Fundamenta Informaticae XX (2018) 1–17
DOI 10.3233/FI-2016-0000
IOS Press

Graph Model Simulation of Human Brain's Functional Activity at Resting State by Means of the FD Model

Paolo Dulio
Dipartimento di Matematica
Politecnico di Milano
Piazza Leonardo da Vinci 32
20133 Milano, Italy
paolo.dulio@polimi.it

Paolo Finotelli
Dipartimento di Matematica
Politecnico di Milano
Piazza Leonardo da Vinci 32
20133 Milano, Italy
paolo.finotelli@polimi.it

Andrea Frosini
Dipartimento di Matematica e Informatica
Universitá degli Studi di Firenze
Viale Morgagni 65
50134, Firenze, Italy
andrea.frosini@unifi.it

Elisa Pergola
Dipartimento di Matematica e Informatica
Universitá degli Studi di Firenze
Viale Morgagni 65
50134, Firenze, Italy
elisa.pergola@unifi.it

Alice Presenti
imec-Vision lab, Dept. of Physics
University of Antwerp
Universiteitsplein 1
2610 Antwerp, Belgium
alice.presenti@uantwerpen.be

Abstract. It is commonly accepted that the various parts of the human brain interact as a network at macroscopic, mesoscopic and microscopic level. Recently, different network models have been proposed to mime the brain behavior both at resting state and during tasks: Our study concerns one of those model that consider both the physical and functional connectivity as well as topological metrics of the brain networks. We provide evidence of the soundness of the model by means of an synthetic dataset based on the existing literature concerning the active cerebral areas at the resting state. Furthermore, we consider Ruzicka similarity measure in order to stress the predictive capability of the model and provide a thresholding criterium. Some network statistics are finally provided.

1. Introduction

The human brain is a complex system composed by several different components which can interact each other as a network structure; the components depend on the scale of analysis, for example, at microscale, the components are *single neurons*. In the last few years, brain network has been of great interest for many researchers, fascinated by its still not fully understood structure and organization. The mathematical structure of graphs revealed to be a valuable tool to model the brain activity, and to detect and analyze most of its properties. Nowadays, graph theory is also used to show how brain interactions change between subjects and, for the same subject, in different life periods in order to reveal similarities and detect neural disorders.

Depending on the analysis' proposal, the brain network can be examined from a macro to a microscopic scale, so that each node can represent a single neuron, a group of neurons or an entire brain region. The meaning of the brain network's edges, instead, depends on which kind of brain connectivity the network models. Regardless of the node's meaning, we can distinguish between three kinds of brain connectivity, which in turn give rise to different graphs:

- *Structural Connectivity*, also called the *human connectome*, is the anatomical connection between the nodes, which can be considered stable at short time intervals, while undergoes changes in the long run, due both to the natural morphological modification of the brain during ages and to possible injuries and neural diseases. The human connectome is commonly represented as an undirected weighted graph, whose arcs' weights are the physical distances among the nodes.

- *Functional Connectivity* describes the temporal correlation of the signals between nodes. Nowadays, it is investigated by means of noninvasive techniques, such as electroencephalography (EEG), positron emission tomography (PET) and functional magnetic resonance imaging ($fMRI$). Functional connectivity can be evaluated both at resting state (when an individual is not performing an explicit task) both submitting the subject to a particular task (observing an image, listening to some music, thinking of something, and so on). It is remarkable that brain regions can be functionally correlated even if they are not physically connected, when they are involved in the same task. The functional connectivity can also represented as an undirected weighted graph, whose arcs' weights are usually the Pearson's r correlation coefficients.

- *Effective Connectivity* denotes the causal (directed) influence between the nodes of the brain network. It is computed with transfer entropy, a standard measure of the directed information flow. The effective connectivity is represented by means of a directed weighted graph, whose arcs' weights are mostly estimated using Granger causal or dynamic causal model coefficients [6].

In this paper the mathematical model of brain functional activity among brain regions, called FD-model, defined by Paolo Dulio and Paolo Finotelli in [9] is considered. The authors propose the use of a graph obtained from the interaction of the functional matrix and the connectome. The most relevant intuitions basically lie in: First, introducing the anatomical distance between couple of nodes (for example the Euclidean distance). Second, in inserting a parameter related to time in order to include the subject's age (it is known that functional connectivity changes during life: The brain of a young subject who highly stimulates his brain studying and acquiring new knowledge everyday is surely

more functionally active, from a functional view, than the one of an old subject). Third, in involving the degree of the nodes.

One of the most relevant insight of our study is the creation of synthetic data to model the functional connectivity in healthy subjects at resting state, based on the state-of-the-art on functional networks. Furthermore, we provide a new method to threshold the data obtained from the model: This need is of primary relevance since, by definition, the brain network presents a fully connected structure that makes hard to detect and visualize those areas most active and interconnected even at resting state. It is frequent in literature to select a threshold either basing on the percentage of arcs to be eliminated or to set an *a priori* value, without considering the global structure of the graph and its weights. The major drawback we may step on is that of eliminating also significant connections; our approach tries to prevent this problem by dynamically choosing a different threshold according to the structure of each computed graph. At the same time, we try to get rid of the weakest connections that may be meaningless and regarded as noise in the dataset.

Finally, the computed models are statistically analyzed upon increasing thresholds and pairwise compared to check their degree of similarities by using Ruzicka measure [8]. The last step is devoted to check the significance of the obtained networks by comparing them to random networks. Interestingly, this check gives, among other things, the opportunity to study their small world property. Other statistics are also investigated, and here mentioned, to provide additional material to a final discussion about the neurobiological interpretation of our results.

2. Details of the FD-model

In this section we will describe the FD-model, a new mathematical model of the brain activity recently proposed by Paolo Dulio and Paolo Finotelli in [9], in order to test its behavior at resting state with a synthetic data simulation based on the recent literature.

In the FD-model, the main idea consists of computing the correlation between any couple of nodes as a function depending on different parameters, and not only on the $fMRI$ collected data. It is suggested that also the structural connectivity should be considered, as well as the topological parameters such as the node degrees. The role of time is also discussed and introduced, distinguishing the contribution related to the duration of a given task, from the modifications caused aging on the involved functional activity.

2.1. A synthetic implementation of the FD-model at the resting state

The aim of our work is to construct and analyze the weighted undirected graph $G = (V, E)$ proposed by Dulio and Finotelli, that models the structure of brain activity, where V is the set of nodes obtained by parcelling the gray matter depending on anatomical and functional differences. There is an arc $e = \{i, j\} \in E$ between the nodes i and j if and only if the two regions are supposed to activate similarly, and the weight of the arc $w(e)$ represents the functional connectivity that describes statistical patterns of dynamic interactions among regions [16].

It is commonly assumed that the functional activity changes during the individual's life (consider that, from the birth to the age of eighteen both the brain physical structure and the functional network

highly develop, then the growth slows down until, in adulthood, it starts to decrease). Therefore, in order to consider the effects of aging, it has been introduced a time parameter t, reaching a final definition for each arc's weight of the form:

$$W(i,j,t) = \beta(i,j,t_{fix})e^{-\eta(t)D(i,j)+\alpha(i,j,t)}. \tag{1}$$

Observe that this model belongs to the class of exponential law models (see [17], [15]), although it incorporates basic information as the functional matrix and the time dependence.

Now, we provide some more details of the function (1):

- $\beta(i,j,t_{fix})$ represents the functional stimulus resulting in the brain by the mutual activity of the areas i and j. Here, the time t_{fix} is an arbitrary instant in the interval of time when the subject is exposed to a specific task. In [9] there is a detailed mathematical approach that justifies the use of the topological parameter $deg(i)deg(j)$ as a first order approximation of the function $\beta(i,j,t_{fix})$ (the degree of the nodes is calculated from the matrix of the functional activity after a suitable thresholding procedure, see for example [10]).

- The matrix D is the matrix of the human brain connectome, whose entries are the physical distances between the centers of the regions after the parcellation (i.e. the procedure used by the scientists to discretize the gray matter). It is a common choice to consider Euclidean distances between centroids of the regions of interest, even though sometimes it turns out to be much shorter than the real fibers [3].

- The function $\eta(t)$ is also related to the matrix D, and it is a positive factor that, due to the minus sign, represents a distance penalty, since two far regions are supposed to have a lower functional interaction than two close ones. The dependence from the time is due to the following consideration: in older subject functional activity is supposed to decrease, so a greater value of η has to be considered.

  At present, no η evaluations have been proposed for humans in order to simulate a network that well fits on brain network's properties of modularity and efficiency [17]. For this reason, Dulio and Finotelli suggest to choose the value (not time dependent) $\eta = 8$ that has been estimated for macaque monkeys, widely used in order to understand the functioning of human brain. Although there are many aspects of humans' abilities present in monkeys, the evolutionary process has contributed to significantly differentiate between the two brains (one for all, humans only can speak and use grammar or reflect on thoughts [12]). Following the study of Vèrtes et al. in [17], this model could be exploited to find a suitable value of η for humans that best fits the properties of clustering, efficiency, modularity and degree distribution of brain networks. An optimization process could provide a value for this parameter that could make the modeled network much more similar to the real one, providing a further step on the understanding of the hidden properties in the human brain behavior.

- The function $\alpha(i,j,t)$ strictly depends on the task which the subject is acting, and it is defined as:

$$\alpha(i,j,t) = F(i,j)g(t), \tag{2}$$

where $F = F(i,j)$ is the matrix of the synthetic functional connectivity, and $g(t)$ is a time dependent function which corrects its evolution during life. In [9] there is a detailed description of how $g(t)$ should be evaluated as a function of various periods of life, according to the natural evolution of neural architecture. In this study we use a simplified approach that does not require to consider this time dependence, even though, for a deeper analysis, it would be very interesting to focus on the role of t to accurately model the functional activity, verifying the age of the subject.

So, after removing time dependence, and taking into account the previous considerations, the model simplifies as follows: $W = [W(i,j)]$ where

$$W(i,j) = deg(i)deg(j)e^{-\eta D(i,j)+F(i,j)}. \tag{3}$$

Of course, in case one aims at comparing the collected $fMRI$ data with the entries provided by the FD-model, Equation (3) must be normalized by $\max(W)$. Since the present work has a different focus, we prefer to avoid normalization in view of a clearer presentation of our results.

Now, the quantities in Equation 3 will be further discussed, and the tools necessary to their definition will be provided.

2.2. Going deeper into the FD-model

We recall that the aim of our work is to construct and analyze the weighted undirected graph $G = (V, E)$ proposed by Dulio and Finotelli, that models the structure of brain activity, where V is the set of nodes obtained by parcelling the gray matter depending on anatomical and functional differences. There is an arc $e = \{i, j\} \in E$ between the nodes i and j if and only if the two regions are supposed to activate similarly, and the weight of the arc $w(e)$ represents the functional connectivity that describes statistical patterns of dynamic interactions among regions [16].

We show how to uniform the values of the matrices in Equation (3): F represents the functional brain connectivity, so its entries are real values in $[-1, 1]$, that we restrict to $[0, 1]$ due to the fact that, at present, there is not an unanimous agreement about the neurobiological interpretation of the negative correlation indices. On the other hand, D contains the Euclidean distances between the nodes that usually are expressed in millimeters. Therefore, to obtain homogeneous values acting as exponent of Equation (3), the suitable value of $\eta = 8$ needs to be conveniently scaled. Moreover, we underline that the two matrices must be acquired from the same subject, and their corresponding elements must refer to the same cerebral area. Only after such a process, it is possible to compute the values of W and find a biological interpretation to the obtained undirected and loopless weighted graph.

From here on, we will refer to the graph and to its adjacency matrix without distinction, since they represent the same mathematical object.

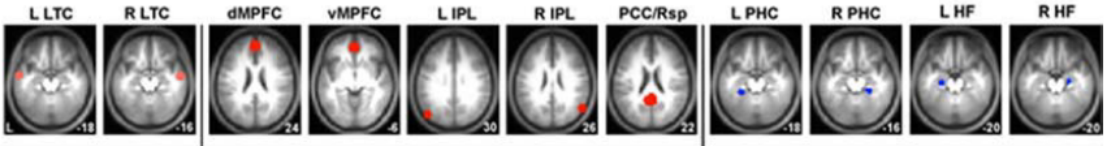
We point out that the connectome matrix D we use for the model implementation is available on the Groupe d'Imagerie Neurofonctionnelle's website (http://www.gin.cnrs.fr/fr/outils/aal-aal2/). In order to find the spatial coordinates of the centers, we evaluate the barycenter of the corresponding region.

On the other hand, the matrix F represents the functional connectivity of a subject under a specific task. In this paper a simulation of the functional activity at resting state will be created and discussed.

As a proposal for further studies, it would be interesting to simulate the functional activity of a specific task using knowledge available on neurobiological literature. However, of primary interest will be to insert a real functional matrix both to estimate the η value for humans and to examine in detail the properties of the model.

The Default Mode Network (DMN) comprehends those regions of the brain that are most active at resting state. Referring to Table 1, where the correlations between the DMN's regions are estimated, we notice that the values range in the interval $[0, 0.6]$. Since those are the mostly interconnected nodes, it is reasonable to assume 0.6 to be the maximum value in the whole network. Therefore, in order to simulate the matrix F, we simulate its random values in the range of $[0, 0.6]$, satisfying all the previously mentioned constraints. Finally, to make F more reliable, we substitute the values on the block of the DMN with the correlations (see Table 1) evaluated by Buckner et al. in [5]. Observe that there are also some negative values: the presence of negative correlation in functional activity is still debated in neuroscientific community, and it has not a unique interpretation so, following the literature, we decide to set the negative values to 0.

Table 1. Interaction between Default Mode Network's areas. The abbreviations L LTC and R LTC stand for left and right lateral temporal cortex; dMPFC and vMPFC refer to the dorsal and ventral medial prefrontal cortex; L IPL and R IPL are the left and the right part of the inferior parietal lobule; PCC/Rsp is the area that includes the posterior cingulate cortex and the retrosplenial cortex; L PHC and R PHC stand for left and right parahippocampal cortex; L HF and R HF are the left and the right part of the hippocampal formation. Image is from [5].



	L LTC	R LTC	dMPFC	vMPFC	L IPL	R IPL	PCC/Rsp	L PHC	R PHC	L HF	R HF
L LTC	1.00	**0.41**	**0.16**	**0.12**	**0.14**	**0.12**	**0.12**	**0.11**	**0.06**	**0.18**	**0.14**
R LTC		1.00	**0.16**	**0.18**	**0.07**	**0.20**	**0.19**	**0.08**	**0.10**	**0.15**	**0.17**
dMPFC			1.00	**0.47**	**0.22**	**0.31**	**0.34**	**-0.06**	**-0.10**	**-0.01**	**-0.04**
vMPFC				1.00	**0.27**	**0.31**	**0.52**	**0.11**	**0.06**	**0.20**	**0.16**
L IPL					1.00	**0.47**	**0.49**	**0.25**	**0.10**	**0.11**	**0.06**
R IPL						1.00	**0.42**	**0.12**	**0.05**	**0.09**	**0.07**
pCC/Rsp							1.00	**0.23**	**0.16**	**0.26**	**0.21**
L PHC								1.00	**0.57**	**0.31**	**0.28**
R PHC									1.00	**0.28**	**0.28**
L HF										1.00	**0.61**
R HF											1.00

Still referring to the resting state, Moussa et al. in [14] distinguish other areas that active weakly, nevertheless significantly: the Sensory Motor Network (SMN), the visual cortex and the basal ganglia. In literature their correlation indices are not yet estimated, anyway, to highlight their importance in resting state activity, we decide to set to zero the entries greater than 0.4 in the corresponding blocks. Finally, in the remaining areas of the matrix we delete all values over 0.2.

In Fig. 1 it is provided a graphic description of the structure of the obtained matrix simulating functional connectivity at resting state.

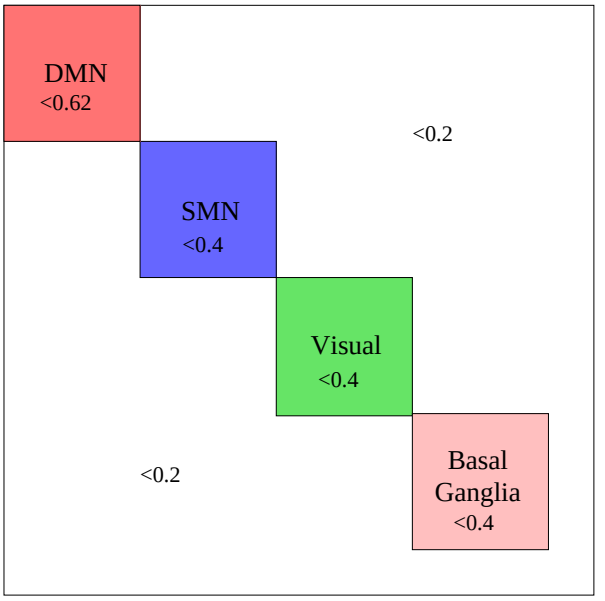


Figure 1. The structure of the simulated functional matrix: the values of the DMN are estimated in Tab. 1, the blocks of the SMN, visual cortex and basal ganglia have values lower than 0.4, while the remaining values of the matrix are lower than 0.2.

3. Description and validation of the chosen adaptive thresholding method

Now, we consider the problem of finding a suitable strategy to define a cutoff threshold, in order to eliminate the weakest connections in W, which are most likely to be associated to experimental noise, and to reduce the density of the graph to make it computationally tractable. This implies to find a value ξ for the entries of W and to set to zero all the values such that $W(i,j) \leq \xi$. Obviously, the higher the threshold, the greater is the probability to eliminate noisy connections, but, if too large, the risk to loose characterizing properties as being small-world becomes much higher [1]. So, the selection of a suitable threshold is a crucial point on the analysis of the data. Many different methods have been proposed during these last years. In [1], Achard et al. examine small-world properties of the network according to different values of the threshold. They observe that, as the threshold grows, the network becomes sparsely connected (the mean degree decreases), fragmented (the largest cluster size decreases), while the small-world ratio increases. Therefore, a method to select a suitable threshold is to choose a value that preserves the brain network properties. A statistical approach can also be used: It is frequent to make a priori choices of the functional connectivity matrix's threshold on the range of $[0.3, 0.5]$. A threshold over 0.5 is not recommended since the graph has shown to increase its randomness [2]. However, such method neglects all entries under a fixed value, without considering the distribution of the matrix's values, with a consequent possible loss of some significant connections. A different statistical method has been proposed by Finotelli et al. in [10], with the aim of reducing the loss of possible meaningful small data related to the specific task. They compare the collected data

with a database concerning a different task, assumed as baseline. Using their method, the threshold is selected by analyzing the distribution of the specific dataset, thus resulting a more accurate and task-dependent choice. It is worth mentioning the approach in [4], where Bordier et al. apply the statistical physics' method of percolation analysis to graph theory, and they obtain a near optimal threshold for brain networks. Proper of their method is the study of the changes on the connected subgraphs according to growing threshold values. Hence, they define the percolation threshold as the break point of the largest connected component, that, in their opinion, is near to the optimal choice in order to eliminate the weakest connections without deleting significant ones.

In the sequel we introduce a new technique to set a threshold to the matrix W as in Equation (3), that bases on the comparison between the matrices generated at different values of the cutoff. To evaluate the differences between the related graphs it is possible to use similarity measures [8], that quantifies the degree of similarity of a matrix with another.

Select a similarity measure between matrices

According to [13] and [7], among the most relevant (and commonly used) similarity measures between two points $\overline{x} = (x_1, \ldots, x_n)$ and $\overline{y} = (y_1, \ldots, y_n)$ in $\mathbb{R}^n$ we have, for $i = 1, \ldots, n$:

1. Jaccard measure
$$d(\overline{x}, \overline{y}) = \frac{\sum x_i y_i}{\sum x_i^2 + \sum y_i^2 - \sum x_i y_i}; \tag{4}$$

2. Cosine measure
$$d(\overline{x}, \overline{y}) = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2}\sqrt{\sum y_i^2}} = \frac{< \overline{x}, \overline{y} >}{\|\overline{x}\|_2 \|\overline{y}\|_2} = \cos\theta, \tag{5}$$
where θ is the angle between $\overline{x}$ and $\overline{y}$;

3. Dice measure
$$d(\overline{x}, \overline{y}) = \frac{\sum x_i y_i}{\sum x_i^2 + \sum y_i^2}; \tag{6}$$

4. Ruzicka measure
$$d(\overline{x}, \overline{y}) = \frac{\sum \min\{x_i, y_i\}}{\sum \max\{x_i, y_i\}}. \tag{7}$$

We decide to use Ruzicka similarity to compare two matrices of dimensions $n \times m$, regarded as two points in $\mathbb{R}^{n \times m}$. Basically, this choice is motivated by the fact that this measure combines quantitative similarity of spatial correspondent points with the detection of structural and topological properties, fitting in this way the purposes of this study. It is of particular relevance for the study of FD-model that each thresholded value does not affect the distance (providing a notion of stability to the model) and that it has a non-linear behavior on the values' range in order to stress similarities on the higher (and more relevant since related to state or task characteristics) ones, and ignoring small noise perturbations.

As previously mentioned, the idea is to evaluate the degree of similarity of couple of matrices obtained two by two from a starting one when varying the thresholds. To establish the values of these

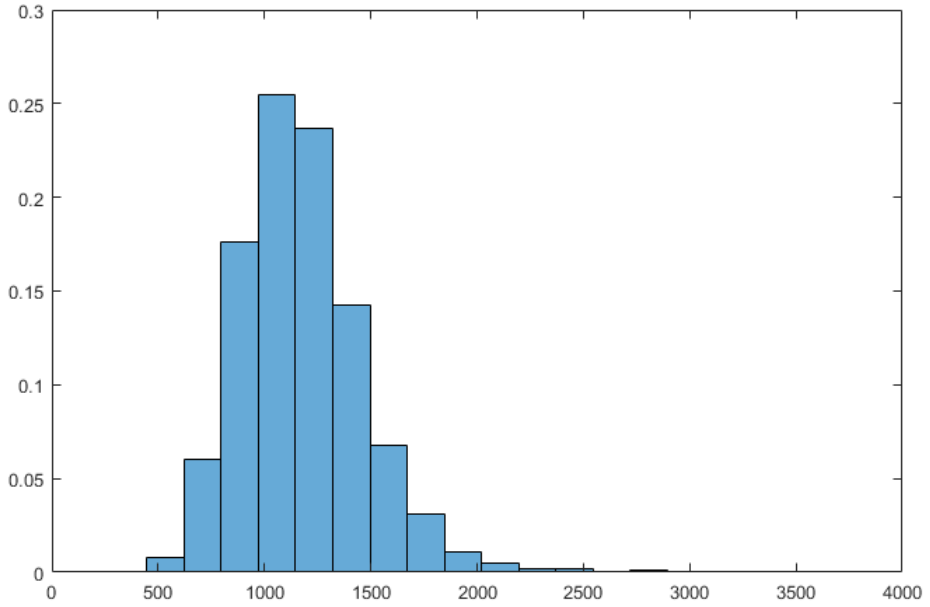


Figure 2. The histogram represents the percentage of arcs with weights in a specific range. Values have been grouped in twenty bins, but, as it can be observed from the histogram, tails have strongly non-significant percentages.

thresholds, the interval of entries of W, obtained form Equation (3), is divided into 20 subintervals of the same width (said *bins*), as showed in Fig. 2.

We create an iterative algorithm to subsequently set the threshold of W on the maximal value of every bin: At each step, it is computed the Ruzicka similarity between the matrix at the previous step and the matrix obtained using the current cut-off, until the graph disconnects.

The presence of more than one connected component, is interpreted as the fragmentation of the brain network into two or more parts that cannot communicate. This situation can model for example ischemia, brain tumor, or brain injuries resulting from a traumatic episode. In this study we are assuming to deal with healthy subject, that's why we suppose the brain network to be connected.

The evolution of Ruzicka similarity as function of the different cut-offs, for 10 simulations, is illustrated in Fig. 3. We observe that all the simulations follow the same evolution, supporting the initial hypothesis that all the matrices generated according to this model must show the same behavior.

Select the threshold

Observing Fig. 3 it is clear that the Ruzicka similarity index remains almost the same for low threshold values, while from a certain point onwards it undergoes a rapid variation, since a substantial part of the arcs are cut. According with this consideration, the threshold point is selected determining the maximum variation on the slope of the Ruzicka similarity's graph. Motivation for this choice relies

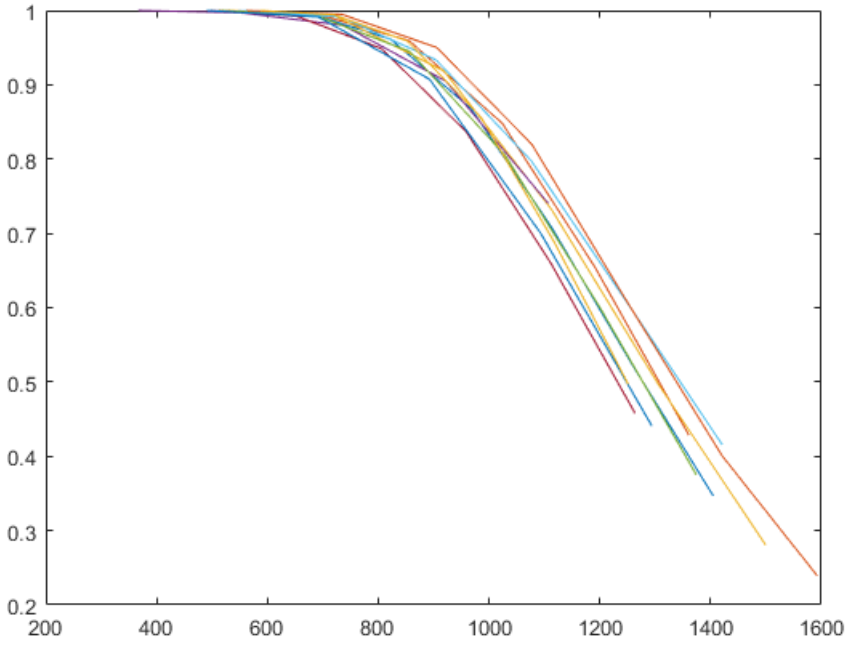


Figure 3. The evolution of Ruzicka similarity on 10 matrices computed on consecutive thresholds. On the x axis there are the values of the thresholds, while on the y one the corresponding Ruzicka similarity index between two consecutive thresholded matrices.

on the fact that it is suppose that to select a previous cutoff could not guarantee to neglect enough elements, and noisy data could still remain. On the other hand, a greater threshold could imply the loss of some relevant values.

Just to give an example, Figure 4 plots one of the graphs after the threshold. The nodes are spatially located using their coordinates (x, y, z); the nodes' colors are assigned based on their degree: A yellow node has higher degree, showing its attitude as potential hub.

NON LEGGIBILE ? IL NUMERO DEGLI ARCHI MOSTRATI E' GIA' STATO FORTEMENTE RIDOTTO, MIGLIORARE ULTERIORMENTE LA FIGURA NON E' POSSIBILE. HO PROVATO ANCHE A METTERE ARCHI PIU' O MENO SPESSI IN BASE AL PESO MA RISULTA COMUNQUE UNA FIGURA POCO LEGGIBILE

LA FIGURA CONTINUA A NON ESSERE LEGGIBILE. SI POTREBBE METTERLA DA SOLA IN UNA PAGINA E CERCARE DI MAGNIFICARLA. FORSE LA SITUAZIONE MIGLIOREREBBE. AD OGNI MODO NELLA CAPTION HO PREMESSO CHE QUEL CHE APPARE E' IL MASSIMO CHE SI POSSA FARE.

In Table 2 is shown the percentage of neglected arc and the threshold value for 20 simulated matrices.

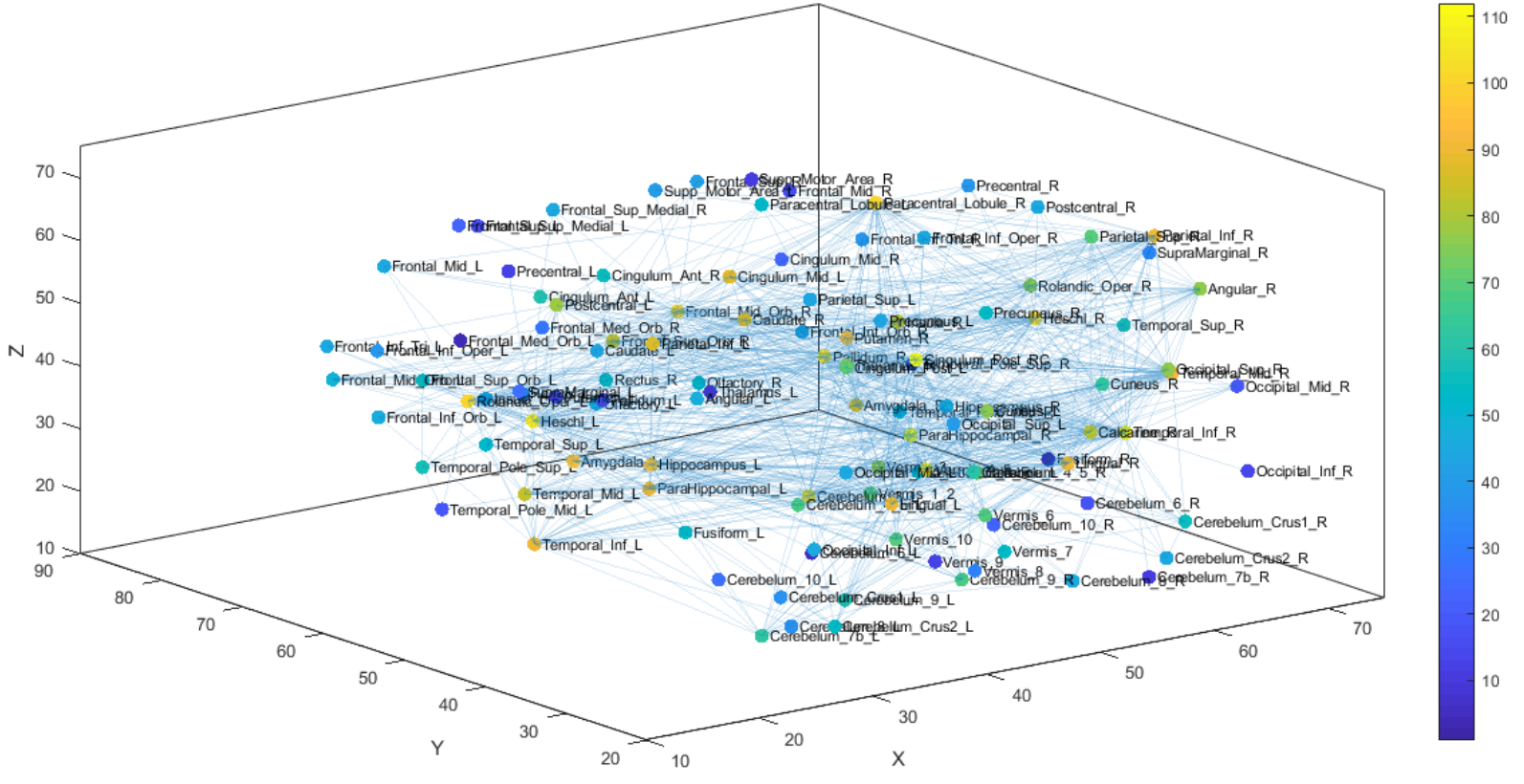


Figure 4. The plot of a simulated graph after the threshold. In order to obtain a better visualization only those arcs having high weights are shown. The nodes' colors indicate their degrees. We are aware that the figure is dense, but what shown is the result of a severe selection of the links originally present.

Table 2. The table contains the percentage of cut edges and the value of the threshold for 20 matrices.

Percentual	**Threshold**	**Percentual**	**Threshold**
0.15	890.41	0.23	1021.55
0.33	1136.395	0.36	1089.99
0.36	1039.67	0.24	956.25
0.28	1075.76	0.24	983.74
0.41	1100.69	0.29	1036.72
0.29	1064.67	0.34	1070.87
0.36	1119.49	0.20	945.29
0.20	1024.05	0.37	1100.70
0.21	928.28	0.26	976.65
0.31	1118.33	0.61	1303.23

4. Discovering similarities

It is natural to expect that different subjects with similar characteristics (e.g. age, lifestyle, pathologies, ...) have similar W matrices.

For this reason, we generate 200 couples of simulated matrices W, apply to them the thresholding procedure and compare the resulting matrices using the Ruzicka similarity measure. The Ruzicka similarity index generally ranges from 0 to 1, while the mean of the values obtained from this experiment is ≈ 0.49, with a standard deviation of ≈ 0.06. We point out that the Ruzicka index is however relevant, showing that the thresholding process, step after step, does not cause heavy changes in the network characteristics. Nevertheless, some values are expected to be higher, or at least ranging in a narrow interval.

To provide evidence to the clustering potential of the similarity measure, we proceed in comparing the model's networks to random graphs, mainly searching for structural differences. To this purpose, each simulated matrix W is compared, again with the Ruzicka index, to a random matrix R_W obtained as follows: Its entries are randomly generated on the same range of those in W; R_W is symmetric with main diagonal null; the same number of deleted element in W after the threshold are set to zero, choosing their position randomly, in order to better reflect the structure of W. The results of 200 of these comparisons, again after thresholding the matrices, show, in general, lower values in comparison to those in the previous comparison, with a mean of ≈ 0.30 and a standard deviation of ≈ 0.09 . From the results obtained it is possible to conclude that the similarity index could not perfectly separate random matrices from those obtained by applying the FD-model, since the intersection between the two intervals of similarity is not void. Though, it is important to observe that the compared matrices are very similar in structure (they have the same number of zeros, the same entries' range and they both are symmetric with null main diagonal) and the clear difference between the average values allows to state that the Ruzicka similarity can be stressed to help in identify their structural differences. This result opens a window onto further studies for detecting brain discrepancies. Indeed, it seems

appropriate to conjecture that the model matrices, computed on real data from healthy subjects, and on data of subjects with mental illness or brain injuries, may substantially differ when compared by using Ruzicka similarity.

5. Brain Networks' properties and their validation

Various graph measures are used to make statistical analysis of the brain network; they can be classified as (see [16]):

- *Segregation*: to distinguish the tendency of the network's elements to group together in clusters or cliques. Among them, the most used is the node's *clustering coefficient*, which measures the density of connections between its topological neighbors. The average of clustering coefficients overall the network's nodes measures the network's degree of segregation. Nodes that are very interconnected are said to compose a *cluster*, and they are most likely to share information;
- *Integration*: global measures to underline the network's property of spreading information. One of the most used is the *characteristic path length*, that is the average of the shortest path lengths among all the couples of nodes. Short paths allow the efficient diffusion of information through the graph, avoiding its dispersion;
- *Influence*: to detect the importance of a specific node on the diffusion of the information. The simplest and most used measure is the node *degree*. Nodes with higher influence measure than the others are said to be *hubs*.

It is remarkable that all studies of human brain networks revealed its *small-world* structure, characterized by a high clustering coefficient, like regular lattices, and low characteristic path length, like random networks [19]. This means that the system presents segregated and integrated characteristics and this important property contributes to a both economical and efficient distribution of the information among the network. Few for all networks showing small world architecture are the World Wide Web, the Kevin Bacon Graph (where nodes are actors and there's a link between to actors if they've acted in the same film) and the electrical power grid (the network that models the electricity distribution from generating stations to consumers).

High clustering coefficient of the network is mainly due to its modularity, namely its organization into *modules*: nodes belonging to the same module are highly connected, while connections between different modules are not predominant. Modularity favors information flow between nodes and areas dealing with the same functional activity, ensures that each region keeps its proper function, and avoids that information spreads throughout the network [16].

Different modules are interconnected by means of *connector hubs* [9], which play a key role on integration. Interconnected hubs are said to be a *rich club*. Even inside a module important and highly connected nodes can be identified, taking the name of *provincial hubs*. There is not a unique way to identify hubs in a network, but surely high degree is one of the principal methods.

Another important property of the brain network model is the *scale-freedom*, i.e. its nodes' degree distribution follows a power-law form [18], that is $P(k) \sim k^{\alpha}$. In a scale-free network most of the nodes have very low degree while only few ones have many link connections.

For a set of 100 artificial data simulated using the formula 3, we calculate the following statistics:

- The *average clustering coefficient*, in order to evaluate the networks' degree of segregation. It is defined as:
$$C = \frac{\sum_{v \in V} c(v)}{|V|}, \tag{8}$$
where $c(v)$ is the clustering coefficient of a node, that is the fraction between the number of triangles on the graph and the number of possible triangles involving the node and it quantifies the number of neighbours of the node that are also mutually neighbours [2] .
- The *average degree* $D = 2\frac{|E|}{|V|}$, which is the average number of edges per node;
- The *characteristic path length* L, namely the average length of the shortest path between each pair of nodes;
- The *density*, that is the fraction between the average degree and $|V| - 1$. It is a measure to investigate the network's segregation too.

From these simulations, the average degree D has a mean of ≈ 71.50 and a standard deviation of ≈ 15.58, the latter information meaning that this value ranges on a wide interval: this depends on the fact that the threshold is adapted to each matrix and the percentage of arcs to be eliminated is not fixed. Moreover, being the matrices all of dimension 116×116, the maximum degree of each node is 115; in the 12% of the cases the average degree was bigger than 90: these are situations in which the algorithm for thresholding has not been successful on deleting a sufficient number of arcs, returning a graph that is still approximately a complete graph. However, the average of the evaluations of D is a reasonable value. Concerning to the clustering coefficient, its mean is ≈ 0.86 with standard deviation of ≈ 0.03. Its value close to 1 can be interpreted as representative of a network strongly interconnected where the information flows efficiently. Regarding to the network density, i.e. the proportion of possible neighbors in the network that are actually connected, the values obtained have its average ≈ 0.62 and a standard deviation of ≈ 0.14. This mean value lies approximately in the middle of the density range, i.e. $[0, 1]$, so the network can not be defined neither sparse either dense. Finally, observing the mean of the average path length, that is $\approx 1,37$, and its standard deviation of ≈ 0.14 , we can once again realize that the network is composed of many cerebral areas well connected.

Observing the nodes' degree, we identified hubs highlighting one node each time that its degree is over 95. As expected, the nodes in the Default Mode Network have greater probability to be a hub, around the 68%; the nodes belonging to the SMN, the Visual Network and the Basal Ganglia, instead, have a smaller, even if still high, probability of the 0.47% to be identified as hubs; for the remaining nodes, the probability remains quite low (around the 25%). So, we can conclude that the network computed from the model still has the structure, shown in Fig. 1, that has been defined for the synthetic functional matrix, and this provides evidence to the good behavior of the model by Dulio and Finotelli for the resting state.

In a final analysis we examine the small- world property of the model matrices. To this aim, we compare the synthetic networks W arising from the model after the threshold, with a random network W_R with the same number of nodes, the same average degree d and the same degree distribution.

Following the literature, to create a random graph with the same degree sequence as in a given graph G, we used the *configuration model*. Being L_W and L_R the characteristic path lengths, C_W and C_R the clustering coefficients respectively of the synthetic matrix and the random matrix, the index σ is defined as:

$$\sigma = \frac{\frac{C_W}{C_R}}{\frac{L_W}{L_R}}. \tag{9}$$

The network's model has the small-world property (see [1]) if $\sigma > 1$; we got this result for all the 100 simulations we performed, proving the small-wordness of the synthetic network, that guarantees the information to flow both locally and globally efficiently, as expected in neural networks [11].

6. Concluding remarks and perspectives

The aim of this paper is the implementation of the model by Dulio and Finotelli in [9] using synthetic data for the resting state functional activity, and the identification of a similarity measure between matrices that allows to group on subjects similar for age, life-style, pathologies, and so on.

Another preeminent goal concerns the investigation on a thresholding method to neglect the weakest connections. Such a thresholding method acts dynamically on each computed matrix by detecting the most suitable value in order to maintain the characteristics of the analyzed matrix, neglecting in this way the entries that can be associated to background noise, very weak functional activity as well as artifacts (in case real data are considered). Ruzicka similarity is applied to compare the same matrix when increasingly thresholded, and, basing on the obtained results, a suitable threshold is selected in presence of the maximal variation of the Ruzicka similarity, i.e. where the matrix shows the greatest change after consecutive thresholds. Such a strategy reveals its innovative power: It prevents to either remove not enough useless elements, or delete meaningful values.

To show the accuracy of the model, pairs of model matrices are compared. Further, another comparison is made with random matrices, by considering once again Ruzicka similarity. As expected, the obtained results show, on average, a higher degree of similarity among the matrices of the model rather than with randomly generated ones. In the end, the most relevant statistics on brain networks, according to literature, are analyzed, showing that Dulio and Finotelli's model allows to generate graphs that follow the expected structural organization.

Further studies will include the time dependence in the model, in particular considering the anatomical matrix changing in time, in order to better model brain changes during lifetime, and to test the statistics introduced in this paper to interpret the differences. Moreover, all the results obtained here are based on synthetic data at resting state: In next studies, it will be of primary relevance to perform the same analysis on synthetic data simulating a specific task or on real data. Ultimately, there are several long term projects to be considered. For instance, one of them consists in using real data to identify a suitable value for the distance penalty η for human subjects since, at present, its exact value has not been identified. Another project should include in the analysis the study of negative functional correlation indices. Indeed, at present only the non negative entries of the functional connectivity matrix are considered, this is due to the fact that, inside the neuroscientific community there is not a complete agreement about the neurobiological meaning of such negative correlations. Finally, the

study of the functional (and anatomical) modularity of the human brain at the resting state, as well as during cognitive tasks, could be a line of research of great interest.

References

[1] Achard S., Salvador R., Whitcher B., Suckling J., Bullmore E. A Resilient, Low-Frequency, Small-World Human Brain Functional Network with Highly Connected Association Cortical Hubs, *The Journal of Neuroscience*, 2016. **26**(1):63-72. doi: 10.1371/journal.pone.0044428.

[2] Alexander-Bloch A.F., Gogtay N., Meunier D., Birn R., Clasen L., Lalonde F., Lenroot R., Giedd J., Bullmore E.T. Disrupted modularity and local connectivity of brain functional networks in childhood-onset schizophrenia. *Frontiers in systems neuroscience*, 2010. **4**(147). doi: 10.3389/fnsys.2010.00147.

[3] Betzel R.F., Avena-Koenigsberger A., Goñi J., He Y., de Reus M.A., Griffa A., Vèrtes P.E., Mišic B., Thiran J.P., Hagmann P., van den Heuvel M., Zuo X.N., Bullmore E.T., Sporns O. Generative models of the human connectome. *Neuroimage*, 2016. **124**(Pt A): 1054-1064. doi: 10.1016/j.neuroimage.2015.09.041.

[4] Bordier C., Nicolini C., Bifone A. Graph Analysis and Modularity of Brain Functional Connectivity Networks: Searching for the Optimal Threshold. *Frontiers in Neuroscience* 2017. **11**(441). doi: 10.3389/fnins.2017.00441.

[5] Buckner R.L., Andrews-Hanna J.R., Schacter D.L.. The Brain's Default Network: Anatomy, Function, and Relevance to Disease. *Annals of the New York Academy of Sciences*, 2008. **1124**: 1-38. doi: 10.1196/annals.1440.011.

[6] Bullmore E., Sporns O. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 2009. **10**(3): 186-198. doi: 10.1038/nrn2575.

[7] Cha S.H. Comprehensive Survey on Distance/Similarity Measures between Probability Density Functions *International Journal of Mathematical Models and Methods in Applied sciences*, 2007. **4**(1). doi: 10.1.1.154.8446.

[8] Deza M.M., and Deza E. Encyclopedia of Distances. Springer-Verlag Berlin Heidelberg, 2010. doi: 10.1007/978-3-662-44342-2. 4^{th} Edition.

[9] Finotelli P., Dulio P. A mathematical model for evaluating the functional connectivity strongness in healthy people. *Archives Italiennes de Biologie*, 2015. **153**(4): 279-300. doi: 10.12871/00039829201544.

[10] Finotelli P., Dulio P., Varotto G., Rotondi F., Panzica F. A statistical proposal for selecting a data-depending threshold in neurobiology. *Archives Italiennes de Biologie*, 2016. **154**(2-3):78-101. doi: 10.12871/00039829201625.

[11] Latora V., Marchiori M. Efficient Behavior of Small-World Networks. *Physical Review Letters* 2001. **87** (19) 198701: 1-4. DOI: 10.1103/PhysRevLett.87.198701

[12] Passingham R. How good is the macaque monkey model of the human brain? *Current Opinion in Neurobiology*, 2009. **19**(1):6-11. doi: 10.1016/j.conb.2009.01.002.

[13] Manjeet Singh J. Relevancy Measurement of Retrieved Webpages Using Ruzicka Similarity Measure. *International Journal of Engineering Research & Technology*, 2014. **3**(7) 1520-1523. ISSN:2278-0181.

[14] Moussa M.N., Steen M.R., Laurienti P.J., Hayasaka S. Consistency of Network Modules in Resting-State fMRI Connectome Data. *PLoS ONE* 2012. **7**(8), doi: 10.1371/journal.pone.0044428.

[15] Obando C., de Vico Fallani F. Graph Models of Brain Connectivity Networks. In: Annual Meeting of the Organization for Human Brain Mapping, 2016.

[16] Sporns O. Structure and function of complex brain networks. *Dialogues in Clinical Neuroscience*, 2013. **15**(3): 247-262.

[17] Vèrtes P.E., Alexander-Bloch A.F., Gogtay N., Gieddb J.N., Rapoport J.L., Bullmorea E.T. Simple models of human brain functional networks. In: Proceedings of the National Academy of Sciences of the United, 2012. **109**(15) 5868-5873. doi: 10.1073/pnas.1111738109.

[18] Xiao W., Guanrong C. Complex networks: Small-world, scale-free and beyond. *Circuits and Systems Magazine, IEEE*, 2003. **3** 6-20.

[19] Watts D.J., Strogatz S.H. Collective dynamics of 'small-world' networks. *Nature* 1998. **393** 440442.