Maximum entropy null models for thresholded undirected weighted networks

Carlo Nicolini[1], Giulia Forcellini[1,2] & Angelo Bifone[1,3]

ISTITUTO ITALIANO DI TECNOLOGIA

CIMeC Center for Mind/Brain Sciences

[1]Center for Neuroscience and Cognitive Systems, Istituto Italiano di Tecnologia, Rovereto (Italy)

[2]Center for Mind and Brain Sciences, University of Trento, Rovereto, (Italy)

[3]Department of Molecular Biotechnology and Health Sciences, University of Torino, Torino, (Italy)

1. We introduce a new null model for networks with continuos weights and a sparsification threshold. **2.** We applied an information theoretic approach to define network distance of empirical networks from their null models. **3.** In the case of brain connectivity networks we demonstrate the existence of an optimal threshold that maximizes the "distance" from their random counterpart.

Background and goals

Complex functional brain networks are obtained by continuous association measures between time series. These measures yield dense weighted matrices, intractable with classical network science methods.

Currently no consensus exists to determine the correct threshold level for network sparsification.

Here we propose a specific null model for positive-weighted networks with real-valued weights, and we apply it to understand the effects of thresholding on dense networks.

We then compare the null model with real networks at different threshold by means of spectral entropies.

Finally we look at the differences between empirical networks and their null models as a function of threshold, to identify the best cutoff level.

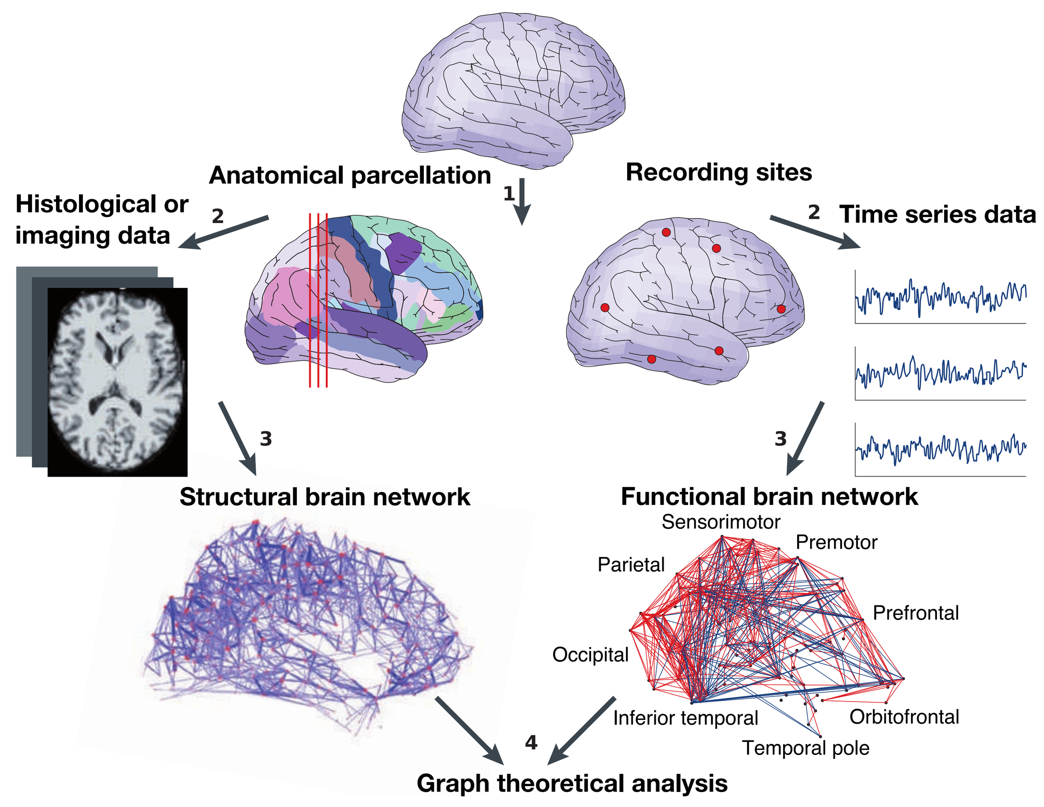

Figure 1: Worflow in graph-theoretical analysis of brain functional connectivity.

Methods: spectral entropy

We employed a scale resolved description of a network, based on network Von Neumann entropy, a quantity guided by the diffusion dynamics over the network [1, 4]:

$$S(\boldsymbol{\rho}) = -\operatorname{Tr}\left[\boldsymbol{\rho} \ln \boldsymbol{\rho}\right] \qquad \boldsymbol{\rho} = \frac{e^{-\beta \mathbf{L}}}{\operatorname{Tr}\left[e^{-\beta \mathbf{L}}\right]}$$

where $\boldsymbol{\rho}$ is the Von Neumann density matrix, based on the graph Laplacian. The density matrix describes the spreading dynamics of a conserved quantity over the network as a function of the hyperparameter β.

Spectral relative entropy is the amount of information lost describing the empirical network with density $\boldsymbol{\rho}$ using the null model with density $\boldsymbol{\sigma}$:

$$S(\boldsymbol{\rho} \| \boldsymbol{\sigma}) = \operatorname{Tr}\left[\boldsymbol{\rho} \ln\left(\boldsymbol{\rho} - \boldsymbol{\sigma}\right)\right]$$

Larger values of relative entropy indicate networks with very different diffusion profiles, while $S(\boldsymbol{\rho}|\boldsymbol{\sigma}) = 0$ only when $\boldsymbol{\rho} \equiv \boldsymbol{\sigma}$ networks.

The spectral entropy $S(\boldsymbol{\rho})$ as a function of β gives a scale resolved description of the diffusion dynamics:

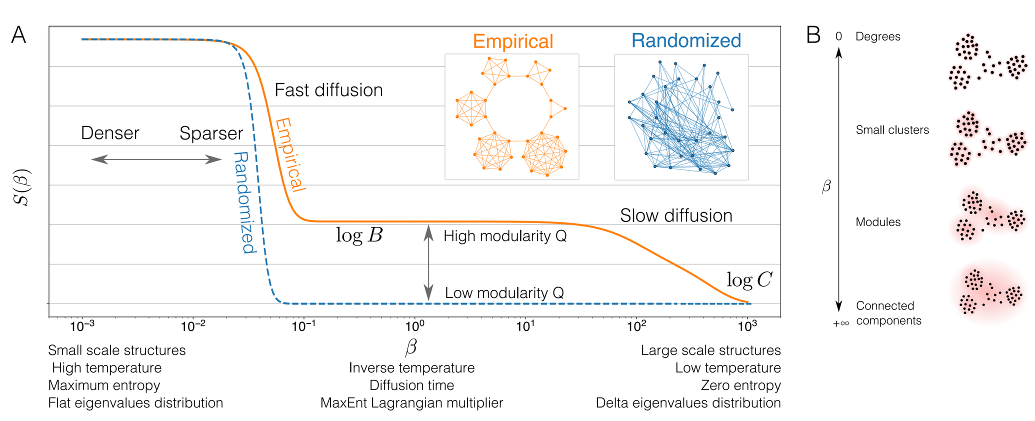

Figure 2: Diagram of spectral entropy of a highly structured network (orange) and its randomized counterpart that maintains the same degree sequence (blue).

Maximum entropy null model(s)

Continuous Weighted Thresholded Enhanced Configuration Model (CWTECM): An analytical null model of empirical networks preserving degree and strenght sequence defined on **positive real weights** plus a threshold hyperparameter t. It is based on the Hamiltonian:

$$H_{\mathrm{CWTECM}}(W|\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{i<j} \alpha_{ij} \Theta(w_{ij} - t) + \beta_{ij} w_{ij} \Theta(w_{ij} - t).$$

Unbiased estimation by **maximum likelihood** method [3, 2]:

$$\operatorname*{argmax}_{x,y} \sum_i s_i \ln y_i + k_i \ln x_i + \sum_{i<j} \ln\left(\frac{\ln(y_i y_j)}{t\left(\ln(y_i y_j)\right) - x_i x_j \left(y_i y_j\right)^t}\right)$$

Example of model fitting shown in Figure 3. Simpler model with $x_i = x$ and $y_i = y$ reproduces fixed link number and total weight, is called **Continuous Weighted Thresholded Enhanced Random Graph model**,**CWTERG**.

Analysis

Thresholding at increasing sparsity levels, we found that maximal difference between empirical and null networks are observed when the network starts breaking apart, around **percolation threshold**. As a proxy of network similarity, we computed the **spectral relative entropy** $S(\boldsymbol{\rho}\|\boldsymbol{\sigma})$ of empirical networks with their random counterparts, as well as the spectral entropy curves as a function of the β parameter. The networks display spectral entropy very close to that of their original counterparts, at very dense threshold levels. With increasing threshold and decreasing density, the difference between the empirical network and the null model increases, a result of the removal of noise and spurious weak connections. Around percolation threshold, the differences are maximal, and the large-scale structure of the network emerges (Figure 4).

Results

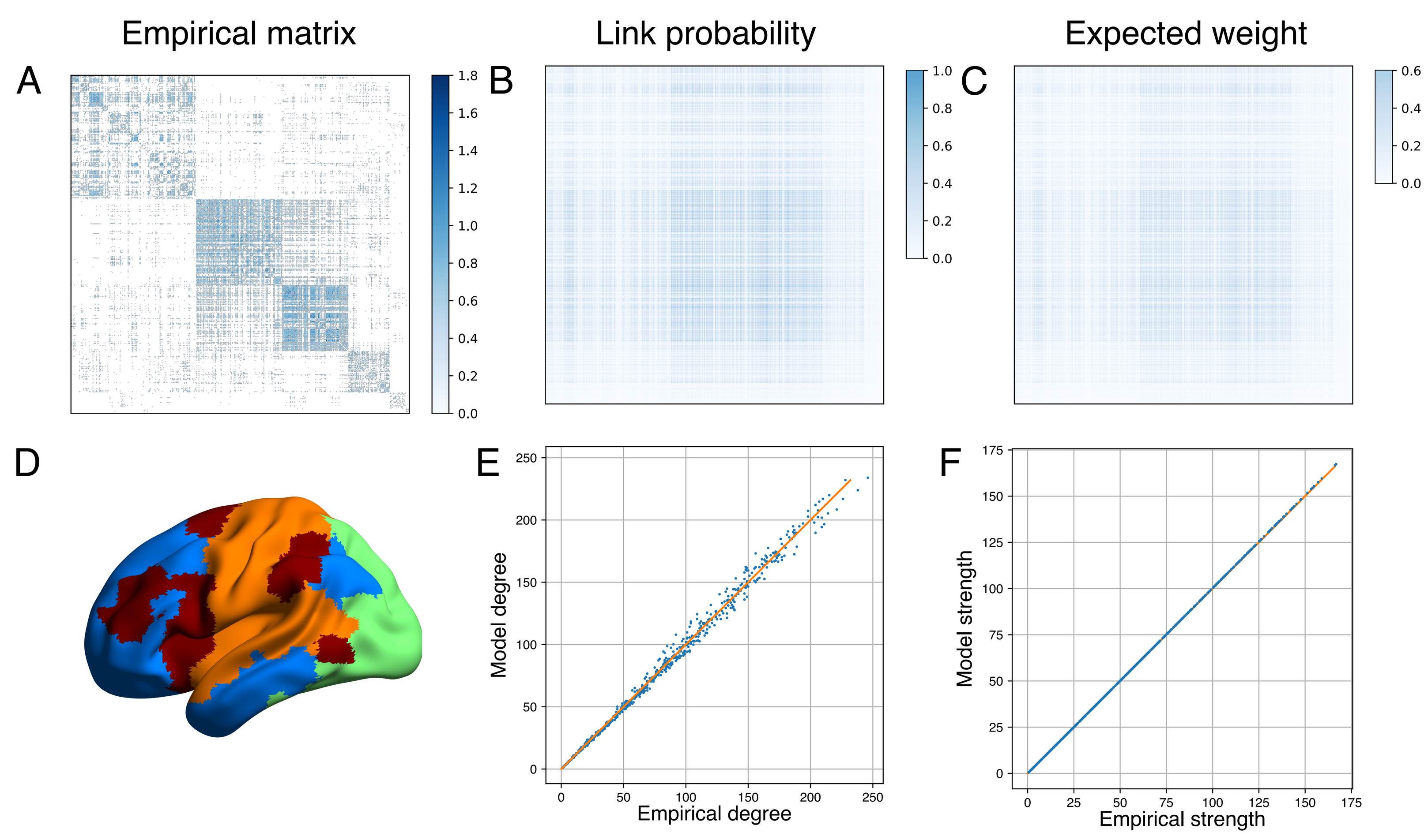

Figure 3: Continuous Weighted Thresholded Enhanced Configuration Model fitted on a real functional network. Panel **A**: The empirical functional connectivity matrix, thresholded at percolation with rows and columns reorganized by the maximum modularity community structure, to highlight the community structure. Panels **B,C**: the link probability and expected link weights predicted by the model. Panels **E,F**: the reconstructed degrees and strength as sums 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). The reconstruction error is very precise with very small deviations from the identity line.

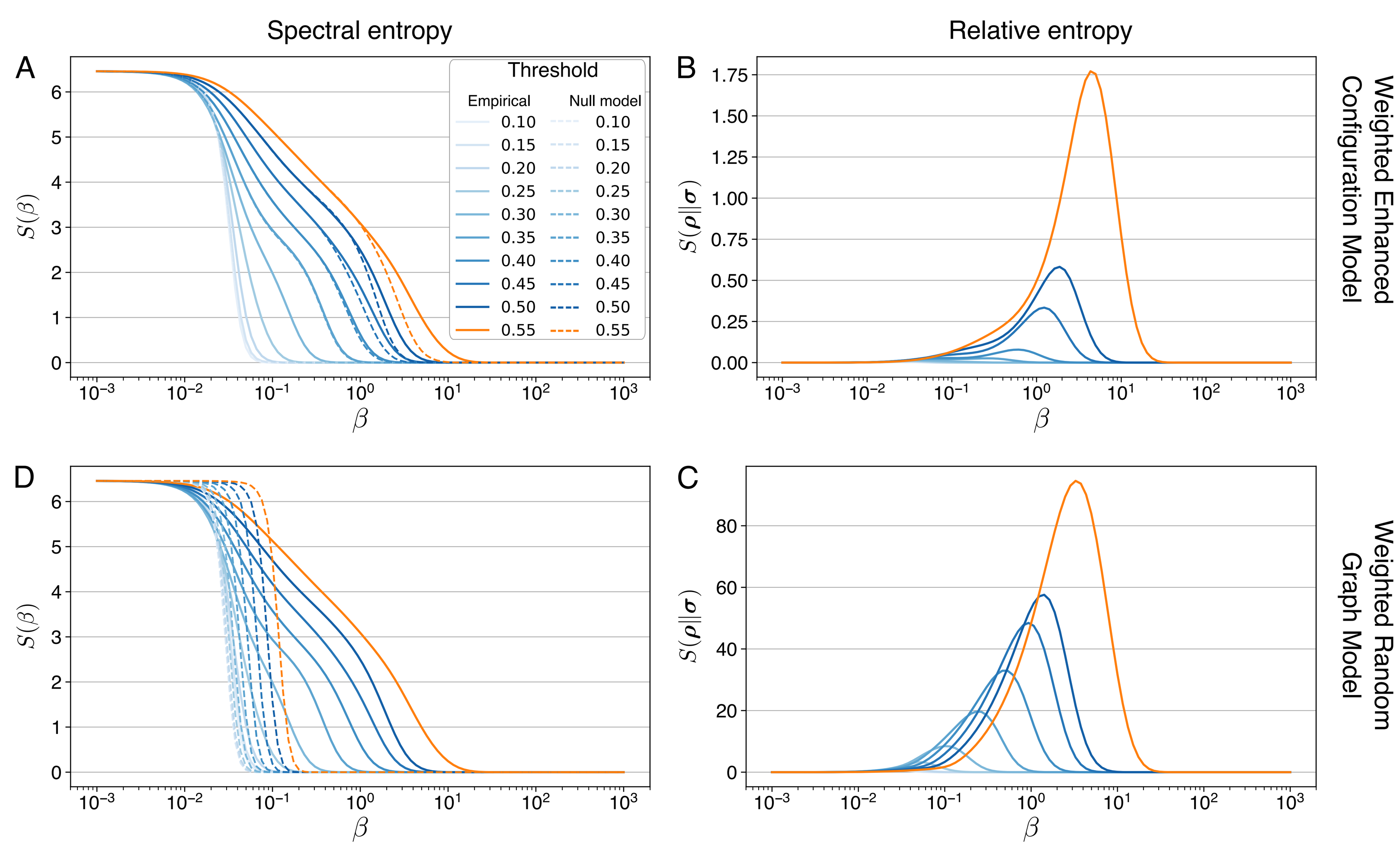

Figure 4: Spectral entropies and relative entropies of a brain functional network compared to its randomized counterpart. Blue shaded lines represent networks thresholded at values from $z = 0.1$ to $z = 0.5$. Orange lines denote the network at percolation threshold. Solid lines are the curves relative to the empirical network. Dashed lines are the curves relative to the randomized networks. Panels **A,B** show the results with respect to the CWTECM model. Panels **C,D** show the variant with fixed number of links and total weight, corresponding to setting all $x_i = x$, $y_i = y$.

References

[1] M. De Domenico and J. Biamonte. *Phys. Rev. X*, 6:041062, Dec. 2016.

[2] D. Garlaschelli and M. I. Loffredo. *Phys. Rev. E*, 78:015101, July 2008.

[3] R. Mastrandrea, et al. *New J. Phys.*, 16(4):043022, 2014.

[4] C. Nicolini, et al. *Phys. Rev. E*, 98:022322, Aug. 2018.

Download this poster!

The analyses in this poster are done with the Python tool **networkqit**. More informations are available at:

networkqit.github.io

Contact me at:

carlo.nicolini@iit.it

carlonicolini.github.io