Adaptive strategy for the statistical analysis of connectomes

MESKALDIJI, Djalel Eddine, et al.

Abstract

We study an adaptive statistical approach to analyze brain networks represented by brain connection matrices of interregional connectivity (connectomes). Our approach is at a middle level between a global analysis and single connections analysis by considering subnetworks of the global brain network. These subnetworks represent either the inter-connectivity between two brain anatomical regions or by the intra-connectivity within the same brain anatomical region. An appropriate summary statistic, that characterizes a meaningful feature of the subnetwork, is evaluated. Based on this summary statistic, a statistical test is performed to derive the corresponding p-value. The reformulation of the problem in this way reduces the number of statistical tests in an orderly fashion based on our understanding of the problem. Considering the global testing problem, the p-values are corrected to control the rate of false discoveries. Finally, the procedure is followed by a local investigation within the significant subnetworks. We contrast this strategy with the one based on the individual measures in terms of power. We show that this [...]
Adaptive Strategy for the Statistical Analysis of Connectomes

Djalel Eddine Meskaldji1*, Marie-Christine Ottet2, Leila Cammoun1, Patric Hagmann3, Reto Meuli3, Stephan Eliez2, Jean Philippe Thiran1, Stephan Morgenthaler4

1 Signal Processing Laboratory (LTSS), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, 2 Office Médico-Pédagogique (OMP), Geneva, Switzerland, 3 Department of Radiology, University Hospital Center and University of Lausanne (CHUV-UNIL), Lausanne, Switzerland, 4 Institute of Mathematics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

Abstract

We study an adaptive statistical approach to analyze brain networks represented by brain connection matrices of interregional connectivity (connectomes). Our approach is at a middle level between a global analysis and single connections analysis by considering subnetworks of the global brain network. These subnetworks represent either the interconnectivity between two brain anatomical regions or by the intra-connectivity within the same brain anatomical region. An appropriate summary statistic, that characterizes a meaningful feature of the subnetwork, is evaluated. Based on this summary statistic, a statistical test is performed to derive the corresponding p-value. The reformulation of the problem in this way reduces the number of statistical tests in an orderly fashion based on our understanding of the problem. Considering the global testing problem, the p-values are corrected to control the rate of false discoveries. Finally, the procedure is followed by a local investigation within the significant subnetworks. We contrast this strategy with the one based on the individual measures in terms of power. We show that this strategy has a great potential, in particular in cases where the subnetworks are well defined and the summary statistics are properly chosen. As an application example, we compare structural brain connection matrices of two groups of subjects with a 22q11.2 deletion syndrome, distinguished by their IQ scores.

Introduction

Understanding the human brain is one of the greatest challenges in science. A vast number and variety of methods have been developed and applied to analyze and study its organization, development and function. In the recent years, the determination of the interregional brain connectivity (connectome [1,2]) has attracted much attention given the advent of new in-vivo imaging techniques. One distinguishes between structural connectivity and functional connectivity. Structural connectivity refers to the existence of axonal fibers that interconnect parts of the brain. For example, recent advances in diffusion imaging and tractography permit the construction of high-resolution connection matrices estimating interregional connectivity of the human brain cortex [3]. Functional connectivity is based on joint activation of brain regions and is derived from BOLD contrast MRI, Magnetic EncephaloGraphy (MEG) or other time series data that represent brain activation while the subject performs certain tasks or the subject is in resting state [4,5]. Structural and functional connection matrices have been used to study properties of brain networks mainly to understand its organization and development [5–7]. These connection matrices have also been used to study differences between groups of individuals based on either single connections or by global measures [8]. When performing such group comparisons based on single connections, a large number of correlated statistical tests are routinely performed and, in order to control the occurrences of false discoveries of pair-wise differences, a correction for multiplicity (e.g. Bonferroni or other procedures) is necessary, which greatly reduces the power of the comparisons. We propose in this article a practical and yet effective strategy to analyze complex networks, in particular, brain networks represented by connection matrices. The proposed strategy exploits the potential of high resolution connection matrices by following the concept of “borrowing strength”. In neuroimaging and in many other fields of applications, measures are often correlated and quite well defined regionally. For instance, when analyzing brain networks, instead of performing the statistical tests on the level of single connections between pairs of regions of interest (ROIs), which represent nodes in the brain network, it may be advantageous to reformulate the question in terms of comparisons based on connections between relevant groups of nodes. This is particularly appropriate in cases, where connectivity between groups of ROIs is of major interest to the researcher. The connectivity between groups of ROIs represents a brain subnetwork and...
corresponds to a block (provided the matrix is properly sorted) in the connection matrix. The proposed strategy is based on the construction of an appropriate statistic in each subnetwork where subnetworks are predefined by the researcher based on prior knowledge. This statistic, which reflects the investigator’s research hypothesis or knowledge of the problem, efficiently summarizes a given feature of the subnetwork and will be used for comparisons instead of all the values observed on the ROI level. This strategy has the common advantages of cluster based methods, namely, the reduction of the number of tests and the reduction of the noise variance. Furthermore, the researcher has the opportunity to use some topological network measures that cannot be defined on the single connections level. Of course, the significance obtained by the proposed strategy is at the level of subnetworks and not at the level of single connections. This means that the proposed strategy can be seen as a first stage of an exploratory procedure or coarse scale analysis where the researcher is interested in finding affected subnetworks, that is, those containing one or more affected connections and hence, the interpretation of the results becomes more complex with larger subnetworks. In this direction, we discuss how the proposed strategy could be followed by a local investigation of the connections inside the significant subnetworks.

The paper is organized as follows. First, we briefly discuss the subject of multiple comparisons. Then, we show the benefit of grouping statistical tests into subsets. A particular summary statistic, the mean of values, is studied in detail. We also discuss the possibility of performing a local investigation within the significant subsets. We will call that procedure a “two stage procedure”. After presenting the statistical part of the paper, we show its applicability on complex networks in general and on brain networks in particular. Finally, we present a real application of comparing two groups of structural connection matrices derived from a population of individuals affected with 22q11.2 deletion syndrome [9,10]. With this application, we will emphasize the structural brain connectivity differences that exist between high (IQ above 70) and low (IQ below 70) cognitive functioning 22q11.2 deletion syndrome [11].

Table 1. The general outcome of a multiple comparisons.

| Number of hypotheses that are | Statistically non-significant | statistically significant | Total |
|------------------------------|-----------------------------|--------------------------|------|
| True                         | TN                          | FP                       | m(0) |
| False                        | FN                          | TP                       | m(1) |
| Total                        | m-R                         | R                        | m    |

A total of m null hypotheses are tested. FP is the number of Type I errors or the number of false positives (rejected true hypotheses). Physical significance as indicated in the first column means the existence of a real effect, whereas statistical significance refers to the detection of such effect by means of measurements. FN is the number of Type II errors or the number of false negatives (false hypotheses not rejected). The number R of rejected hypotheses is an observable random variable, while FP, FN, TP and TN are unobservable random variables. The number of true null hypotheses m(0) is also unknown in practice. The empirical Type I error rate is defined by FP/m(0), while the empirical Type II error rate is defined by FN/m(1) and the estimated average power is TP/m(0). See [12 and 13].

doi:10.1371/journal.pone.0023009.t001

Grouping methods

Grouping tests or cluster based methods are relying to the concept called “borrowing strength” which is well summarized by [19] who wrote ‘A more explicit use of the dependence structure should result in a powerful method’. This concept was adopted for example in [20] in a geographical application where clusters were defined to be geographical regions. Following this concept, [21] used a cluster based approach to analyze fMRI data to detect activations. They argued by the fact that voxels of a neurological type belonging to a unique anatomical region will usually exhibit positively correlated behavior whether the measure is physiological or functional [21,22]. A quite different cluster based method was proposed in [23] to analyze fMRI data. In their proposal, the choice of clusters is defined beforehand using prior information. In general, the grouped tests do not have to correspond to units in the same vicinity. For this reason, we say that tests are grouped into subsets instead of using the term of clusters. For example, a subnetwork could be the connections (edges) between two clusters of ROIs (a bi-subnetwork).

We give a general formulation of grouping tests into subsets, and then explore further the case of using the mean as a summary statistic.

General formulation of grouping methods. Consider a set \( \{ h_j \}_{j = 1, \ldots, m} \) of m hypotheses to be tested. Each hypothesis is set to test a certain assessment about a single unit that we call atom. The set of all atoms is called the global set of interest S. Each single

Platin and Scheich
hypothesis $h_j (j = 1, \ldots, m)$ is tested based on the observation of a vector $x_{jk}$ (of dimension $q$) measured in the atom $a_j$ for each individual $k (k = 1, \ldots, n)$. The global data (the design matrix) is of dimension $m \times q \sim n$. Note that if $n = 1$, we are reduced to a single-subject experiment. We group the $m$ atoms into $s$ disjointed subsets $S_1, \ldots, S_s$ such that $\bigcup_j S_j = S$ and $S_j \cap S_i = \emptyset$ for $i \neq j$, where a subset contains a set of atoms that are linked by attributes associated with the problem at hand. The number of atoms in the subset $S_i$ is $m_i (i = 1, \ldots, s)$, that is, $\sum_i m_i = m$.

For each subset $S_i$ we consider a univariate summary statistic $t_i$, a function of all observed data within the corresponding subset of all individuals, that is $t_i = t_i (\{x_{jk} | j \in S_i, k = 1, \ldots, n\})$ for all $i = 1, \ldots, s$. The generalization to a multivariate summary statistic is straightforward. Note that if $m_i = 1$, we are reduced to the multivariate atom-wise analysis.

The main advantages of dividing the global set of interest into subsets of atoms are the following: the reduction of the number of tests and the reduction of the noise variance. Effectively, the number of subsets $s$ can be much reduced compared to the number $m$ of atom-wise tests. For example, using the Bonferroni correction, the global test level $\alpha$ is divided by $s$ instead of $m$ for each single test. This considerably increases the power of comparisons. In addition, aggregating will reduce the variance of the outcome, which facilitates the detection of significant structures, while isolated significant atoms are rarely considered. This leads to a desirable robustness and is the second reason for increased power. The main disadvantage of grouping methods is the potential loss of information. If the effect is concentrated on a single atom for example, diluting the atom inside a subset weakens the effect. Furthermore, if the subset contains atoms with positive effect and atoms with negative effect, using the mean as a summary statistic clearly disadvantages the grouping methods.

**Properties of the subset-wise analysis using the mean as a summary statistic.** Using the mean of the values observed in all the atoms within a subset is a natural choice of summary statistics. It is also the simplest choice to derive analytical expressions that afford the comparison between the different strategies. However, other summary statistics with a contextual meaning would be preferable such as those we propose in the application to brain networks.

We now, contrast the strategy based on subsets (Subset-Wise-Analysis, SWA) to the approach that separately tests atoms (Atom-Wise-Analysis, AWA) in terms of power. We restrict the comparison to the case of positive effect only.

SWA and AWA solve different problems. While SWA tests the main advantages of the different subsets, AWA tests the significance of atoms. This has to be kept in mind when comparing the power of the two different strategies.

Consider the problem of detecting atoms with positive effect when comparing two groups of individuals. Denote by $x_{jk}$ for all $j = 1, \ldots, m$ and for all $k = 1, \ldots, n = n_1 + n_2$ the measurements associated with the atom $a_j$ of the individual $k$, where $n_1$ and $n_2$ are the sizes of the two groups. For simplification, suppose that $n_1 = n_2$ and that the measurements are univariate. For the first group, the observations are of the form $x_{jk} = \mu_j + \varepsilon_{jk}$ for all $j = 1, \ldots, m$ and for all $k = 1, \ldots, n_1$. For the second group, in the non-affected atoms (where there is no effect), the observed values are of the form $x_{jk} = \mu_j + \varepsilon_{jk}$ and in the affected atoms (that contain the positive effect) the observations are of the form $x_{jk} = \mu_j + \Delta + \varepsilon_{jk}$ for all $k = n_1 + 1, \ldots, n$. For both cases, $\varepsilon_{jk} (j = 1, \ldots, m$ and $k = 1, \ldots, n)$ are independent realizations of a normal random white noise, that is, $\varepsilon_{jk} \sim N(0,1)$. $\Delta$ is the raw effect in the atom $a_j$. The number of non-affected atoms and affected atoms within the global set of interest are $m(0)$ and $m(1) = m - m(0)$ respectively. We assume that the variance $\sigma^2$ is known and includes both the noise variance and the intra-subject variability. When proceeding according to the AWA to detect the atoms with positive effect, we perform one sided tests $h_j^{(0)} : \Delta_j = 0$ vs. $h_j^{(1)} : \Delta_j > 0$, for all $j = 1, \ldots, m$. In all the computations, we use the simple case $\Delta = \Delta_j$.

If the aim is to strongly control the FWER of atoms $\leq \alpha$ using the Bonferroni procedure, each single test is performed at level $\alpha/m$. The null hypothesis $h_j^{(0)}$ is thus rejected if the decision, $d_j = (1/m) \sum_{k = n_1 + 1}^{n_1 + n_2} x_{jk} - (1/n_1) \sum_{j = 1}^{m(1)} x_{jk}$ satisfies $d_j > c_j$, and $c_j$ is given by

$$c_j = \frac{\sigma}{\sqrt{n_1 m}} \Phi^{-1} \left( 1 - \frac{\alpha}{m} \right).$$

where $\Phi$ is the cumulative distribution function of the standard normal distribution.

Now, the global set of interest is divided into $s$ subsets. The number of non-affected subsets, which contain only non-affected atoms, is $s(0)$ and the number of affected subsets, which contain at least one affected atom, is $s(1) = s - s(0)$. Consider an affected subset of size $m_i$, where only $m_i(1) = m_i(\{1, \ldots, m_i\})$ of the atoms are affected with positive effect $\Delta = \Delta_j$. In fact, it is more realistic to consider partially affected subsets than considering completely affected subsets. This could happen for example if the segmentation that defines the subsets does not exactly match the true limits of the anatomical or functional regions. It could happen also, if only a proportion of atoms in the anatomical or functional region are affected.

For each subset $(m/s)$ of size $m_i (i = 1, \ldots, s)$, we construct a summary statistic $t_i$, based on the means of the differences between the two groups for each atom in the subset $S_i$, that is, $t_i = t_i (\{\sum_{j \in S_i} x_{jk} \})$ for all $i = 1, \ldots, s$. The generalization to a multivariate summary statistic is straightforward.

For the non-affected subsets, the statistic $t_i$ is a realization of a random variable $T_i \sim N \left( 0, \frac{\sigma^2}{n} \right)$, while for the affected subsets, $t_i$ is a realization of a random variable $T_i \sim N \left( \pi_i \Delta, \frac{\sigma^2}{n} \right)$ where $\pi_i = m_i(1)/m_i$ is the proportion of affected atoms in the subset $S_i$. The distribution of $T_i$ depends on the distribution of $\varepsilon_{jk}$. In the non-normal case, if the size $m_i$ or if the number $n$ of subjects are large enough, the central limit theorem (CLT) leads to an approximation of the distribution of $T_i$ by a normal distribution. Still, we have to consider that the $p$-values are sensitive to the distribution of the summary statistic in the right tail of the approximated distribution.

We define the FWER of subsets as the probability of having at least one false positive subset, that is, declaring as significant a subset that contains no affected atoms. To control the FWER of subsets at level $\alpha$ and again, using the Bonferroni procedure on subsets, the null hypothesis $H_j^{(0)} : \pi_i = 0$ is rejected for the subset $S_i$ if the corresponding observed summary statistic $t_i > c_i$, and $c_i$ is given by

$$c_i = \frac{\sigma}{\sqrt{n_i m_i}} \Phi^{-1} \left( 1 - \frac{\alpha}{s} \right).$$

This relation shows that the critical values for the SWA do not only depend on the ratio $(m/s)$ which represents the reduction of
the number of tests, but also on \( \frac{1}{\sqrt{m_i}} \) which decreases as long as the subset size increases and is due to the reduction in the noise variance.

We can also compare the power curves corresponding to the two different strategies. In the case of the AWA, the power \( P \ weight of detecting an affected atom is given by

\[
P_{\text{pow}j} = 1 - \Phi \left( \Phi^{-1} \left( 1 - \frac{2}{m_i} \right) - \sqrt{m_i} \pi_1 \Delta \right),
\]

whereas the power of detecting an affected subset that contains a proportion \( \pi_i \) of affected atoms is

\[
P_{\text{pow}i} = 1 - \Phi \left( \Phi^{-1} \left( 1 - \frac{2}{m_i} \right) - \sqrt{m_i} \pi_i \Delta \right).
\]

Given this two power functions, we want to know when \( P_{\text{pow}i} > P_{\text{pow}j} \), that is, under what conditions we have a greater chance to detect an affected subset \( S_i \) than detecting the affected atoms \( a_j \in S_i \). We have

\[
P_{\text{pow}i} > P_{\text{pow}j} \Leftrightarrow \Phi^{-1} \left( 1 - \frac{2}{m_i} \right) - \sqrt{m_i} \pi_i \Delta > \Phi^{-1} \left( 1 - \frac{2}{m_j} \right) - \sqrt{m_j} \pi_j \Delta,
\]

that is,

\[
P_{\text{pow}i} > P_{\text{pow}j} \Leftrightarrow \pi_i < \pi_j,
\]

where

\[
\pi_i = \frac{1}{\sqrt{m_i}} \left( \frac{1}{\sqrt{m_i}} \right) \Phi^{-1} \left( 1 - \frac{2}{m_i} \right) - \Phi^{-1} \left( 1 - \frac{2}{m_j} \right).
\]

If the proportion of affected atoms is less than \( \pi_i \), the power of detecting an affected subset is reduced since the effect is diluted. However, when the proportion of affected atoms \( \pi_i \) exceeds the threshold value \( \pi_i \), the SWA is more powerful than the AWA in the sense that detecting subsets which contains at least a proportion \( \pi_i \) of affected atoms is more powerful than detecting the individual affected atoms within such a subset. The threshold \( \pi_i \) depends on the subset size \( m_i \). Larger subsets give more advantage than smaller subsets, but of course, this has an influence only when different resolutions are available since we supposed that the segmentation that defines subsets is predefined. Note that the threshold \( \pi_i \) depends also on the variance \( \sigma^2 \), the raw effect \( \Delta \) and on the number of subjects \( n \) in the sense that the advantage of the SWA becomes more pronounced when the variance increases (when the Signal to Noise Ratio (SNR) decreases) and when the raw effect and the number of subjects decrease.

Note also that \( \Phi^{-1} \left( 1 - \frac{2}{m_i} \right) - \Phi^{-1} \left( 1 - \frac{2}{m_j} \right) \) is positive since \( 1 - \frac{2}{m} < 1 - \frac{2}{m} \) and \( \Phi \) is an increasing function. Then,

\[
\sqrt{m_i} \pi_i > 1 \Rightarrow m_i \geq \pi_i \Rightarrow P_{\text{pow}i} > P_{\text{pow}j} \Leftrightarrow P_{\text{pow}j} > P_{\text{pow}i}.
\]

So, a sufficient condition for having \( P_{\text{pow}i} > P_{\text{pow}j} \) is that the number of affected atoms in that subset is greater than the square root of the size of the subset. A similar result was found by simulations in [23] using an adaptive FDR procedure.

We simulated tests when the observed values in the non-affected and in the affected atoms are independent realizations of \( x_k \sim N(0, n) \) and \( x_k \sim N(\Delta, n) \) respectively. This is a particular case where \( m_j = 0, \sigma^2 = n, \Delta = \Delta \). The number of affected atoms \( m_i = 1, \ldots, m_i \) for all affected subsets. In this case, the power of detecting a significant effect in an affected subset given by equation 4 becomes

\[
P_{\text{pow}i} = 1 - \Phi \left( \Phi^{-1} \left( 1 - \frac{2}{m_i} \right) - \pi_i \sqrt{m_i} \Delta \right).
\]

For simplicity, the \( s \) subsets are chosen to have the same size \( m_i \), for all \( i \) in \( \{1, \ldots, s\} \). To control the rate of false positives on atoms and subsets, we used the Bonferroni procedure to control the FWER and the Benjamini-Hochberg procedure (BH95) introduced in [12] to control the FDR.

In Figures 1 and 2, we see the behavior of the power of detecting partially affected subsets, depending on the proportion of truly affected atoms \( \pi_i \) using the SWA, compared to the power of detecting affected atoms using the AWA. In figure 1, we used the Bonferroni procedure, and in figure 2 we used the BH95 procedure. The raw effect was set to be \( \Delta = 1 \) or 2 and the subset sizes are \( m_i = 4, 8 \) or 16. The power plotted is the average of \( \frac{TP}{m_i} \) across the simulations (average power defined in [24]). Since \( \Delta = \Delta \) in all affected atoms, the average power is equal to the per-pair power defined in [25]. This holds also for the power of detecting affected subsets since they have the same size, where we obviously have to replace \( n \) by \( s \) and \( m_i \) by \( \Delta \). In addition, the power has a different meaning in the two cases. For the AWA we want to detect significance on the atom level, while for the SWA we seek significant subsets.

In all cases, for small values of the proportion \( \pi_i \), the mean as a summary statistic reduces the power of detecting a significant effect within such subsets because the effect is diluted. However, when the proportion of affected atoms \( \pi_i \) exceeds the threshold value \( \pi_i \), the SWA is more powerful than the AWA in the sense that detecting subsets which contains at least a proportion \( \pi_i \) of affected atoms is easier than detecting the individual affected atoms. In both cases, by using the Bonferroni procedure or the BH95 procedure, the threshold \( \pi_i \) depends on the size of subsets \( m_i \) and the raw effect \( \Delta \). This is in accordance with the result derived analytically for the Bonferroni case.

A two-stage atom-wise analysis. In the previous section, we showed that detecting affected subsets using the mean as a summary statistic is more powerful than trying to detect individual affected atoms when a subset contains more than \( \sqrt{m_i} \) affected atoms. We expect that this condition would be satisfied in neuroimaging, because we assume the existence of positive correlations between atoms within the same subset. This was the principal motivation of our work. We want to shed light in this section, on an interesting question related to the grouping tests strategy. Consider the following two stage AWA procedure. First, we apply the SWA to detect affected subsets. Then, in each subset declared as affected, we perform a local investigation, that is, we apply a multiple comparisons procedure inside that subset to detect affected atoms. Does this have an advantage over the classic AWA procedure? As the question has a complex answer, we restrict our discussion to a simple special case.

**Proposition 1.** The subset \( S_i \) contains \( m_i \geq 1 \) atoms and is considered along with \( s \) subsets. The corresponding observations are either \( x_k \sim N(0, n) \) or \( x_k \sim N(\Delta, n) \). The summary statistic is
Figure 1. Power of detecting affected atoms and partially affected subsets depending on the proportion $\pi_i$. For the multiplicity correction, the Bonferroni procedure is used. Three different values of the subsets’ size $m_i$ (4, 8 or 16) and two different values of the raw effect $\Delta$ (1 or 2). The other parameters are: $m = 4096$, $s^{(1)} = s / 5$, $\alpha = 0.05$.

doi:10.1371/journal.pone.0023009.g001
Figure 2. Power of detecting affected atoms and partially affected subsets depending on the proportion \( \pi_i \). For the multiplicity correction, the BH95 procedure is used. Three different values of the subsets' size \( m_i \) (4, 8 or 16) and two different values of the raw effect \( \Delta \) (1 or 2) are used. The other parameters are:

- \( m = 4096 \),
- \( s^{(i)} = s/5 \),
- \( \alpha = 0.05 \).

doi:10.1371/journal.pone.0023009.g002
the subset mean difference $t_i$, which is significant in the Bonferroni sense, that is, $t_i \geq c_i = \frac{1}{\sqrt{m_i}} \Phi^{-1}\left(1 - \frac{\alpha}{s}\right)$. If $m_i \leq s$, it follows that at least one individual atom will also pass the second-stage Bonferroni test.

**Proof.** Suppose that all atoms in the subset $S_i$ satisfy $d_j < c_m = \frac{1}{\sqrt{m_i}} \Phi^{-1}\left(1 - \frac{2}{m_i}\right)$, that is, no atom pass the second-stage Bonferroni test. We have

$$d_j < c_m \Rightarrow t_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d_j < \frac{1}{m_i} \sum_{j=1}^{m_i} \frac{1}{\sqrt{m_i}} \Phi^{-1}\left(1 - \frac{2}{m_i}\right).$$

Given that $m_i \leq s \Rightarrow \Phi^{-1}\left(1 - \frac{2}{m_i}\right) \leq \Phi^{-1}\left(1 - \frac{2}{s}\right)$, we have

$$t_i < \frac{1}{\sqrt{m_i}} \left(\frac{1}{\sqrt{m_i}} \Phi^{-1}\left(1 - \frac{2}{s}\right)\right) = \frac{1}{\sqrt{m_i}} c_i < c_i.$$

**Proposition 2.** In a significant subset, the power of detecting affected atoms is larger when using the two-stage AWA than using the classic AWA if the Bonferroni procedure is used in both cases.

**Proof.** Obviously this is true, because

$$c_j = \Phi^{-1}\left(1 - \frac{2}{m_i}\right) > c_m = \Phi^{-1}\left(1 - \frac{2}{m_i}\right).$$

This proposition does not assure that the two-stage AWA is more powerful than the global AWA. In fact, the global AWA may detect other affected atoms, which do not belong to the declared significant subsets. However, the two-stage AWA is more powerful in the subsets that contain more than $\sqrt{m_i}$. Hence, the two-stage AWA tends to detect affected structures, which is often more interesting for the researcher. See [23].

In addition, simulations show that the rate of the false positives is not controlled in the strong sense by the two-stage AWA even if the Bonferroni procedure is used in the two stages. However, the weak control of false positives can easily be proved in this case.

### The SWA in the context of connection matrices and brain networks

Suppose that a global region of interest in the human brain is subdivided into small ROIs. A connection matrix is a weighted symmetric matrix $A$ where rows/columns correspond to ROIs and each cell $A(r,r')$ of the matrix represents a certain measure of the connectivity between the two ROIs $r$ and $r'$. A connection matrix defines a network where the nodes correspond to the ROIs and the weighted edges correspond to the measure of the connectivity between the corresponding ROIs. A subnetwork of the human brain network corresponds to a block of the connection matrix. We consider separately two kinds of subnetworks. The first type represents the intra-connection within the same group of ROIs and whose corresponding blocks are localized on the diagonal of the global connection matrix. The second type corresponds to the networks that represent the interconnections between two groups of ROIs. These are bi-subnetworks and their corresponding blocks are localized out of the diagonal in the global connection matrix. See Figure 3 for an illustration.

#### How to choose subsets or subnetworks.

As mentioned in the introduction, the subsets or subnetworks are chosen beforehand by the researcher based on an a priori knowledge. This prior information is obtained either by an investigation of independent data sets or by using a segmentation atlas. Both the AWA and the SWA are hampered by a common difficulty, due to the fact that atoms or subsets have to correspond between the different subjects. This important difficulty was emphasized in [23]. In practice, the ROIs do not exactly match geometrically between individuals, in particular for higher resolutions. Consequently, the effect of affected atoms will be scattered and will be diluted. However, the mismatched ROIs have more chance to be located in the same group of ROIs used to define a subnetwork, in particular, if they are geometrically not close to the frontiers of the subregions. In this case, the advantage of the SWA is even more pronounced. The mismatching between atoms causes problems for any method that define subsets on the basis of the estimated correlations between atoms.

#### Summary statistics and statistical analysis.

As mentioned in the introduction, three levels of analysis are considered to analyze the brain complex network. The first level is the atom level where we perform an AWA. Atoms could be an ROI (a node) or a

![Figure 3. Illustration of the different types of subnetworks within a brain network.](image-url)
connection between two ROIs (an edge). In the first case, one could use some network measures that can be evaluated for a node. For example, one can evaluate for each node its degree, its clustering coefficient, etc. In the second case, the weight of the edge can be used. We call the AWA performed with network based measures the Atom Network Based Analysis (ANBA). The second level considered is the global level that corresponds to the global network. We call the analysis at that level the Global Network Based Analysis (GNBA). In the third level, we consider the analysis at the level of subnetworks by considering each subnetwork as a complex network. We call the analysis at that level the Sub-Netowrk Based Analysis (SNBA). In the two last levels, a variety of summary statistics could be chosen, but of course, the number of tests performed is not the same at the two levels. The choice of summary statistics depends on the subnetwork. One can construct summary statistics by averaging the network based measures on atoms within each subnetwork. This has the properties of the subset mean discussed in the previous sections. For example, one can use the mean of the edges’ weights or the mean of the nodes’ degrees, centrality, efficiency, modularity, etc. On the other hand, several summary statistics which do not necessarily use the mean of the values and which cannot be defined on the level of atoms could be used as a summary statistic. For example, one can use the node-degree distribution, the small world properties, etc. of each sub-network as a summary statistic. See [26,27] for a list of relevant network measures that could be used in the brain network analysis. It is difficult to use network based summary statistics that do not use the mean of atom values such as small world properties, as a summary statistic since the control of false positives depends on the distribution of the summary statistics, in particular, when using the two-stage procedure, where conditional distributions come into play. In addition, the network based measures are evaluated on noisy connection matrices. It is then, worth to investigate the impact of this factor on the statistical analysis. This two important aspects need to be investigated and they are parts of our future work.

Application to structural connection matrices of the human brain

The purpose of this application section is to give a simple example of a real application of the proposed strategy to compare normalized whole-brain structural connection matrices derived from diffusion MRI tractography.

The construction of structural brain connection matrices. The processing pipeline used to derive connection matrices compared in this application is basically divided into two pathways. See [3] for more details. On one hand, the cortical surface is extracted from a high resolution T1-weighted Magnetic Resonance (MR) image and subdivided into \( N = 83 \) anatomical parcels by matching the most important sulci using atlas-based segmentation. On the other hand, a whole brain tractography is performed on diffusion MR images, which results in millions of virtual fibers spread over the brain. The combination of these two procedures allows the construction of connection matrices by computing the connection density for each pair of ROIs. Considering the cortical parcellation and the white matter tractography described above, the fiber bundle \( F(r,r') \) connecting the pair of ROIs \( (r,r') \) could be identified. The value of the connection matrix cell \( A(r,r') \) is the connection density between these ROIs and is defined as follows:

\[
A(r,r') = \frac{1}{\int \mathbf{f}(r,r')/l(r)}
\]

where \( l(r) \) is the length of the fiber along its trajectory.

The correction term \( l(r) \) in the denominator is needed to eliminate the linear bias towards longer fibers introduced by the tractography algorithm. We obtain at the end of the application of the pipeline an \( N \times N \) symmetric matrix \( \mathbf{A} \). The pipeline is summarized in Figure 4.

Description of the data. The clinical group used for comparing the AWA and the SWA or equivalently, for comparing ANBA and SNBA, is a group of subjects with a 22q11.2 deletion syndrome (22q11DS) [9,10]. Amongst other manifestations, this syndrome shows a mild cognitive impairment (MCI) resulting from a loss of IQ performance [11,28].

Amongst the 22q11DS population, there is an existing discrepancy in the cognitive abilities between patients with a relatively high IQ level (above 70) and patients with a low IQ level (below 70). Delineating the structural brain connectivity that sustains this discrepancy may provide useful clues to understand how brain connections are involved in the loss of intellectual functioning in the 22q11DS population.

The high IQ group (\( >70 \)) is composed of \( n_h = 14 \) patients, 7 girls and 7 boys (mean age = 14.5±2.9 years, ranged from 7.4 to 17.6 years and mean IQ = 80.4±6.7). The low IQ group (\( \leq 70 \)) is composed of \( n_l = 12 \) patients, 5 girls and 7 boys (mean age 14.8±3.9 years old ranged from 7.2 to 19.8 years and mean IQ = 60±6.8).

In addition of the connection density matrices, we have Fractional Anisotropy (FA) connection matrices where the weight represents the mean FA along a fiber tract connecting a pair of ROIs.

Data analysis. To detect atoms (connections) or subnetworks that differ between the two groups, we consider different strategies:

First, we consider the ANBA analysis using all the available variables (density of fibers, FA, density of fibers truncated by the FA) and the combinations of variables as univariate and multivariate statistics.

Second, we consider the SNBA analysis with different summary statistics.

- The SWA using the mean of the density of fibers \( \tau_j \) in each subnetwork as a summary statistic.
- The truncated proportion, where the summary statistic in each subnetwork is defined by \( \bar{\tau}_j = \frac{1}{m_i} \sum_{\{j \in S_i\}} I_{\{x_j > \mu_j\}} \), where \( I \) is the indicator function. This variable is interpreted as the number of interconnections between ROIs within the subnetwork \( S_i \) (number of edges). We use here, a truncation threshold based on the FA values. One can choose other thresholds for the truncation depending on the nature of the problem.
- The effective mean, where the summary statistic is defined by

\[
\frac{\sum_{\{j \in S_i\}} (x_j)\{I_{\{x_j > \mu_j\}}\}}{\sum_{\{j \in S_i\}} I_{\{x_j > \mu_j\}}} \]

which is interpreted as the mean of edges’ weights in the subnetwork \( S_i \). Since the distribution of the two later statistics is unknown, we use the Wilcoxon-Mann-Whitney rank test [29] instead of the Student test to derive the p-values.
- The summary test statistic is a p-variate statistic that includes a combination of \( p \) univariate statistics among the summary statistics defined above. In this case, the test follows by computing the statistic \( f = \frac{n_h n_l}{n_h + n_l - p - 1} \frac{(\bar{\tau}_l - \bar{\tau}_l)^T C^{-1} (\bar{\tau}_l - \bar{\tau}_l)}{p(n_h + n_l - 1)} \), where \( C \) is the estimated covariance matrix of the data given by

\[
C = \frac{1}{m_h + m_l - 2} \left( \begin{array}{cc}
(n_h - 1) C_h + (n_l - 1) C_l
\end{array} \right)
\]

and \( C_h \) and \( C_l \) are the estimated covariance matrices of the high IQ group and the low IQ group respectively. The statistic \( f \) follows a Fisher distribution \( F_{p(n_h + n_l - 1) - p} \).
We apply the SNBA to the data by defining subnetworks as the interconnections between brain cortex lobes. The brain cortex is divided into 13 lobes. By considering the interconnections between these 13 lobes and the intra-connections within these 13 lobes, we define \( s = \binom{13}{2} + 13 = 91 \) subnetworks. The SNBA is followed by a local investigation in the subnetworks detected as significant.

For the multiplicity correction of p-values we use the Bonferroni correction with level FWER = 0.05 and the BH95 procedure with FDR = 0.1 for all tests.

**Results and Discussion**

First, using the AWA (or the ANBA), no significant results were found with any statistic (univariate or multivariate). This is due first, to the small number of individuals in each group and second, to the fact that the two compared groups belong to the same community of patients and so, no big differences could exist between the two groups (small raw effect) and of course, this is due to the multiplicity correction since \( \binom{83}{2} + 83 = 3486 \) tests are performed simultaneously.

Using the SWA (or the SNBA), three subnetworks were detected as significant after the BH95 correction of the p-values. Only one of the three subnetworks passed the Bonferroni correction.

Within these three subnetworks, only 5 connections are detected as significantly different within the three detected subnetworks.

The application results presented here are in accordance with the performance showed in the theoretical part either by analytical formulas or by simulations. We are in a situation where the raw effect and the sample size are both small. So, the advantage of the SWA over the AWA is pronounced.

Amongst the five significant connections detected as significant, four were found to be connectively reduced in 22q11DS patients with a low IQ compared with the ones with a high IQ. Right caudate intra-connectivity and the number of fibers connecting it with the putamen were reduced. Right accumbens nucleus inter-connectivity with sub-thalamic nucleus, and left lingual intra-connectivity were also found reduced. The last connection referring to the number of fibers connecting the left putamen and the left superior temporal cortex were increased.

Morphological alterations of the caudate, putamen, left superior temporal gyrus (STG) and lingual area have frequently been found in 22q11DS [30–32].

Here, we show that there is a specific alteration of the connectivty of the striatal structure (composed of the caudate and the putamen) affecting the cortico-striatonigral-thalamocortical circuit [33] and therefore may impairs the cognitive functioning [34] in the 22q11DS with the low IQ.

**Conclusion and future directions**

We proposed a statistical network based strategy to analyze and compare brain networks with the aim of increasing the power of detections. The strategy is based on grouping tests into subsets that define brain subnetworks in order to reduce the number of tests. We showed in the simulation examples and in the real application the relevance in neuroimaging, in particular, when the number of
tests is very large and when the raw effect and the sample size are small. We proposed as well the use of the local investigation in the significant subnetworks in a second stage. We can summarize the subnetwork based analysis (SNBA) as follows:

1. Define subnetworks based on the prior knowledge.
2. Choose an appropriate summary statistic for each subnetwork, which has a contextual interpretation.
3. Apply a multiple comparisons procedure (that controls the FWER, the FDR or other measures).
4. If desired, a local investigation in significant subnetworks may help to interpret the results.

We showed that if a subset of size $m_i$ contains more than $m_i \tilde{p}_i$ affected atoms where $\tilde{p}_i$ is given by equation 5, the power of detecting such a subset is greater than the power of detecting each affected atom within the subset. The threshold $\tilde{p}_i$ is always smaller than $\frac{1}{\sqrt{m_i}}$ and we expect that this condition i.e. $\left(\sqrt{m_i} \tilde{p}_i > 1\right)$ is often satisfied in neuroimaging, that is, a part of a subset or a subnetwork behaves coherently.

It should be emphasized that the proposed strategy as presented in the application uses some particular examples of summary statistics and shows a real advantage over the atom-wise comparisons. A part of our future work will be focused on how to estimate more complex network based summary statistics on noisy connection matrices and to show the control of false positives under their use.

Acknowledgments

We thank the Academic Editor and the reviewers for their insightful and constructive comments.

Author Contributions

Conceived and designed the experiments: DEM SM. Performed the experiments: DEM M-CO. Analyzed the data: DEM. Contributed reagents/materials/analysis tools: DEM M-CO LC PH RM SE JPT. Wrote the paper: DEM M-CO SM.

References

1. Hagmann P (2005) From diffusion MRI to brain connectomics. PhD Thesis (Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland).
2. Sporns O, Tononi G, Kotter R (2005) The human connectome: A structural description of the human brain. PLOS Comput Biol 1: e42.
3. Hagmann P, Cammoun L, Gigandet X, Meuli R, Honey CJ, et al. (2008) Mapping the Structural Core of Human Cerebral Cortex. PLOS Biology 6: e139.
4. Biwali B, Yetkin FZ, Haughton VM, Hyde JS (1995) Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. Magn Reson Med 34: 537–541.
5. Raichle ME, MacLeod AM, Snyder AZ, Powers WJ, Gusnard DA, et al. (2001) A default mode of brain function. Proc Natl Acad Sci USA 98: 676–682.
6. De Smedt B, Devriendt K, Fryns JP, Vogels A, Gewillig M, et al. (2007) A new syndrome involving cleft palate, cardiac anomalies, typical facies, and learning disabilities: velo-cardio-facial syndrome. Cleft Palate J 15(1): 56–62.
7. Fair DA, Cohen AL, Dosenbach NUF, Church JA, Miezin FM, et al. (2008) The maturing architecture of the brain’s default network. PNAS USA 105: 4028–4032.
8. Fornito A, Zalesky A, Bullmore ET (2010) Network scaling effects in graph analytic studies of human resting-state fMRI data. Front Syst Neurosci 4: 22.
9. Shprintzin RJ, Goldberg RB, L ’emis ML, Sidoti EJ, Berkman MD, et al. (1978) Deletions and microdeletions of 22q11.2 in velo-cardio-facial syndrome. Am J Med Genet 44(2): 261–8.
10. Driscoll DA, Spinner NB, Budarf ML, McDonald-McGinn DM, Zackai EH (1992) Deletions and microdeletions of 22q11.2 in velo-cardio-facial syndrome. Am J Med Genet 44(2): 261–8.
11. Horn M, Dunnett CW (2004) Power and sample size comparisons of stepwise FWE and FDR controlling test procedures in the normal many-one case. Lecture Notes-Monograph Series 47: 48–64.
12. Logan BR, Rowe DB (2004) An evaluation of thresholding techniques in fMRI analysis. Neuroimage 22: 95–108.
13. Simes RJ (1986) An improved Bonferroni procedure for multiple tests of significance. Biometrika 73: 751–754.
14. Horn M, Dunnett CW (2004) Power and sample size comparisons of stepwise FWE and FDR controlling test procedures in the normal many-one case. Lecture Notes-Monograph Series 47: 48–64.
15. Shprintzin RJ, Goldberg RB, Lewin ML, Sidoti EJ, Berkman MD, et al. (1978) Deletions and microdeletions of 22q11.2 in velo-cardio-facial syndrome. Am J Med Genet 44(2): 261–8.
16. Fair DA, Cohen AL, Dosenbach NUF, Church JA, Miezin FM, et al. (2008) The maturing architecture of the brain’s default network. PNAS USA 105: 4028–4032.
17. Fornito A, Zalesky A, Bullmore ET (2010) Network scaling effects in graph analytic studies of human resting-state fMRI data. Front Syst Neurosci 4: 22.
18. Logan BR, Rowe DB (2004) An evaluation of thresholding techniques in fMRI analysis. Neuroimage 22: 95–108.
19. Horn M, Dunnett CW (2004) Power and sample size comparisons of stepwise FWE and FDR controlling test procedures in the normal many-one case. Lecture Notes-Monograph Series 47: 48–64.
20. Logan BR, Rowe DB (2004) An evaluation of thresholding techniques in fMRI analysis. Neuroimage 22: 95–108.
21. Horn M, Dunnett CW (2004) Power and sample size comparisons of stepwise FWE and FDR controlling test procedures in the normal many-one case. Lecture Notes-Monograph Series 47: 48–64.
22. Logan BR, Rowe DB (2004) An evaluation of thresholding techniques in fMRI analysis. Neuroimage 22: 95–108.
23. Logan BR, Rowe DB (2004) An evaluation of thresholding techniques in fMRI analysis. Neuroimage 22: 95–108.
24. Logan BR, Rowe DB (2004) An evaluation of thresholding techniques in fMRI analysis. Neuroimage 22: 95–108.
25. Logan BR, Rowe DB (2004) An evaluation of thresholding techniques in fMRI analysis. Neuroimage 22: 95–108.
26. Logan BR, Rowe DB (2004) An evaluation of thresholding techniques in fMRI analysis. Neuroimage 22: 95–108.