Self-similarity of human brain connectomes

Author: Carlos Arnedo Joya
*Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.**

Advisor: M. Ángeles Serrano

Abstract: The human brain is extremely complex, with multiple spatial scales interacting simultaneously. The connectome is a comprehensive map of the neural connections between brain regions that is useful to delve deeper into the human brain. We studied the multiscale organization of human brain connectomes and found two symmetries using two datasets from healthy subjects. First, we measured the multiscale properties at five different hierarchical resolutions and recovered the previously reported result that they remain self-similar as the resolution is decreased. Second, we performed a quantitative analysis that shows that the structural features of the connectomes remain self-similar when using a degree-thresholding renormalization method on the highest resolution network layer. Our results suggest that the human brain may be working near a critical point, and that brain regions could be mapped to a hyperbolic hidden metric space, defining distances that explain the structure of brain connectomes and helping in the development of simulation and brain reconstruction tools.

I. INTRODUCTION

The human brain, consisting of eighty six billion of interlinked neurons, is one of the least-understood networks from the perspective of connectivity and functional properties. The reason is simple, we lack maps telling us which neurons are linked together. Efforts to study the brain in the field introduced the concept of the connectome. The connectome [1] represents a comprehensive map of neural connections between regions in the brain, capturing both the structural connections formed by groups of neurons or the functional connections that control activity and information flow.

In recent years, research efforts directed to characterize connectivity in the brain have been focused on detecting properties found in most real networks at a single spatial resolution. However, the human brain is extremely complex with multiple scales interacting with one another. In this study, we investigate the network features of the multiscale human (MH) connectome at five anatomical resolutions using two datasets of healthy subjects obtained by different imaging techniques.

First, we measured topological properties of the human connectome at different length scales for different subjects in each dataset, and found statistical self-similarity as the resolution decreased [2], resembling a fractal. Second, the degree-thresholding renormalization [3] technique was applied to the connectome of the same subjects with the highest resolution scale to show that certain network properties, particularly clustering, exhibit self-similarity. To demonstrate that these results were not trivial, the same measures were performed on randomized versions of these networks.

All the code implemented in this study was developed using the FORTRAN 90 programming language [10]

*Electronic address: carlosarnedo19@gmail.com

II. DATA AND METHODS

A. Datasets

We used the two datasets analyzed in [2]. One of the datasets is from the Human Connectome Project (HCP) which contains multiscale connectomes, obtained from weighted magnetic resonance imaging (MRI), of 44 healthy subjects. The other dataset is from the University of Lausanne (UL), consisting of diffusion spectrum MRI data from 40 subjects. The neural fibers, responsible for transmitting information throughout the nervous system, and the connectomes were reconstructed using different techniques indicated in [2].

The nodes in the connectomes represent parcels of cortical and subcortical regions (excluding the brainstem), and the connections between them indicate the presence of neural fibers. The layers in the multiscale reconstruction have 1014, 462, 233, 128, and 82 nodes, which are labeled as 0, 1, 2, 3, and 4, respectively. The multiscale parcellation was produced by iterating a coarse-graining operation to layer 0 and producing successive layers with lower resolution. Each node corresponds to a larger portion of the brain as the resolution decreases. These networks are undirected and unweighted, which means that the connections have no defined direction and equal intensity.

To provide a comprehensive description of the MH connectome, the adjacency matrix, A_{ij}, is used. The adjacency matrix of a N node network has N rows and N columns, with A_{ij}=1 if nodes i and j are connected and A_{ij}=0 otherwise. Because the datasets represent sparse networks, meaning that the average number of connections per node is low, only a small percentage of A_{ij} elements are nonzero. As an alternative, networks can be displayed in the form of an edgelist, which is a list of the network's links. We used a list of neighbors and a list of the number of neighbors for each node to avoid

storing the zero elements of the adjacency matrix.

B. Degree distribution

The most fundamental local property of a network's node is its degree: the number of total edges attached to a vertex, meaning the number of neighbors that it has in the network. The degree distribution $P(k)$ is the probability of a node having k neighbors. Sometimes the complementary cumulative degree distribution $P_C(k)$ is used, which is the probability of a node having a degree larger than k. Usually, researchers try to fit empirical data with power-law degree distributions [4] of the type $P(k) \propto k^{-\gamma}$ with $\gamma > 2$, known as fat-tailed degree distributions. Power-law distribution exhibits the heterogeneous character of the network [5], meaning that a few nodes have a high number of nearest neighbors, while most nodes have a relatively low number of links. The fact that the degree distribution can be approximated by a power law is a first indication of self-similar organization [6]. It is worth noting that, unlike many real networks, the human brain connectome may not exhibit a scale-free property with a power-law exponent in the range of [2, 3]. Recent studies suggest that the lack of scale-freeness in human brain connectomes makes it less susceptible to random attacks [7].

C. Correlations

To measure correlations between the degrees of connected nodes, known as degree-degree correlations, the conditional probability $P(k'|k)$ could be calculated, which can be challenging with real data. Instead, it is more practical to measure the average nearest neighbor degree, $\overline{k}_{nn}(k)$, which captures the tendency of nodes to connect to peers based on their degree [8]. The expression for $\overline{k}_{nn}(k)$ is given by Eq. (1) , where N_k represents the total number of nodes with degree k, $\Upsilon(k)$ is the set of such nodes, and a_{ij} denotes the adjacency matrix element connecting node i with node j.

$$\overline{k}_{nn}(k) = \frac{1}{N_k} \sum_{i \in \Upsilon(k)} \frac{1}{k_i} \sum_j a_{ij} k_j \tag{1}$$

The clustering coefficient is a scalar feature used to calculate three-point correlations that measures the likelihood that two nodes with a common neighbor are also connected to one another. It was introduced in social networks to determine whether a friend of your friend was also your friend [8]. From a local standpoint, it can be calculated as the number of triangles (loops of length three) passing through node i, T_i, divided by the total number of possible triangles given the number of neighbors:

$$c_i = \frac{2T_i}{k_i(k_i - 1)} \tag{2}$$

The degree-dependent clustering coefficient [8] can be measured by defining an average of c_i over the set of vertices of a given degree class:

$$\overline{c}(k) = \frac{1}{N_k} \sum_{i \in \Upsilon(k)} c_i = \frac{2}{k(k-1)N_k} \sum_{i \in \Upsilon(k)} T_i \tag{3}$$

Because most real-world networks have high clustering, it is an important measure for capturing correlations.

D. Degree-thresholding renormalization

The degree-thresholding renormalization method [3] consists of generating a hierarchy of nested graphs from a given network graph G. This is done by extracting a subgraph $G(k_T)$ for each degree threshold $k_T = 0, 1, 2,...$ which includes nodes with degrees greater than k_T.

The degree of each node in $G(k_T)$ is then recalculated based on its connections in the subgraph, and the new neighbors of each node are computed. The degrees of the nodes are then rescaled by dividing them by the average internal degree $\langle k_i(k_T) \rangle$ in $G(k_T)$, resulting in $k_i / \langle k_i(k_T) \rangle$.

This process is repeated for each nested graph. This renormalization, which was applied in [3] to scale-free networks, provides a hierarchical view of the network, revealing structural properties at different scales, based on the degree thresholds chosen.

E. Random Networks

In order to check that the obtained network properties are not trivial, null models are used. These models generate randomized versions of our graph while preserving certain features. In our study, we used a random rewiring process to create a randomized network that preserved the degree distribution of the connectome [9], while simultaneously destroying all correlations except for the structural ones. To generate the randomized network, we begin by randomly selecting two distinct links from the edgelist. Next, the endpoints of these links are exchanged while ensuring that self-loops and multiple connections are avoided, considering the undirected and unweighted nature of our network. This process is repeated for at least the number of links in the original network.

III. RESULTS AND DISCUSSION

A. Self-similarity in the MH connectome

For subject No.22 from the HCP dataset and No.10 from the UL dataset we measured the next properties at each layer l of the MH connectome: the complementary cumulative degree distribution $P_C^{(l)}(k_{res}^{(l)})$, the normalized average nearest neighbor degree calculated using Eq. (1),

denoted as $\overline{k}^{(l)}_{nn,n}(k^{(l)}_{res}) = \overline{k}^{(l)}_{nn}(k^{(l)}_{res})\langle k^{(l)}\rangle / \langle (k^{(l)})^2\rangle$, the degree-dependent clustering coefficient calculated using Eq. (3), denoted as $\overline{c}^{(l)}(k^{(l)}_{res})$, average degree and average clustering coefficient. To account for the variation in average degree across layers and make them more comparable, these features were calculated as a function of $k^{(l)}_{res} = k^{(l)}/\langle k^{(l)}\rangle$. Examining Table I, we see that the connectomes in the UL dataset have a lower density than the connectomes in HCP dataset, as evidenced by the significantly fewer links and the average degrees. Despite this sparsity, the average local clustering coefficient in both datasets remains relatively consistent for all subjects.

Dataset	Subject	N	L	$\langle k\rangle$	$k_{\max}$	$\langle c\rangle$
HCP	No. 0	1014	37910	74.8±1.5	529	0.411±0.004
	No. 2	1014	39609	78.1±1.5	566	0.422±0.004
	No. 4	1014	43276	85.4±1.7	678	0.414±0.004
	No. 22	1014	38855	76.6±1.5	598	0.399±0.004
	No. 34	1014	44274	87.3±1.7	643	0.418±0.003
UL	No. 3	1011	12991	25.7±0.6	277	0.410±0.005
	No. 5	1014	14340	28.3±0.7	323	0.407±0.005
	No. 8	1002	13910	27.8±0.7	243	0.405±0.005
	No. 10	1014	15222	30.0±0.7	294	0.415±0.005
	No. 15	1007	16208	32.2±0.7	334	0.431±0.005

TABLE I: Overview of the layer 0 of five subjects from each dataset. The number of nodes (N), the number of links (L), its average degree ($\langle k\rangle = 2L/N$), its average local clustering coefficient ($\langle c\rangle$), and ±1 the standard error (SE) interval around the mean, and the maximum degree (k_{max}).

Plots in Fig. 1 exhibit an important overlap, suggesting self-similarity across layers of the MH connectome for the two selected subjects. $P^{(l)}_C(k^{(l)}_{res})$ in Fig. 1 A and B displays some degree of heterogeneity across the connectomes. In Fig. 1 C $\overline{k}^{(l)}_{nn,n}(k^{(l)}_{res})$ is almost constant as a function of the rescaled degree, while in Fig.1 D $\overline{k}^{(l)}_{nn,n}(k^{(l)}_{res})$ has assortative behavior, meaning that nodes tend to connect to others with similar degree. In the *Insets* of Fig. 1 C and D, we observe a slight increase in the average degree $\langle k\rangle$ followed by a decrease as the resolution decreases. Instead, in the *Insets* of Fig. 1 E and F, we observe an increasing average local clustering coefficient $\langle c\rangle$, which explains the observed shifts in Fig. 1 E and F. It is noteworthy that layers 3 and 4 are sensitive to finite-size effects given their small number of nodes.

To summarize, our results support the self-similarity of the MH connectome for the selected subject in each dataset. Despite the fact that both datasets were obtained using different image techniques, the network features of the two datasets are strikingly similar. However, as in [2], we should measure all of the remaining 82 subjects' characteristics to validate our findings.

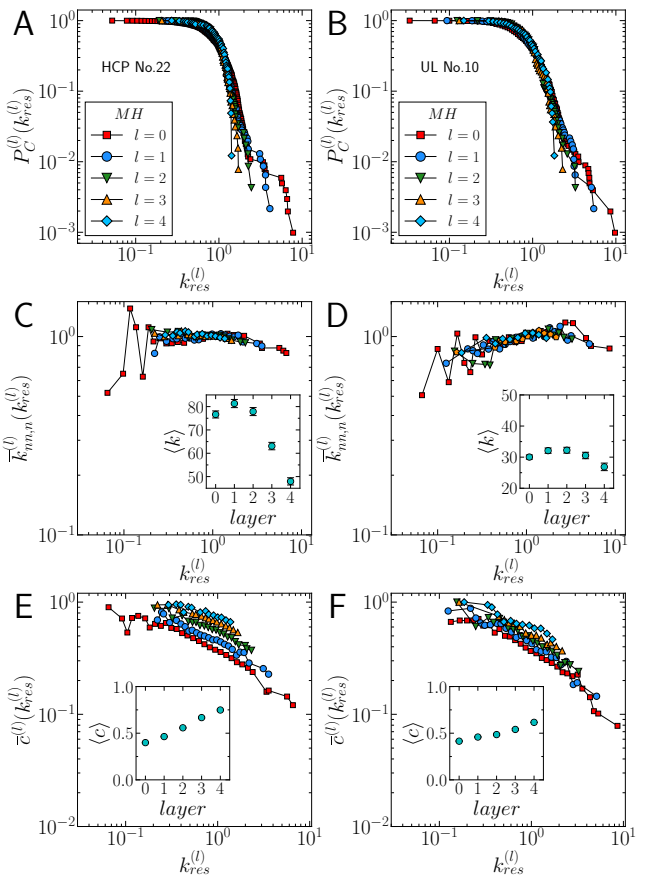

FIG. 1: Self-similarity of the MH connectome across different resolutions. (A, C, E) Results for HCP subject No.22. (B, D, F) Results for UL subject No.10. (A, B) Complementary cumulative degree distribution $P^{(l)}_C(k^{(l)}_{res})$. (C, D) Normalized average nearest neighbor degree $\overline{k}^{(l)}_{nn,n}(k^{(l)}_{res})$. (C & D, *Insets*) Average degree $\langle k\rangle$ across layers. (E, F) Degree-dependent clustering coefficient $\overline{c}^{(l)}(k^{(l)}_{res})$. (E & F, *Insets*) Average clustering coefficient $\langle c\rangle$ across layers. In C, D, E, F, *Insets*, error bars indicate the two-SE interval around the mean.

B. Self-similarity applying degree-thresholding renormalization

The main objective of this part of the study is to detect self-similarity of the nested graphs by applying a degree-thresholding renormalization procedure to the highest resolution layer 0 of five subjects from each dataset. All the network features described in section III A are now measured, except for the normalized average nearest neighbor degree, because clustering is more relevant.

For the sake of brevity, the degree distribution is not shown here. When degree-thresholding renormalization was applied to human brain connectomes, the curves $P_C(k_{res})$ as a function of k_{res} overlapped, suggesting self-similarity. Finite-size effects appeared above a certain threshold, so they had to be carefully chosen. As UL datasets are sparser than HCP ones, thresholds are

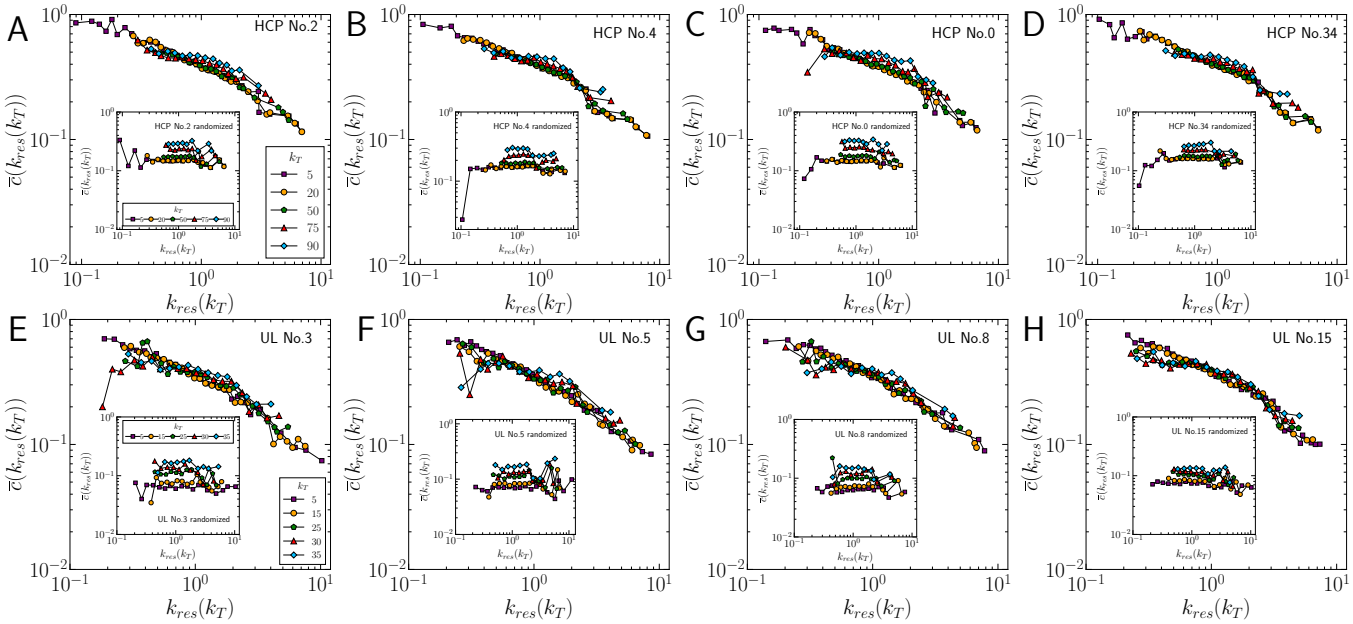

FIG. 2: Self-similarity of layer 0 applying degree thresholding renormalization. (A, B, C, D) Results from HCP subjects. (E, F, G, H) Results from UL subjects. The degree-dependent clustering coefficient as a function of the rescaled degree for different thresholds. (*Insets*) Degree-dependent clustering coefficient as a function of the rescaled degree for different thresholds from the respective randomized versions.

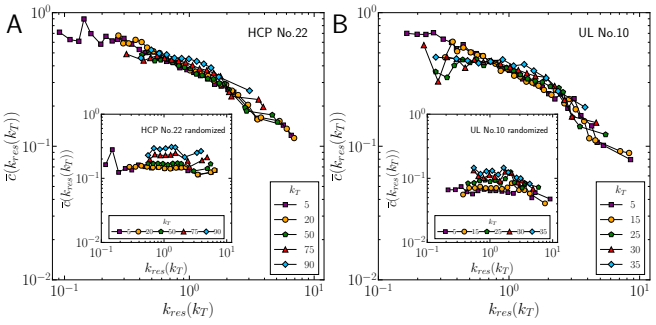

FIG. 3: Self-similarity of layer 0 applying degree thresholding renormalization. (A, B) The degree-dependent clustering coefficient as a function of the rescaled degree for different thresholds of the subject No.22 from HCP and the subject No.10 from UL, respectively. (A, B, *Insets*) Degree-dependent clustering coefficient as a function of the rescaled degree for different thresholds from the respective randomized versions.

smaller because statistical fluctuations manifest earlier. This also indicates that there is a value above which self-similarity disappears. The degree-dependent clustering coefficient as a function of k_{res}, represented in Fig. 2 and 3, presents an overlap for the different degree thresholds k_T for all the analyzed subjects from each dataset, displaying self-similarity. In order to show that this was not a trivial property, we generated randomized versions of the connectomes while preserving the degree distribution through the random rewiring process. The overlapping patterns in clustering indicate that self-similarity observed in the original networks is not present in their randomized versions, as shown in the *Insets* of Fig. 2 and 3. Because these random networks preserve the degree distribution, and the degree distribution overlaps in the original brain connectomes, the degree distribution of the randomized versions overlaps as well. With respect to the average degree and average clustering coefficient shown in Fig. 4, for $\langle k \rangle$ there is not a discernible difference between the real networks and their randomized counterparts, very likely because they share the same degree distribution. In Fig. 4 A and B, as well as Fig. 1 *Insets* C and D, $\langle k \rangle$ has the same behavior, slightly increasing at first and then dropping to a lower value due to finite-size effects. Looking at $\langle c \rangle$ in Fig. 4 C and D, while it remains almost constant as a function of k_T in the original networks, it grows as k_T increases in the randomized versions. The $\langle c \rangle$ of the randomized networks from UL dataset is smaller than that of HCP probably because UL networks are sparser. Furthermore, when comparing clustering, we can see that human brain connectomes from both datasets have higher clustering than their randomized counterparts, which is one of the fundamental properties observed in most real networks.

In a nutshell, observing the effects on clustering after applying the degree-thresholding renormalization for the human brain connectomes and their randomized counterparts, we can say that self-similarity is present across all subjects, and they all exhibit high clustering. As observed in the MH connectomes, despite the fact that both datasets were obtained using different techniques, the results are very similar, and there is a threshold value below

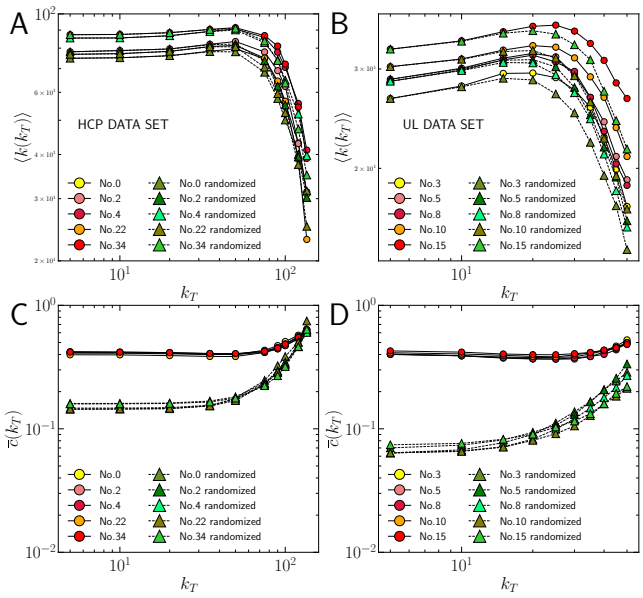

FIG. 4: (A, C) Results for HCP subjects. (B, D) Results for UL subjects. (A, B) Average degree as a function of the threshold degree k_T for the renormalized networks and their randomized counterparts. (C, D) Average clustering coefficient as a function of the threshold degree k_T for the renormalized networks and their randomized counterparts.

which self-similarity disappears, probably due to finite-size effects.

IV. CONCLUSIONS

The architecture of the human brain fuels human behavior and is extremely complex, with multiple spatial scales interacting with one another. We attempt to capture the essence of this complexity by simplifying it and identifying two symmetries.

On the one hand, when the resolution was reduced by a hierarchical anatomical approach, the MH connectome remained self-similar for different subjects from two datasets, despite being obtained using different methodologies. The presence of finite-size effects in layers 3 and 4 suggests that there is a limit beyond which self-similarity on the MH connectome vanishes, but other scales should be explored. This self-similarity indicates that the connectivity rules in the connectomes are independent of the resolution scale (at least with the ones studied), which could indicate that the brain is working near a critical point of a phase transition and may be helpful in the development of tools for brain reconstruction and simulation.

When we applied the degree-thresholding renormalization procedure to the layer 0 of the MH connectome for different subjects from each dataset, we found that the degree distribution and clustering remained self-similar as the degree threshold increased, whereas in randomized versions that preserved the degree distribution, the clustering did not. When the original connectomes were compared to their randomized counterparts, high clustering was observed in the original ones, which is a fundamental property of many real networks. The presence of self-similarity under this renormalization agrees with properties displayed by geometric network models in hyperbolic space, which could indicate that brain regions can be mapped to a hyperbolic hidden metric space defining distances that explain the structure of brain connectomes.

Datasets with higher connectome resolution are required to extract more conclusions, and we hope that one day we will be able to track the brain neuron by neuron to obtain the greatest real network ever built.

Acknowledgments

I would like to thank my advisor, M. Ángeles Serrano, for all the guidance and for introducing me to the amazing field of networks.

[1] Sporns, O. et al. (2005). "The human connectome: a structural description of the human brain." PLoS Comput Biol, 1(4), e42.

[2] Zheng, M., Allard, A., Hagmann, P., Alemán-Gómez, Y., & Serrano, M. Á. (2020). "Geometric renormalization unravels self-similarity of the multiscale human connectome." Proc. Natl. Acad. Sci., 117(33), 20244-20253.

[3] Serrano, M. Á., Krioukov, D., & Boguná, M. (2008). "Self-similarity of complex networks and hidden metric spaces." Phys. Rev. Lett., 100(7), 078701.

[4] Clauset, A. et al. (2009). "Power-law distributions in empirical data." SIAM Rev., 51(4), 661-703.

[5] Dorogovtsev, S. N. et al. (2004). "The shortest path to complex networks." arXiv:cond-mat/0404593.

[6] Serrano, M.,& Boguñá, M. (2022). "The Shortest Path to Network Geometry: A Practical Guide to Basic Models and Applications." Cambridge: Cambridge University Press.

[7] Achard, S. et al. (2006). "A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs." J. Neurosci., 26(1), 63-72.

[8] Serrano, M. et al. (2007). "Correlations in complex networks." In: Large scale structure and dynamics of complex networks. (pp. 35-66).

[9] Newman, M. E. et al. (2004). "Finding and evaluating community structure in networks." Phys. Rev. E, 69(2), 026113.

[10] Code can be found at: `https://github.com/arnedito/Self-similarity-in-human-brain-connectomes.git`