NeuroImage 211 (2020) 116603

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

# Scale-resolved analysis of brain functional connectivity networks with spectral entropy

Carlo Nicolini [a,1,*], Giulia Forcellini [a,b,1], Ludovico Minati [b,c,d], Angelo Bifone [a,e]

[a] Center for Neuroscience and Cognitive Systems, Istituto Italiano di Tecnologia, Rovereto, TN, Italy
[b] Center for Mind/Brain Sciences (CIMeC), University of Trento, Rovereto, TN, Italy
[c] Tokyo Tech World Research Hub Initiative (WRHI), Institute of Innovative Research, Tokyo Institute of Technology, Yokohama, 226-8503, Japan
[d] Complex Systems Theory Dept., Institute of Nuclear Physics, Polish Academy of Sciences, Kraków 31-342, Poland
[e] Department of Molecular Biotechnology and Health Sciences, University of Torino, Torino, Italy

ARTICLE INFO

*Keywords:*
Resting-state
Null models
Threshold
Motion correction
Graph theory
Spectral entropy

ABSTRACT

Functional connectivity is derived from inter-regional correlations in spontaneous fluctuations of brain activity, and can be represented in terms of complete graphs with continuous (real-valued) edges. The structure of functional connectivity networks is strongly affected by signal processing procedures to remove the effects of motion, physiological noise and other sources of experimental error. However, in the absence of an established ground truth, it is difficult to determine the optimal procedure, and no consensus has been reached on the most effective approach to remove nuisance signals without unduly affecting the network intrinsic structural features. Here, we use a novel information-theoretic approach, based on von Neumann entropy, which provides a measure of information encoded in the networks at different scales. We also define a measure of distance between networks, based on information divergence, and optimal null models appropriate for the description of functional connectivity networks, to test for the presence of nontrivial structural patterns that are not the result of simple local constraints. This formalism enables a scale-resolved analysis of the distance between a functional connectivity network and its maximally random counterpart, thus providing a means to assess the effects of noise and image processing on network structure. We apply this novel approach to address a few open questions in the analysis of brain functional connectivity networks. Specifically, we demonstrate a strongly beneficial effect of network sparsification by removal of the weakest links, and the existence of an optimal threshold that maximizes the ability to extract information on large-scale network structures. Additionally, we investigate the effects of different degrees of motion at different scales, and compare the most popular processing pipelines designed to mitigate its deleterious effect on functional connectivity networks. We show that network sparsification, in combination with motion correction algorithms, dramatically improves detection of large scale network structure.

## 1. Introduction

Complex networks theory provides a robust framework to study the structural and functional organization of brain connectivity, which can be naturally represented as a graph (Newman, 2010), a collection of nodes (anatomical brain regions) and edges (functional or structural coupling between nodes), as in Fig. 1.

Several foundational concepts have been borrowed from network theory and have become part of the neuroscience parlance. By way of example, properties like small-worldness (Bassett and Bullmore, 2006), scale-freeness (van den Heuvel et al., 2008) and modularity (Sporns and Betzel, 2016; Nicolini and Bifone, 2016; Nicolini et al., 2017) have been demonstrated in brain networks, providing insight into their complex topological organization and its bearing on the dynamical processes underlying brain function in health and disease (Sporns, 2010).

However, despite its increasing popularity, this approach is still the subject of debate, and a number of seemingly simple, yet critically important questions remain open. For example, a widely accepted measure of the distance between graphs, e.g. to compare a network with an appropriate null model, is still missing, and so is a measure of the information encoded in a network. This problem is exacerbated by the lack of a ground truth structure for brain networks.

* Corresponding author.
*E-mail addresses:* nicolinidisi@gmail.com, angelo.bifone@iit.it (C. Nicolini).
[1] These two authors contributed equally.

https://doi.org/10.1016/j.neuroimage.2020.116603
Received 10 January 2020; Accepted 1 February 2020
Available online 7 February 2020
1053-8119/© 2020 Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

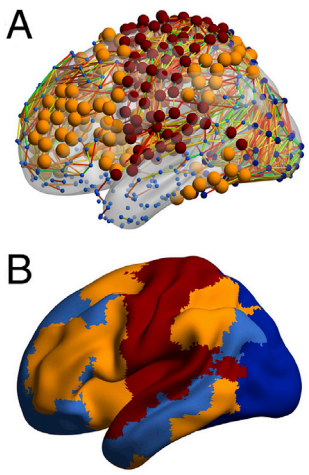


**Fig. 1.** Examples of a resting state functional brain network. In Panel A., nodes represent brain regions, connected by links whose value is obtained by averaging Fisher transformed Pearson correlations over many subjects. Radius of nodes indicates the degree, while their colors indicate the modular membership as overlaid on a surface template in Panel B. The partition entails four communities and is found with the Louvain algorithm (Blondel et al., 2008), modularity value $Q \approx 0.4$.

Neuroimaging methods, like functional MRI (fMRI), electroencephalography and magnetoencephalography (EEG, MEG) are often used to assess brain functional connectivity; the resulting networks depend tremendously on data processing and experimental parameters, but the lack of an established reference makes it challenging to determine the optimal procedures univocally. Glaring examples of this problem include the ongoing debate regarding the use of Global Signal Regression (Murphy et al., 2009) in resting-state functional connectivity, or the application of thresholds for network sparsification (Bordier et al., 2017; van den Heuvel et al., 2017; Hallquist and Hillary, 2019), as well as the physical meaning of thresholding weighted networks (Robinson, 2019) or the spatial effects of cortex geometry in determining brain network's properties (Henderson and Robinson, 2011).

Specific network metrics (such as node degree distribution or modularity) have often been used for network comparison, but they do not capture the intrinsically multiscale structure of brain networks, focusing only on specific local or global features. Moreover, such measures describe properties that may also be present in random networks. By way of example, large values of modularity have been observed in random networks (Guimera et al., 2004) as well as in natural networks. Hence, a measure of the distance between a given brain network and an appropriate null model, i.e., a random counterpart that satisfies certain constraints, would be essential to address the fundamental question: how far from random are brain networks?

Information-theoretic measures based on the Shannon entropy would appear to be a natural choice to address this issue, but an extension of this formalism to complex networks has proven challenging, although useful in specific contexts (De Domenico et al., 2015; Anand and Bianconi, 2009).

Recently, pioneering work by De Domenico and Biamonte (De Domenico and Biamonte, 2016) has demonstrated the use of spectral entropies to define distances between pairs of complex networks. Specifically, those authors recognized that the Laplacian of the adjacency matrix describing a given network can be used to construct a matrix that satisfies the same mathematical properties of quantum mechanical density matrices, thus enabling the extension of von Neumann entropy to complex networks. Several implications of this elegant development may be useful for the analysis of brain networks. Differently from other methods, this approach does not rely on a subset of network descriptors (Squartini and Garlaschelli, 2017), but provides an information-based measure that takes into account the entire network structure at all scales. The strength of this formalism lays in the dynamical description of a diffusion process taking place over the network, with the characteristic time described by a parameter $\beta$ that determines the diffusion scale. Hence, spectral entropies provide a scale-resolved, information-based metric to define and optimize network models.

The same framework enables measuring the distance between networks that can be rigorously defined in terms of quantum relative entropy, or information divergence (Wilde, 2013). This quantifies the information gain when a model is used to explain an empirical observation. Moreover, minimization of relative entropy can be used to optimize model parameters (De Domenico and Biamonte, 2016; Nicolini et al., 2018), or to select different models based on their ability to reproduce the data.

Here, we extend the novel formalism to the study of brain functional connectivity, and demonstrate its potential to address a few open problems in the analysis of resting-state connectivity networks.

Firstly, we implement two new models of maximally-random networks with specific local or global properties to evaluate their deviation from their original counterparts at different scales. The simpler model describes a class of networks where the total number of links and total weight are constrained to match those of another network. The latter model addresses the more general case in which both degree and strength sequence are preserved upon randomization of the edges.

At variance with previous null models based on integer link weights (Squartini et al., 2015; Mastrandrea et al., 2014), and as an important point of novelty, our models are applicable to networks with continuous weights, like those derived from functional connectivity, and can be optimized fixing their density to that of the original network, by including an external threshold parameter.

A most contentious methodological issue in graph analysis, as applied to the study of brain connectivity, is the one of network sparsification. Functional connectivity networks are generally derived from pairwise correlations of spontaneous fluctuations extracted from each pair of brain regions, resulting, by definition, in a fully connected weighted matrix. However, dense networks are computationally demanding, and weak links, which represent the overwhelming majority of edges, might contain spurious correlations. Network sparsification is then an essential step to recover the network structure.

A number of different thresholding techniques have been proposed (Hallquist and Hillary, 2019; van den Heuvel et al., 2017; Bordier et al., 2017; Santarnecchi et al., 2014; Lohse et al., 2014; Schwarz and McGonigle, 2011; de Vico Fallani et al., 2017) but the choice of the threshold, which strongly affects the topological features extracted from the network, such as the highly specific intermodular links (Zalesky et al., 2016), remains somewhat arbitrary, and according to some authors, ill-posed (Robinson, 2019). Adding to this, it is now argued that the sparsification procedure itself might actually inject artifactual structures within the network of interest. A recent study, indeed, revealed the introduction of some complex features in random networks as a sole result of thresholding (Cantwell et al., 2019). Altogether, given the lack of agreement on the best thresholding level, and the uncertain significance of the weakest and negative links, there is a trend to completely avoid the application of a threshold, and to work with fully connected networks (Schlesinger et al., 2017; Goulas et al., 2015; Bassett et al., 2011; Rubinov and Sporns, 2011). Starting from the above considerations, here we apply the novel formalism of spectral entropies to investigate the effects of thresholding on resting-state networks, with the aim to assess whether this practice is beneficial for the extraction of information on the large-scale organization of the network, and whether an optimal threshold that maximizes separation between a network and its random counterpart exists.

Other open issues in the functional connectivity field relate to more basic aspects of image processing. For example, it is well known that small head movements during the resting-state fMRI scan can substantially impact the subsequent functional connectivity analysis (Power

et al., 2011; Van Dijk et al., 2012; Satterthwaite et al., 2012). Motion artifacts are a significant cause of spurious correlations that can substantially affect the structure of functional connectivity networks derived from functional MR. The search for an optimal strategy for the correction of motion-related noise has become a center of attention in the field (Ciric et al., 2017; Satterthwaite et al., 2019; Parkes et al., 2018). Moreover, non-neural physiological activity, like cardiovascular pulsation, respiratory cycle and autonomic fluctuations can inject spurious correlations across multiple frequencies (Siegel et al., 2014; Marchitelli et al., 2016). A plethora of different noise correction techniques have been introduced, all aiming at the reduction or removal of the impact of in-scanner motion effects (Ciric et al., 2017; Burgess et al., 2016). However, a consensus on the most effective approach to removing motion effects without substantially affecting the network intrinsic structural features is still lacking. For example, a major debate revolves around the application of a global signal regression (GSR) (Fox et al., 2009; Saad et al., 2012; Aquino et al., 2019). The aggressiveness of approaches based on GSR, and the subsequent introduction of negative correlations in the network following its application, has made it the subject of discussion in the neuroscientific community, despite its apparent efficacy in removing spatially correlated spurious fluctuations. Here, we seek to demonstrate the potential use of the spectral entropy formalism to study the effects of some of the most popular motion correction procedures on network structure. To this end, we assess the effects of motion on the information contained in the network at different scales, and compare the efficacy of various data processing pipelines for the recovery of structural information in the presence of different degrees of motion.

## 2. Materials and methods

### 2.1. Null models for continuous (real-valued) thresholded networks

In network science, and in the study of statistical properties of graphs, a null-model is a mathematical entity representing a family of graphs that match some of the properties of a network, while remaining maximally noncommittal with regard to properties not explicitly specified (Jaynes, 1957; Park and Newman, 2004; Squartini and Garlaschelli, 2017). Null models provide a powerful way to test whether nontrivial, high-order network patterns are genuine or just the outcome of simpler local properties. The configuration model of Newman's modularity (Newman, 2006), for example, is a null model in which the link probability $p_{ij}$ is only based on the degrees of each edge endpoints $i$ and $j$. A large deviations from the predicted link probability in a subset of edges, is an indication that non-trivial relations exist between those links, which cannot be explained by the degrees alone.

Desirable null models do not trade complexity for sufficiency or redundancy (Betzel and Bassett, 2017): two constraints leading to the same ensemble should be merged into one. Moreover, null models should be neither too complicated nor too simple. Too many parameters and the null model no longer represents the state of maximal agnosticism as overfitting the data precisely displays all its features. Conversely, too few parameters and the picture it conveys is oversimplified, hence lacking explanatory power.

In this sense, we are looking for a model of rs-fMRI networks that is complex enough to match simple features of the network, but remaining completely uninformative over higher-order patterns. A similar model should at most deal with local properties of networks, i.e. quantities which are only functions of the immediate neighbors of the nodes, or more precisely, linear functions of the adjacency matrix (Squartini and Garlaschelli, 2017). In some cases, local constraints such as the number of links or the degree sequence may already fully contain all the information conveyed by the network. Networks of this kind have no statistical patterns or regularities (Betzel and Bassett, 2017; Peixoto, 2015) beyond those described by local properties: local features explicitly enforced represent a null hypothesis that we can use as a reference to quantify significant deviations or patterns.

As an another design choice for a proper null model for resting-state fMRI networks, we should take into account the continuous nature of link weights, the binary backbone, and weighted structural patterns. Importantly, here we only deal with positive link weights, as in the literature most of the graph-theoretical quantities have no simple interpretations in networks with negative links. Specific null models of Pearson correlation matrices, where one retain both positive and negative correlations, have been presented by MacMahon and Garlaschelli (2015) and Masuda et al. (2018).

Here we extend previous random graph models (Squartini and Garlaschelli, 2011; Mastrandrea et al., 2014; Garlaschelli and Loffredo, 2008; Cimini et al., 2019) to networks with real-valued links distributed over a connectivity backbone modeled by the degree and strength sequence. These local variables are the optimal trade-off to shape the irreducible and unavoidable complexity needed to accommodate the heterogeneous structure of real networks. Nonetheless, we also describe a simpler model with only two global constraints, namely the total number of links and weight.

In the next section we will show that the former model captures most of the long-range connectivity and mesoscopic structures, while the latter adequately describes local features but remains uninformative of larger-scale structure.

Hereby, we embrace the Exponential Random Graph model (ERGM) (Park and Newman, 2004; Newman, 2010; Obando and De Vico Fallani, 2017) and analytically build the maximally random counterparts of real networks where only specific features are reproduced, on average. The details of the mathematical formalism of the Exponential Random Graph model are given in Appendix A.1. Due to its nature, the ERGM defines ensembles with *soft* constraints: little fluctuations around the average have smaller but non-zero probability. This is at variance with micro-canonical approaches, also known as rewiring algorithms, like the one of Maslov and Sneppen (2002), where zero probability is assigned to networks with even the slightest violation of the desired properties (Cimini et al., 2019). However, the analytical tractability, together with unbiasedness, flexibility and easiness of sampling methods (Squartini and Garlaschelli, 2017), make Exponential Random Graphs a very powerful tool to define null models.

We first describe a model where the number of links and total weight are constrained and then move to the more general case where both the degree and the strength sequences of the network are considered. Most importantly, both models include a hard thresholding procedure (Esfahlani and Sayama, 2018; van den Heuvel et al., 2017; Schwarz and McGonigle, 2011) that is often used in analysis pipelines.

#### 2.1.1. Random networks with fixed links number and weight

Here we introduce the *Continuous Weighted Thresholded Enhanced Random Graph Model*, or simply CWTERG, as the random graph model that fixes the average total number of links $L^\star$ and the average total weight $W^\star$, together with an external threshold parameter $t$. The CWTERG is obtained within the ERGM formalism described in Appendix A.2, by a Hamiltonian that explicitly enforces these two properties and reads:

$$H_{\mathrm{CWTERG}}(\mathbf{G}|u,v) = uL(\mathbf{G}) + vW(\mathbf{G}), \tag{1}$$

where the Lagrangian multipliers $\boldsymbol{\theta}$ of the problem are the two scalars, $u$ and $v$. This Hamiltonian is designed to weight the contribution of binary links with the term $u$ and the contribution of weighted links with the term $v$. The role of the threshold parameter $t$ becomes clear if a dense network is fed in the model, and its null network is sought for as a function of the threshold.

Fitting the CWTERG model to a network requires to solve a system of two nonlinear equations, hence finding the values of the Lagrangian multipliers $\hat{u}, \hat{v}$ such that the observed number of links and total weights $L^\star, W^\star$ are:

$$\begin{cases} L^{\star} = \binom{N}{2} \dfrac{1}{\hat{v}te^{\widehat{u+vt}} + 1} \\ W^{\star} = \binom{N}{2} \dfrac{\hat{v}t + 1}{\hat{v}(\hat{v}te^{\widehat{u+vt}} + 1)} \end{cases} \tag{2}$$

Alternatively, one can find the optimal estimates of $\hat{u}, \hat{v}$ by maximization of the log-likelihood of the model $\mathscr{L}_{\text{CWTERG}}$:

$$\mathscr{L}_{\text{CWTERG}}(\mathbf{G}|u,v) = -uL(\mathbf{G}) - vW(\mathbf{G}) - \binom{N}{2}\log\left(t + \frac{e^{-u-vt}}{v}\right). \tag{3}$$

We report in Appendix A.2 the entire theoretical derivation of the expressions for the ensemble averages of the number of links and total weight, as well as the probability density function of the model.

#### 2.1.2. Random networks with fixed degrees and strengths

The CWTERG model describes the ensemble of networks whose total weight and number of links are constrained to reflect those of the observed graph. Hence it can be considered an extension of the Erdős-Renyi random graph model to thresholded weighted networks. However, this model only describes networks with uniform connectivity patterns, and it ignores the heterogeneity of the degrees and strengths. To overcome this issue, here we introduce the *Continuous Weighted Thresholded Enhanced Configuration Model*, or CTWECM, by defining a Hamiltonian that simultaneously considers the degree and strength sequences:

$$H_{\text{CWTECM}}(\mathbf{G}|\mathbf{u},\mathbf{v}) = \sum_{i<j}(u_i + u_j)\Theta(w_{ij} - t) + (v_i + v_j)w_{ij}\Theta(w_{ij} - t), \tag{4}$$

where $u_i, v_i$ are the Lagrangian multipliers associated to node degrees and strength, respectively. The node degrees are obtained as the sum of binary variables, hence the Heaviside function $\Theta$ is exactly centered at $t$, taking values one or zero if the edge weight exceeds the cut-off threshold. Similarly, the threshold $t$ shapes the sequence of nodes strength, by contributing with a factor $\sum_j w_{ij}$ for weights greater than the cut-off $t$. Although it may appear different, the structural form of the Hamiltonian of the CWTECM is the same as the one of the previously introduced CWTERG: the detailed analytical expression of the probability density function is reported in Appendix A.3.

Fitting the CWTECM to a network requires solving a system of $2N$ non-linear equations, $N$ of them for the degrees and $N$ for the strengths, to find the values of the Lagrangian multipliers $\hat{\mathbf{u}}, \hat{\mathbf{v}}$ such that the original degrees $k_i^{\star}$ and strengths $s_i^{\star}$ are:

$$\begin{cases} k_i^{\star} = \sum_{j\neq i} \dfrac{1}{1 + (v_i + v_j)te^{(u_i+u_j)+(v_i+v_j)t}} \\ s_i^{\star} = \sum_{j\neq i} \dfrac{(v_i + v_j)t + 1}{(v_i + v_j)t\left((v_i + v_j)te^{(u_i+u_j)+(v_i+v_j)t} + 1\right)}. \end{cases} \tag{5}$$

In the same spirit of the CWTERG, the optimal parameters $\hat{\mathbf{u}}, \hat{\mathbf{v}}$ can either be found by solution of the above system of $2N$ equations, or by maximization of the model log-likelihood that reads:

$$\mathscr{L}_{\text{CWTECM}}(\mathbf{G}|\mathbf{u},\mathbf{v}) = -\sum_i u_i k_i(\mathbf{G}) - \sum_i v_i s_i(\mathbf{G}) - \sum_{i<j}\log\left(t + \frac{e^{-u_i-u_j-t(v_i+v_j)}}{v_i + v_j}\right). \tag{6}$$

In the following we will study the properties of these two null models at different scales using a powerful information theoretical tool, based on the analogy with quantum statistical mechanical systems.

### 2.2. Spectral entropies

Classical maximum entropy methods are a tool for describing the ensemble of networks that show on average the same desired descriptors. However, when considering the problem of comparing two networks at all scales on a wide variety of metrics, classical maximum entropy methods fail in providing a statistically reliable tool, as one should design and solve a specific maximum entropy problem for every specific descriptor (Park and Newman, 2004; De Domenico and Biamonte, 2016).

It is possible to extend the maximum entropy approach to networks represented as quantum mechanical systems, by replacing the Shannon entropy with the von Neumann entropy (De Domenico and Biamonte, 2016; Nicolini et al., 2018):

$$S(\boldsymbol{\rho}) = -\text{Tr}[\boldsymbol{\rho}\log\boldsymbol{\rho}], \tag{7}$$

where $\boldsymbol{\rho}$ is the von Neumann density matrix, a Hermitian and positive definite matrix with unitary trace, that admits a spectral decomposition as:

$$\boldsymbol{\rho} = \sum_{i=1}^{n} \lambda_i(\boldsymbol{\rho})\varphi_i\varphi_i^T \tag{8}$$

for an orthonormal basis $\{\varphi_i\}$ and eigenvalues $\lambda_i(\boldsymbol{\rho})$.

By application of the same maximum entropy principle discussed in the previous sections, but constraining the networks to have, on average, the same Laplacian (see Appendix A) one finds that the role of classical probability in this case is replaced by the following density matrix:

$$\boldsymbol{\rho} = \frac{e^{-\beta\mathbf{L}}}{\text{Tr}}[e^{-\beta\mathbf{L}}] \tag{9}$$

that describes the result of constraining the diffusion properties on the networks, and is in the form of a quantum Gibbs-Boltzmann distribution. Intuitively, the elements of $\boldsymbol{\rho}$ represent the normalized amount of information transferred in a diffusion process between two nodes. The denominator $Z = \text{Tr}[e^{-\beta\mathbf{L}}]$ is the so-called partition function of the system (to be distinguished from $Z$ in Eq. (A.6) of the null models), which can also be expressed as the sum of the eigenvalues of the matrix $e^{-\beta\mathbf{L}}$ (Golub and Van Loan, 1996) as follows:

$$Z = \text{Tr}[e^{-\beta\mathbf{L}}] = \sum_{i=1}^{n} e^{-\beta\lambda_i(\mathbf{L})} \tag{10}$$

with $\lambda_i(\mathbf{L})$ the eigenvalues of the Laplacian $\mathbf{L}$. The von Neumann density matrix defined in Eq. (9) is based on exponentially scaled eigenvalues of the graph Laplacian, and contains contributions at different scales of patterns in the network, tuned by the scalar β. The role of β is far from trivial (Nicolini et al., 2018): it can be interpreted as an inverse temperature (in classical statistical mechanics) or a normalized time (Nicolini et al., 2018) in modeling heat diffusion over the network. In the $\beta \to 0$ limit, the density matrix can be expanded linearly as $\boldsymbol{\rho} \sim \mathbf{I} - \mathbf{L}$ and carries information about local connectivity patterns. On the other hand, for $\beta \to \infty$, the diffusive behaviour is governed by the smallest non-zero eigenvalue of the Laplacian $\lambda_2$, hence $\boldsymbol{\rho} \sim e^{-\beta\lambda_2}\varphi_2\varphi_2^T$, where $\varphi_2$ is the eigenvector associated to $\lambda_2$. This eigenvector is also called the Fiedler eigenvector and embodies the large scale structure of the graph (Xiao et al., 2005).

The Laplacian spectrum encloses several critical topological properties of graphs (Estrada, 2011; Anderson and Morley, 1985; Merris, 1994; de Lange et al., 2014). For instance, the multiplicity of the zero eigenvalue corresponds to the number of connected components, while the multiplicity of each eigenvalue is related to graph symmetries (Merris, 1994; de Lange et al., 2014, 2016), and the concept of expanders and isoperimetric number are connected to the first and second-largest eigenvalues (Cheeger, 1969; Donetti et al., 2006). Moreover, the graph Laplacian appears often in the study of random walkers (Lovász, 1993;

Masuda et al., 2017), diffusion (Bray and Rodgers, 1988), combinatorics (Mohar, 1991), and a large number of other applications (Merris, 1994; Mohar, 1991). For this reason, the spectral entropy, which is ultimately based on Laplacian eigenvalues, describes a large number of typical properties of the network, aggregated in a single quantity.

For the sake of comparing two different networks represented by the density matrices $\boldsymbol{\rho}$ and $\boldsymbol{\sigma}$ here we use the notion of von Neumann relative entropy (Wilde, 2013; De Domenico and Biamonte, 2016) that reads:

$$S(\boldsymbol{\rho} \parallel \boldsymbol{\sigma}) = \mathrm{Tr}\,[\boldsymbol{\rho}(\log\boldsymbol{\rho} - \log\boldsymbol{\sigma})] \tag{11}$$

This definition encloses the concept of network similarity, as the relative entropy is a positive quantity, and is zero if and only if $\boldsymbol{\rho} = \boldsymbol{\sigma}$. For this reason, and in the rest of this manuscript, we quantify the similarity between a rs-fMRI network represented by density matrix $\boldsymbol{\rho}$ and its randomized counterpart denoted by $\boldsymbol{\sigma}$, by means of von Neumann relative entropy $S(\boldsymbol{\rho} \parallel \boldsymbol{\sigma})$. Importantly, we evaluate $S(\boldsymbol{\rho} \parallel \boldsymbol{\sigma})$ and not $S(\boldsymbol{\sigma} \parallel \boldsymbol{\rho})$ as we are interested to compute the amount of information that is lost when using the model density $\boldsymbol{\sigma}$ to describe the data density matrix $\boldsymbol{\rho}$, and not viceversa. Additionally, it is straightforward (Nicolini et al., 2018; De Domenico and Biamonte, 2016) to show that with this order the minimum of relative entropy over all density matrices $\boldsymbol{\sigma}$ corresponds to the maximum of a log-likelihood functional, $\log\mathscr{L}(\boldsymbol{\sigma})$, defined as

$$\log\mathscr{L}(\boldsymbol{\sigma}) = \mathrm{Tr}[\boldsymbol{\rho}\log\boldsymbol{\sigma}]. \tag{12}$$

Similarly to Eq. (A.7), in the spectral framework it is possible to decompose the log-likelihood into the sum of Hamiltonian and free energy. This happens in equilibrium conditions where the density matrices $\boldsymbol{\rho}$ and $\boldsymbol{\sigma}$ are in the form of a Gibbs distribution like specified in Eq. (9). In particular, denoted $\boldsymbol{\rho}^\star$ and $\boldsymbol{\sigma}_{\mathrm{rand}}$ as the density matrices of a network $\mathbf{G}^\star$ and of its randomized counterpart $\mathbf{G}_{\mathrm{rand}}$ with Laplacian $\mathbf{L}_{\mathrm{rand}}$, respectively, the resulting spectral log-likelihood from Eq. (12) takes the following form:

$$\log\mathscr{L}(\boldsymbol{\sigma}_{\mathrm{rand}}) = -\,\mathrm{Tr}\,[\boldsymbol{\rho}^\star\mathbf{L}_{\mathrm{rand}}] - \log\mathrm{Tr}\,[e^{-\beta\mathbf{L}_{\mathrm{rand}}}]. \tag{13}$$

Here we compare networks with their randomized counterparts sampled from the maximum entropy models CWTERG and CWTECM using spectral entropies and relative spectral entropies. We leverage these tools in order to resolve differences between networks at different scales governed by the diffusion parameter β. Indeed, graph models that preserve the local features (like nodal strength and degree) exhibit comparatively similar spectral entropies to those of the original counterpart in the regime $\beta \to 0$. Conversely, mesoscopic structures that cannot be modeled by solely constraining local features, will result in different von Neumann entropy $S$ in the large β regime.

This concept is illustrated in Fig. 2, where a highly regular network and its randomized counterpart are depicted, together with their von Neumann entropies, over a large range of the β parameter. The orange line in Fig. 2A, describes the entropy $S$ of the modular network. A large plateau indicates the tendency of a random walker to remain trapped in medium-size highly dense subset of nodes: its height is indicative of the logarithm of the number of modules $B$. On the other hand in the two extreme β regimes, entropy tends to its maximum or minimum attainable values, log$N$ or log$C$ respectively, where $C$ is the number of weakly connected components. Highly related to this observation is the study of Lambiotte et al. (2014), that however only measured the number of detected communities as a function of diffusion time. A degree-preserving randomized network has more uniform diffusion properties instead, as testified by the sharply falling blue dashed curve, where no specific mesoscopic patterns can be found. In the $\beta \to 0$ limit, diffusion is only limited to the local neighborhood of nodes, hence, the orange and blue curves look similar. With a physical metaphor, reminiscent of the Infomap algorithm (Rosvall and Bergstrom, 2008), random walkers spend more time at increasingly larger and isolated structures of the modular network, from small clusters to larger modules (Fig. 2B). Finally, at steady-state for $\beta \to \infty$ both the two curves converge to the global structure, only specified by the number of connected components.

The concepts and models exposed in this section are implemented in the open source package *networkqit*, written in Python.

## 3. Data and preprocessing

For the purpose of this study we selected publicly available resting-state networks as benchmarks to test this new theoretical framework.

### 3.1. Resting-state network

To evaluate the effects of different thresholding on the network structure, we have chosen a resting-state network computed as a group average of 27 healthy volunteers (mean age 24 yrs) and described in

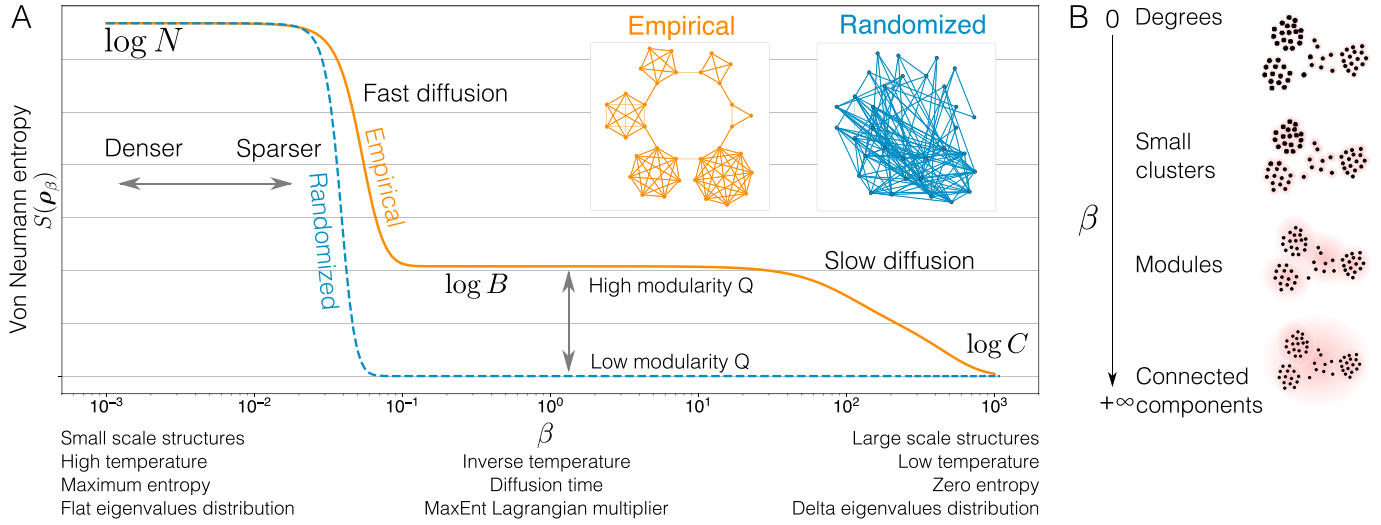


**Fig. 2.** Von Neumann spectral entropy $S(\boldsymbol{\rho}_\beta)$ as a function of β for a highly-ordered network $\boldsymbol{\rho}^\star$ (orange), and its randomized, degree preserving counterpart (blue), $\boldsymbol{\rho}_{\mathrm{rand}}$. Panel A. For small values of β, the spectral entropy reaches its maximum value $S = \log N$, while in the large β limit it tends to the logarithm of the number of connected components $C$. Intermediate values of β highlight mesoscopic structures. The height of the plateau is related to the overall modularity of the network, while its positioning on the horizontal axis depends on network links density. Differently from the highly regular ring of cliques (orange), the randomized network (blue) shows no structure at all scales, hence its von Neumann entropy decreases rapidly. Panel B shows that low β correspond to local features while large β describes large scale features.

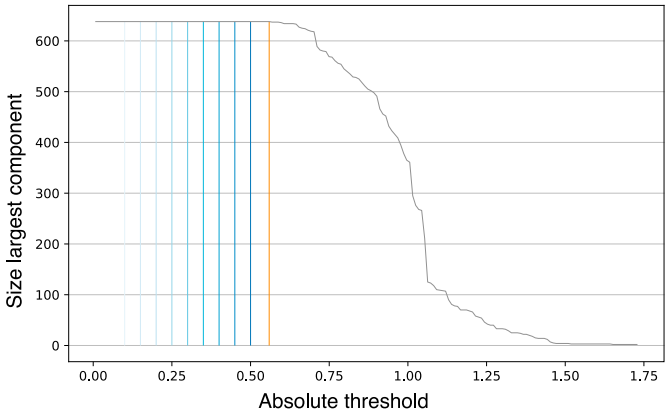


**Fig. 3.** Percolation analysis for the healthy resting state dataset of 638 nodes, described in Crossley et al. (Crossley et al., 2013). The gray curve indicates the size of the largest connected component as a function of the threshold. The blue lines correspond to threshold values from 0.1 to 0.5, the orange line is the percolation threshold, where the largest connected component starts breaking apart.

Ref. (Crossley et al., 2013), alongside with the ethical statements. This functional connectivity network is a popular benchmark for testing graph-theoretical methods, and was chosen to enable direct comparison with previous literature. The connectivity matrix is available at Ref (Crossley et al., 2016). Functional data were acquired with a Siemens Tim Trio 3T scanner, with a TR = 2 s, TE = 31 ms, voxel size 3.5× 3.5× 3 mm, for a total of 153 vol recorded for 5 min and 6 s. Regional time-series were extracted for 638 nodes using Crossley's parcellation scheme (Crossley et al., 2013), head rotations and translations together with their derivatives and mean cerebrospinal fluid time series were regressed and band-passed (0.01–0.1 Hz). The functional connectivity matrix was derived computing pairwise Pearson correlations, normalized by the Fisher transform, and finally across subjects. The network corresponds to the unthresholded version made publicly available through the Brain Connectivity Toolbox (BCT (Rubinov and Sporns, 2010),). To assess the effect of thresholding, we applied a range of different absolute thresholds, from $t = 0.1$ corresponding to a link density $d = 98\%$, to the point where the network remains fully connected.

Above this threshold, nodes start detaching from the main largest connected component, reflecting the hierarchy of modules comprised in the network (Gallos et al., 2012). Here, absolute thresholding corresponds to the removal of all edges with weight $w_{ij} < t$ where $t$ is a real positive number. The point where the network breaks apart is dubbed percolation point. Thus, with the term percolation threshold we mean the highest value of absolute threshold $t^{\star}$ such that the undirected network remains connected, i.e. it comprises a single connected component. The percolation analysis of the Crossley network is shown in Fig. 3.

### 3.2. Motion and motion correction

To study the effects of motion, we selected neuroimaging data from the MPI-Leipzig Study for Mind-Body-Emotion Interactions project (LEMON (Mendes et al., 2019),), obtained from the OpenfMRI database, accession number *ds000221*. Ethical statements are present in the original references by the groups who performed the experiments. Given the growing concern in the neuroimaging literature regarding the effects of head motion on resting-state functional connectivity data (Power et al., 2012), and the impact of different motion-correction techniques (Ciric et al., 2017), we decided to explore the performance of the two null models over matrices containing different degrees of motion and different preprocessing pipelines. Indeed, according to recent findings, even very small head movements (0.2 mm) can substantially affect functional connectivity networks, increasing spurious correlations and altering its underlying topology (Power et al., 2012; Van Dijk et al., 2012). The use of this large data-base has enabled us to select subgroups of subjects with different levels of motion and adequate size for a significant between group comparison.

From this dataset, participants were selected according to the age range; only participants ranging from 20 to 30 y.o. were included in our study, to avoid age effects in subsequent analyses. All MRI data were acquired with a 3T scanner (Magnetom Verio, Siemens Healthcare, Erlangen, Germany), with the following parameters TR = 1.4 s, TE = 39.4 ms, for a total of 657 vol, resulting in 15 min and 30 s of recording. A total of 117 subjects were selected. Structural and functional images were preprocessed with FSL (v 5.0) (Jenkinson et al., 2012). High-resolution structural images were registered to the MNI template and segmented (fast segmentation), separating white matter and ventricles masks. Functional preprocessing included motion correction and realignment (mcflirt), coregistration to the structural image using boundary based registration (BBR) and then normalized to the MNI template. For each participant, we extracted regional mean time series from 638 parceled areas, based on the same template employed for the other networks already addressed in the present study. A Butterworth bandpass filter between 0.01 and 0.1 Hz was applied to all the time series.

For the purposes of the study, all participants were divided into three groups according to their degree of motion (Low, Medium, and high-motion), measured as the proportion of outlier volumes present within the time series. To evaluate the motion level of each subject, Framewise Displacement (FD) was computed according to Power (Power et al., 2012). Timepoints were flagged as outliers affected by motion when FD $> 0.3$ mm. Criteria for group subdivision, decided after careful inspection of the data, were the following:

- Low-motion ($N = 39$): less than 1% data affected.
- Medium-motion ($N = 39$): data affected between 1% and 5%.
- High-motion ($N = 39$): more than 5% data affected.

The three groups were balanced for age and sex, but different for in-scanner motion.

Based on the growing debate related to the best noise-correction technique to apply on resting-state data, we tested two different pipelines, plus one pipeline where no de-noising strategy was applied. We selected and analyzed the results on the following pipelines:

- P0: no motion-correction technique applied beside rigid image realignment carried out with *mcflirt* (Jenkinson et al., 2002);
- FIX: based on the FMRIB trained classifier of Independent Component Analysis, components related to noise (FIX (Salimi-Khorshidi et al., 2014),), extracted from single-subjects time series;
- 9P: a common method that requires the regression of different factors, such as 6 movement parameters, the average signal extracted from white matter and cerebrospinal fluid, plus the regression of the global signal (GSR), measured as the average of all the voxels of the brain extracted from subject-specific brain masks.

Before the regression of all the confound parameters from subjects' time series, a Butterworth bandpass filter of 0.01 and 0.1 Hz was applied to all the regressors, avoiding reintroduction of signal related to nuisance covariates (Lindquist et al., 2019).

Altogether, we specifically selected pipelines based on different principles. One strategy relies on independent components classification (FIX), the second includes the regression of the global signal (9P), a controversial practice. As a reference, for the simple evaluation of pure effects of motion over the architecture of the functional network, we considered a pipeline where only the mandatory image preprocessing steps (realignment, normalization, coregistration, filtering) have been applied (P0).

Differences among groups in terms of connectivity strength were measured by means of simple t-tests. Overall functional connectivity

strength in every network was addressed as the mean of all positive links (van den Heuvel et al., 2017). From an effective pipeline we would expect a reduction in the differences induced by motion in the three groups. At the same time, we would expect that the attenuation of these differences would not alter the topological structure of the functional network. Conversely, an excessively aggressive motion-correction approach may also remove genuine correlations, thus erasing large-scale network structure.

## 4. Results and discussion

### 4.1. Null model fitting

As a first application of the models and framework exposed in the previous sections, in Fig. 4 we report the results of the maximum likelihood estimation of the CWTECM and CWTERG models in the network of healthy subjects obtained from Crossley et al. (2013) and described in section 3.1.

The fully-dense network was thresholded at percolation level, corresponding to a links cut-off value of $t = 0.55$ (12% link-density) as shown in Fig. 4A. The maximization of log-likelihood resulted in the optimal parameters $\hat{\mathbf{u}}, \hat{\mathbf{v}}$ that define the link probability matrix $p_{ij}$ (Eq. (A.29)) in Fig. 4B, and the expected link weights matrix $\langle w_{ij} \rangle$ (Eq. (A.30)) in Fig. 4C. Importantly, panels B and C of Fig. 4 show the ensemble analytical averages, and not numerical averages over a set of samples from the ensemble. In panels B and C of Fig. 4 the model degrees and strengths are computed by summing over the rows of the model link probability matrix and model expected link weights matrix, and are plotted against the real degrees and strengths.

Even though, theoretically, one should find exact correspondence between real and model degree sequences, practical consequences of numerical optimization make the problem hard to solve exactly as a system of $N$ non-linear equations must be solved with very high precision (Garlaschelli and Loffredo, 2008). Hence, as a model convergence diagnostic, we fitted two linear regressions on both degrees and strengths of the real network versus those of the model. The results of the linear regressions are shown in the insets of Panels E and F of Fig. 4. The regression slope of the degrees is close to identity, $a = 0.99$ with a $R^2$ coefficient close to 1 but an intercept $b = -0.232$. On the other hand, the strength reconstruction is more accurate with a regression slope $c \approx 1$, a smaller intercept $d = 0.016$ and a $R^2$ coefficient of one. Ideally, when the log-likelihood maximization is perfectly converging to the global maximum $u_i^\star, v_i^\star$, these linear regression should return slope 1 and intercept 0 with $R^2 \equiv 1$.

Panels G and H of Fig. 4 demonstrate the difference between the optimal models (CTWERG and CWTECM) in terms of spectral entropies curves and relative entropies as a function of the scale parameter β. The spectral entropy of the CTWERG model fits that of the original network only at local scale, and drops rapidly for larger betas. Conversely, the spectral entropy of the optimal CTWECM closely matches its empirical counterpart for a wide range of beta, with the exception of the largest scales. This behaviour indicates that fixing the degree and strength sequence represents a strong constraint, which determines the network structure at the meso-scale. Spectral entropy at large scale reflects the network's modular organization, which cannot be accounted for by local constraints. For visualization purposes, Panel H has two different vertical scales: the relative entropy of the CWTERG model is sixty times larger than the one of the CWTECM.

### 4.2. Effects of thresholding

Here we use these null models within the theoretical framework of spectral entropies to explore the effects of network sparsification on the structure of functional connectivity networks. Specifically, we applied different levels of absolute thresholds to the empirical network and to the models (from $t = 0.10$ to the point where the network breaks apart). Hence, we computed the spectral entropies of thresholded networks and corresponding null models, for different values of β. We then used relative entropy to quantify the information-theoretic distance as a function of threshold. The hypothesis is that the distance between the empirical

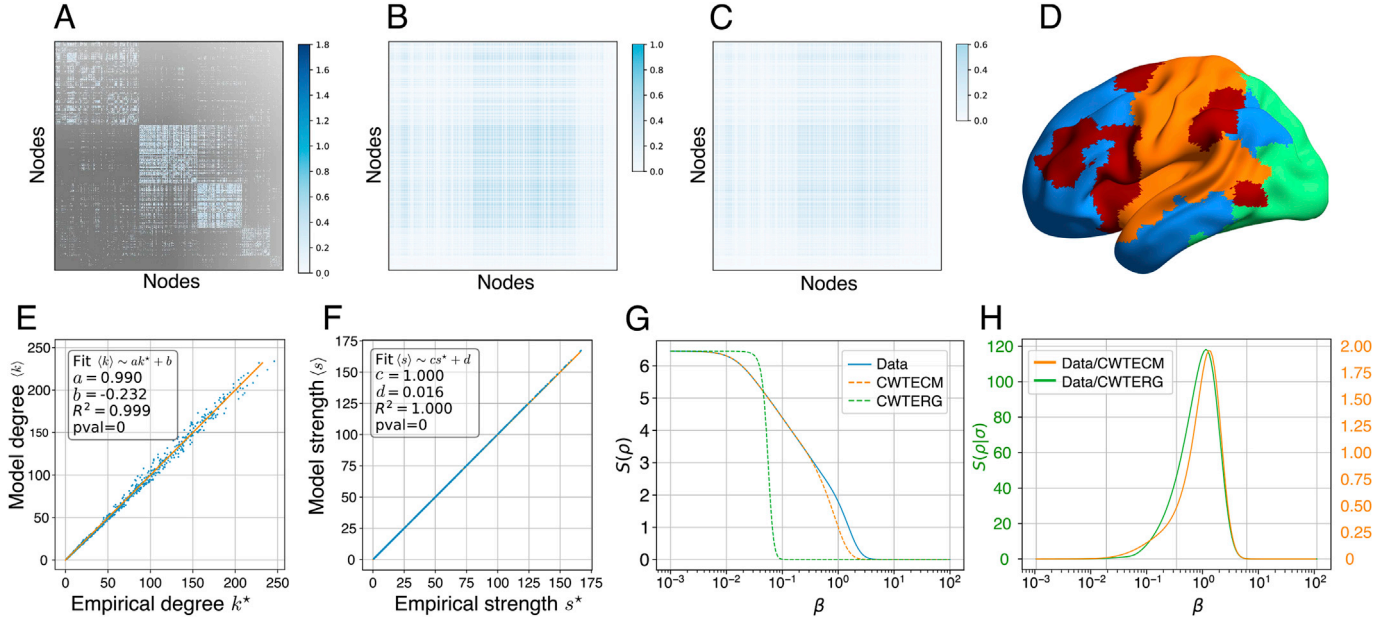


**Fig. 4.** Continuous Weighted Thresholded Enhanced Configuration Model fitted on a real functional network. Panel A: The empirical functional connectivity matrix, thresholded at percolation with rows and columns reorganized by the maximum modularity community structure, to highlight the community structure as overlaid on a brain in Panel D. Panels B,C: the link probability and expected link weights as from Eq. (6). Panels E,F: the reconstructed degrees and strengths as sum over the rows of the link probabilities and expected weights matrices. On the horizontal axis the empirical degrees (strengths), on the vertical axis the model degrees (strengths): $R^2$ and regression slopes are shown as inset. Panel G: the spectral entropy of the network (blue line), CWTECM null model (orange dashed line) and CWTERG null model. Panel H: relative entropy of the network with respect to CWTECM (orange line) and CWTERG (green line). Link probability and expected weights for the CWTERG are not shown, as they are scalar numbers, not matrices.

network and its maximally random counterpart depends on the sparsification threshold, and that there may exist an optimal threshold value that maximizes this distance, striking the optimal balance between removal of spurious correlations and undesirable suppression of structural information.

The results are summarized in Fig. 5, which shows spectral and relative entropies for the empirical network and both null models at various threshold levels. Firstly, we observe that at lower threshold levels (depicted in light blue, Fig. 5) the relative entropy decreases sharply as a function of β, a result of faster diffusion time. Indeed, lower threshold levels correspond to denser networks, and consequently faster diffusion. Decreasing network density results in a right-shift of spectral entropy curves for all networks, as shown in Fig. 5A,C.

However, the effects of thresholding are different in the empirical network and in its null models, a result of the structure of the functional connectivity network that is not accounted for by its randomized counterparts.

Indeed, large-scale structures of the empirical network emerge at higher thresholds (darker blue), as reflected in Fig. 5A by the presence of "information shoulders" i.e., plateaus in the spectral entropy curve. This phenomenon is not equally present in the two null models. Should the thresholding procedure highlight mesoscopic structures only accounted by local constraints, we would expect similar high values of $S$ on both the thresholded random counterparts of the empirical graph. However, the CWTERG shows no indications of a high-level organization at any threshold, as seen by the sharply falling entropy within a very small range of β (Fig. 5C). Indeed, as previously demonstrated, the CWTERG destroys network structure by completely shuffling nodes' neighborhoods. As a result, the diffusion process rapidly spans the whole network, as every node has uniform probability of being connected to every other node.

On the other hand, the spectral entropy of the CWTECM closely corresponds to the one of the empirical network over a broad range of β values. Significant differences only appear at large scales for increasing thresholds. This result shows that degree and strength sequence constrain local and medium-scale structures. In accordance with the results of reference (Cantwell et al., 2019), we observed that the large-scale community structure is the only feature that is not accounted for by local properties.

This same phenomenon is also reported in Fig. 5B,D where the relative entropies are shown for both models. The relative entropy for the CWTERG attains a higher maximum at slightly smaller values of β than for the CWTECM. Intuitively, it takes less time for a random walker to explore a random network than a complex network where modules and local structures may hamper the diffusion process. Moreover, for both cases, relative entropies accentuate the effects of thresholding, as they increase with increasing sparsity level, peaking around the percolation point, just before the network starts breaking apart.

The results in Fig. 5 demonstrate that, in both cases, the maximal spectral entropy difference of the empirical network from its null model is found around the percolation threshold.

Taken together, these results demonstrate the beneficial effects of sparsification to enhance and retrieve the network's modular structure. Importantly, we find that this effect is maximal at percolation threshold, above which the network starts breaking apart into unconnected subunits. At low threshold levels, the empirical network is remarkably close to a random network with similar local features. Only at higher thresholds its meso- and large-scale structure emerges, and difference from the null models become apparent. Interestingly, the presence of an optimal threshold appears to be independent of the specific null model. These results are consistent with previous empirical findings in model networks (Bordier et al., 2018) and provide a theoretical foundation to the use of sparsification methods to study the large-scale topology of functional connectivity networks.

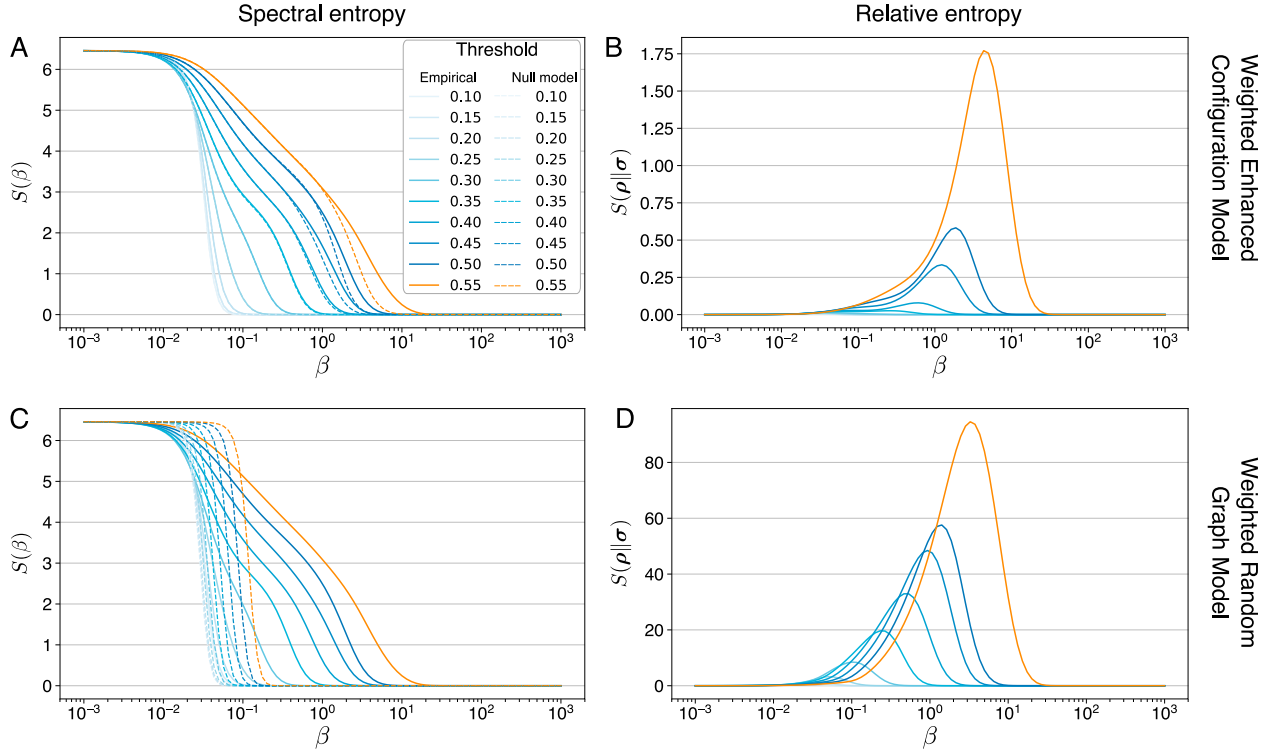


**Fig. 5.** Spectral entropies and relative entropy of the Crossley's functional connectivity network compared to its randomized counterparts. Blue shaded lines represent networks thresholded at absolute values from 0.1 to 0.5. Orange lines denote the network at percolation. Solid lines are referred to the empirical network. Dashed lines are referred to the randomized models. Panels A,B show the results with respect to the CWTECM model. Panels C,D show the results with respect to the CWTERG model.

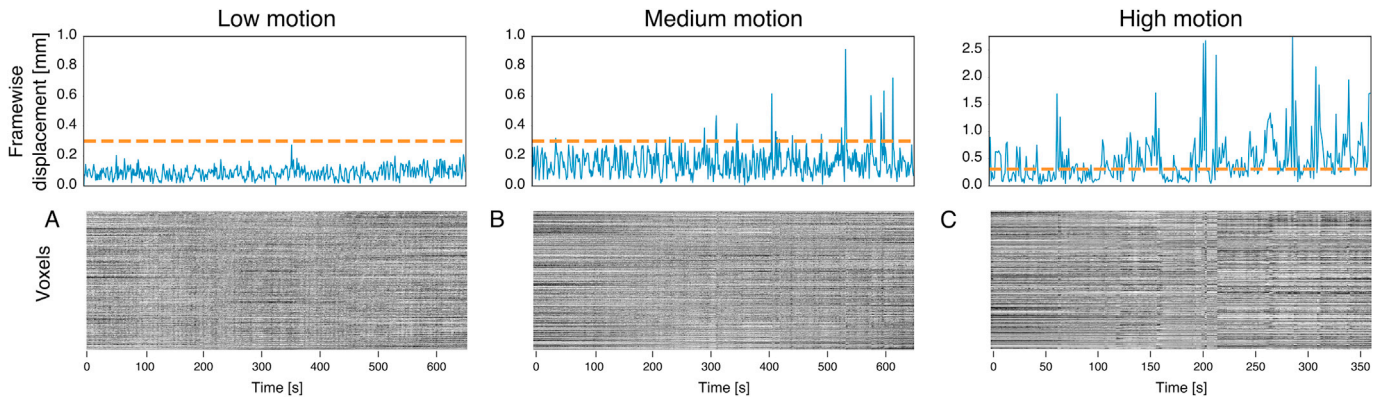


**Fig. 6.** Examples of the effects of head movements over resting-state fMRI signal. Upper panels show the Framewise Displacement (FD) of the three motion groups. Lower panels show the so-called gray plots of corresponding fMRI scans (see Power et al. (2012)), representing the signal intensity of all voxels in the brain (randomly ordered) over time. The orange line at FD 0.3 mm represents the limit of outliers.

### 4.3. Effects of motion and motion correction

We have also applied the concept of an information-based comparison with null models to investigate the effects of motion and motion correction pipelines typically used in fMRI studies. A major debate in the resting-state functional connectivity community concerns the effects that different preprocessing pipelines can have over the functional time series (Power et al., 2012; Ciric et al., 2017). Here, our goal is to investigate to what extent motion could render the network more or less similar to its random counterpart, and the efficacy of different pipelines and thresholding procedures in mitigating the effects of motion.

We used the same atlas with 638 parcels that was used in the previous examples, and we applied the same model fitting and comparison techniques to resting-state dataset with different degrees of motion and different motion-correction techniques (see Data and Preprocessing, section 3). Specifically, we considered three pipelines: P0 with no motion correction, a second pipeline based on FIX (Salimi-Khorshidi et al., 2014), and 9P, a pipeline that includes global signal regression (Ciric et al., 2017). We applied these pipelines on three motion groups: low, medium, and high-motion (see Data and Preprocessing, section 3). These three groups are defined on the basis of Framewise Displacement (FD), a metric commonly used to evaluate the amount of head motion in rsFC (Power et al., 2011), that is computed as the sum of the absolute values of the derivatives of the six motion parameters. Power and colleagues (Power et al., 2014) showed that even small head movements (FD $>$ 0.15 mm) could cause significant changes affecting all voxels in the brain. These movements can be visually identified utilizing the so called *grayplots*, as shown in Fig. 6. Grayplots depict the signal intensity of every voxel in the brain over time. Here, we report three examples of how the magnitude of head movements impacts the whole time series. In Fig. 6 one can appreciate the effects of motion as abrupt changes of voxel intensities in correspondence of head movements. These artifacts give rise to spurious correlations at different scales.

The three groups here evaluated are perfectly balanced for age and gender, with exactly the same acquisition procedures, but different for in-scanner motion. Thus, at the group level, we would expect these participants to share same global functional connectivity characteristics. Possible differences should be driven by the presence of motion.

Firstly, we evaluated the effects of motion on functional connectivity strength. In line with previous reports (Ciric et al., 2017), we observed a substantial increase in functional connectivity induced by motion in P0 (Fig. 7). Fig. 7A shows that the distribution of link weights for the medium and high-motion groups is right-shifted compared to the low groups, reflecting higher correlation strength. At the subject level this shift is highly significant across all groups (medium $>$ low: $p < 10^{-5}$; high $>$ low: $p < 10^{-4}$). Through the application of specific denoising strategies we sought to investigate to what degree this spurious difference in functional strength between groups can be reduced. The histograms depicted in panels B and C of Fig. 7 show that both pipelines appear to significantly decrease the differences in edge-weight distribution at the group level across different motion conditions. Specifically, the pipeline based on independent components classification (FIX) substantially reduces the right shift of the medium and high-motion groups that was apparent with P0. Importantly, in this condition the edge-weight distribution of the medium-motion group almost completely overlaps with the low-motion curve. Yet, at the individual level the functional connectivity strength, measured as the mean of all positive edges in the graph, shows statistical difference (medium $>$ low: $p = 0.007$). In contrast, the histogram representing the edge-weight distribution extracted from the high-motion group still presents a highly significant right shift reflecting higher functional strength (high$>$low: $p < 10^{-3}$). A different pattern is revealed in the strength distribution after the application of GSR. In this case, all curves are highly overlapping, indicating similar functional connectivity across the three groups. This is revealed by the lack of significant differences at the individual level in edge strength (high $>$ low: $p = 0.8$; medium $>$ low: $p = 0.48$). As a matter of concern, all three distributions are centered around zero. Indeed, after GSR the number of negative correlations dramatically increases, involving almost half of the edges within the network. These observations are in line with previous reports and concerns related to the controversial application of this denoising approach (Ciric et al., 2017).

In Fig. 8 we show the spectral entropy curves and relative entropies for the three pipelines considered. In light of the previous findings, we present here only the results related to the CWTERG, as the constraints imposed by the CWTECM also affect larger scales, and may reduce sensitivity to the effects of motion at a mesoscopic level. Panels A,B and C show the von Neumann entropy curves of the differently pre-processed resting-state networks across three degrees of motion. In this specific case, we applied a single threshold, namely the lowest absolute threshold that guarantees connectedness in all three motion groups within the same pipeline ($t = 0.44$, $d = 20\%$ for P0; $t = 0.29$, $d = 7\%$ for GSR; $t = 0.25$, $d = 15\%$ for FIX). We can observe that, when considered at the same absolute threshold, the low-motion group always shows higher spectral entropy across the entire $\beta$ range. This is especially evident in the P0 pipeline. It appears, indeed, that movement artifacts significantly affect the mesoscopic patterns within the empirical network. This trend is confirmed by the smaller entropy values of the high-motion group compared to the medium and low-motion groups, which is observed across all analysis pipelines. In Fig. 8A, this point is further highlighted by the lack of a clear-cut modular structure in both the medium and high-motion groups, whereas a shoulder present at medium scales for the low-motion group denotes a different degree of inter-modular density. A similar trend suggests that head movements tend to make the network closer to its null model, i.e. more random. Popular correction techniques

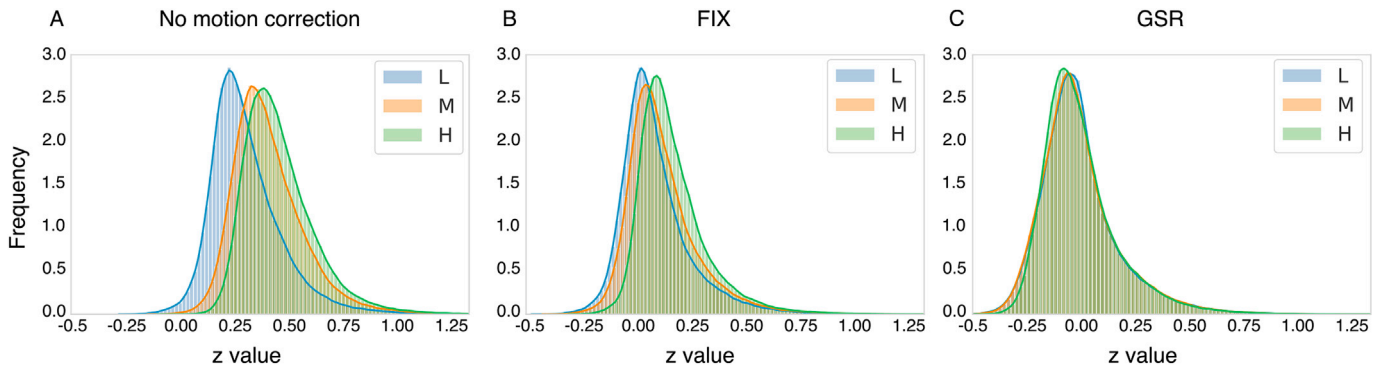


**Fig. 7.** Effects of motion and commonly applied motion-correction techniques over the distribution of functional connectivity strength. Panel A depicts effects of motion as assessed by means of a pipeline where no motion correction strategies has been applied (P0). As a consequence of motion, we observe strong changes in the functional connectivity strength across the three groups (medium > low: $p < 10^{-4}$; high ≥ low: $p < 10^{-5}$). Panel B represents the effects of the application of FIX over edge-weight distribution. Differences among groups are still present but attenuated (medium > low: $p = 0.007$; high > low = $p < 10^{-3}$). Panel C shows the effects of application of GSR. Differences among groups are not present (medium > low: $p = 0.48$; high > low: p = 0.8).

mitigate this confounding effect, decoupling functional connectivity and motion.

Panels B,C of Fig. 8 show spectral entropy curves for the pipelines FIX and GSR, respectively. As already discussed, both pipelines importantly reduce the difference in spectral entropy between the three groups. It is noteworthy that the application of the FIX pipeline in Fig. 8B highlights the manifest presence of more prominent shoulders in the entropy curve in all groups, again a signature of mesoscopic organization. From this view, the cleanup of resting-state data through independent component analysis appears to emphasize the global structure of the network, in presence of head movements. The same cannot be appreciated in the groups preprocessed with a GSR pipeline: differences in spectral entropies are reduced, but no clear large-scale structure seems to emerge from these curves.

From the relative entropies generated over several different absolute thresholds, we can appreciate a strong effect related to the sparsification procedure. In Panels D, E, F of Fig. 8, the relative entropies of the high-motion groups for all pipelines and their respective null models are presented. Here, we report the high-motion group, which is more affected by head movements and shows more evidently the beneficial effects of the application of different preprocessing pipelines and thresholds. Additional analyses are reported in the supplementary materials.

With the application of increasing thresholds, the distance of the empirical network from its random counterpart with same density gets larger, and it reaches maximum at percolation, despite the presence of motion and independently of the pipeline applied. Specifically, we can observe a higher relative entropy at percolation for the pipeline based on GSR (Fig. 8F). This preprocessing technique notably benefits from the thresholding procedure, considering the substantial difference between the maximum relative entropy attained at percolation and its values for denser networks. In line with previous studies (Ciric et al., 2017), the main effect of GSR is an increase in network modularity, mirrored by greater values of relative entropies at large scales, suggesting a well organized high-order architecture. Yet, the lack of an "information shoulder" in the spectral entropy curve suggests the presence of a more uniform structure, with similar intra-modular density across different communities, and similar size of the modules. Importantly, we note that thresholding emphasizes meso- and large-scale structure in combination with FIX (Fig. 8F), but also in the absence of any motion correction (Fig. 8D). Indeed, network sparsification appears to have a large effect per se, even for P0, in separating the empirical network form its null model.

Interestingly, relative entropy between empirical network and null model increases with the threshold, and reaches its maximum at percolation. This last observation further supports the application of a

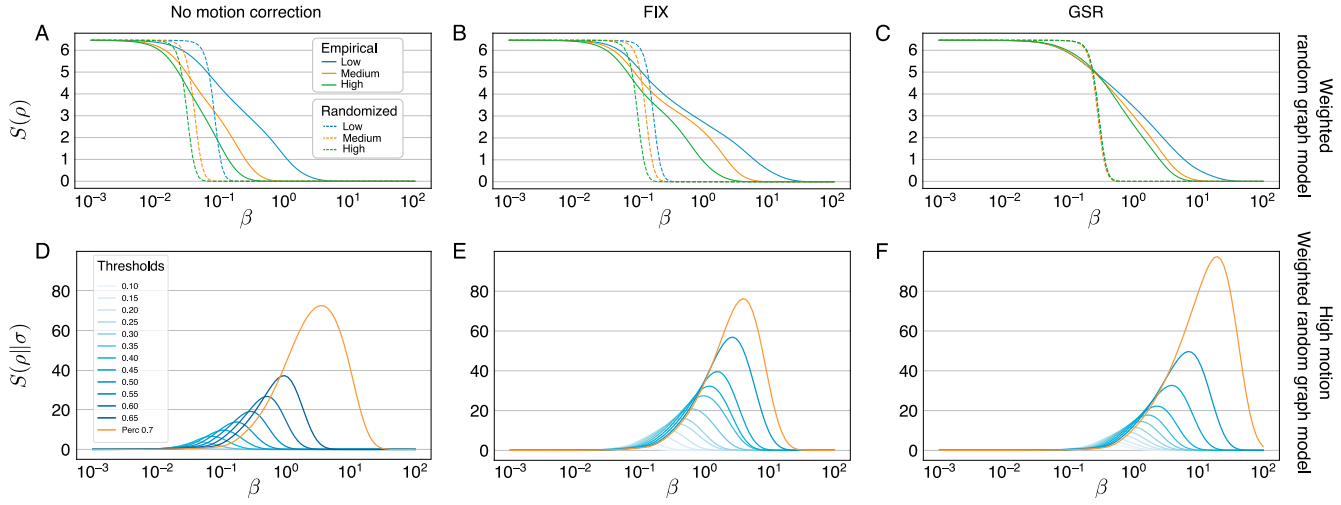


**Fig. 8.** Panels A,B,C show the spectral entropies of networks for the pipelines 0,FIX and GSR (solid lines), together with their randomized counterpart (CWTERG, dashed lines) over all motion groups. The relative entropies of networks from the high-motion group are shown in panels D, E, F where the blue shades correspond to increasing absolute thresholds, while the orange lines correspond to percolation threshold, which has the maximum relative entropy at large scales.

thresholding procedure, in contrast with recent literature that suggests that sparsification should be avoided (Schlesinger et al., 2017; Goulas et al., 2015; Bassett et al., 2011; Rubinov and Sporns, 2011). Our results demonstrate the importance of application of a threshold, irrespectively of the pre-processing pipeline. While we do not draw any strong conclusion on the relative efficacy of GSR and FIX, we note that network sparsification is advantageous regardless of the choice of denoising pipeline.

## 5. Conclusion

The nature of resting state functional MRI networks based on pairwise association measures, like the Pearson correlation, is of a dense square matrix. Several experimental factors are involved in shaping the properties of these matrices and no consensus exists in the literature on the best practice for the definition and processing of these matrices and the associated connectivity graphs. In the present work we have introduced a novel theoretical framework that contributes to shed light on several open issues in the analysis of brain functional connectivity networks. Firstly, we define the CWTERG and CWTECM null models, which enable extension of the maximum entropy random graph formalism to networks with threshold and real positive weights, as those encountered in fMRI. Secondly, we have shown that the spectral entropies framework can be applied to the differences of networks with respect to their random versions from local to global scales.

Leveraging this new approach, we studied the effects of thresholding procedures and motion-correction pipelines. The application of a threshold to resting-state networks is a contentious step debated in the field.

Here, by means of advanced information theory tools, we found that complete functional connectivity networks present a high degree of randomness, due to the contribution of spurious correlations to weak links, that conceals their large scale structure. Sparsification of the network is an essential step to differentiate networks from their null model and is highly beneficial to study the large scale architecture of real-world networks.

Further, we demonstrated from first principles the existence of an optimal thresholding point, where the empirical network is maximally distant from random. Specifically, we found that application of a percolation threshold strikes the optimal balance between the removal of spurious connections and genuine information, thus maximizing the information that can be extracted from the system. The importance of sparsification can also be appreciated through the evaluation of the effects of motion and different preprocessing pipelines.

Motion increases randomness and reduces spectral entropies across the whole $\beta$ domain, bringing the network closer to its random counterpart. The effects of motion are mitigated by popular motion-correction approaches. However, we found that network sparsification has a beneficial effect irrespectively of the specific denoising strategy applied, and that the percolation threshold maximizes the distance of the empirical network from its randomized counterpart.

Hence, as an important and practical remark, we suggest that the application of a percolation threshold is critical for the extraction of the large-scale structure from the network.

Finally, the novel framework of spectral maximum entropy networks provides a new and powerful approach that significantly extends the repertoire of tools for the study of functional connectivity networks at multiple scales.

## Author contributions

C.N. developed the mathematical methods and software. C.N., G.F analyzed the data. G.F. collected and processed the neuroimaging data. C.N., G.F., L.M., A.B. wrote the paper. All authors reviewed the manuscript.

## Declaration of competing interests

The authors declare no competing financial interests.

## Acknowledgments

The authors wish to thank Dr. Tiziano Squartini and Prof. Diego Garlaschelli for useful methodological discussions. The authors also thank Prof. Edward Bullmore and Prof. Nicholas Crossley for providing data and templates. This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under grant agreement 668863, SyBil-AA. L. Minati gratefully acknowledges funding by the World Research Hub Initiative (WRHI), Institute of Innovative Research (IIR), Tokyo Institute of Technology, Tokyo, Japan.

## Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2020.116603.

## Appendix A. Theoretical framework

In this theoretical section we introduce the notation and describe the formalism of maximum entropy random graph models that is central to this manuscript. We summarize here a few definitions that are necessary to make this paper self-contained.

We consider undirected weighted graphs $G = (V, E)$ with $|V| = N$ number of nodes, $|E| = L$ number of links and $W$ total edge weight. We denote the weighted adjacency matrix as $\mathbf{W} = \{w_{ij}\}$, the binary adjacency matrix $\mathbf{A} = a_{ij}$ and the weighted graph Laplacian as $\mathbf{L} = \mathbf{D} - \mathbf{W}$, where $\mathbf{D}$ is a diagonal matrix of the node strengths. We indicate the degree and strength as $k_i = \sum_{i \neq j} a_{ij}$ and $s_i = \sum_{i \neq j} w_{ij}$, respectively. The Heaviside step function is indicated as $\Theta(x)$.

### A.1. Exponential Random Graph Models

We let $\mathbf{G}$ denote a network in a random graph ensemble $\mathscr{G}$, and $\mathbf{G}^\star$ an observed empirical network. The ensemble $\mathscr{G}$ consists of all networks with the same number of nodes $N$ and of the same type (undirected, weighted etc.) as $\mathbf{G}^\star$, including $\mathbf{G}^\star$ itself. Our goal is to find an analytical description of the random graphs $\mathbf{G}$ that share the same network descriptors of $\mathbf{G}^\star$, and to eventually be able to sample networks from the ensemble. In other words, we look for the functional form of the probability distribution $P(\mathbf{G})$ over the ensemble $\mathscr{G}$, for which the values of descriptors are on average as close as possible to those of the empirical network.

We denote the chosen descriptors by an array $\mathbf{C}^\star = \mathbf{C}(\mathbf{G}^\star)$ with $c$ elements. The array contains network-related quantities, like the number of links, the total weight, or the node and strength sequences, and is instrumental in shaping the analytic form of the ensemble described by $P(\mathbf{G})$. Depending on the constraints set by $\mathbf{C}^\star$ one gives rise to different random graph models. By standard probability arguments, the expected value of the descriptors $\mathbf{C}(\mathbf{G})$

over the ensemble $\mathscr{G}$ are found as:

$$\langle \mathbf{C}(\mathbf{G}) \rangle = \int_{\mathbf{G} \in \mathscr{G}} \mathbf{C}(\mathbf{G}) P(\mathbf{G}). \tag{A.1}$$

The functional form of $P(\mathbf{G})$ can be obtained by Shannon entropy maximization subjected to the constraints represented by $\mathbf{C}$, where Shannon entropy is defined as

$$S(P) = -\int_{\mathbf{G} \in \mathscr{G}} P(\mathbf{G}) \log P(\mathbf{G}) \tag{A.2}$$

as we are integrating over a continuous probability distribution function, rather than on a discrete one. The procedure to obtain $P(\mathbf{G})$ is rooted in Jaynes's Maximum Entropy formalism (Jaynes, 1957), a very general statistical mechanics principle that here leads to exact expressions for the probability of occurrence of a vast majority of graph models. The general idea is that, in order to maximize Eq. (A.2) with a series of constraints, which read $\mathbf{C}^{\star} = \langle \mathbf{C} \rangle$, one has to build a Lagrangian functional $\mathfrak{L}$ of the type that follows:

$$\mathfrak{L}(P) = S(P) - \alpha \left( 1 - \int_{\mathbf{G} \in \mathscr{G}} P(\mathbf{G}) \right) - \boldsymbol{\theta} \cdot (\mathbf{C}^{\star} - \langle \mathbf{C} \rangle) \tag{A.3}$$

where the term multiplied by α is the probability normalization condition, and the remaining terms measure the deviation of the imposed ensemble properties $\langle \mathbf{C} \rangle$ from the desidered ones $\mathbf{C}^{\star}$, each weighted by the lagrange multiplier $\theta_i$ for $i = 1, \ldots, c$. A standard derivation based on finding the stationary point of the Lagrangian functional of Eq. (A.3) (see Squartini and Garlaschelli (2017); Park and Newman (2004)), shows that the solution of constrained entropy maximization problem is found is the resulting conditional probability:

$$P(\mathbf{G}|\boldsymbol{\theta}) = \frac{e^{-H(\mathbf{G},\, \boldsymbol{\theta})}}{Z(\boldsymbol{\theta})} \tag{A.4}$$

where $H(\mathbf{G}, \boldsymbol{\theta})$ is the graph Hamiltonian, defined as a linear combination of constraints:

$$H(\mathbf{G}, \boldsymbol{\theta}) = \sum_{i=1}^{c} \theta_i C_i(\mathbf{G}) = \boldsymbol{\theta} \cdot \mathbf{C}(\mathbf{G}) \tag{A.5}$$

and the denominator $Z(\boldsymbol{\theta})$ is a normalizing quantity called *partition function*, defined by marginalization over all networks $\mathbf{G}$ in the ensemble $\mathscr{G}$:

$$Z(\boldsymbol{\theta}) = \int_{\mathbf{G} \in \mathscr{G}} e^{-H(\mathbf{G}, \boldsymbol{\theta})}. \tag{A.6}$$

The above results show that the graph probability $P(\mathbf{G}|\boldsymbol{\theta})$ depends on the Lagrange multipliers $\boldsymbol{\theta}$, and that it is a function of the constraints considered.

For model fitting purpose, it can be shown (Squartini and Garlaschelli, 2017) that the log-likelihood, obtained as the logarithm of $P(\mathbf{G})$, that reads

$$\mathscr{L}(\boldsymbol{\theta}) = \log P(\mathbf{G}^{\star}|\boldsymbol{\theta}) = -H(\mathbf{G}^{\star}|\boldsymbol{\theta}) - \log Z(\boldsymbol{\theta}) \tag{A.7}$$

is maximized by the particular value $\widehat{\boldsymbol{\theta}}$ such that the ensemble average $\langle \mathbf{C} \rangle_{\boldsymbol{\theta}^{\star}}$ of each constraint equals the empirical value $\mathbf{C}(\mathbf{G}^{\star})$ measured on the real network:

$$\langle \mathbf{C} \rangle^{\star} = \int_{\mathbf{G} \in \mathscr{G}} \mathbf{C}(\mathbf{G}) P(\mathbf{G}|\boldsymbol{\theta}^{\star}) = \mathbf{C}(\mathbf{G}^{\star}) \tag{A.8}$$

and the gradients of $\mathscr{L}(\boldsymbol{\theta})$ are proportional to the difference between $\mathbf{C}^{\star}$ and $\langle \mathbf{C} \rangle$. For maximum-entropy ensembles, the maximum likelihood principle indicates the choice of parameters that meet the constraints, and defines a procedure for model fitting: either by maximizing the log-likelihood from Eq. (A.7) by means of gradient based numerical optimization methods (Nocedal and Wright, 2006), or alternatively by solving the system of nonlinear equations defined by Eq. (A.8).

Importantly, the average value of the constraints at the optimal parameters estimate $\widehat{\boldsymbol{\theta}}$ can be obtained by taking the gradient with respect of the associated Lagrange multiplier, of the negative logarithm of the partition function. In other words, one has:

$$\langle \mathbf{C} \rangle = -\nabla_{\boldsymbol{\theta}} \log Z(\boldsymbol{\theta})| \tag{A.9}$$

In the following, we show a practical application of this approach to the null models in the main text and used for the description of resting-state brain connectivity networks.

*Appendix A.2 Fixing average number of links and total weight: CWTERG*

Here we build the ensemble of networks with the same average number of links, the same average total weight, and a threshold parameter $t$, which, differently from other methods, considers positive real link weights.

Following Eq. (A.5), we commence by the definition of the constraints defining the Hamiltonian relevant to this model. The number of links of an undirected network, with an external control parameter $t$, is obtained by summing all elements of the upper-diagonal part of the binary adjacency matrix $\Theta(w_{ij} - t)$, as follows:

$$L = \sum_{i<j}^{N} \Theta\left(w_{ij} - t\right), \tag{A.10}$$

and similarly we obtain to the total weight:

$$W = \sum_{i<j}^{N} w_{ij}\Theta\left(w_{ij} - t\right). \tag{A.11}$$

The two network metrics that we want to constrain on the ensemble $P(\mathbf{G})$ are specified by the two above equations. Additionally, the factor $\Theta(w_{ij} - t)$ makes it possible to incorporate the threshold into the network metrics, ignoring all weights below $t$. By looking at Eq. (A.5), it is easy to write the graph Hamiltonian as the scalar product of the lagrangian multipliers $\boldsymbol{\theta} = (u, v)$ with the associated properties $(L, W)$. This reads:

$$H_{\mathrm{CWTERG}} = uL(\mathbf{G}) + vW(\mathbf{G}) = \sum_{i<j} u\Theta\left(w_{ij} - t\right) + vw_{ij}\Theta\left(w_{ij} - t\right). \tag{A.12}$$

In order to compute the partition function, we plug the Hamiltonian $H_{\mathrm{CWTERG}}$ in the exponential and integrate, as in Eq. (A.6), over all weights $w'_{ij}$ from 0 to $+\infty$:

$$Z_{\mathrm{CWTERG}} = \int_{\mathbf{G}\in\mathscr{G}} e^{-H_{\mathrm{CWTERG}}(\mathbf{G},\boldsymbol{\theta})} \tag{A.13}$$

$$= \int_{0}^{\infty} e^{-\left[\sum_{i<j} u\Theta\left(w_{ij}'-t\right)+vw_{ij}'\Theta\left(w_{ij}'-t\right)\right]} dw'_{ij} \tag{A.14}$$

$$= \int_{0}^{\infty} \prod_{i<j} e^{-u\Theta\left(w'_{ij}-t\right)-vw'_{ij}\Theta\left(w'_{ij}-t\right)} dw'_{ij},$$

$$= \prod_{i<j} \int_{0}^{\infty} e^{-u\Theta(w'-t)-vw'\Theta(w'-t)} dw'$$

$$= \left(t + e^{-u}\int_{0}^{\infty} e^{-vw'\Theta(w'-t)dw'}\right)\binom{n}{2} \tag{A.15}$$

$$= \left(t + \frac{e^{-u-vt}}{v}\right)\binom{n}{2}$$

Following Equation (A.9) we then obtain two analytical formulae that describe, in terms of the lagrangian multipliers $u$ and $v$, the expected number of links and total weights as:

$$(\langle L\rangle, \langle W\rangle) = -\nabla_{(u,v)} \log Z_{\mathrm{CWTERG}} \tag{A.16}$$

with the following result:

$$\frac{\partial F}{\partial u} = \langle L\rangle = \binom{n}{2}\frac{1}{vte^{u+vt}+1} \tag{A.17}$$

$$\frac{\partial F}{\partial v} = \langle W\rangle = \binom{n}{2}\frac{vt+1}{v(vte^{u+vt}+1)} \tag{A.18}$$

where we have defined $F = -\log Z_{\mathrm{CWTERG}}$, as in Eq. (A.9). Based on these two results, we indicate the link density and the expected link weights of the CWTERG model respectively as:

$$p_{\mathrm{CWTERG}} = \frac{1}{vte^{u+vt}+1} \tag{A.19}$$

$$\langle w\rangle_{\mathrm{CWTERG}} = \frac{vt+1}{v(vte^{u+vt}+1)}. \tag{A.20}$$

Finally, the probability density function of the CWTERG model is obtained as:

$$P(\mathbf{G}|u,v) = \frac{e^{-uL(\mathbf{G})-vW(\mathbf{G})}}{\left(t+\frac{e^{-u-vt}}{v}\right)\binom{n}{2}}. \tag{A.21}$$

This probability distribution function completely defines the model.

*Appendix A.3 Fixing degree and strength sequence: CWTECM*

Here we look for the ensemble of random networks with the same degree sequence and strength sequence. The degree of node $i$ is obtained from the adjacency matrix with an external threshold parameter $t$ as:

$$k_i = \sum_{j\neq i}\Theta(w_{ij}-t) \tag{A.22}$$

and similarly for the strenght of node $i$:

$$s_i = \sum_{j\neq i} w_{ij}\Theta(w_{ij}-t). \tag{A.23}$$

The appropriate Hamiltonian for this problem is then:

$$\begin{aligned} H_{\mathrm{CWTECM}} &= \sum_{i=1} u_i k_i + v_i s_i \\ &= \sum_{i,j} u_i\Theta(w_{ij}-t) + v_i w_{ij}\Theta(w_{ij}-t) \\ &= \sum_{i<j}(u_i+u_j)\Theta(w_{ij}-t) + (v_i+v_j)w_{ij}\Theta(w_{ij}-t). \end{aligned} \tag{A.24}$$

With this Hamiltonian, the CWTECM is completely described by the following probability density function $P(\mathbf{G}|\mathbf{u},\mathbf{v})$:

$$P(\mathbf{G}|\mathbf{u},\mathbf{v}) = \prod_{i<j}\frac{e^{-\left[(u_i+u_j)+w_{ij}(v_i+v_j)\right]\theta(w_{ij}-t)}}{Z_{\mathrm{CWTECM}}} \tag{A.25}$$

The partition function is calculated similarly to the one of the CWTERG: every occurrence of $u$ and $v$ is replaced by $(u_i+u_j)$ and $(v_i+v_j)$, respectively, and the product is done over all the undirected pairs $i<j$:

$$Z_{\mathrm{CWTECM}} = \prod_{i<j}\left(t+\frac{e^{-(u_i+u_j)-(v_i+v_j)t}}{v_i+v_j}\right). \tag{A.26}$$

Following the same line as for the CWTERG model, we also compute the expected degrees and strengths by taking the derivatives with respect to the corresponding lagrange multipliers $(u_i,v_i)$ of the negative log-partition function. We find:

$$\frac{\partial F_{\mathrm{CWTECM}}}{\partial u_i} = \langle k_i\rangle = \sum_{j\neq i} p_{ij} \tag{A.27}$$

$$\frac{\partial F_{\mathrm{CWTECM}}}{\partial v_i} = \langle s_i\rangle = \sum_{j\neq i}\langle w_{ij}\rangle. \tag{A.28}$$

where the link-probability matrix $p_{ij}$ and the expected link weight matrix $\langle w_{ij}\rangle$ are defined as:

$$p_{ij} = \frac{1}{1+(v_i+v_j)te^{(u_i+u_j)+(v_i+v_j)t}} \tag{A.29}$$

$$\langle w_{ij}\rangle = \frac{(v_i+v_j)t+1}{(v_i+v_j)\left[(v_i+v_j)te^{(u_i+u_j)+(v_i+v_j)t}+1\right]} \tag{A.30}$$

Finally, the log-likelihood of the model is simply obtained by the logarithm of the probability density function $P(\mathbf{G}|\mathbf{u},\mathbf{v})$ as:

$$\log\mathscr{L}(\mathbf{G}|\mathbf{u},\mathbf{v}) = -\left[\sum_i u_i k_i(\mathbf{G}) + v_i s_i(\mathbf{G})\right] - \sum_{i<j}\log\left(t+\frac{e^{-(u_i+u_j)-(v_i+v_j)t}}{v_i+v_j}\right) \tag{A.31}$$

## References


Anand, K., Bianconi, G., 2009. Entropy measures for networks: toward an information theory of complex topologies. Phys. Rev. E 80 (4), 045102. https://doi.org/10.1103/PhysRevE.80.045102.

Anderson Jr., W.N., Morley, T.D., 1985. Eigenvalues of the laplacian of a graph. Linear Multilinear Algebra 18 (2), 141–145.

Aquino, K.M., Fulcher, B.D., Parkes, L., Sabaroedin, K., Fornito, A., 2019. Identifying and removing widespread signal deflections from fmri data: rethinking the global signal regression problem. bioRxiv, 662726.

Bassett, D.S., Bullmore, E., 2006. Small-world brain networks. Neuroscientist 12 (6), 512–523.

Bassett, D.S., Wymbs, N.F., Porter, M.A., Mucha, P.J., Carlson, J.M., Grafton, S.T., 2011. Dynamic reconfiguration of human brain networks during learning. Proc. Natl. Acad. Sci. U.S.A. 108 (18), 7641–7646. https://doi.org/10.1073/pnas.1018985108/-/DCSupplemental.

Betzel, R.F., Bassett, D.S., 2017. Generative models for network neuroscience: prospects and promise. J. R. Soc. Interface 14 (136), 20170623.

Blondel, V.D., Guillaume, J.l., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of communities in large networks. J. Stat. Mech. Theor. Exp. (10), P10008 https://doi.org/10.1088/1742-5468/2008/10/P10008, 2008.

Bordier, C., Nicolini, C., Bifone, A., 2017. Graph analysis and modularity of brain functional connectivity networks: searching for the optimal threshold. Front. Neurosci. 11 https://doi.org/10.3389/fnins.2017.00441.

Bordier, C., Nicolini, C., Forcellini, G., Bifone, A., 2018. Disrupted modular organization of primary sensory brain areas in schizophrenia. Neuroimage: Clinica 18, 682–693. https://doi.org/10.1016/j.nicl.2018.02.035.

Bray, A.J., Rodgers, G.J., 1988. Diffusion in a sparsely connected space: a model for glassy relaxation. Phys. Rev. B 38 (16), 11461–11470.

Burgess, G.C., Kandala, S., Nolan, D., Laumann, T.O., Power, J.D., Adeyemo, B., Harms, M.P., Petersen, S.E., Barch, D.M., 2016. Evaluation of denoising strategies to address motion-correlated artifacts in resting-state functional magnetic resonance imaging data from the human connectome project. Brain Connect. 6 (9), 669–680.

Cantwell, G., Liu, Y., Maier, B.F., Schwarze, A.C., Serván, C.A., Snyder, J., St-Onge, G., 2019. Thresholding Normally Distributed Data Creates Complex Networks arXiv preprint arXiv:190208278.

Cheeger, J., 1969. A lower bound for the smallest eigenvalue of the laplacian. In: Proceedings of the Princeton Conference in Honor of Professor S. Bochner, pp. 195–199.

Cimini, G., Squartini, T., Saracco, F., Garlaschelli, D., Gabrielli, A., Caldarelli, G., 2019. The statistical physics of real-world networks. Nat Rev Phys 1 (1), 58.

Ciric, R., Wolf, D.H., Power, J.D., Roalf, D.R., Baum, G.L., Ruparel, K., Shinohara, R.T., Elliott, M.A., Eickhoff, S.B., Davatzikos, C., et al., 2017. Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. Neuroimage 154, 174–187.

Crossley, N.A., Mechelli, A., Vertes, P.E., Winton-Brown, T.T., Patel, A.X., Ginestet, C.E., McGuire, P., Bullmore, E.T., 2013. Cognitive relevance of the community structure of the human brain functional coactivation network. Proc. Natl. Acad. Sci. U.S.A. https://doi.org/10.1073/pnas.1220826110.

Crossley, N.A., Mechelli, A., Vertes, P.E., Winton-Brown, T.T., Patel, A.X., Ginestet, C.E., McGuire, P., Bullmore, E.T., 2016. https://github.com/CarloNicolini/communitya lg/raw/master/data/GroupAverage_rsfMRI_unthr.adj.

Van Dijk, K.R., Sabuncu, M.R., Buckner, R.L., 2012. The influence of head motion on intrinsic functional connectivity mri. Neuroimage 59 (1), 431–438.

De Domenico, M., Biamonte, J., 2016. Spectral entropies as information-theoretic tools for complex network comparison. Phys. Rev. X 6, 041062. https://doi.org/10.1103/PhysRevX.6.041062.

De Domenico, M., Nicosia, V., Arenas, A., Latora, V., 2015. Structural reducibility of multilayer networks. Nat. Commun. 6, 6864. https://doi.org/10.1038/ncomms7864.

Donetti, L., Neri, F., Muñoz, M.A., 2006. Optimal network topologies: expanders, cages, ramanujan graphs, entangled networks and all that. J Stat Mech-Theory E, P08007, 2006(08).

Esfahlani, F.Z., Sayama, H., 2018. A percolation-based thresholding method with applications in functional connectivity analysis. In: International Workshop on Complex Networks. Springer, pp. 221–231.

Estrada, E., 2011. The Structure of Complex Networks: Theory and Applications. Oxford University Press, Inc., New York, NY, USA.

Fox, M.D., Zhang, D., Snyder, A.Z., Raichle, M.E., 2009. The global signal and observed anticorrelated resting state brain networks. J. Neurophysiol. 101 (6), 3270–3283.

Gallos, L.K., Makse, H.a., Sigman, M., 2012. A small world of weak ties provides optimal global integration of self-similar modules in functional brain networks. Proc. Natl. Acad. Sci. U. S. A. 109 (8), 2825–2830.

Garlaschelli, D., Loffredo, M.I., 2008. Maximum likelihood: extracting unbiased information from complex networks. Phys. Rev. E 78, 015101. https://doi.org/10.1103/PhysRevE.78.015101.

Golub, G.H., Van Loan, C.F., 1996. Matrix Computations, third ed. Johns Hopkins University Press, Baltimore, MD, USA.

Goulas, A., Schaefer, A., Margulies, D.S., 2015. The strength of weak connections in the macaque cortico-cortical network. Brain Struct. Funct. 220 (5), 2939–2951. https://doi.org/10.1007/s00429-014-0836-3.

Guimera, R., Sales-Pardo, M., Amaral, L.A.N., 2004. Modularity from fluctuations in random graphs and complex networks. Phys. Rev. E 70 (2), 025101. https://doi.org/10.1103/PhysRevE.70.025101.

Hallquist, M.N., Hillary, F.G., 2019. Graph theory approaches to functional network organization in brain disorders: a critique for a brave new small-world. Netw Neurosci 3 (1), 1–26. https://doi.org/10.1162/netn_a_00054.

Henderson, J.A., Robinson, P.A., 2011. Geometric effects on complex network structure in the cortex. Phys. Rev. Lett. 107 (1) https://doi.org/10.1103/PhysRevLett.107.018102.

van den Heuvel, M.P., Stam, C.J., Boersma, M., Pol, H.H., 2008. Small-world and scale-free organization of voxel-based resting-state functional connectivity in the human brain. Neuroimage 43 (3), 528–539.

van den Heuvel, M., de Lange, S., Zalesky, A., Seguin, C., Yeo, T., Schmidt, R., 2017. Proportional thresholding in resting-state fmri functional connectivity networks and consequences for patient-control connectome studies: issues and recommendations. Neuroimage 152, 437–449. https://doi.org/10.1016/j.neuroimage.2017.02.005.

Jaynes, E.T., 1957. Information theory and statistical mechanics. Phys. Rev. 106 (4), 620–630.

Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17 (2), 825–841.

Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M., 2012. Fsl. Neuroimage 62 (2), 782–790.

Lambiotte, R., Delvenne, J.C., Barahona, M., 2014. Random walks, Markov processes and the multiscale modular organization of complex networks. IEEE Trans. Netw. Sci. Eng. 1 (2), 76–90. https://doi.org/10.1109/TNSE.2015.2391998 arXiv:1502.04381.

de Lange, S.C., de Reus, M.A., van den Heuvel, M.P., 2014. The laplacian spectrum of neural networks. Front. Comput. Neurosci. 7 (January), 1–12.

de Lange, S.C., van den Heuvel, M.P., de Reus, M.A., 2016. The role of symmetry in neural networks and their laplacian spectra. Neuroimage 141, 357–365.

Lindquist, M.A., Geuter, S., Wager, T.D., Caffo, B.S., 2019. Modular preprocessing pipelines can reintroduce artifacts into fmri data. Hum. Brain Mapp. 40 (8), 2358–2376. https://doi.org/10.1002/hbm.24528.

Lohse, C., Bassett, D.S., Lim, K.O., Carlson, J.M., 2014. Resolving anatomical and functional structure in human brain organization: identifying mesoscale organization in weighted network representations. PLoS Comput. Biol. 10 (10) https://doi.org/10.1371/journal.pcbi.1003712.

Lovász, L., 1993. Random walks on graphs: a survey. Bolyai Math Stud 2 (2), 1–46.

MacMahon, M., Garlaschelli, D., 2015. Community detection for correlation matrices. Phys. Rev. X 5 (2), 1–34 arXiv:1311.1924.

Marchitelli, R., Minati, L., Marizzoni, M., Bosch, B., Bartrés-Faz, D., Müller, B.W., Wiltfang, J., Fiedler, U., Roccatagliata, L., Picco, A., et al., 2016. Test-retest reliability of the default mode network in a multi-centric f mri study of healthy elderly: effects of data-driven physiological noise correction techniques. Hum. Brain Mapp. 37 (6), 2114–2132.

Maslov, S., Sneppen, K., 2002. Specificity and stability in topology of protein networks. Science 296 (5569), 910–913.

Mastrandrea, R., Squartini, T., Fagiolo, G., Garlaschelli, D., 2014. Enhanced reconstruction of weighted networks from strengths and degrees. New J. Phys. 16 (4), 043022.

Masuda, N., Porter, M.A., Lambiotte, R., 2017. Random walks and diffusion on networks. Phys. Rep. 716–717, 1–58. https://doi.org/10.1016/j.physrep.2017.07.007. URL:, (random walks and diffusion on networks). http://www.sciencedirect.com/science/article/pii/S0370157317302946.

Masuda, N., Kojaku, S., Sano, Y., 2018. Configuration model for correlation matrices preserving the node strength. Phys. Rev. 98 (1), 012312.

Mendes, N., Oligschläger, S., Lauckner, M.E., Golchert, J., Huntenburg, J.M., Falkiewicz, M., Ellamil, M., Krause, S., Baczkowski, B.M., Cozatl, R., Osoianu, A., Kumral, D., Pool, J., Golz, L., Dreyer, M., Haueis, P., Jost, R., Kramarenko, Y., Engen, H., Ohrnberger, K., Gorgolewski, K.J., Farrugia, N., Babayan, A., Reiter, A., Schaare, H.L., Reinelt, J., Röbbig, J., Uhlig, M., Erbey, M., Gaebler, M., Smallwood, J., Villringer, A., Margulies, D.S., 2019. A functional connectome phenotyping dataset including cognitive state and personality measures. Sci Data 6. https://doi.org/10.1038/sdata.2018.307.

Merris, R., 1994. Laplacian matrices of graphs: a survey. Linear Algebra Appl 197–198 (C), 143–176.

Mohar, B., 1991. The laplacian spectrum of graphs. Graph Theor. Appl. 871–898. Wiley.

Murphy, K., Birn, R.M., Handwerker, D.A., Jones, T.B., Bandettini, P.A., 2009. The impact of global signal regression on resting state correlations: are anti-correlated networks introduced? Neuroimage 44 (3), 893–905. https://doi.org/10.1016/j.neuroimage.2008.09.036.

Newman, M.E.J., 2006. Modularity and community structure in networks. Proc. Natl. Acad. Sci. U.S.A. 103 (23), 8577–8582.

Newman, M., 2010. Networks: an Introduction. OUP Oxford.

Nicolini, C., Bifone, A., 2016. Modular structure of brain functional networks: breaking the resolution limit by surprise. Sci. Rep. 6, 19250.

Nicolini, C., Bordier, C., Bifone, A., 2017. Community detection in weighted brain connectivity networks beyond the resolution limit. Neuroimage 146, 28–39 arXiv:1609.04316.

Nicolini, C., Vlasov, V., Bifone, A., 2018. Thermodynamics of network model fitting with spectral entropies. Phys. Rev. E 98, 022322. https://doi.org/10.1103/PhysRevE.98.022322.

Nocedal, J., Wright, S., 2006. Numerical Optimization. Springer Science & Business Media.
Obando, C., De Vico Fallani, F., 2017. A statistical model for brain networks inferred from large-scale electrophysiological signals. J. R. Soc. Interface 14 (128), 20160940.
Park, J., Newman, M.E.J., 2004. Statistical mechanics of networks. Phys. Rev. E 70 (6), 066117.
Parkes, L., Fulcher, B., Yücel, M., Fornito, A., 2018. An evaluation of the efficacy, reliability, and sensitivity of motion correction strategies for resting-state functional mri. Neuroimage 171, 415–436.
Peixoto, T.P., 2015. Model selection and hypothesis testing for large-scale network models with overlapping groups. Phys. Rev. X 5 (1), 011033 arXiv:1409.3059.
Power, J.D., Cohen, A.L., Nelson, S.M., Wig, G.S., Barnes, K.A., Church, J.A., Vogel, A.C., Laumann, T.O., Miezin, F.M., Schlaggar, B.L., Petersen, S.E., 2011. Functional network organization of the human brain. Neuron 72 (4), 665–678.
Power, J.D., Barnes, K.A., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., 2012. Spurious but systematic correlations in functional connectivity mri networks arise from subject motion. Neuroimage 59 (3), 2142–2154. https://doi.org/10.1016/j.neuroimage.2011.10.018.
Power, J.D., Mitra, A., Laumann, T.O., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., 2014. Methods to detect, characterize, and remove motion artifact in resting state fmri. Neuroimage 84, 320–341.
Robinson, P.A., 2019. Physical brain connectomics. Phys. Rev. 99 (1), 23–25. https://doi.org/10.1103/PhysRevE.99.012421. URL:
Rosvall, M., Bergstrom, C.T., 2008. Maps of random walks on complex networks reveal community structure. Proc. Natl. Acad. Sci. U.S.A. 105 (4), 1118–1123.
Rubinov, M., Sporns, O., 2010. Complex network measures of brain connectivity: uses and interpretations. Neuroimage 52 (3), 1059–1069.
Rubinov, M., Sporns, O., 2011. Weight-conserving characterization of complex functional brain networks. Neuroimage 56 (4), 2068–2079.
Saad, Z.S., Gotts, S.J., Murphy, K., Chen, G., Jo, H.J., Martin, A., Cox, R.W., 2012. Trouble at rest: how correlation patterns and group differences become distorted after global signal regression. Brain Connect. 2, 25–32. https://doi.org/10.1089/brain.2012.0080.
Salimi-Khorshidi, G., Douaud, G., Beckmann, C.F., Glasser, M.F., Griffanti, L., Smith, S.M., 2014. Automatic denoising of functional mri data: combining independent component analysis and hierarchical fusion of classifiers. Neuroimage 90, 449–468.
Santarnecchi, E., Galli, G., Polizzotto, N.R., Rossi, A., Rossi, S., 2014. Efficiency of weak brain connections support general cognitive functioning. Hum. Brain Mapp. 35 (9), 4566–4582. https://doi.org/10.1002/hbm.22495.
Satterthwaite, T.D., Wolf, D.H., Loughead, J., Ruparel, K., Elliott, M.A., Hakonarson, H., Gur, R.C., Gur, R.E., 2012. Impact of in-scanner head motion on multiple measures of functional connectivity: relevance for studies of neurodevelopment in youth. Neuroimage 60 (1), 623–632.
Satterthwaite, T.D., Ciric, R., Roalf, D.R., Davatzikos, C., Bassett, D.S., Wolf, D.H., 2019. Motion artifact in studies of functional connectivity: characteristics and mitigation strategies. Hum. Brain Mapp. 40 (7), 2033–2051.
Schlesinger, K.J., Turner, B.O., Grafton, S.T., Miller, M.B., Carlson, J.M., 2017. Improving resolution of dynamic communities in human brain networks through targeted node removal. PloS One 12 (12). https://doi.org/10.1371/journal.pone.0187715.
Schwarz, A.J., McGonigle, J., 2011. Negative edges and soft thresholding in complex network analysis of resting state functional connectivity data. Neuroimage 55 (3), 1132–1146. https://doi.org/10.1016/J.NEUROIMAGE.2010.12.047.
Siegel, J.S., Power, J.D., Dubis, J.W., Vogel, A.C., Church, J.A., Schlaggar, B.L., Petersen, S.E., 2014. Statistical improvements in functional magnetic resonance imaging analyses produced by censoring high-motion data points. Hum. Brain Mapp. 35 (5), 1981–1996.
Sporns, O., 2010. Networks of the Brain. MIT press. https://doi.org/10.1017/CBO9781107415324.004.
Sporns, O., Betzel, R.F., 2016. Modular brain networks. Annu. Rev. Psychol. 67 (1), 613–640 arXiv:15334406.
Squartini, T., Garlaschelli, D., 2011. Analytical maximum-likelihood method to detect patterns in real networks. New J. Phys. 13 (8), 083001.
Squartini, T., Garlaschelli, D., 2017. Maximum-Entropy Networks: Pattern Detection, Network Reconstruction and Graph Combinatorics. Springer International Publishing. https://doi.org/10.1007/978-3-319-69438-2.
Squartini, T., Mastrandrea, R., Garlaschelli, D., 2015. Unbiased sampling of network ensembles. New J. Phys. 17 (2), 023052.
de Vico Fallani, F., Latora, V., Chavez, M., 2017. A topological criterion for filtering information in complex brain networks. PLoS Comput. Biol. 13 (1), e1005305.
Wilde, M.M., 2013. Quantum Information Theory. Cambridge University Press.
Xiao, B., Wilson, R.C., Hancock, E.R., 2005. Characterising graphs using the heat kernel. In: BMVC 2005 - Proceedings of the British Machine Vision Conference 2005, pp. 1–10. https://doi.org/10.5244/C.19.92.
Zalesky, A., Fornito, A., Cocchi, L., Gollo, L.L., van den Heuvel, M.P., Breakspear, M., 2016. Connectome sensitivity or specificity: which is more important? Neuroimage 142, 407–420. https://doi.org/10.1016/j.neuroimage.2016.06.035.