


# SCIENTIFIC REPORTS



OPEN

## An ANOVA approach for statistical comparisons of brain networks

Daniel Fraiman <sup>1,2</sup> & Ricardo Fraiman<sup>3,4</sup>

The study of brain networks has developed extensively over the last couple of decades. By contrast, techniques for the statistical analysis of these networks are less developed. In this paper, we focus on the statistical comparison of brain networks in a nonparametric framework and discuss the associated detection and identification problems. We tested network differences between groups with an analysis of variance (ANOVA) test we developed specifically for networks. We also propose and analyse the behaviour of a new statistical procedure designed to identify different subnetworks. As an example, we show the application of this tool in resting-state fMRI data obtained from the Human Connectome Project. We identify, among other variables, that the amount of sleep the days before the scan is a relevant variable that must be controlled. Finally, we discuss the potential bias in neuroimaging findings that is generated by some behavioural and brain structure variables. Our method can also be applied to other kind of networks such as protein interaction networks, gene networks or social networks.

Understanding how individual neurons, groups of neurons and brain regions connect is a fundamental issue in neuroscience. Imaging and electrophysiology have allowed researchers to investigate this issue at different brain scales. At the macroscale, the study of brain connectivity is dominated by MRI, which is the main technique used to study how different brain regions connect and communicate. Researchers use different experimental protocols in an attempt to describe the true brain networks of individuals with disorders as well as those of healthy individuals. Understanding resting state networks is crucial for understanding modified networks, such as those involved in emotion, pain, motor learning, memory, reward processing, and cognitive development, among others. Comparing brain networks accurately can also lead to the precise early diagnosis of neuropsychiatric and neurological disorders<sup>1,2</sup>. Rigorous mathematical methods are needed to conduct such comparisons.

Currently, the two main techniques used to measure brain networks at the whole brain scale are Diffusion Tensor Imaging (DTI) and resting-state functional magnetic resonance imaging (rs-fMRI). In DTI, large white-matter fibres are measured to create a connectional neuroanatomy brain network, while in rs-fMRI, functional connections are inferred by measuring the BOLD activity at each voxel and creating a whole brain functional network based on functionally-connected voxels (i.e., those with similar behaviour). Despite technical limitations, both techniques are routinely used to provide a structural and dynamic explanation for some aspects of human brain function. These magnetic resonance neuroimages are typically analysed by applying network theory<sup>3,4</sup>, which has gained considerable attention for the analysis of brain data over the last 10 years.

The space of networks with as few as 10 nodes (brain regions) contains as many as  $10^{13}$  different networks. Thus, one can imagine the number of networks if one analyses brain network populations (e.g. healthy and unhealthy) with, say, 1000 nodes. However, most studies currently report data with few subjects, and the neuroscience community has recently begun to address this issue<sup>5-7</sup> and question the reproducibility of such findings<sup>8-10</sup>. In this work, we present a tool for comparing samples of brain networks. This study contributes to a fast-growing area of research: network statistics of network samples<sup>11-14</sup>.

We organized the paper as follows: In the Results section, we first present a discussion about the type of differences that can be observed when comparing brain networks. Second, we present the method for comparing brain networks and identifying network differences that works well even with small samples. Third, we present an example that illustrates in greater detail the concept of comparing networks. Next, we apply the method to resting-state fMRI data from the Human Connectome Project and discuss the potential biases generated by some behavioural and brain structural variables. Finally, in the Discussion section, we discuss possible improvements, the impact of sample size, and the effects of confounding variables.

<sup>1</sup>Departamento de Matemática y Ciencias, Universidad de San Andrés, Buenos Aires, Argentina. <sup>2</sup>Consejo Nacional de Investigaciones Científicas y Tecnológicas, Buenos Aires, Argentina. <sup>3</sup>Centro de Matemática, Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay. <sup>4</sup>Instituto Pasteur de Montevideo, Montevideo, Uruguay. Correspondence and requests for materials should be addressed to D.F. (email: [dfracman@udesa.edu.ar](mailto:dfracman@udesa.edu.ar))

Received: 13 November 2017

Accepted: 6 March 2018

Published online: 16 March 2018

## Results

**Preliminars.** Most studies that compare brain networks (e.g., in healthy controls vs. patients) try to identify the subnetworks, hubs, modules, etc. that are affected in the particular disease. There is a widespread belief (largely supported by data) that the brain network modifications induced by the factor studied (disease, age, sex, stimulus) are *specific*. This means that the factor will similarly affect the brains of different people.

On the other hand, labeled networks can be modified in many different ways while preserving the nodes, and these modifications can be categorized into three. In the first category, called here *localized modifications*, some particular identified links suffer changes by the factor. In the second, called *unlocalized modifications*, some links change, but the changed links differ among subjects. For example, the degree of interconnection of some nodes may decrease/increase by 50%, but in some individuals, this happens in the frontal lobe, in others in the right parietal lobe or the occipital lobe, and so on. In this case, the localization of the links/nodes affected by the factor can be considered random. In the third category, called here *global modifications*, some links (not the same across subjects) are changed, and these changes produce a global alteration of the network. For example, they can notably decrease/increase the average path length, the average degree, or the number of modules, or just produce more heterogeneous networks in a population of homogeneous ones. This last category is similar to the unlocalized modifications case, but in this case, an important global change in the network occurs.

In all cases, there are changes in the links influenced by the “factor”, while nodes are fixed. How to detect if any of these changes have occurred (hereinafter called detection) is one of the main challenges of this work. And, once their occurrence has been determined, we aim to identify where they occurred (hereinafter called identification). The difficulty lies in statistically asserting that the factor produced true modifications in the huge space of labeled networks. We aim to detect all three types of network modifications. Clearly, as is always true in statistics, more precise methods can be proposed when hypotheses regarding the data are more accurate (e.g., that the differences belong to the global modifications category). However, this last approach requires one to make many more assumptions about the brain’s behaviour. The assumptions are generally unverifiable; for this reason, we use a nonparametric approach, following the adage “less is more”, which is often very useful in statistics. For the detection problem, we developed an analysis of variance (ANOVA) test specifically for networks. As is well known, ANOVA is designed to test differences among the means of the subpopulations, and one may observe that equal means have different distributions. However, we propose a definition of means that will differ in the presence of any of the three modification categories mentioned above. As is well known, the identification stage is computationally far more complicated, and we address it partially looking at the subset of links or a subnetwork that present the highest network differences between groups.

**Network Theory Framework.** A network (or graph), denoted by  $G = (V, E)$ , is an object described by a set  $V$  of nodes (vertices) and a set  $E \subset V \times V$  of links (edges) between them. In what follows, we consider families of networks defined over the same fixed finite set of  $n$  nodes (brain regions). A network is completely described by its adjacency matrix  $A \in \{0, 1\}^n \times n$ , where  $A(i, j) = 1$  if and only if the link  $(i, j) \in E$ . If the matrix  $A$  is symmetric, then the graph is undirected; otherwise, we have a directed graph.

Let us suppose we are interested in studying the brain network of a given population, where most likely brain networks differ from each other to some extent. If we randomly choose a person from this population and study his/her brain network, what we obtain is a random network. This random network,  $\mathbf{G}$ , will have a given probability of being network  $G_1$ , another probability of being network  $G_2$ , and so on until  $G_{\tilde{n}}$ . Therefore, a random network is completely characterized by its probability law,

$$p_k := \mathbb{P}(\mathbf{G} = G_k) \text{ for all } k \in \{1, \dots, \tilde{n}\}. \quad (1)$$

Likewise, a random variable is also completely characterized by its probability law. In this case, the most common test for comparing many subpopulations is the analysis of variance test (ANOVA). This test rejects the null hypothesis of equal means if the averages are statistically different. Here, we propose an ANOVA test designed specifically to compare networks.

To develop this test, we first need to specify the null assumption in terms of some notion of mean network and a statistic to base the test on. We only have at hand two main tools for that: the adjacency matrices of the networks and a notion of distance between networks.

The first step for comparing networks is to define a distance or metric between them. Given two networks  $G_1, G_2$  we consider the most classical distance, the edit distance<sup>15</sup> defined as

$$d(G_1, G_2) = \sum_{i < j} |A_{G_1}(i, j) - A_{G_2}(i, j)|. \quad (2)$$

This distance corresponds to the minimum number of links that must be added and subtracted to transform  $G_1$  into  $G_2$  (i.e. the number of different links), and is the  $L_1$  distance between the two matrices. We will also use equation (2) for the case of weighted networks, i.e. for matrices with  $A(i, j)$  taking values between 0 and 1. It is important to mention that the results presented here are still valid under other metrics<sup>16–18</sup>.

Next, we consider the average weighted network - hereinafter called the average network - defined as the network whose adjacency matrix is the average of the adjacency matrices in the sample of networks. More precisely, we consider the following definitions.

**Definition 1** Given a sample of networks  $\{G_1, \dots, G_l\}$  with the same distribution

- (a) The *average network*  $\mathcal{M}$  that has as adjacency matrix the average of the adjacency matrices

$$A_{\mathcal{M}}(i, j) = \frac{1}{l} \sum_{k=1}^l A_{G_k}(i, j), \tag{3}$$

which in terms of the population version corresponds to the *mean matrix*  $\mathcal{M}(i, j) = \mathbb{E}(A_G(i, j)) =: p_{ij}$ .  
 (b) The average distance around a graph  $H$  is defined as

$$\bar{d}_G(H) = \frac{1}{l} \sum_{k=1}^l d(G_k, H), \tag{4}$$

which corresponds to the mean population distance

$$\tilde{d}_G(H) = \sum_{i=1}^{\bar{n}} d(G_i, H) p_i. \tag{5}$$

With these definitions in mind, the natural way to define a measure of network variability is

$$\sigma := \bar{d}_G(\mathcal{M}), \quad \bar{\sigma} = \bar{d}_G(\tilde{\mathcal{M}}), \tag{6}$$

which measures the average distance (variability) of the networks around the average weighted network.

Given  $m$  subpopulations  $G^1, \dots, G^m$  the null assumption for our ANOVA test will be that the means of the  $m$  subpopulations  $\tilde{\mathcal{M}}_1, \dots, \tilde{\mathcal{M}}_m$  are the same. The test statistic will be based on a normalized version of the sum of the differences between  $\bar{d}_{G^i}(\mathcal{M}_i)$  and  $\bar{d}_G(\mathcal{M}_i)$ , where  $\bar{d}_{G^i}$  and  $\bar{d}_G$  are calculated according to (4) using the  $i$ -sample and the pooled sample respectively. This is developed in more detail in the next section.

*Detecting and identifying network differences.* DetectionNow we address the testing problem. Let  $G_1^1, G_2^1, \dots, G_{n_1}^1$  denote the networks from subpopulation 1,  $G_1^2, G_2^2, \dots, G_{n_2}^2$  the ones from subpopulation 2, and so on until  $G_1^m, G_2^m, \dots, G_{n_m}^m$  the networks of subpopulation  $m$ . Let  $G_1, G_2, \dots, G_n$  denote, without superscript, the complete pooled sample of networks, where  $n = \sum_{i=1}^m n_i$ . And finally, let  $\mathcal{M}_i$  and  $\sigma_i$  denote the average network and the variability of the  $i$ -subpopulation of networks. We want to test ( $H_0$ )

$$H_0: \tilde{\mathcal{M}}_1 = \tilde{\mathcal{M}}_2 = \dots = \tilde{\mathcal{M}}_m \tag{7}$$

that all the subpopulations have the same mean network, under the alternative that at least one subpopulation has a different mean network.

It is interesting to note that for objects that are networks, the average network ( $\mathcal{M}$ ) and the variability ( $\sigma$ ) are not independent summary measures. In fact, the relationship between them is given by

$$\sigma = 2 \sum_{i < j} A_{\mathcal{M}}(i, j) (1 - A_{\mathcal{M}}(i, j)). \tag{8}$$

Therefore, the proposed test can also be considered a test for equal variability. The proposed statistic for testing the null hypothesis is:

$$T := \frac{\sqrt{m}}{a} \sum_{i=1}^m \sqrt{n_i} \left( \frac{n_i}{n_i - 1} \bar{d}_{G^i}(\mathcal{M}_i) - \frac{n}{n - 1} \bar{d}_G(\mathcal{M}_i) \right), \tag{9}$$

where  $a$  is a normalization constant given in Supplementary Information 1.3. This statistic measures the difference between the network variability of each specific subpopulation and the average distance between all the populations and the specific average network. Theorem 1 states that under the null hypothesis (items (i) and (ii))  $T$  is asymptotically Normal(0, 1), and if  $H_0$  is false (item (iii))  $T$  will be smaller than some negative constant  $c$ . This specific value is obtained by the following theorem (see the Supplementary Information 1 for the proof).

**Theorem 1.** Under the null hypothesis, the T statistic fulfills (i) and (ii), while  $T$  is sensitive to the alternative hypothesis, and (iii) holds true.

- (i)  $\mathbb{E}(T) = 0$
- (ii)  $T$  is asymptotically ( $K = \min\{n_1, n_2, \dots, n_m\} \rightarrow \infty$ ) Normal(0, 1).
- (iii) Under the alternative hypothesis,  $T$  will be smaller than any negative value if  $K$  is large enough (The test is consistent).

This theorem provides a procedure for testing whether two or more groups of networks are different. Although having a procedure like the one described is important, we not only want to detect network differences, we also want to identify the specific network changes or differences. We discuss this issue next.

**Identification.** Let us suppose that the ANOVA test for networks rejects the null hypothesis, and now the main goal is to identify network differences. Two main objectives are discussed:

- Identification of all the links that show statistical differences between groups.
- Identification of a set of nodes (a subnetwork) that present the highest network differences between groups.

The identification procedure we describe below aims to eliminate the noise (links or nodes without differences between subpopulations) while keeping the signal (links or nodes with differences between subpopulations).

Given a network  $G = (V, E)$  and a subset of links  $\tilde{E} \subset E$ , let us generically denote  $G_{\tilde{E}}$  the subnetwork with the same nodes but with links identified by the set  $\tilde{E}$ . The rest of the links are erased. Given a subset of nodes  $\tilde{V} \subset V$  let us denote  $G_{\tilde{V}}$  the subnetwork that only has the nodes (with the links between them) identified by the set  $\tilde{V}$ . The  $T$  statistic for the sample of networks with only the set of  $\tilde{E}$  links is denoted by  $T_{\tilde{E}}$ , and the  $T$  statistic computed for all the sample networks with only the nodes that belong to  $\tilde{V}$  is denoted by  $T_{\tilde{V}}$ .

The procedure we propose for identifying all the links that show statistical differences between groups is based on the minimization for  $\tilde{E} \subset E$  of  $T_{\tilde{E}}$ . The set of links,  $\bar{E}$ , defined by

$$\bar{E} \equiv \arg \min_{\tilde{E} \subset E} T_{\tilde{E}} \quad (10)$$

contain all the links that show statistical differences between subpopulations. One limitation of this identification procedure is that the space  $E$  is huge ( $\#E = 2^{n(n-1)/2}$  where  $n$  is the number of nodes) and an efficient algorithm is needed to find the minimum. That is why we focus on identifying a group of nodes (or a subnetwork) expressing the largest differences.

The procedure proposed for identifying the subnetwork with the highest statistical differences between groups is similar to the previous one. It is based on the minimization of  $T_{\tilde{V}}$ . The set of nodes,  $N$ , defined by

$$N \equiv \arg \min_{\tilde{V} \subset V} T_{\tilde{V}} \quad (11)$$

contains all relevant nodes. These nodes make up the subnetwork with the largest difference between groups. In this case, the complexity is smaller, since the space  $V$  is not so big ( $\#V = 2^n - n - 1$ ).

As in other well-known statistical procedures such as cluster analysis or selection of variables in regression models, finding the size  $\tilde{j} := \#N$  of the number of nodes in the true subnetwork is a difficult problem due to possible overestimation of noisy data. The advantage of knowing  $\tilde{j}$  is that it reduces the computational complexity for finding the minimum to an order of  $n^{\tilde{j}}$  instead of  $2^n$  if we have to look for all possible sizes. However, the problem in our setup is less severe than other cases since the objective function ( $T_{\tilde{V}}$ ) is not monotonic when the size of the space increases. To solve this problem, we suggest the following algorithm.

Let  $V_{\{j\}}$  be the space of networks with  $j$  distinguishable nodes,  $j \in \{2, 3, \dots, n\}$  and  $V = \bigcup_j V_{\{j\}}$ . The nodes  $N_j$

$$N_j \equiv \arg \min_{\tilde{V} \in V_{\{j\}}} T_{\tilde{V}}, \quad \text{with} \quad T_j \equiv \min_{\tilde{V} \in V_{\{j\}}} T_{\tilde{V}} \quad (12)$$

define a subnetwork. In order to find the true subnetwork with differences between the groups, we now study the sequence  $T_2, T_3, \dots, T_n$ . We continue with the search (increasing  $j$ ) until we find  $\tilde{j}$  fulfilling

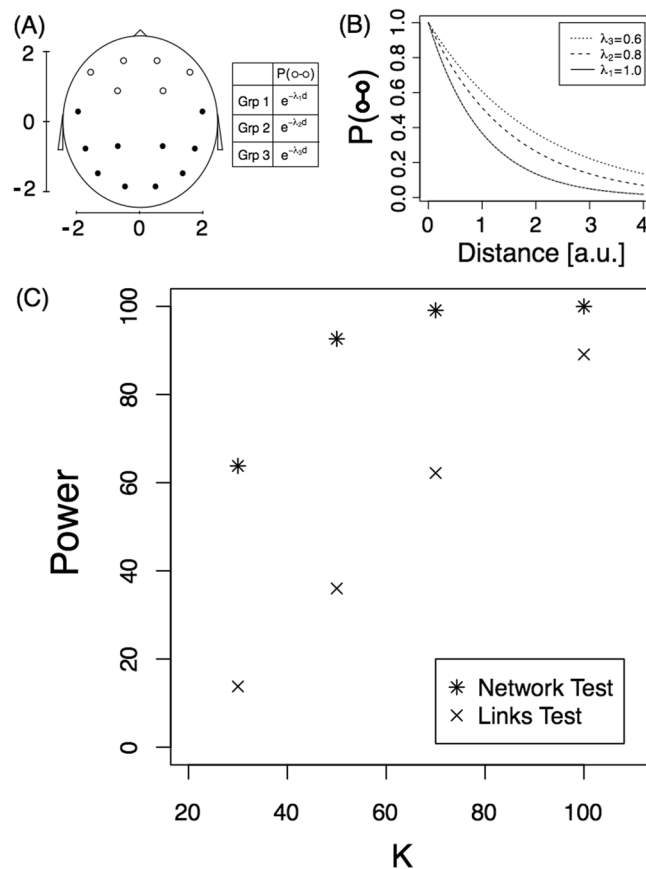
$$\tilde{j} \equiv \max \{j \in \{3, 4, \dots, n\}: T_j - T_{j-1} < -g(\text{sample size})\}, \quad (13)$$

where  $g$  is a positive function that decreases together with the sample size (in practice, a real value).  $N_{\tilde{j}}$  are the nodes that make up the subnetwork with the largest differences among the groups or subpopulations studied.

It is important to mention that the procedures described above do not impose any assumption regarding the real connectivity differences between the populations. With additional hypotheses, the procedure can be improved. For instance, in<sup>14,19</sup> the authors proposed a methodology for the edge-identification problem that is powerful only when the real difference connection between the populations form a large unique connected component.

**Examples and Applications.** A relevant problem in the current neuroimaging research agenda is how to compare populations based on their brain networks. The ANOVA test presented above deals with this problem. Moreover, the ANOVA procedure allows the identification of the variables related to the brain network structure. In this section, we show an example and application of this procedure in neuroimaging (EEG, MEG, fMRI, eCoG). In the example we show the robustness of the procedures for testing and identification of different sample sizes. In the application, we analyze fMRI data to understand which variables in the dataset are dependent on the brain network structure. Identifying these variables is also very important because any fair comparison between two or more populations requires these variables be controlled (similar values).

**Example.** Let us suppose we have three groups of subjects with equal sample size,  $K$ , and the brain network of each subject is studied using 16 regions (electrodes or voxels). Studies show connectivity between certain brain regions is different in certain neuropathologies, in aging, under the influence of psychedelic drugs, and more recently, in motor learning<sup>20,21</sup>. Recently, we have shown that a simple way to study connectivity is by what the physics community calls “the correlation function”<sup>22</sup>. This function describes the correlation between regions as a function of the distance between them. Although there exist long range connections, on average, regions (voxels or electrodes) closer to each other interact strongly, while distant ones interact more weakly. We have shown that



**Figure 1.** Detection problem. (A) Diagram of the scalp (each node represent an EEG electrode) on an x-y scale and the link probability. The three groups confirm the equation  $P(\circ \leftrightarrow \bullet) = P(\bullet \leftrightarrow \bullet) = e^{-d}$ . (B) Link probability of frontal electrodes,  $P(\circ \leftrightarrow \circ)$ , as a function of the distance for the three subpopulations. (C) Power of the tests as a function of sample size,  $K$ . Both tests are presented.

the way in which this function decays with distance is a marker of certain diseases<sup>23–25</sup>. For example, patients with a traumatic brachial plexus lesion with root avulsions revealed a faster correlation decay as a function of distance in the primary motor cortex region corresponding to the arm<sup>24</sup>.

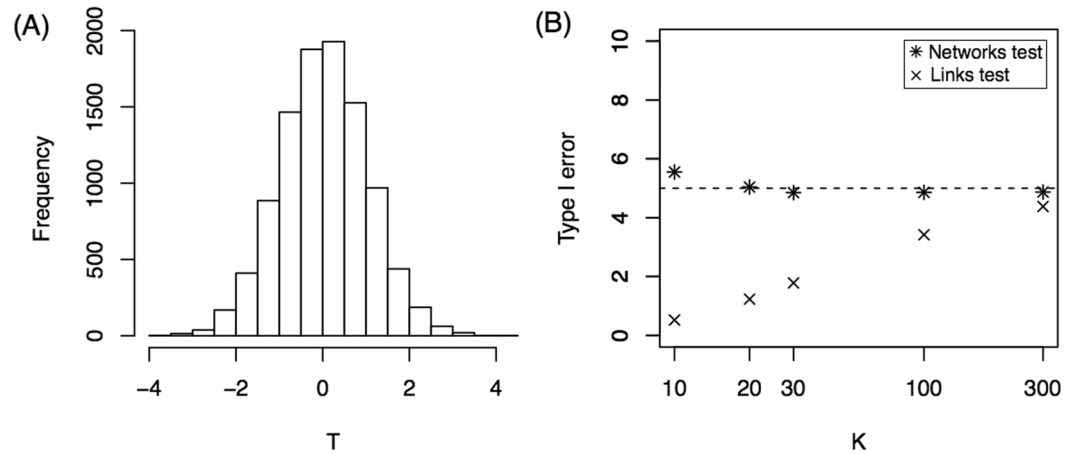
Next we present a toy model that analyses the method's performance. In a network context, the behaviour described above can be modeled in the following way: since the probability that two regions are connected is a monotonic function of the correlation between them (i.e. on average, distant regions share fewer links than nearby regions) we decided to skip the correlations and directly model the link probability as an exponential function that decays with distance. We assume that the probability that region  $i$  is connected with  $j$  is defined as

$$P(i \leftrightarrow j) = e^{-\lambda_1 d(i, j)}, \quad (14)$$

where  $d(i, j)$  is the distance between regions  $i$  and  $j$ . For the alternative hypothesis, we consider that there are six frontal brain regions (see Fig. 1 Panel A) that interact with a different decay rate in each of the three subpopulations. Figure 1 panel (A) shows the 16 regions analysed on an x-y scale. Panel (B) shows the link probability function for all electrodes and for each subpopulation. As shown, there is a slight difference between the decay of the interactions between the frontal electrodes in each subpopulation ( $\lambda_1 = 1$ ,  $\lambda_2 = 0.8$  and  $\lambda_3 = 0.6$  for groups 1, 2 and 3, respectively). The aim is to determine whether the ANOVA test for networks detects the network differences that are induced by the link probability function.

Here we investigated the power of the proposed test by simulating the model under different sample sizes ( $K$ ).  $K$  networks were computed for each of the three subpopulations and the  $T$  statistic was computed for each of 10,000 replicates. The proportion of replicates with a  $T$  value smaller than  $-1.65$  is an estimation of the power of the test for a significance level of 0.05 (unilateral hypothesis testing). Star symbols in Fig. 1C represent the power of the test for the different sample sizes. For example, for a sample size of 100, the test detects this small difference between the networks 100% of the time. As expected, the test has less power for small sample sizes, and if we change the values  $\lambda_2$  and  $\lambda_3$  in the model to 0.66 and 0.5, respectively, power increases. In this last case, the power changed from 64% to 96% for a sample size of 30 (see Supplementary Fig. S1 for the complete behaviour).

To the best of our knowledge, the  $T$  statistic is the first proposal of an ANOVA test for networks. Thus, here we compare it with a naive test where each individual link is compared among the subpopulations. The procedure is as follows: for each link, we calculate a test for equal proportions between the three groups to obtain a p-value



**Figure 2.** Null hypothesis. (A) Histogram of  $T$  statistics for  $K = 30$ . (B) Percentage of Type I Error as a function of sample size,  $K$ . Both tests are presented.

for each link. Since we are conducting multiple comparisons, we apply the Benjamini-Hochberg procedure controlling at a significance level of  $\alpha = 0.05$ . The procedure is as follows:

1. Compute the p-value of each link comparison,  $pv_1, pv_2, \dots, pv_m$ .
2. Find the  $j$  largest p-value such that  $pv_{(j)} \leq \frac{j}{m}\alpha$ .
3. Declare that the link probability is different for all links that have a p-value  $\leq pv_{(j)}$ .

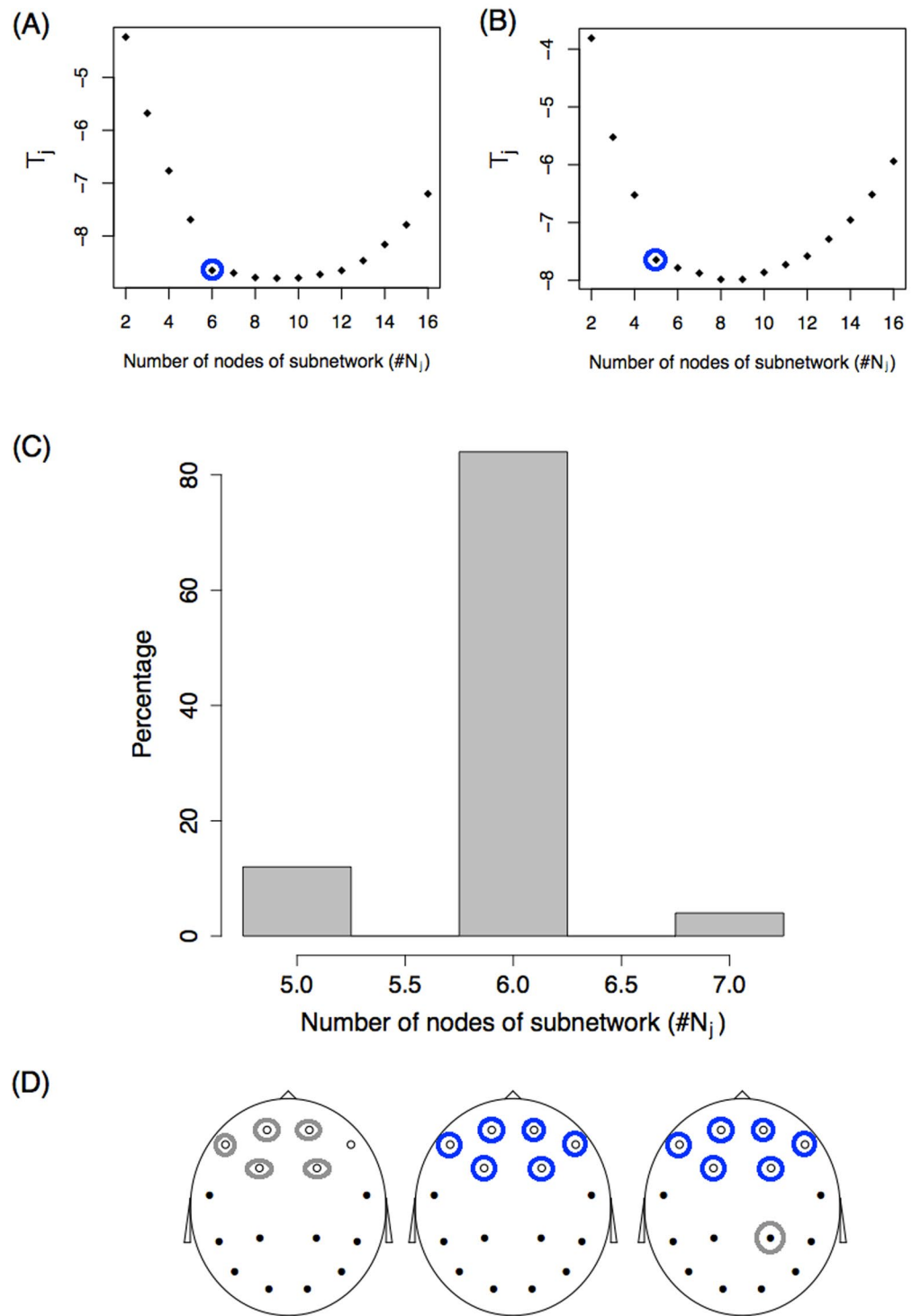
This procedure detects differences in the individual links while controlling for multiple comparisons. Finally, we consider the networks as being different if at least one link (of the 15 that have real differences) was detected to have significant differences. We will call this procedure the “Links Test”. Crosses in Fig. 1C correspond to the power of this test as a function of the sample size. As can be observed, the test proposed for testing equal mean networks is much more powerful than the previous test.

Theorem 1 States that  $T$  is asymptotically (sample size  $\rightarrow \infty$ ) Normal(0, 1) under the Null hypothesis. Next we investigated how large the sample size must be to obtain a good approximation. Moreover, we applied Theorem 1 in the simulations above for  $K = \{30, 50, 70, 100\}$ , but we did not show that the approximation is valid for  $K = 30$ , for example. Here, we show that the normal approximation is valid even for  $K = 30$  in the case of 16-node networks. We simulated 10,000 replicates of the model considering that all three groups have exactly the same probability law given by group 1, i.e. all brain connections confirm the equation  $P(i \leftrightarrow j) = e^{-\lambda_1 d(i,j)}$  for the three groups ( $H_0$  hypothesis). The  $T$  value is computed for each replicate of sample size  $K = 30$ , and the distribution is shown in Fig. 2(A). The histogram shows that the distribution is very close to normal. Moreover, the Kolmogorov-Smirnov test against a normal distribution did not reject the hypothesis of a normal distribution for the  $T$  statistic (p-value = 0.52). For sample sizes smaller than 30, the distribution has more variance. For example, for  $K = 10$ , the standard deviation of  $T$  is 1.1 instead of 1 (see Supplementary Fig. S2). This deviation from a normal distribution can also be observed in panel B where we show the percentage of Type I errors as a function of the sample size ( $K$ ). For sample sizes smaller than 30, this percentage is slightly greater than 5%, which is consistent with a variance greater than 1. The Links test procedure yielded a Type I error percentage smaller than 5% for small sample sizes.

Finally, we applied the subnetwork identification procedure described before to this example. Fifty simulations were performed for the model with a sample size of  $K = 100$ . For each replication, the minimum statistic  $T_j$  was studied as a function of the number of  $j$  nodes in the subnetwork. Figure 3A and B show two of the 50 simulation outcomes for the  $T_j$  function of ( $j$ ) number of nodes. Panel A shows that as nodes are incorporated into the subnetwork, the statistic sharply decreases to six nodes, and further incorporating nodes produces a very small decay in  $T_j$  in the region between six and nine nodes. Finally, adding even more nodes results in a statistical increase. A similar behaviour is observed in the simulation shown in panel B, but the “change point” appears for a number of nodes equal to five. If we define that the number of nodes with differences,  $\tilde{j}$ , confirms

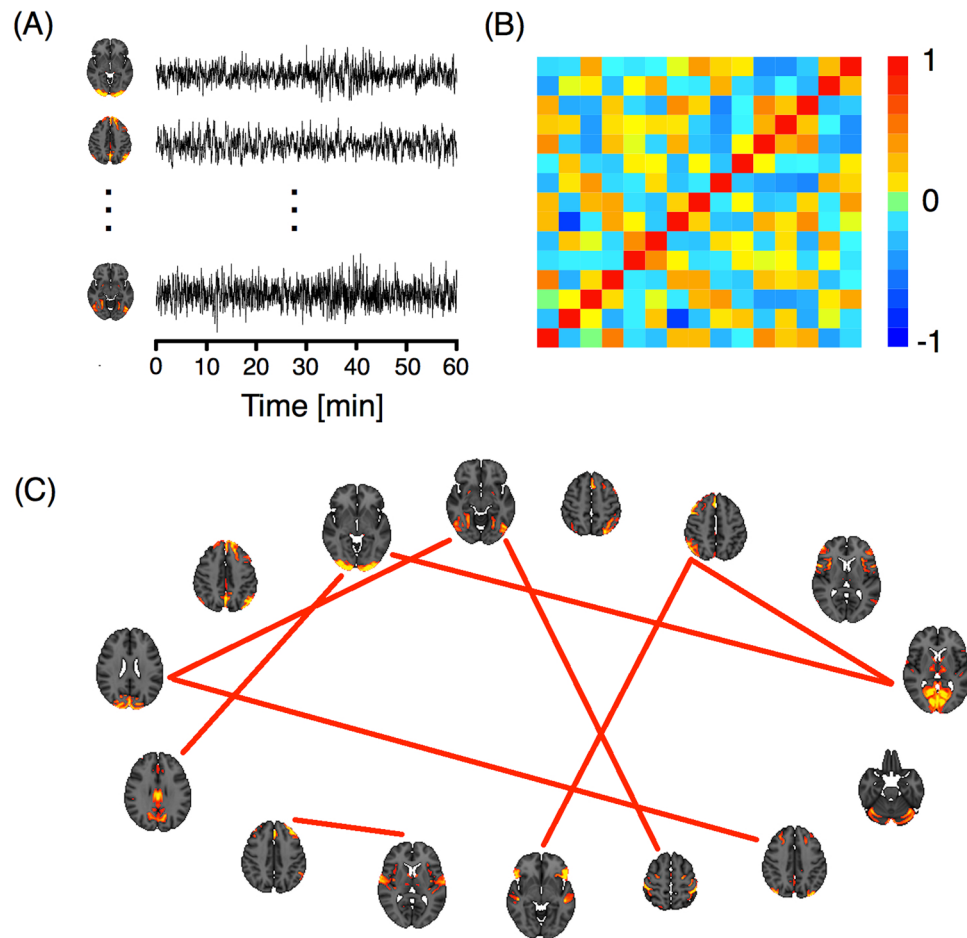
$$\tilde{j} \equiv \max \{j \in \{3, 4, \dots, n\}: T_j - T_{j-1} < -0.25\}, \quad (15)$$

we obtain the values circled. For each of the 50 simulations, we studied the value  $\tilde{j}$  and a histogram of the results is shown in Panel C. With the criteria defined, most of the simulations (85%) result in a subnetwork of 6 nodes, as expected. Moreover, these 6 nodes correspond to the real subnetwork with differences between subpopulations (white nodes in Fig. 1A). This was observed in 100% of simulations with  $\tilde{j} = 6$  (blue circles in Panel D). In the simulations where this value was 5, five of the six true nodes were identified, and five of the six nodes with differences vary between simulations (represented with grey circles in Panel D). For the simulations where  $\tilde{j} = 7$ , all six



**Figure 3.** Identification problem. (A,B) Statistic  $T_j$  as a function of the number of nodes of the subnetwork ( $j$ ) for two simulations. Blue circles represent the value  $\tilde{j}$  following the criteria described in the text. (C) Histogram of the number of subnetwork nodes showing differences,  $\tilde{j}$ . (D) Identification of the nodes. Blue and grey circles represent the nodes identified from the set  $N_j$ . Circled blue nodes are those identified 100% of the time. Grey circles represent nodes that are identified some of the time. On the left, grey circles alternate between the six white nodes. On the right, the grey circle alternates between the six black nodes.

real nodes were identified and a false node (grey circle) that changed between simulations was identified as being part of the subnetwork with differences.



**Figure 4.** (A) ICA components and their corresponding time series. (B) Correlation matrix of the time series. (C) Network representation. The links correspond to the nine highest correlations.

The identification procedure was also studied for a smaller sample size of  $K = 30$ , and in this case, the real subnetwork was identified only 28% of the time (see Supplementary Fig. S3 for more details). Identifying the correct subnetwork is more difficult (larger sample sizes are needed) than detecting global differences between group networks.

**Resting-state fMRI functional networks.** In this section, we analysed resting-state fMRI data from the 900 participants in the 2015 Human Connectome Project (HCP<sup>26</sup>). We included data from the 812 healthy participants who had four complete 15-minute rs-fMRI runs, for a total of one hour of brain activity. We partitioned the 812 participants into three subgroups and studied the differences between the brain groups. Clearly, if the participants are randomly divided into groups, no brain subgroup differences are expected, but if the participants are divided in an intentional way, differences may appear. For example, if we divided the 812 by the amount of hours slept before the scan ( $G_1$  less than 6 hours,  $G_2$  between 6 and 7 hours, and  $G_3$  more than 7) it might be expected<sup>27,28</sup> to observe differences in brain connectivity on the day of the scan. Moreover, as a by-product, we obtain that this variable is an important factoring variable to be controlled before the scan. Fortunately, HCP provides interesting individual socio-demographic, behavioural and structural brain data to facilitate this analysis. Moreover, using a previous release of the HCP data (461 subjects), Smith *et al.*<sup>29</sup>, using a multivariate analysis (canonical correlation), showed that a linear combination of demographics and behavior variables highly correlates with a linear combination of functional interactions between brain parcellations (obtained by Independent Component Analysis). Our approach has the same spirit, but has some differences. In our case, the main objective is to identify variables that “explain” (that are dependent with) the individual brain network. We do not impose a linear relationship between non-imaging and imaging variables, and we study the brain network as a whole object without different “loads” in each edge. Our method does not impose any kind of linearity, and it also detects linear and non-linear dependence structures.

Data were pre-processed by HCP<sup>30–32</sup> (details can be found in<sup>30</sup>), yielding the following outputs:

1. Group-average brain regional parcellations obtained by means of group-Independent Component Analysis (ICA<sup>33</sup>). Fifteen components are described.
2. Subject-specific time series per ICA component.



Figure 4(A) shows three of the 15 ICA components with the specific one hour time series for a particular subject. These signals were used to construct an association matrix between pairs of ICA components per subject. This matrix represents the strength of the association between each pair of components, which can be quantified by different functional coupling metrics, such as the Pearson correlation coefficient between the signals of the component, which we adopted in the present study (panel (B)). For each of the 812 subjects, we studied functional connectivity by transforming each correlation matrix,  $\Sigma$ , into binary matrices or networks,  $G$ , (panel (C)). Two criteria for this transformation were used<sup>34–36</sup>: a fixed correlation threshold and a fixed number of links criterion. In the first criterion, the matrix was thresholded by a value  $\rho$  affording networks with varying numbers of links. In the second, a fixed number of link criteria were established and a specific threshold was chosen for each subject.

As we have already mentioned, HCP provides interesting individual socio-demographic, behavioural and structural brain data. Variables are grouped into seven main categories: alertness, motor response, cognition, emotion, personality, sensory, and brain anatomy. Volume, thickness and areas of different brain regions were computed using the T1-weighted images of each subject in Free Surfer<sup>37</sup>. Thus, for each subject, we obtained a brain functional network,  $G$ , and a multivariate vector  $X$  that contains this last piece of information.

The main focus of this section is to analyse the “impact” of each of these variables ( $X$ ) on the brain networks (i.e., on brain activity). To this end, we first selected a variable such as  $k$ ,  $X_k$ , and grouped each subject according to his/her value into only one of three categories (Low, Medium, or High) just by placing the values in ascending and using the 33.3% percentile. In this way, we obtained three groups of subjects, each identified by its correlation matrix  $\Sigma_1^L, \dots, \Sigma_{n_L}^L, \Sigma_1^M, \dots, \Sigma_{n_M}^M$  and  $\Sigma_1^H, \dots, \Sigma_{n_H}^H$ , or by its corresponding network (once the criteria and the parameter are chosen)  $G_1^L, \dots, G_{n_L}^L, G_1^M, \dots, G_{n_M}^M$ , and  $G_1^H, \dots, G_{n_H}^H$ . The sample size of each group ( $n_L$ ,  $n_M$ , and  $n_H$ ) is approximately 1/3 of 812, except in cases where there were ties. Once we obtained these three sets of networks, we applied the developed test. If differences exist between all three groups, then we are confirming an interdependence between the factoring variable and the functional networks. However, we cannot yet elucidate directionality (i.e., different networks lead to different sleeping patterns or vice versa?).

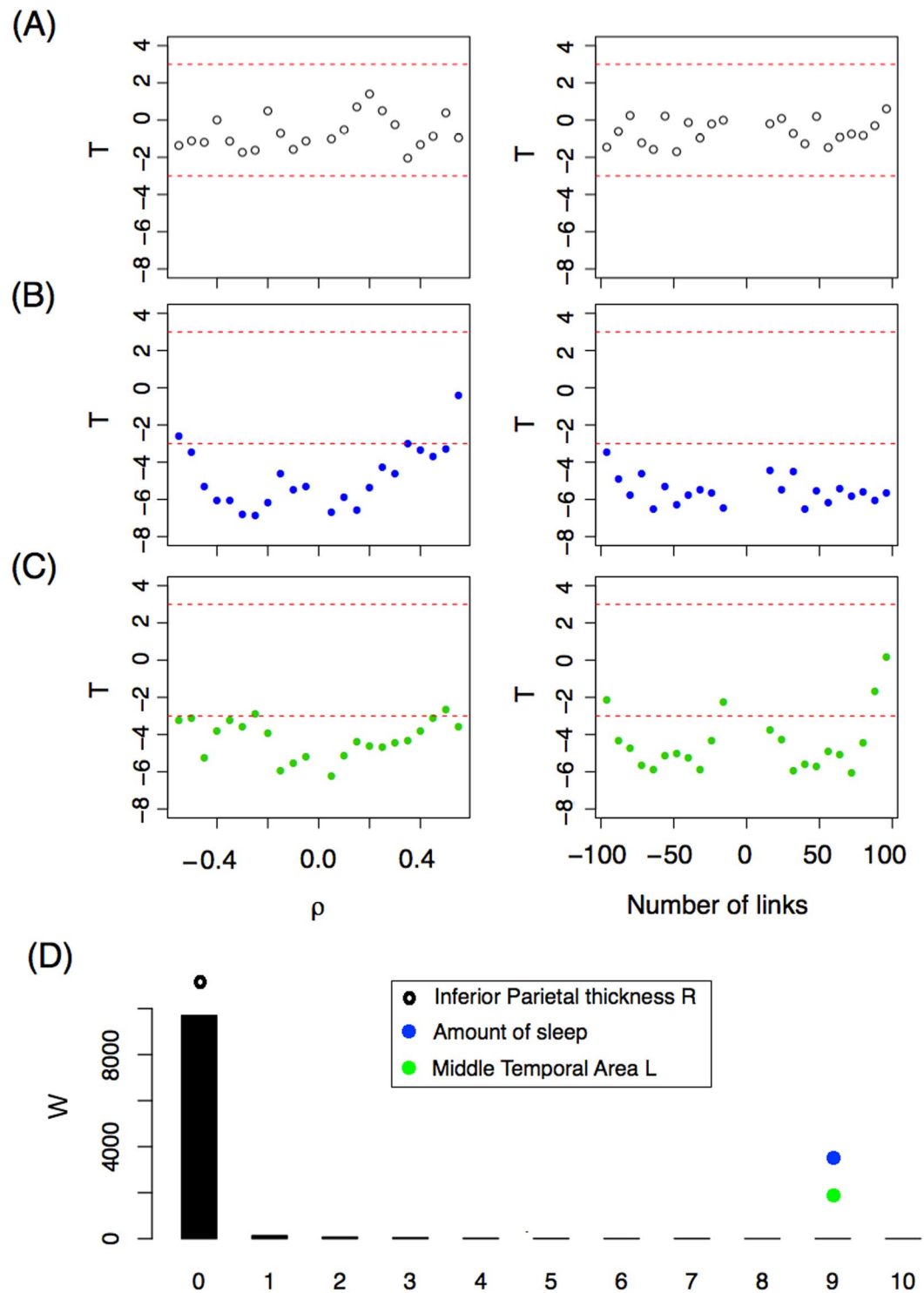
After filtering the data, we identified 221 variables with 100% complete information for the 812 subjects, and 90 other variables with almost complete information, giving a total of 311 variables. We applied the network ANOVA test for each of these 311 variables and report the  $T$  statistic. Figure 5(A) shows the  $T$  statistic for the variable *Thickness of the right Inferior Parietal* region. All values of the  $T$  statistic are between  $-2$  and  $2$  for all  $\rho$  values using the fixed correlation criterion (left panel) for constructing the networks. The same occurs when a fixed number of link criteria is used (right panel). According to Theorem 1, when there are no differences between groups,  $T$  is asymptotically normal  $(0, 1)$ , and therefore a value smaller than  $-3$  is very unlikely ( $p$ -value = 0.00135). Since all  $T$  values are between  $-2$  and  $2$ , we assert that *Thickness of the right Inferior Parietal* region is not associated with the resting-state functional interactions. In panel (B), we show the  $T$  statistic for the variable *Amount of hours spent sleeping on the 30 nights prior to the scan* (“During the past month, how many hours of actual sleep did you get at night? (This may be different than the number of hours you spent in bed.)”) which corresponds to the alertness category. As one can see, most  $T$  values are much lower than  $-3$ , rejecting the hypothesis of equal mean network. Importantly, this shows that the number of hours a person sleeps is associated with their brain functional networks (or brain activity). However, as explained above, we do not know whether the number of hours slept the nights before represent these individuals’ habitual sleeping patterns, complicating any effort to infer causation. In other words, six hours of sleep for an individual who habitually sleeps six hours may not produce the same network pattern as six hours in an individual who normally sleeps eight hours (and is likely tired during the scan). Alternatively, different activity observed during waking hours may “produce” different sleep behaviours. Nevertheless, we know that the amount of hours slept before the scan should be measured and controlled when scanning a subject. In Panel (C), we show that brain volumetric variables can also influence resting-state fMRI networks. In that panel, we show the  $T$  value for the variable *Area of the left Middle temporal* region. Significant differences for both network criteria are also observed for this variable.

Under the hypothesis of equal mean networks between groups, we expect not to obtain a  $T$  statistic less than  $-3$  when comparing the sample networks. We tested several different thresholds and numbers of links in order to present a more robust methodology. However, in this way, we generate sets of networks that are dependent on each criterion and between criteria, similarly to what happens when studying dynamic networks with overlapping sliding windows. This makes the statistical inference more difficult. To address this problem, we decided to define a new statistic based on  $T$ ,  $W_3$ , and study its distribution using the bootstrap resampling technique. The new statistic is defined as,

$$W_3 = \min \{ \Delta_+^\rho, \Delta_-^\rho, \Delta_+^L, \Delta_-^L \}, \quad (16)$$

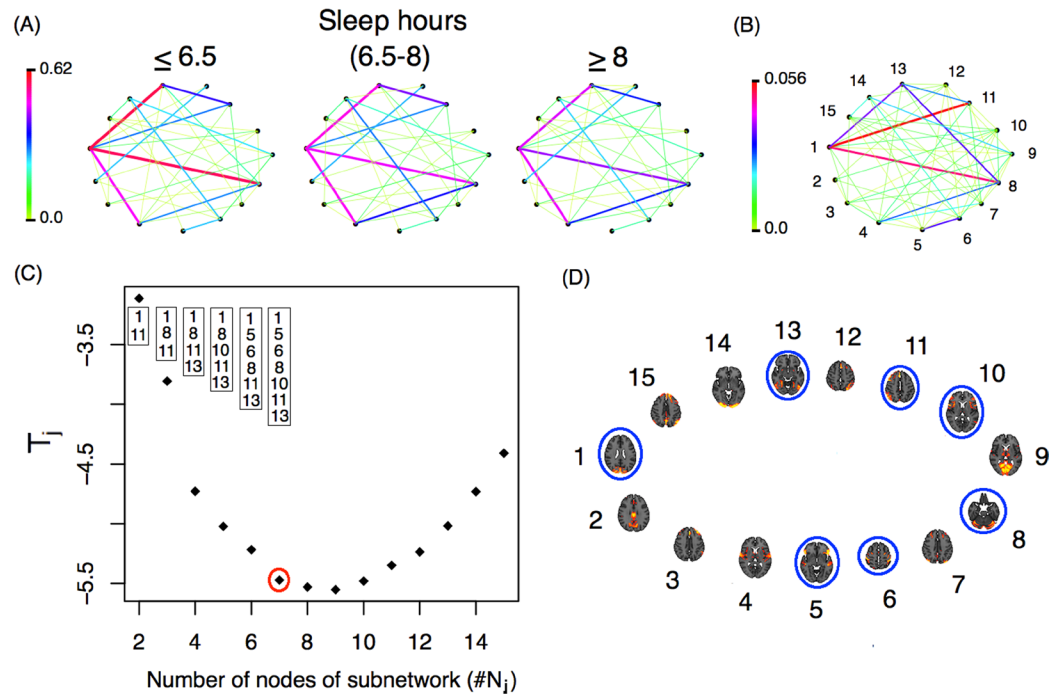
where  $\Delta$  is the number of values of  $T$  that are lower than  $-3$  for the resolution (grid of thresholds) studied. The supraindex in  $\Delta$  indicates the criteria (correlation threshold,  $\rho$  or number of links fixed,  $L$ ) and the subindex indicates whether it is for positive or negative parameter values ( $\rho$  or number of links). For example, Fig. 5(C) reveals that the variable *Area of the left Middle temporal* confirms having  $\Delta_+^\rho = 10$ ,  $\Delta_-^\rho = 10$ ,  $\Delta_+^L = 9$ , and  $\Delta_-^L = 9$ , and therefore  $W_3 = 9$ . The distribution of  $W_3$  under the null hypothesis is studied numerically. Ten thousand random resamplings of the real networks were selected and the  $W_3$  statistic was computed for each one. Figure 5(D) shows the  $W$  empirical distribution (under the null hypothesis) with black bars. Most  $W_3$  values are zero, as expected. In this figure, the  $W_3$  values of the three variables described are also represented by dots. The extreme values of  $W_3$  for the variables Amount of Sleep and Middle Temporal Area L confirm that these differences are not a matter of chance. Both variables are related to brain network connectivity.

So far we have shown, among other things, that functional networks differ between individuals who get more or fewer hours of sleep, but how do these networks differ exactly? Fig. 6(A) shows the average networks for the



**Figure 5.** (A–C)  $T$ -statistics as a function of (left panel)  $\rho$  and (right panel) the number of links for three variables: (A) Right Intraparietal Thickness, (B) Number of hours slept the nights prior to the scan. (C) Left Middle Temporal Area. (D)  $W$ -statistic distribution (black bars) based on a bootstrap strategy. The  $W$ -statistic of the three variables studied is depicted with dots.

three groups of subjects. There are differences in connectivity strength between some of the nodes (ICA components). These differences are more evident in panel (B), which presents a weighted network  $\Psi$  with links showing the variability among the subpopulation's average networks. This weighted network is defined as



**Figure 6.** (A) Average network for each subgroup defined by hours of sleep (B) Weighted network with links that represent the differences among the subpopulation mean networks. (C)  $T_j$ -statistic as a function of the number of nodes in each subnetwork ( $j$ ). The nodes identified by the minimum  $T_j$  are presented in the boxes, while the number of nodes identified by the procedure are represented with a red circle. (D) Nodes from the identified subnetwork are circled in blue. The nodes identified in (D) correspond to those in panel (B).

$$\Psi(i, j) = \frac{1}{3} \sum_{s=1}^3 |\mathcal{M}^{\text{grp } s}(i, j) - \overline{\mathcal{M}}(i, j)|, \quad (17)$$

where  $\overline{\mathcal{M}}(i, j) = \frac{1}{3} \sum_{s=1}^3 \mathcal{M}^{\text{grp } s}$ . The role of  $\Psi$  is to highlight the differences between the mean networks. The greatest difference is observed between nodes 1 and 11. Individuals that sleep 6.5 hours or less show the strongest connection between ICA component number 1 (which corresponds to the occipital pole and the cuneal cortex in the occipital lobe) and ICA component number 11 (which includes the middle and superior frontal gyri in the frontal lobe, the superior parietal lobule and the angular gyrus in the parietal lobe). Another important connection that differs between groups is the one between ICA components 1 and 8, which corresponds to the anterior and posterior lobes of the cerebellum. Using the subnetwork identification procedure previously described (see Fig. 6C) we identified a 7-node subnetwork as the most significant for network differences. The nodes that make up that network are presented in panel D.

The results described above refer to only three of the 311 variables we analysed. In terms of the remaining variables, we observed more variables that partitioned the subjects into groups presenting statistical differences between the corresponding brain networks. Two more behavioral variables were identified the variable Dimensional Change Card Sort (CardSort\_AgeAdj and CardSort\_Unadj) which is a measure of cognitive flexibility, and the variable motor strength (Strength\_AgeAdj and Strength\_Unadj). Also 20 different brain volumetric variables were identified, the complete list of these variables is shown in Suppl. Table S1. It is important to note that these brain volumetric variables are largely dependent on each other; for example, individuals with larger inferior-temporal areas often have a greater supratentorial volume, and so on (see Suppl. Fig. S4).

We have reported only those variables for which there is very strong statistical evidence in favor of the existence of dependence between the functional networks and the “behavioral” variables, irrespectively of the threshold used to build up the networks. There are other variables that show this dependence only for some levels of the threshold parameter, but we do not report these to avoid reporting results that may not be significant. Our results complement those observed in<sup>29</sup>. In particular, Smith *et al.* report that the variable *Picture Vocabulary test* is the most significant. With a less restrictive criterion, this variable can also be considered significant with our methodology. In fact, the  $W_3$  value equals 3 (see Supplementary Fig. S5 for details), which supports the notion (see panel D in Fig. 5) that the variable *Picture Vocabulary test* is also relevant for explaining the functional networks. On the other hand, the variable we found to vary significantly ( $W_3 = 9$ ) the *Amount of sleep* is not reported by Smith *et al.* Perhaps the canonical correlation cannot find the variable because it looks for linear correlations in a high dimensional space. It is well known that non-linearities appear typically in high dimensional statistical problems (See for instance<sup>38</sup>). To capture nonlinear associations, a kernel CCA method was introduced, see<sup>39,40</sup> and the references therein. By contrast, our method does not impose any kind of linearity, and detects linear as well as

non-linear dependence structures. The variable “Cognitive flexibility” (*Card Sort*) found here was also reported in<sup>38</sup>. Finally, the brain volumetric variables we found to be relevant here were not analyzed in<sup>29</sup>.

So far, we apply the methodology presented here to analyse brain data by using only 15 brain ICA dimensions (provided by HCP). But, what is the impact of working with more ICA components? Does we identify more covariables? Fortunately, we can respond these questions since more ICA dimensions were recently made available on HCP webpage. Three new cognitive variables, *Working memory*, *Relational processing* and *Self-regulation/Impulsivity* were identified for higher network dimension (50 and 300 ICA dimensions, see Suppl. Table S2 for details).

## Discussion

Performing statistical inference on brain networks is important in neuroimaging. In this paper, we presented a new method for comparing anatomical and functional brain networks of two or more subgroups of subjects. Two problems were studied: the detection of differences between the groups and the identification of the specific network differences. For the first problem, we developed an ANOVA test based on the distance between networks. This test performed well in terms of detecting existing differences (high statistical power). Finally, based on the statistics developed for the testing problem, we proposed a way of solving the identification problem. Next, we discuss our findings.

**Identification.** Based on the minimization of the  $T$  statistic, we propose a method for identifying the subnetwork that differs among the subgroups. This subnetwork is very useful. On the one hand, it allows us to understand which brain regions are involved in the specific comparison study (neurobiological interpretation), and on the other, it allows us to identify/diagnose new subjects with greater accuracy.

The relationship between the minimum  $T$  value for a fixed number of nodes as a function of the number of nodes ( $T_j$  vs.  $j$ ) is very informative. A large decrease in  $T_j$  incorporating a new node into the subnetwork ( $T_{j+1} \ll T_j$ ) means that the new node and its connections explain much of the difference between groups. A very small decrease shows that the new node explains only some of the difference because either the subgroup difference is small for the connections of the new node, or because there is a problem of overestimation.

The correct number of nodes in each subnetwork must verify

$$\tilde{j} \equiv \max \{j \in \{3, 4, \dots, n\}: T_j - T_{j-1} < -g(\text{sample size})\}. \quad (18)$$

In this paper, we present ad hoc criteria in each example (a certain constant for  $g(\text{sample size})$ ) and we do not give a general formula for  $g(\text{sample size})$ . We believe that this could be improved in theory, but in practice, one can propose a natural way to define the upper bound and subsequently identify the subnetwork, as we showed in the example and in the application by observing  $T_j$  as a function of  $j$ . Statistical methods such as the one developed for change-point detection may be useful in solving this problem.

**Sample size.** What is the adequate sample size for comparing brain networks? This is typically the first question in any comparison study. Clearly, the response depends on the magnitude of the network differences between the groups and the power of the test. If the subpopulations differ greatly, then a moderate number of networks in each group is enough. On the other hand, if the differences are not very big, then a larger sample size is required to have a reasonable power of detection. The problem gets more complicated when it comes to identification. We showed in Example 1 that we obtain a good identification rate when a sample size of 100 networks is selected from each subgroup. Thus, the rate of correct identification is small for a sample size of for example 30.

**Confounding variables in Neuroimaging.** Humans are highly variable in their brain activity, which can be influenced, in turn, by their level of alertness, mood, motivation, health and many other factors. Even the amount of coffee drunk prior to the scan can greatly influence resting-state neural activity. What variables must be controlled to make a fair comparison between two or more groups? Certainly age, gender, and education are among those variables, and in this study we found that the amount of hours slept the nights prior to the scan is also relevant. Although this might seem pretty obvious, to the best of our knowledge, most studies do not control for this variable. Five other variables were identified, each one related with some dimensions of cognitive flexibility, self-regulation/impulsivity, relational processing, working memory or motor strength. Finally, we identified as being relevant a set of 20 highly interdependent brain volumetric variables. In principle, the role of these variables is not surprising, since comparing brain activity between individuals requires one to pre-process the images by realigning and normalizing them to a standard brain. In other words, the relevance of specific area volumes may simply be a by-product of the standardization process. However, if our finding that brain volumetric variables affect functional networks is replicated in other studies, this poses a problem for future experimental designs. Specifically, groups will not only have to be matched by variables such as age, gender and education level, but also in terms of volumetric variables, which can only be observed in the scanner. Therefore, several individuals would have to be scanned before selecting the final study groups.

In sum, a large number of subjects in each group must be tested to obtain highly reproducible findings when analysing resting-state data with network methodologies. Also, whenever possible, the same participants should be tested both as controls and as the treatment group (paired samples) in order to minimize the impact of brain volumetric variables.

## References

1. Deco, G. & Kringelbach, M. L. Great expectations: using whole-brain computational connectomics for understanding neuropsychiatric disorders. *Neuron* **84**, 892–905 (2014).
2. Stephan, K. E., Iglesias, S., Heinzle, J. & Diaconescu, A. O. Translational perspectives for computational neuroimaging. *Neuron* **87**, 716–732 (2015).

3. Bullmore, E. & Sporns, O. Complex brain networks: network theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience* **10**, 186–196 (2009).
4. Fornito, A., Zalesky, A. & Bullmore, E. Fundamentals of Brain Network Analysis. *Elsevier*.
5. Anonymous Focus on human brain mapping. *Nat. Neurosci.* **20**, 297–298 (2017).
6. Button, K. S. *et al.* Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* **14**, 365–376 (2013).
7. Poldrack, R. Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nat. Rev. Neurosci.* **18**, 115–126 (2017).
8. Nichols, T. E. *et al.* Best Practices in Data Analysis and Sharing in Neuroimaging using MRI. *Nat. Neurosci.* **20**, 299–303 (2016).
9. Bennett, C. M. & Miller, M. B. How reliable are the results from functional magnetic resonance imaging? *Annals of the New York Academy of Sciences* **1191**, 133–155 (2010).
10. Brown, E. N. & Behrmann, M. Controversy in statistical analysis of functional magnetic resonance imaging data. *Proc Natl Acad Sci USA* **114**, E3368–E3369 (2017).
11. Fraiman, D., Fraiman, N. & Fraiman, R. Non Parametric Statistics of Dynamic Networks with distinguishable nodes. *Test* **26**, 546–573 (2017).
12. Cerqueira, A., Fraiman, D., Vargas, C. & Leonardi, F. A test of hypotheses for random graph distributions built from EEGdata. *IEEE Transactions on Network Science and Engineering* **4**, 75–82 (2017).
13. Kolar, M., Song, L., Ahmed, A. & Xing, E. *Estimating Time-varying networks*. *Ann. Appl. Stat. Estimating Time-varying networks*. **4**, 94–123 (2010).
14. Zalesky, A., Fornito, A. & Bullmore, E. Network-based statistic: identifying differences in brain networks. *Neuroimage* **53**, 1197–1207 (2010).
15. Sanfeliu, A. & Fu, K. A distance measure between attributed relational graphs. *IEEE T. Sys. Man. Cyb.* **13**, 353–363 (1983).
16. Schieber, T. *et al.* Quantification of network structural dissimilarities. *Nature communications* **8**, 13928 (2017).
17. Shimada, Y., Hirata, Y., Ikeguchi, T. & Aihara, K. Graph distance for complex networks. *Scientific reports* **6**, 34944 (2016).
18. Gao, X., Xiao, B., Tao, D. & Li, X. A survey of graph edit distance. *Pattern Anal Appl* **13**, 113–129 (2010).
19. Zalesky, A., Cocchi, L., Fornito, A., Murray, M. & Bullmore, E. Connectivity differences in brain networks. *Neuroimage* **60**, 1055–1062 (2012).
20. Della-Maggiore, V., Villalta, J. I., Kovacevic, N. & McIntosh, A. R. Functional Evidence for Memory Stabilization in Sensorimotor Adaptation: A 24-h Resting-State fMRI Study. *Cerebral Cortex* **27**, 1748–1757 (2015).
21. Mawase, F., Bar-Haim, S. & Shmuelof, L. Formation of Long-Term Locomotor Memories Is Associated with Functional Connectivity Changes in the Cerebellar/Thalamic/Cortical Network. *Journal of Neuroscience* **37**, 349–361 (2017).
22. Fraiman, D. & Chialvo, D. What kind of noise is brain noise: anomalous scaling behavior of the resting brain activity fluctuations. *Frontiers in Physiology* **3**, 307 (2012).
23. Garcia-Cordero, I. *et al.* Stroke and neurodegeneration induce different connectivity aberrations in the insula. *Stroke* **46**, 2673–2677 (2015).
24. Fraiman, D. *et al.* Reduced functional connectivity within the primary motor cortex of patients with brachial plexus injury. *Neuroimage Clinical* **12**, 277–284 (2016).
25. Dottori, M. *et al.* Towards affordable biomarkers of frontotemporal dementia: A classification study via network's information sharing. *Scientific Reports* **7**, 3822 (2017).
26. Human Connectome Project. <http://www.humanconnectomeproject.org/>
27. Kaufmann, T. *et al.* The brain functional connectome is robustly altered by lack of sleep. *NeuroImage* **127**, 324–332 (2016).
28. Krause, A. *et al.* The sleep-deprived human brain. *Nature Reviews Neuroscience* **18**, 404–418 (2017).
29. Smith, S. *et al.* A positive-negative mode of population covariation links brain connectivity, demographics and behavior. *Nature neuroscience* **18**, 1565–1567 (2015).
30. Human Connectome Project. WU-Minn HCP 900 Subjects Data Release: Reference Manual. 67–87 (2015).
31. Griffanti, L. *et al.* ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. *Neuroimage* **95**, 232–247 (2014).
32. Smith, S. M. *et al.* Resting-state fMRI in the Human Connectome Project. *Neuroimage* **80**, 144–168 (2013).
33. Beckmann, C., DeLuca, M., Devlin, J. & Smith, S. Investigations into resting-state connectivity using independent component analysis. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **360**, 1001–1013 (2005).
34. Fraiman, D., Saunier, G., Martins, E. & Vargas, C. Biological Motion Coding in the Brain: Analysis of Visually Driven EEG Functional Networks. *PLoS One*, 0084612 (2014).
35. Amoruso, L. *et al.* Brain network organization predicts style-specific expertise during Tango dance observation. *Neuroimage* **146**, 690–700 (2017).
36. van den Heuvel, M. P. *et al.* Proportional thresholding in resting-state fMRI functional connectivity networks and consequences for patient-control connectome studies: Issues and recommendations. *Neuroimage* **152**, 437–449 (2017).
37. Fischl, B. FreeSurfer. *Neuroimage* **62**, 774–781 (2012).
38. Buhlmann, P. & van der Geer, S. Statistics for High-Dimensional Data: Methods, Theory and Applications. Springer (2011).
39. Yoshida, K., Yoshimoto, J. & Doya, K. Sparse kernel canonical correlation analysis for discovery of nonlinear interactions in high-dimensional data. *BMC Bioinformatics* **18**, 108 (2017).
40. Yamanishi, Y., Vert, J. P., Nakaya, A. & Kanehisa, M. Extraction of correlated gene clusters from multiple genomic data by generalized kernel canonical correlation analysis. *Bioinformatics* **19**, 323–330 (2003).

## Acknowledgements

We thank two anonymous reviewers for extensive comments that helped improve the manuscript significantly. Data were provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University. This paper was produced as part of the activities of FAPESP Research, Innovation and Dissemination Center for Neuromathematics (Grant No. 2013/07699-0, S. Paulo Research Foundation). This work was partially supported by PAI UdeSA.

## Author Contributions

D.F. and R.F. conceived the research, analysed the data and wrote the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-23152-5>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018