme known as parametric empirical Bayes, PEB)(K. J. Friston et al., 2016, 2016; Zeidman et al., 2019). Since PEB formulation relies on BMR, naïve BMR can be specifically useful for performing group analysis over large networks - by avoiding matrix multiplications. Note that the whole top-down parcellation scheme is also generic and can be used for different data types (other than neuroimaging) and different DCMs developed for other applications (K. J. Friston et al., 2022). In future work, we intend to generalize this technique for conducting multi-scale parcellation at the group level, within a random effects framework.

Finally, in clinical neuroscience, this approach might be of interest to researchers inspecting the hierarchical organization of the brain in neurological and psychiatric disorders. While most previous studies have focused on alterations in the hierarchies of *functional* and *structural* networks, the present method facilitates multi-scale parcellation of (potentially very large) *effective* networks - to inspect the hierarchical arrangement of causal networks in the brain and any pathological alterations therein. One could inspect, for instance, how the scaling behavior (such as scale invariance of different measures) differs in normal and patient populations; how the spatial maps of the parcels differ between two groups; whether certain disorders result in abnormal fragmentation of the causal networks; which scale(s) are the most affected; etc. We hope that the proposed multi-scale parcellation scheme, based on dynamic causal modeling, will be able to contribute to our understanding of the brain's integration and segregation principles, in health and disease.

Conclusions

This technical note introduced a multi-scale top-down parcellation scheme for dynamic causal models, with application to neuroimaging data. This method is a generalization of the ΔBIC method, when formulated as Bayesian model comparison. In practice, because the partitioned model is nested inside the full model, ΔF could be computed using Bayesian model reduction, which circumvented the need to fit separate sub-models to the data. Theoretically, once we transformed the original graph (into precision weighted squared connections) an intuitive relationship emerged under the naïve (statistical independence) assumption: $\Delta F = -Cut$,. As such, solving the causal model partitioning problem came down to solving the minCut problem in graph theory, which we tackled using spectral clustering. The statistical significance of the partitions was further assessed using the null distribution of minCut for (matched) randomized graphs. This method was applied to simulated and empirical causal models, for face and construct validity. Consequently, the empirical network showed scale invariant behavior in multiple measures. Moreover, we studied whether the partitions were statistically independent, conditioned upon the boundary states. To investigate this, first we identified the Markov blanket states of the cyclic causal model as the boundary states of the (recurrent) effective connectome; then, we introduced a hypothesis testing procedure to inspect the evidence for additional MB states based on the functional connectome – which were mostly nonexistent in the current dataset. Future generalizations of the parcellation scheme (to perform multi-partitioning, group analysis with random effects, parametric significance testing, etc.) were discussed, and potential applications in systems and clinical neuroscience were articulated.

Software and Data availability:

The MATLAB codes to reproduce the results in this paper can be found at: https://github.com/tszarghami/Parcellation . Linearized DCM, PEB and BMR procedures are MATLAB routines (spm_dcm_J.m, spm_dcm_peb.m, spm_dcm_bmr.m) in SPM12 software: https://www.fil.ion.ucl.ac.uk/spm/. Spectral clustering is implemented in the Compressive Spectral Clustering toolbox: http://cscbox.gforge.inria.fr/. Graph randomization is part of the Brain Connectivity toolbox: https://sites.google.com/site/bctnet/. The attention to visual motion fMRI dataset is publicly available at: https://www.fil.ion.ucl.ac.uk/spm/data/attention/.

References

Aguilera, M., Poc-López, Á., Heins, C., & Buckley, C. L. (2022). *Knitting a Markov blanket is hard when you are out-of-equilibrium: Two examples in canonical nonequilibrium models* (No. arXiv:2207.12914). arXiv. https://doi.org/10.48550/arXiv.2207.12914

Alexander-Bloch, A. F., Gogtay, N., Meunier, D., Birn, R., Clasen, L., Lalonde, F., Lenroot, R., Giedd, J., & Bullmore, E. T. (2010). Disrupted modularity and local connectivity of brain functional networks in childhood-onset schizophrenia. *Frontiers in Systems Neuroscience*, *4*, 147.

Ans, B., Hérault, J., Jutten, C., & others. (1985). Architectures neuromimétiques adaptatives: Détection de primitives. *Proceedings of Cognitiva*, *85*, 593–597.

Ao, P. (2004). Potential in stochastic differential equations: Novel construction. *Journal of Physics A: Mathematical and General*, *37*(3), L25.

Ashourvan, A., Telesford, Q. K., Verstynen, T., Vettel, J. M., & Bassett, D. S. (2019). Multi-scale detection of hierarchical community architecture in structural and functional brain networks. *PLOS ONE*, *14*(5), e0215520. https://doi.org/10.1371/journal.pone.0215520

Axer, M., & Amunts, K. (2022). Scale matters: The nested human connectome. *Science*, *378*(6619), 500–504. https://doi.org/10.1126/science.abq2599

Baba, K., Shibata, R., & Sibuya, M. (2004). Partial correlation and conditional correlation as measures of conditional independence. *Australian & New Zealand Journal of Statistics*, *46*(4), 657–664.

Baba, K., & Sibuya, M. (2005). Equivalence of partial and conditional correlation coefficients. *Journal of the Japan Statistical Society*, *35*(1), 1–19.

Bastiani, M., & Roebroeck, A. (2015). Unraveling the multiscale structural organization and connectivity of the human brain: The role of diffusion MRI. *Frontiers in Neuroanatomy*, *9*. https://doi.org/10.3389/fnana.2015.00077

Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, *7*(6), 1129–1159.

Benozzo, D., Baggio, G., Baron, G., Chiuso, A., Zampieri, S., & Bertoldo, A. (2024). Analyzing asymmetry in brain hierarchies with a linear state-space model of resting-state fMRI data. *Network Neuroscience*, *8*(3), 965–988. https://doi.org/10.1162/netn_a_00381

Betzel, R. F., & Bassett, D. S. (2017). Multi-scale brain networks. *NeuroImage*, *160*, 73–83. https://doi.org/10.1016/j.neuroimage.2016.11.006

Bhattacharya, A., Linero, A., & Oates, C. J. (2024). *Grand Challenges in Bayesian Computation* (No. arXiv:2410.00496). arXiv. https://doi.org/10.48550/arXiv.2410.00496

Bonifazi, P., Erramuzpe, A., Diez, I., Gabilondo, I., Boisgontier, M. P., Pauwels, L., Stramaglia, S., Swinnen, S. P., & Cortes, J. M. (2018). Structure–function multi-scale connectomics reveals a major role of the fronto-striato-thalamic circuit in brain aging. *Human Brain Mapping*, *39*(12), 4663–4677. https://doi.org/10.1002/hbm.24312

Büchel, C., & Friston, K. J. (1997). Modulation of connectivity in visual pathways by attention: Cortical interactions evaluated with structural equation modelling and fMRI. *Cerebral Cortex (New York, NY: 1991)*, *7*(8), 768–778.

Chang, Y.-T., Pantazis, D., & Leahy, R. M. (2014). To cut or not to cut? Assessing the modular structure of brain networks. *NeuroImage*, *91*, 99–108. https://doi.org/10.1016/j.neuroimage.2014.01.010

Chen, J., & Gupta, A. K. (1997). Testing and Locating Variance Changepoints with Application to Stock Prices. *Journal of the American Statistical Association*, *92*(438), 739–747. https://doi.org/10.1080/01621459.1997.10474026

Chung-Hsien Wu, Yu-Hsien Chiu, Chi-Jiun Shia, & Chun-Yu Lin. (2006). Automatic segmentation and identification of mixed-language speech using delta-BIC and LSA-based GMMs. *IEEE Transactions on Audio, Speech and Language Processing*, *14*(1), 266–276. https://doi.org/10.1109/TSA.2005.852992

Costa, M., Gonçalves, A. M., & Teixeira, L. (2016). Change-point detection in environmental time series based on the informational approach. *Electronic Journal of Applied Statistical Analysis*, *9*(2).

D'Acunto, G., Bonchi, F., Morales, G. D. F., & Petri, G. (2024). Extracting the Multiscale Causal Backbone of Brain Dynamics. In F. Locatello & V. Didelez (Eds.), *Proceedings of the Third Conference on Causal Learning and Reasoning* (Vol. 236, pp. 265–295). PMLR. https://proceedings.mlr.press/v236/d-textsc-char13acunto24a.html

D'Angelo, E., & Jirsa, V. (2022). The quest for multiscale brain modeling. *Trends in Neurosciences*, *45*(10), 777–790. https://doi.org/10.1016/j.tins.2022.06.007

Dauwels, J. (2007). On Variational Message Passing on Factor Graphs. *2007 IEEE International Symposium on Information Theory*, 2546–2550. https://doi.org/10.1109/ISIT.2007.4557602

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, *39*(1), 1–22.

Doucet, G., Naveau, M., Petit, L., Delcroix, N., Zago, L., Crivello, F., Jobard, G., Tzourio-Mazoyer, N., Mazoyer, B., Mellet, E., & Joliot, M. (2011). Brain activity at rest: A multiscale hierarchical functional organization. *Journal of Neurophysiology*, *105*(6), 2753–2763. https://doi.org/10.1152/jn.00895.2010

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). Simoudis, Evangelos; Han, Jiawei; Fayyad, Usama M., eds. *A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise*, 226–231.

Farahibozorg, S. R., Harrison, S. J., Bijsterbosch, J. D., Woolrich, M. W., & Smith, S. M. (2024). *Multiscale Modes of Functional Brain Connectivity*. Neuroscience. https://doi.org/10.1101/2024.05.28.596120

Ferrarini, L., Veer, I. M., Baerends, E., Van Tol, M., Renken, R. J., Van Der Wee, N. J. A., Veltman, Dirk. J., Aleman, A., Zitman, F. G., Penninx, B. W. J. H., Van Buchem, M. A., Reiber, J. H. C., Rombouts, S. A. R. B., & Milles, J. (2009). Hierarchical functional modularity in the resting-state human brain. *Human Brain Mapping*, *30*(7), 2220–2231. https://doi.org/10.1002/hbm.20663

Frässle, S., Lomakina, E. I., Kasper, L., Manjaly, Z. M., Leff, A., Pruessmann, K. P., Buhmann, J. M., & Stephan, K. E. (2018). A generative model of whole-brain effective connectivity. *NeuroImage*, *179*, 505–529. https://doi.org/10.1016/j.neuroimage.2018.05.058

Frässle, S., Lomakina, E. I., Razi, A., Friston, K. J., Buhmann, J. M., & Stephan, K. E. (2017). Regression DCM for fMRI. *NeuroImage*, *155*, 406–421. https://doi.org/10.1016/j.neuroimage.2017.02.090

Frässle, S., Manjaly, Z. M., Do, C. T., Kasper, L., Pruessmann, K. P., & Stephan, K. E. (2021). Whole-brain estimates of directed connectivity for human connectomics. *NeuroImage*, *225*, 117491. https://doi.org/10.1016/j.neuroimage.2020.117491

Friston, K. J., Fagerholm, E. D., Zarghami, T. S., Parr, T., Hipólito, I., Magrou, L., & Razi, A. (2021). Parcels and particles: Markov blankets in the brain. *Network Neuroscience*, *5*(1), 211–251. https://doi.org/10.1162/netn_a_00175

Friston, K. J., Flandin, G., & Razi, A. (2022). Dynamic causal modelling of COVID-19 and its mitigations. *Scientific Reports*, *12*(1), 12419. https://doi.org/10.1038/s41598-022-16799-8

Friston, K. J., Litvak, V., Oswal, A., Razi, A., Stephan, K. E., Van Wijk, B. C. M., Ziegler, G., & Zeidman, P. (2016). Bayesian model reduction and empirical Bayes for group (DCM) studies. *NeuroImage*, *128*, 413–431. https://doi.org/10.1016/j.neuroimage.2015.11.015

Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J., & Penny, W. (2007). Variational free energy and the Laplace approximation. *NeuroImage*, *34*(1), 220–234. https://doi.org/10.1016/j.neuroimage.2006.08.035

Friston, K., Parr, T., & Zeidman, P. (2019). *Bayesian model reduction* (No. arXiv:1805.07092). arXiv. https://doi.org/10.48550/arXiv.1805.07092

Friston, K., & Penny, W. (2011). Post hoc Bayesian model selection. *NeuroImage*, *56*(4), 2089–2099. https://doi.org/10.1016/j.neuroimage.2011.03.062

Gao, Z., Xiao, X., Fang, Y.-P., Rao, J., & Mo, H. (2024). A Selective Review on Information Criteria in Multiple Change Point Detection. *Entropy*, *26*(1), 50. https://doi.org/10.3390/e26010050

Geiger, D., & Meek, C. (2005). Structured variational inference procedures and their realizations. *International Workshop on Artificial Intelligence and Statistics*, 104–111.

Hansen, J. Y., Shafiei, G., Voigt, K., Liang, E. X., Cox, S. M. L., Leyton, M., Jamadar, S. D., & Misic, B. (2023). Integrating multimodal and multiscale connectivity blueprints of the human cerebral cortex in health and disease. *PLOS Biology*, *21*(9), e3002314. https://doi.org/10.1371/journal.pbio.3002314

He, Y., Wang, J., Wang, L., Chen, Z. J., Yan, C., Yang, H., Tang, H., Zhu, C., Gong, Q., Zang, Y., & others. (2009). Uncovering intrinsic modular organization of spontaneous brain activity in humans. *PloS One*, *4*(4), e5226.

He, Y., Zeng, D., Li, Q., Chu, L., Dong, X., Liang, X., Sun, L., Liao, X., Zhao, T., Chen, X., Lei, T., Men, W., Wang, Y., Wang, D., Hu, M., Pan, Z., Zhang, H., Liu, N., Tan, S., … Li, S. (2025). The multiscale brain structural re-organization that occurs from childhood to adolescence correlates with cortical morphology maturation and functional specialization. *PLOS Biology*, *23*(4), e3002710. https://doi.org/10.1371/journal.pbio.3002710

Henzinger, M., Li, J., Rao, S., & Wang, D. (2024). Deterministic near-linear time minimum cut in weighted graphs. *Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 3089–3139.

Hu, Y., Nie, Y., Yang, H., Cheng, J., Fan, Y., & Di, Z. (2010). Measuring the significance of community structure in complex networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, *82*(6), 066106.

Hyvärinen, A. (2013). Independent component analysis: Recent advances. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *371*(1984), 20110534.

Iraji, A., Fu, Z., Faghiri, A., Duda, M., Chen, J., Rachakonda, S., DeRamus, T., Kochunov, P., Adhikari, B. M., Belger, A., Ford, J. M., Mathalon, D. H., Pearlson, G. D., Potkin, S. G., Preda, A., Turner, J. A., Van Erp, T. G. M., Bustillo, J. R., Yang, K., … Calhoun, V. D. (2023).

Identifying canonical and replicable multi-scale intrinsic connectivity networks in 100k+ RESTING-STATE FMRI datasets. *Human Brain Mapping*, *44*(17), 5729–5748. https://doi.org/10.1002/hbm.26472

Jaakkola, T. S., & Jordan, M. I. (1998). Improving the mean field approximation via the use of mixture distributions. In *Learning in graphical models* (pp. 163–173). Springer.

Jeffreys, H. (1961). *The theory of probability*. OuP Oxford.

Jordan, M. I., Ghahramani, Z., Jaakkola, T. S., & Saul, L. K. (1998). An Introduction to Variational Methods for Graphical Models. In M. I. Jordan (Ed.), *Learning in Graphical Models* (pp. 105–161). Springer Netherlands. https://doi.org/10.1007/978-94-011-5014-9_5

Karrer, B., Levina, E., & Newman, M. E. (2008). Robustness of community structure in networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, *77*(4), 046119.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795.

Keysers, C., Gazzola, V., & Wagenmakers, E.-J. (2020). Using Bayes factor hypothesis testing in neuroscience to establish evidence of absence. *Nature Neuroscience*, *23*(7), 788–799. https://doi.org/10.1038/s41593-020-0660-4

Kucharský, Š., Wagenmakers, E.-J., Van Den Bergh, D., & Ly, A. (2023). *Analytic Posterior Distribution and Bayes Factor for Pearson Partial Correlations*. PsyArXiv. https://doi.org/10.31234/osf.io/6muwy

Lancichinetti, A., Radicchi, F., & Ramasco, J. J. (2010). Statistical significance of communities in networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, *81*(4), 046110.

Latifi, S., & Carmichael, S. T. (2024). The emergence of multiscale connectomics-based approaches in stroke recovery. *Trends in Neurosciences*, *47*(4), 303–318. https://doi.org/10.1016/j.tins.2024.01.003

Lawrance, A. (1976). On conditional and partial correlation. *The American Statistician*, *30*(3), 146–149.

Li, S., Zhang, J., Huang, K., & Gao, C. (2014). A graph partitioning approach for Bayesian Network structure learning. *Proceedings of the 33rd Chinese Control Conference*, 2887–2892. https://doi.org/10.1109/ChiCC.2014.6897098

Li, Y., & Qi, Y. (2020). Asymptotic distribution of modularity in networks. *Metrika*, *83*(4), 467–484. https://doi.org/10.1007/s00184-019-00740-7

Litvak, V., Jafarian, A., Zeidman, P., Tibon, R., Henson, R. N., & Friston, K. (2019). There's no such thing as a 'true' model: The challenge of assessing face validity. *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, 4403–4408. https://doi.org/10.1109/SMC.2019.8914255

Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, *28*(2), 129–137.

Lu, M., Guo, Z., Gao, Z., Cao, Y., & Fu, J. (2022). Multiscale Brain Network Models and Their Applications in Neuropsychiatric Diseases. *Electronics*, *11*(21), 3468. https://doi.org/10.3390/electronics11213468

Maksimenko, V. A., Runnova, A. E., Frolov, N. S., Makarov, V. V., Nedaivozov, V., Koronovskii, A. A., Pisarchik, A., & Hramov, A. E. (2018). Multiscale neural connectivity during

human sensory processing in the brain. *Physical Review E*, *97*(5), 052405. https://doi.org/10.1103/PhysRevE.97.052405

Marković, D., Friston, K. J., & Kiebel, S. J. (2024). Bayesian sparsification for deep neural networks with Bayesian model reduction. *IEEE Access*, *12*, 88231–88242. https://doi.org/10.1109/ACCESS.2024.3417219

Meng, X., Iraji, A., Fu, Z., Kochunov, P., Belger, A., Ford, J., McEwen, S., Mathalon, D. H., Mueller, B. A., Pearlson, G., & others. (2022). Multimodel order independent component analysis: A data-driven method for evaluating brain functional network connectivity within and between multiple spatial scales. *Brain Connectivity*, *12*(7), 617–628.

Meunier, D. (2009). Hierarchical modularity in human brain functional networks. *Frontiers in Neuroinformatics*, *3*. https://doi.org/10.3389/neuro.11.037.2009

Meunier, D., Lambiotte, R., & Bullmore, E. T. (2010). Modular and Hierarchically Modular Organization of Brain Networks. *Frontiers in Neuroscience*, *4*. https://doi.org/10.3389/fnins.2010.00200

Mirshahvalad, A., Beauchesne, O. H., Archambault, E., & Rosvall, M. (2013). Resampling effects on significance analysis of network clustering and ranking. *PloS One*, *8*(1), e53943.

Mirshahvalad, A., Lindholm, J., Derlén, M., & Rosvall, M. (2012). Significant communities in large sparse networks. *PloS One*, *7*(3), e33721.

Mirzaeian, S., Jensen, K. M., Ballem, A. R., Calhoun, V. D., & Iraji, A. (2025). Ultra-High Order Independent Component Analysis for Intrinsic Connectivity Networks in Resting-State Functional Magnetic Resonance Imaging Data. *2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI)*, 1–4. https://doi.org/10.1109/ISBI60581.2025.10981312

Morante, M., Frølich, K., & Rehman, N. ur. (2024). *Multiscale Functional Connectivity: Exploring the brain functional connectivity at different timescales* (No. arXiv:2406.19041). arXiv. https://doi.org/10.48550/arXiv.2406.19041

Morante, M., Frølich, K., Shahzaib, M., Shakil, S., & Rehman, N. U. (2024). Multiscale Functional Connectivity analysis of episodic memory reconstruction. *Frontiers in Cognition*, *3*, 1433234. https://doi.org/10.3389/fcogn.2024.1433234

Munn, B. R., Müller, E. J., Favre-Bulle, I., Scott, E., Lizier, J. T., Breakspear, M., & Shine, J. M. (2024). Multiscale organization of neuronal activity unifies scale-dependent theories of brain function. *Cell*, *187*(25), 7303-7313.e15. https://doi.org/10.1016/j.cell.2024.10.004

Newman, M. E. J. (2004). Analysis of weighted networks. *Physical Review E*, *70*(5), 056131. https://doi.org/10.1103/PhysRevE.70.056131

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

Oizumi, M., Tsuchiya, N., & Amari, S. (2016). Unified framework for information integration based on information geometry. *Proceedings of the National Academy of Sciences*, *113*(51), 14817–14822.

Park, B., Paquola, C., Bethlehem, R. A. I., Benkarim, O., Neuroscience in Psychiatry Network (NSPN) Consortium, Mišić, B., Smallwood, J., Bullmore, E. T., Bernhardt, B. C., Bullmore, E., Dolan, R., Goodyer, I., Fonagy, P., Jones, P., Moutoussis, M., Hauser, T., Neufeld, S., Romero-Garcia, R., St Clair, M., … Chamberlain, S. (2022). Adolescent

development of multiscale structural wiring and functional interactions in the human connectome. *Proceedings of the National Academy of Sciences*, *119*(27), e2116673119. https://doi.org/10.1073/pnas.2116673119

Penny, W. D. (2012). Comparing Dynamic Causal Models using AIC, BIC and Free Energy. *NeuroImage*, *59*(1), 319–330. https://doi.org/10.1016/j.neuroimage.2011.07.039

Penny, W. D., Stephan, K. E., Mechelli, A., & Friston, K. J. (2004). Comparing dynamic causal models. *NeuroImage*, *22*(3), 1157–1172. https://doi.org/10.1016/j.neuroimage.2004.03.026

Pines, A. R., Larsen, B., Cui, Z., Sydnor, V. J., Bertolero, M. A., Adebimpe, A., Alexander-Bloch, A. F., Davatzikos, C., Fair, D. A., Gur, R. C., Gur, R. E., Li, H., Milham, M. P., Moore, T. M., Murtha, K., Parkes, L., Thompson-Schill, S. L., Shanmugan, S., Shinohara, R. T., … Satterthwaite, T. D. (2022). Dissociable multi-scale patterns of development in personalized brain networks. *Nature Communications*, *13*(1), 2647. https://doi.org/10.1038/s41467-022-30244-4

Prando, G., Zorzi, M., Bertoldo, A., Corbetta, M., Zorzi, M., & Chiuso, A. (2020). Sparse DCM for whole-brain effective connectivity from resting-state fMRI data. *NeuroImage*, *208*, 116367. https://doi.org/10.1016/j.neuroimage.2019.116367

Presigny, C., & Fallani, F. D. V. (2022). Multiscale modeling of brain network organization. *Reviews of Modern Physics*, *94*(3), 031002. https://doi.org/10.1103/RevModPhys.94.031002

Qian, H., & Beard, D. A. (2005). Thermodynamics of stoichiometric biochemical networks in living systems far from equilibrium. *Biophysical Chemistry*, *114*(2–3), 213–220.

Reichardt, J., & Bornholdt, S. (2006). When are networks truly modular? *Physica D: Nonlinear Phenomena*, *224*(1–2), 20–26.

Reichardt, J., & Bornholdt, S. (2007). Partitioning and modularity of graphs with arbitrary degree distribution. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, *76*(1), 015102.

Rousseau, P.-N., Bazin, P.-L., & Steele, C. J. (2025). Multiscale gradients of corticopontine structural connectivity. *Scientific Reports*, *15*(1), 16399. https://doi.org/10.1038/s41598-025-00886-7

Sala, A., Caminiti, S. P., Presotto, L., Premi, E., Pilotto, A., Turrone, R., Cosseddu, M., Alberici, A., Paghera, B., Borroni, B., Padovani, A., & Perani, D. (2017). Altered brain metabolic connectivity at multiscale level in early Parkinson's disease. *Scientific Reports*, *7*(1), 4256. https://doi.org/10.1038/s41598-017-04102-z

Schreiber, G. R., & Martin, O. C. (1999). Cut Size Statistics of Graph Bisection Heuristics. *SIAM Journal on Optimization*, *10*(1), 231–251. https://doi.org/10.1137/S1052623497321523

Seifi, M., Junier, I., Rouquier, J.-B., Iskrov, S., & Guillaume, J.-L. (2013). Stable community cores in complex networks. In *Complex Networks* (pp. 87–98). Springer.

Sevi, H., Jonckheere, M., & Kalogeratos, A. (2022). Generalized spectral clustering for directed and undirected graphs. *arXiv Preprint arXiv:2203.03221*.

Shine, J. M., Müller, E. J., Munn, B., Cabral, J., Moran, R. J., & Breakspear, M. (2021). Computational models link cellular mechanisms of neuromodulation to large-scale neural dynamics. *Nature Neuroscience*, *24*(6), 765–776. https://doi.org/10.1038/s41593-021-00824-6

Tavara, S., & Schliep, A. (2021). Effects of network topology on the performance of consensus and distributed learning of SVMs using ADMM. *PeerJ Computer Science*, *7*, e397. https://doi.org/10.7717/peerj-cs.397

Theodorou, T., Mporas, I., & Fakotakis, N. (2014). An Overview of Automatic Audio Segmentation. *International Journal of Information Technology and Computer Science*, *6*(11), 1–9. https://doi.org/10.5815/ijitcs.2014.11.01

Toker, D., & Sommer, F. T. (2019). Information integration in large brain networks. *PLOS Computational Biology*, *15*(2), e1006807. https://doi.org/10.1371/journal.pcbi.1006807

Tremblay, N., Puy, G., Gribonval, R., & Vandergheynst, P. (2016). Compressive spectral clustering. *International Conference on Machine Learning*, 1002–1011.

Tritschler, A., & Gopinath, R. A. (1999). Improved speaker segmentation and segments clustering using the bayesian information criterion. *6th European Conference on Speech Communication and Technology*, 679–682. https://doi.org/10.21437/Eurospeech.1999-174x

Van Den Heuvel, M. P., Scholtens, L. H., & Kahn, R. S. (2019). Multiscale Neuroscience of Psychiatric Disorders. *Biological Psychiatry*, *86*(7), 512–522. https://doi.org/10.1016/j.biopsych.2019.05.015

Vatsavai, R. R., Symons, C. T., Chandola, V., & Jun, G. (2011). GX-Means: A model-based divide and merge algorithm for geospatial image clustering. *Procedia Computer Science*, *4*, 186–195. https://doi.org/10.1016/j.procs.2011.04.020

Von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and Computing*, *17*(4), 395–416. https://doi.org/10.1007/s11222-007-9033-z

Wang, C., & Shanechi, M. M. (2019). Estimating Multiscale Direct Causality Graphs in Neural Spike-Field Networks. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *27*(5), 857–866. https://doi.org/10.1109/TNSRE.2019.2908156

Wang, Y., Huang, H., Feng, C., & Liu, Z. (2015). Community detection based on minimum-cut graph partitioning. *Web-Age Information Management: 16th International Conference, WAIM 2015, Qingdao, China, June 8-10, 2015. Proceedings 16*, 57–69.

Weatherburn, C. E. (1949). *A first course mathematical statistics* (Vol. 158). CUP Archive.

Wright, D., Igel, C., & Selvan, R. (2024). *BMRS: Bayesian Model Reduction for Structured Pruning* (No. arXiv:2406.01345). arXiv. https://doi.org/10.48550/arXiv.2406.01345

Wu, G. R., Chen, F., Kang, D., Zhang, X., D. Marinazzo, & Chen, H. (2011). Multiscale Causal Connectivity Analysis by Canonical Correlation: Theory and Application to Epileptic Brain. *IEEE Transactions on Biomedical Engineering*, *58*(11), 3088–3096. https://doi.org/10.1109/TBME.2011.2162669

Xing, E. (2004). Graph partition strategies for generalized mean field inference. *Proc Uncertainty in Artificial Intelligence 20 (UAI2004)*.

Xing, E. P., Jordan, M. I., & Russell, S. (2002). A generalized mean field algorithm for variational inference in exponential families. *Proceedings of the Nineteenth Conference on Uncertainty in Artificial Intelligence*, 583–591.

Xue, H., Li, H., Gao, C., & Shi, Z. (2010). Computationally efficient audio segmentation through a multi-stage BIC approach. *2010 3rd International Congress on Image and Signal Processing*, 3774–3777. https://doi.org/10.1109/CISP.2010.5646687

Zeidman, P., Friston, K., & Parr, T. (2023). A primer on Variational Laplace (VL). *NeuroImage*, *279*, 120310. https://doi.org/10.1016/j.neuroimage.2023.120310

Zeidman, P., Jafarian, A., Seghier, M. L., Litvak, V., Cagnan, H., Price, C. J., & Friston, K. J. (2019). A guide to group effective connectivity analysis, part 2: Second level analysis with PEB. *NeuroImage*, *200*, 12–25. https://doi.org/10.1016/j.neuroimage.2019.06.032

Zhao, Q., Hautamaki, V., & Fränti, P. (2008). Knee Point Detection in BIC for Detecting the Number of Clusters. In J. Blanc-Talon, S. Bourennane, W. Philips, D. Popescu, & P. Scheunders (Eds.), *Advanced Concepts for Intelligent Vision Systems* (Vol. 5259, pp. 664–673). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-88458-3_60

Zhou, Z., Li, H., Srinivasan, D., Abdulkadir, A., Nasrallah, I. M., Wen, J., Doshi, J., Erus, G., Mamourian, E., Bryan, N. R., Wolk, D. A., Beason-Held, L., Resnick, S. M., Satterthwaite, T. D., Davatzikos, C., Shou, H., & Fan, Y. (2023). Multiscale functional connectivity patterns of the aging brain learned from harmonized rsfMRI data of the multi-cohort iSTAGING study. *NeuroImage*, *269*, 119911. https://doi.org/10.1016/j.neuroimage.2023.119911

Zou, T., Chen, C., Chen, H., Wang, X., Gan, L., Wang, C., Gao, Q., Zhang, C., Liao, W., Cheng, J., & Li, R. (2024). Structural-functional connectivity decoupling in multiscale brain networks in Parkinson's disease. *BMC Neuroscience*, *25*(1), 78. https://doi.org/10.1186/s12868-024-00918-4