MOX-Report No. 39/2019

Multiscale null hypothesis testing for network-valued data: analysis of brain networks of patients with autism

Lovato, I.; Pini, A.; Stamm, A.; Taquet, M.; Vantini, S.

MOX, Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox-dmat@polimi.it http://mox.polimi.it

Multiscale null hypothesis testing for network-valued data: analysis of brain networks of patients with autism

Ilenia Lovato, Department of Mathematics, University of Pavia
C.so Strada Nuova 65, 27100 Pavia, Italy
Alessia Pini, Department of Statistical Sciences,
Università Cattolica del Sacro Cuore, Largo A. Gemelli 1, 20123, Milano, Italy[1]
Aymeric Stamm, Laboratoire de Mathmatiques Jean Leray, CNRS UMR 6629,
Nantes, Pays de la Loire, 44322, France[2]
Maxime Taquet, Computational Radiology Laboratory, Boston Children's Hospital – Harvard Medical School, Boston, MA, 02115, USA
Simone Vantini, MOX - Department of Mathematics, Politecnico di Milano
P.za Leonardo da Vinci 32, 20133, Milano, Italy

Abstract

Networks are a natural way of representing the human brain for studying its structure and function and, as such, have been extensively used. In this view, case-control studies for understanding autism pertain to comparing samples of healthy and autistic brain networks. In order to understand the biological mechanisms involved in the pathology, it is key to localize the differences on the brain network. Motivated by this question, we hereby propose a general non-parametric finite-sample exact statistical framework that allows to test for differences in connectivity within and between pre-specified areas inside the brain network, with strong control of the family-wise error rate. We demonstrate unprecedented ability to differentiate children with non-syndromic autism from children with both autism and tuberous sclerosis complex using EEG data. The implementation of the method is available in the `R` package `nevada`.

Keywords: Network-valued data, multiple comparisons, family-wise error rate, neuroscience, brain electroencephalography networks, autism.

[1]Alessia Pini and Aymeric Stamm were also supported by MOX - Department of Mathematics, Politecnico di Milano.

[2]Aymeric Stamm was also supported by Center for Analysis, Decisions, and Society, Human Technopole
Palazzo Italia, Via Cristina Belgioioso, 20157 Milano, Italy, by Computational Radiology Laboratory - Boston Children's Hospital, Harvard Medical School and by *2014 Polimi International Fellowship*

1 Introduction

Understanding the human brain is a scientific objective that has received a lot of attention over the past decades. Non-invasive imaging techniques have been devised such as electroencephalography (EEG) or magnetic resonance imaging (MRI), which make it possible to study the brain *in-vivo* in a non-invasive fashion. The brain is naturally organized into functional centers [Fox et al., 2005] interconnected by a wiring of axons. As a result, it is increasingly represented as a *network*, as shown by the tremendous amount of review articles on the topic [Sporns et al., 2005, Bassett and Bullmore, 2006, Bullmore and Sporns, 2009, Rubinov and Sporns, 2010, Van Den Heuvel and Pol, 2010, Bassett and Gazzaniga, 2011, Bullmore and Sporns, 2012, Smith et al., 2013, Preti et al., 2017] and of publications endorsed by the National Academy of Sciences [Greicius et al., 2003, Mantini et al., 2007]. A network is a combinatorial object defined by means of a set V of *vertices* (or nodes) and a set E of *edges* (or links). An existing interaction between two vertices can be expressed either by a weighted edge (strength or distance between two vertices) or an unweighted binary edge (presence or absence). One important application is improved diagnosis, understanding and follow-up of neurodegenerative disorders, which can be achieved through case-control studies of brain networks. In this context, the goal is to assess whether the disease process has affected the connections/edges by statistically comparing a sample of networks from healthy subjects against a sample of networks from patients. In statistical terms, this is often referred to as the local two-sample testing problem, applied to *network-valued data* (NVD), which includes two main difficulties.

First, the statistical unit is not a number nor a vector as in traditional multivariate data analysis but a network. While there is a huge body of existing works in the literature that focus on representing, describing and modelling a single network in order to analyze the relationships (edges) between a set of entities (vertices) [Newman, 2003], little attention has been put towards network-valued two-sample testing. To the best of our knowledge, there are only three recent papers that tackle this issue. Ginestet et al. [2017] use asymptotic inference theory to provide a suitable test statistic along with its distribution for large samples. The other two works alternatively propose non-parametric testing procedures: either via a flexible generative probabilistic model for the data with subsequent Bayesian inference [Durante et al., 2017] or via model-free permutation inference [Lovato et al., 2018]. The interested reader can find a thorough review of statistical inference for network-valued data in this latter work.

The second difficulty for case-control studies pertains to making local inference, i.e. to identifying the specific edges responsible for a globally detected difference in distributions between the two samples. This is of critical importance from a clinical perspective since the diagnosis, understanding and treatments of a neurological condition are radically different when the whole brain network is affected rather than when only a subnetwork is, which is an informa-

tion that only local inferential procedures can provide. On the one hand, if local inference reveals that many edges are altered, it may suggest that the disease process involves a global disruption in the neurogenesis (i.e., the building up of connections) or pruning (i.e., the elimination of connections) processes. For instance, recent evidence have suggested that schizophrenia is caused by a disrupted pruning process [Sekar et al., 2016] which may explain the high number of local differences in brain networks observed in patients [Lynall et al., 2010]. On the other hand, if local inference reveals that only a few edges are altered, it may suggest that only a part of the brain is involved in the disease process and may therefore elicit which neuronal subsystem is responsible for the emergence of symptoms. For example, in bipolar disorder, alterations in a subnetwork connecting the amygdala and the prefrontal cortex have been found which may explain the impaired emotional regulation in these patients [Phillips and Swartz, 2014].

Local inference for NVD is an unsolved problem that poses great challenges, both methodologically and computationally. An edge-by-edge analysis leads to too many comparisons and applying traditional p-value correction schemes that aim at controlling either the family-wise error rate (FWER) or the false discovery rate (FDR) may substantially decrease the overall statistical power of the underlying test. Besides, any other form of summary statistic at the local level (e.g. degree centrality of the vertices) necessarily truncates the information contained in the original networks. Pioneering in this direction, Zalesky et al. [2010] propose Network-Based Statistics (NBS). This is a comprehensive statistical framework for identifying differences in brain networks by controlling the FWER over a data-driven partition. This partition is defined as the set of connected components of a brain network in which the strength of a connection is measured by the value of a relevant test statistic that compares the two samples at this edge and connections with strength below a user-specified threshold are considered inexistant. Later, Ginestet et al. [2017] exploit the fact that the test statistic they propose to make global inference for NVD is actually a linear combination of contributions from each single edge and they identify, through asymptotic inference, which contributions are statistically relevant to the test statistic. Differently, Durante and Dunson [2018] integrate multiple local tests in their analysis, where the differences are explored in terms of a Bayesian nonparametric approach on each edge, while controlling for multiple comparisons. These works suffer from a number of drawbacks. NBS is computationally very demanding, only provides a weak control of the FWER and is not able to detect differences that would lie outside the connected components as acknowledged by the authors themselves (e.g. on isolated edges). The control is actually only guaranteed on a data-driven partition obtained from a surrogate network that may therefore not have any biological meaning with respect to the original networks that one wants to compare. In particular, no control of the FWER is provided for any connection between vertices that belong to two distinct connected components of the surrogate network, although they might very well

be connected in the original networks. Overall, NBS therefore provides weak control of the FWER conditionally to the data at hand, which makes power analysis difficult since the partition would change at every simulation run and raises concerns about the adequacy of the method in terms of reproducibility of results (e.g. test-retest studies). Asymptotic inference as used by Ginestet et al. [2017] is out of the question for case-control studies of neurological disorders in which the sample size of the patient population is always rather small and the Bayesian approach proposed by Durante and Dunson [2018] is limited to unweighted binary networks only.

In response to these limitations, we propose a fully non-parametric approach to locally compare two samples of networks (both in the weighted and in the unweighted case). In details, we start with defining a partition of the vertex set into regions of interest (usually provided by experts in each specific applied field). Each element of the partition therefore consists of a subset of vertices from which an *intra*-subnetwork can naturally be extracted by keeping only edges connecting those vertices. Similarly, every pair of elements in the partition naturally gives rise to an *inter*-subnetwork by keeping only edges that connect vertices between the two elements. The major contribution of the present paper is a hypothesis testing framework that enables local inference at the granularity of these subnetworks, with finite-sample strong control of the family-wise error rate (FWER) and in a fully non-parametric setting that can accommodate complex non-trivial data generating processes. The paper is organized as follows. In Section 2, we mathematically formalize the problem that we propose to address in this work. In particular, we give proper definitions of intra– and inter–subnetworks and we formally define the concept of local inference for network-valued data. The mathematical apparatus required to achieve finite-sample strong control of the FWER is next described in Section 3, together with two procedures that effectively achieve this control. The first generates properly adjusted p-values at the cost of lengthier computations while the second one cuts down computation time by only adjusting the p-values that are below a predefined significance level. Section 4 is dedicated to simulations, in which specific simulated scenarios are described in order to highlight the benefits of our approach with respect to more conventional practices. Finally, we study the EEG data set in Section 5 in the light of our flexible local inferential procedure for network-valued data. We demonstrate the usefulness of our methodology in providing an unprecedented characterization of the neurobiological mechanisms driving non-syndromic autism with respect to autism combined with tuberous sclerosis complex.

2 Local inference for network-valued data

Let us consider two samples of independent and identically distributed network-valued random variables $\mathcal{G}_1 = \{G_{11}, \ldots, G_{1n_1}\}$ and $\mathcal{G}_2 = \{G_{21}, \ldots, G_{2n_1}\}$ of sizes

n_1 and n_2 respectively. Each statistical unit is therefore a network G that we shall define by means of its vertex set V and edge set E. We therefore use the notation $G := (V, E)$. We assume that all the networks in both samples have the same set V of N vertices. Our aim is to identify which *parts* of the network structure (i.e., which vertices and/or edges) are responsible for a statistically significant difference between the data generating processes that led to the two observed samples. The key ingredient for a proper definition of a part of a network is a **partition** of its vertices. In details, let $\mathcal{V} = \{V_i\}_{i \in I}$ be a partition of the common vertex set V of networks in both samples, that is a collection of subsets of V such that:

1. $V = \bigcup_{i \in I} V_i$,
2. $V_i \cap V_j = \emptyset \quad \forall i \neq j$.

The elements of the partition will be the ones over which local inference is performed. Hence, the partition characterizes the regions of the network which are of interest for the study. Such a partition will in general be provided by experts in every applied discipline where a network-valued data analysis approach could be relevant and provide insights into the scientific hypotheses under query. The granularity (i.e. richness or coarseness) of the partition can be arbitrarily chosen. Assuming the availability of such a partition, two classes of subnetworks can be defined as follows:

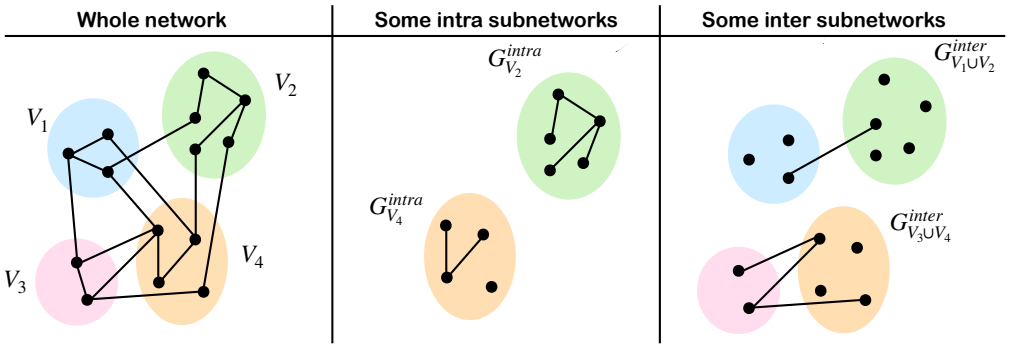
Definition 2.1. *Each element V_i of the partition is a subset of the vertex set V and thus, in turn, a vertex set itself that can be used to define an* ***intra–subnetwork*** $G^{\text{intra}}_{V_i} = \left(V_i, E^{\text{intra}}_{V_i}\right)$*, by keeping in the edge set $E^{\text{intra}}_{V_i}$ only those edges with both endpoints in the same element V_i of the partition. We have in total m intra–subnetworks, where $m := |I|$ is the cardinality of the partition of the initial common vertex set.*

The union $V_i \cup V_j$ of each pair of elements of the partition is a subset of the vertex set V and thus, in turn, a vertex set itself that can be used to define an ***inter–subnetwork*** $G^{\text{inter}}_{V_i \cup V_j} = \left(V_i \cup V_j, E^{\text{inter}}_{V_i \cup V_j}\right)$*, by keeping in the edge set $E^{\text{inter}}_{V_i \cup V_j}$ only those edges with one endpoint in the element V_i and the other endpoint in the element V_j. We have in total $\binom{m}{2}$ inter–subnetworks.*

Figure 1 provides an example of a partition of the vertex set of a network into $m = 4$ elements. Two out of the four possible *intra*–subnetworks and two out of the six possible *inter*–subnetworks are represented for an easier understanding of the above definitions.

We propose to search for local differences in the network structure between the two samples by comparing the intra– and inter–subnetworks induced by the partition of the common vertex set. Consequently, the partition is critical in our analysis as it determines how deep we are willing to go into the network structure in order to localize those differences. An analysis that follows this approach

Figure 1: Example of partition of the vertex set of a network into 4 elements (left), with corresponding $G^{\text{intra}}_{V_i}$ for $i = 2, 4$ (middle) and $G^{\text{inter}}_{V_i \cup V_j}$ for $(i, j) = (1, 2)$ and $(3, 4)$ (right).



provides a very rich output. In effect, not only it is able to identify which subnetworks are responsible for a global statistically significant difference between the two samples, it also provides insights into the nature of the difference, which can be *within* one (or more) elements of the partition or *between* elements of the partition. This is a valuable information for interpretation, especially in medical applications. Accordingly with this aim, we shall formally state the two families of hypothesis tests for intra– and inter–subnetworks as follows:

Definition 2.2. *Let V_i be an element of the partition $\mathcal{V}$ of the common vertex set V. The intra–subnetworks induced by V_i in each sample are governed by two distributions that we denote $\mathbf{F}_1^{G_{V_i}}$ and $\mathbf{F}_2^{G_{V_i}}$ respectively. The associated* ***intra hypothesis test*** *is defined as:*

$$H_0^{\text{intra},i} : \mathbf{F}_1^{G_{V_i}} = \mathbf{F}_2^{G_{V_i}} \quad \textit{against} \quad H_1^{\text{intra},i} : \mathbf{F}_1^{G_{V_i}} \neq \mathbf{F}_2^{G_{V_i}}. \tag{2.1}$$

There is thus a total of m intra hypothesis tests to perform.

Let $V_i \cup V_j$ be the union of a pair of elements of the partition $\mathcal{V}$ of the common vertex set V. The inter–subnetworks induced by $V_i \cup V_j$ in each sample are governed by two distributions that we denote $\mathbf{F}_1^{G_{V_i \cup V_j}}$ and $\mathbf{F}_2^{G_{V_i \cup V_j}}$ respectively. The associated ***inter hypothesis test*** *is defined as:*

$$H_0^{\text{inter},ij} : \mathbf{F}_1^{G_{V_i \cup V_j}} = \mathbf{F}_2^{G_{V_i \cup V_j}} \quad \textit{against} \quad H_1^{\text{inter},ij} : \mathbf{F}_1^{G_{V_i \cup V_j}} \neq \mathbf{F}_2^{G_{V_i \cup V_j}}. \tag{2.2}$$

There is thus a total of $\binom{m}{2} = \frac{m(m-1)}{2}$ inter hypothesis tests to perform.

We aim at defining a statistical methodology that provides a finite-sample strong control of the FWER on the joint family of intra– and inter– hypothesis tests. There are two main challenges to achieve this goal. The first issue is that each single hypothesis involves distributions of networks and, thus, testing these hypotheses requires unconventional statistical methods that are able to deal

with such complex data. In particular, in this work, we will use the flexible non-parametric permutation framework for two-sample testing with populations of networks proposed by Lovato et al. [2018]. The second issue lies in the complexity of any procedure for controlling the FWER. This is because (i) the number of intra– and inter–hypothesis tests to be performed simultaneously can be very large and (ii) the FWER control requires the introduction of auxiliary hypothesis tests, which represents both a methodological (how to properly define them) and a practical (how to handle the computational burden generated by so many tests) problem. Section 3 proposes an answer to this second challenge.

3 Control of the family-wise error rate

3.1 Auxiliary hypothesis tests

Providing a strong control of the FWER over a family of hypothesis tests pertains to ensuring that the probability of making at least one type I error – over all tests that were performed and irrespectively of whether the underlying null hypotheses are true – remains upper-bounded by some predefined significance level α. Many procedures for controlling the FWER have been devised in the literature [see Keppel and Wickens, 2004, for a review on this topic.]. Among the many existing approaches, the closed testing procedure proposed by Marcus et al. [1976] is particularly appealing due to its great flexibility and easy extension to any kind of complex data. Starting from the family of individual null hypotheses to be tested with strong control of the FWER at level α, the key idea is to build up all possible combinations of composite null hypotheses by intersections of the individual ones and subsequently reject an individual null hypothesis if and only if all composite null hypotheses in which it appears can be rejected at significance level α. This procedure guarantees that the probability of making no type I error over the original individual hypothesis tests is at least $1 - \alpha$. The closed testing procedure can therefore be viewed as a hierarchy of auxiliary hypothesis tests at the bottom of which the hypothesis tests of interest lie.

In our framework for NVD, recall that we defined a partition of the common vertex set and that the objective is to perform simultaneously all intra– and inter–hypothesis tests induced by this partition as defined in eqs. (2.1) and (2.2). At the bottom of the hierarchy, there are therefore the m intra– and the $\binom{m}{2}$ inter–hypothesis tests. Moving up in the hierarchy, we must properly define composite null hypotheses, i.e. null hypotheses that are intersections of intra– and inter–hypothesis tests. For this purpose, we resort to the concept of σ-algebra over some non-empty set S, which is a collection of subsets of S that (i) includes the empty set, (ii) is closed under complement and (iii) is closed under countable unions and countable intersections. Furthermore, if $\mathcal{P}(V)$ denotes the power set of some set V and $K \subseteq \mathcal{P}(V)$, we call *$\sigma$-algebra generated by K* the smallest σ-algebra that contains K, i.e. the intersection of all the σ-algebras that contain K. In particular, the σ-algebra generated by the partition $\mathcal{V} = \{V_i\}_{i \in I}$

of the common vertex set is $\sigma(\mathcal{V}) = \{\bigcup_{j\in J} V_j, J \subseteq I\}$, since $V_i \cap V_j = \emptyset$ whenever $i \neq j$ by definition of a partition. When a σ-algebra is generated by a partition, it is possible to define a concept of *dimensionality* for any of its elements. In effect, we say that $A \in \sigma(\mathcal{V})$ – which is the union of some elements of the partition $\mathcal{V}$ – has dimension d when it is the union of exactly d elements V_i of the partition $\mathcal{V}$. We will, from now on, use the notation $\dim(A) = d$.

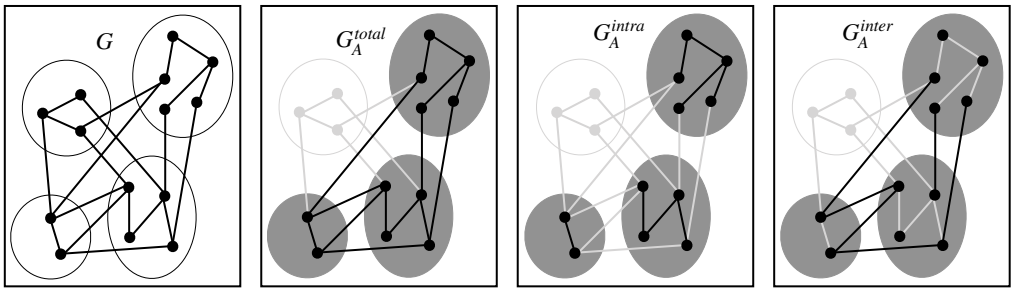
Definition 3.1. *Each element A of the σ-algebra generated by the partition $\mathcal{V}$ of the vertex set is a subset of the vertex set V and thus, in turn, a vertex set itself that can be used to define three classes of subnetworks, which therefore share the same vertex set.*

*The **full subnetwork** induced by A is defined as $G_A^{\text{full}} = \left(A, E_A^{\text{full}}\right)$, in which the edge set contains edges that have their endpoints in any partition elements contained in A. A network in this class is the subgraph of G induced by A [see Diestel, 2018, chapt. 1]. An example of such a network is shown in Figure 2, 2nd column.*

*The **intra subnetwork** induced by A is defined as $G_A^{\text{intra}} = \left(A, E_A^{\text{intra}}\right)$, in which the edge set contains only edges that have both endpoints within a same partition element contained in A. A network in this class features edges that do not exit each single partition element $V_i \subseteq A$. An example of such a network is shown in Figure 2, 3rd column.*

*The **inter subnetwork** induced by A is defined as $G_A^{\text{inter}} = \left(A, E_A^{\text{inter}}\right)$, in which the edge set contains only edges that have endpoints in two distinct partition elements contained in A. A network in this class features edges that connect vertices in two different partition elements $V_i \neq V_j$, $V_i, V_j \subseteq A$. An example of such a network is shown in Figure 2, 4th column.*

Figure 2: Example of partition of the vertex set of a network into 4 elements (1st column), along with the full–subnetwork G_A^{full} (2nd column), the intra–subnetwork G_A^{intra} (3rd column) and the inter–subnetwork G_A^{inter} (4th column) induced by an element A of dimension 3 (dark gray areas) of the σ-algebra generated by the partition. In each panel, vertices and edges that define the corresponding subnetworks are displayed in black.



Remark 3.1. *If A is an element of the partition, i.e. $A \equiv V_i$, then G_A^{inter} is not defined and, by convention, we write $G_A^{\text{inter}} = \emptyset$ and $G_A^{\text{intra}} \equiv G_A^{\text{full}}$.*

Remark 3.2. *In some applications, vertices might be divided into groups in such a way that a single vertex belongs to more than one group. Our framework also covers these cases. In effect, for any subdivision of the vertices into groups, it is always possible to define a partition by isolating overlapping vertices into elements of the partition itself.*

Remark 3.3. *There are two extreme choices for the partition of the vertex set. On the one hand, one could use a partition made of a unique element, i.e. the common vertex set itself. In this case, the σ-algebra generated by the partition coincides with the partition itself and the local testing procedure boils down to performing a global test as in Lovato et al. [2018]. On the other hand, one could think of assigning each vertex to its own element of the partition. In this case, the intra–subnetworks are not defined and the procedure boils down to testing inter–differences on every single edge, which is the most commonly adopted strategy.*

Similarly to what we did for the two classes of subnetworks induced by the elements of the partition $\mathcal{V}$, we can define auxiliary hypothesis tests for the three classes of subnetworks induced by a generic element of the σ-algebra $\sigma(\mathcal{V})$ generated by the partition $\mathcal{V}$. In details, we introduce the following three families of auxiliary hypothesis tests:

Definition 3.2. *Let A be an element of the σ-algebra $\sigma(\mathcal{V})$ generated by the partition $\mathcal{V}$ of the common vertex set V.*

*The full–subnetworks induced by A in each sample are governed by two distributions that we denote $\mathbf{F}_1^{G_A^{\text{full}}}$ and $\mathbf{F}_2^{G_A^{\text{full}}}$ respectively. The associated **full hypothesis test** pertains to testing the following hypotheses:*

$$H_0 : \mathbf{F}_1^{G_A^{\text{full}}} = \mathbf{F}_2^{G_A^{\text{full}}} \quad \textit{against} \quad H_1 : \mathbf{F}_1^{G_A^{\text{full}}} \neq \mathbf{F}_2^{G_A^{\text{full}}}. \tag{3.1}$$

Let us denote by p_A^{full} the p-value of the resulting test.

*The intra–subnetworks induced by A in each sample are governed by two distributions that we denote $\mathbf{F}_1^{G_A^{\text{intra}}}$ and $\mathbf{F}_2^{G_A^{\text{intra}}}$ respectively. The associated **intra hypothesis test** pertains to testing the following hypotheses:*

$$H_0 : \mathbf{F}_1^{G_A^{\text{intra}}} = \mathbf{F}_2^{G_A^{\text{intra}}} \quad \textit{against} \quad H_1 : \mathbf{F}_1^{G_A^{\text{intra}}} \neq \mathbf{F}_2^{G_A^{\text{intra}}}. \tag{3.2}$$

Let us denote by p_A^{intra} the p-value of the resulting test. This test is a generalization of Equation (2.1) to a generic element of the σ–algebra $\sigma(\mathcal{V})$ instead of only an element of the partition $\mathcal{V}$ that generated the σ-algebra.

*The inter–subnetworks induced by A in each sample are governed by two distributions $\mathbf{F}_1^{G_A^{\text{inter}}}$ and $\mathbf{F}_2^{G_A^{\text{inter}}}$ respectively. The associated **inter hypothesis test** pertains to testing the following hypotheses:*

$$H_0 : \mathbf{F}_1^{G_A^{\text{inter}}} = \mathbf{F}_2^{G_A^{\text{inter}}} \quad \textit{against} \quad H_1 : \mathbf{F}_1^{G_A^{\text{inter}}} \neq \mathbf{F}_2^{G_A^{\text{inter}}}. \tag{3.3}$$

Let us denote by p_A^{inter} the p-value of the resulting test. This test is a generalization of Equation (2.2) to a generic element of the σ–algebra $\sigma(\mathcal{V})$ instead of only an element of the partition $\mathcal{V}$ that generated the σ-algebra.

In the next two sections, we will describe two procedures that show how to take advantage of these auxiliary hypothesis tests for providing a finite-sample strong control of the FWER when simultaneously testing intra– and inter–hypothesis tests defined by a given partition. For an easier description of these procedures, we introduce the following families of subnetworks:

$$\mathcal{G}^{\text{full}} := \left\{ G_A^{\text{full}} : A \in \sigma(\mathcal{V}) \right\} \tag{3.4}$$

$$\mathcal{G}^{\text{intra}} := \left\{ G_A^{\text{intra}} : A \in \sigma(\mathcal{V}) \right\} \tag{3.5}$$

$$\mathcal{G}^{\text{inter}} := \left\{ G_A^{\text{inter}} : A \in \sigma(\mathcal{V}), \dim(A) > 1 \right\}. \tag{3.6}$$

3.2 Complete multiscale testing procedure

For the sake of clarity, recall that we aim at finding relevant differences between two samples $\mathcal{G}_1$ and $\mathcal{G}_2$ of networks of size n_1 and n_2 respectively. Networks share the same vertex set V of cardinality N. Let $\mathcal{V} = \{V_i\}_{i \in I}$ be a partition of the vertex set of cardinality m and $\sigma(\mathcal{V})$ be the σ–algebra generated by $\mathcal{V}$.

In this section, we describe a first procedure for providing finite-sample strong control of the FWER over the family of tests induced by intra– and inter–subnetworks generated by the partition. We refer to this procedure as the *complete multiscale testing procedure* (CMTP). This procedure is built on top of the close testing procedure, which can be applied, in the context of network-valued data, on the hierarchy of auxiliary hypothesis tests defined in eqs. (3.1) to (3.3). In details, we perform all three auxiliary hypothesis tests defined in eqs. (3.1) to (3.3) on all subnetworks induced by every single element of the σ–algebra $\sigma(\mathcal{V})$. Each of these tests is a two-sample test for network-valued data. For this purpose, we use the approach proposed in Lovato et al. [2018]. In summary, it is a model-free two-sample test for network-valued data based on a flexible permutation framework that achieves consistency and finite-sample exactness without making any distributional assumption on the compared probability distributions. The user chooses a matrix representation for the networks as well as an appropriate distance. Subsequently, two test statistics based on inter-point distances are computed and combined through the non-parametric combination methodology [Pesarin and Salmaso, 2010, chap. 4] in order to provide a permutation p-value computed according to Phipson and Smyth [2010]. Further details can be found in Lovato et al. [2018]. Finally, we define adjusted p-values for the intra– and inter–hypothesis tests given by eqs. (2.1) and (2.2) as follows:

$$p_{V_i} := \max_{A \in \sigma(\mathcal{V}) : V_i \in A} p_A^{\text{intra}}, p_A^{\text{full}} \qquad p_{V_i \cup V_j} := \max_{A \in \sigma(\mathcal{V}) : V_i, V_j \in A} p_A^{\text{inter}}, p_A^{\text{full}}. \tag{3.7}$$

The CMTP is summarized by Algorithm 1 in Appendix B. Finite-sample strong control of the FWER is guaranteed in the following way:

Lemma 3.1. *The complete multiscale testing procedure (CMTP) guarantees strong control of the FWER over the entire family of tests performed on the union set of networks* $\mathcal{G}^{\text{full}} \cup \mathcal{G}^{\text{intra}} \cup \mathcal{G}^{\text{inter}}$*, where* $\mathcal{G}^{\text{full}}$*,* $\mathcal{G}^{\text{intra}}$ *and* $\mathcal{G}^{\text{inter}}$ *are defined in eqs. (3.4) to (3.6). This means that, if* $G \in \mathcal{G}^{\text{full}} \cup \mathcal{G}^{\text{intra}} \cup \mathcal{G}^{\text{inter}}$ *is the largest subnetwork where* H_0 *is true, then we have that:*

$$\mathbb{P}\left[\left(\exists G_{V_i}^{\text{intra}} \subseteq G : p_{V_i} \leq \alpha\right) \vee \left(\exists G_{V_i \cup V_j}^{\text{inter}} \subseteq G : p_{V_i \cup V_j} \leq \alpha\right)\right] \leq \alpha.$$

Proof. Let G be the largest subnetwork where H_0 is true and A be the vertex set of G. Without loss of generality, assume that G is a subnetwork of type *intra*. Hence we know that $\mathbb{P}[p_A^{\text{intra}} \leq \alpha] = \alpha$ because a single test performed according to Lovato et al. [2018] features finite-sample exactness. Now, since A is an element of the σ-algebra generated by the partition $\{V_i\}$ and G is an intra–subnetwork, there exists at least one index i_0 such that $V_{i_0} \subseteq A$. Given the definition of adjusted *intra* p-value in eq. (3.7), we have that $p_{V_{i_0}} \geq p_A^{\text{intra}}$. As a result, we have that $\{p_{V_{i_0}} \leq \alpha\} \subseteq \{p_A^{\text{intra}} \leq \alpha\}$ and thus $\mathbb{P}\{p_{V_{i_0}} \leq \alpha\} \leq \mathbb{P}\{p_A^{\text{intra}} \leq \alpha\} = \alpha$. The proof follows exactly the same lines if G is a subnetwork of type *inter* or *full*. □

Remark 3.4. *Lemma 3.1 guarantees the control of the FWER on the largest subnetwork* G *in* $\mathcal{G}^{\text{full}} \cup \mathcal{G}^{\text{intra}} \cup \mathcal{G}^{\text{inter}}$ *where* H_0 *is true. It immediately follows from the proof that this type of control is also provided for any other subnetwork in* $\mathcal{G}^{\text{full}} \cup \mathcal{G}^{\text{intra}} \cup \mathcal{G}^{\text{inter}}$ *where* H_0 *is true.*

Remark 3.5. *The finite-sample control of the FWER is guaranteed over the family of tests performed on the union set of networks* $\mathcal{G}^{\text{full}} \cup \mathcal{G}^{\text{intra}} \cup \mathcal{G}^{\text{inter}}$ *but does not extend to tests involving subnetworks that are not induced by the partition. In a sense, this makes the control of the FWER dependent upon the choice of the initial partition of the common vertex set.*

3.3 Truncated multiscale testing procedure

The complete multiscale testing procedure guarantees finite-sample control of the FWER over the family of tests performed on the union set of networks $\mathcal{G}^{\text{full}} \cup \mathcal{G}^{\text{intra}} \cup \mathcal{G}^{\text{inter}}$ by performing a total of $2^m - 1$ tests, where m is the number of elements in the chosen partition. Hence, for moderate to large partitions, the CMTP becomes impractical. This is an issue inherent to the closed testing procedure that provides properly adjusted p-values only when all auxiliary p-values are computed. A number of solutions have been devised in the literature to alleviate this problem. For instance, iterative adjustments of the p-values such as Holm's correction [Holm, 1979] or Hommel's correction [Hommel, 1988] are very popular for their simplicity. More complex procedures including approaches

that rely upon logical restrictions on the hypotheses [e.g. Shaffer, 1986, Royen, 1989] have been proposed as well. The interested reader can refer to Westfall and Tobias [2007] for an extensive review on these topics.

In this paper, we propose a simple strategy by truncation for reducing the computational burden when the ultimate goal after p-value adjustment is to take a binary decision at a pre-specified significance level α. In the closed testing philosophy, given a set of null hypotheses to be tested simultaneously, the p-value associated with a specific null hypothesis is adjusted by taking the maximum among p-values associated with all composite hypotheses involving that hypothesis. The order in which the auxiliary p-values are computed is in general irrelevant because all of them have to be computed for properly adjusting the p-values of the original hypotheses. However, if these p-values are to be confronted to a pre-specified significance level α for rejecting the associated null hypothesis whenever the p-value exceeds α, there is a clear benefit in testing first composite hypotheses that include the largest number of original hypotheses. In effect, if, during the adjustment process, a p-value exceeds α for some composite hypothesis (defined as union of some original hypotheses), then there is no need for computing auxiliary p-values associated with any composite hypothesis included in that one, because the resulting adjusted p-value would only be higher. Figure 8 in Appendix B illustrates the differences between the complete and truncated closed testing procedures on an example in which four hypotheses are tested simultaneously at a significance level $\alpha = 5\%$. In the truncated approach, starting from the top, the first p-value that exceeds α occurs when H_0^1, H_0^2 and H_0^4 are tested simultaneously in $H_0^{1,2,4}$. At this point, all subsequent auxiliary p-values associated with sub-hypotheses of $H_0^{1,2,4}$, that is $H_0^{1,2}$, $H_0^{1,4}$, $H_0^{2,4}$, H_0^1, H_0^2 and H_0^4, become unnecessary. This clearly reduces the computational burden at the cost of possibly underestimating the adjusted p-values for H_0^1, H_0^2 and H_0^4 which are greater than the significance level.

This truncation principle naturally extends to our framework for local inference for populations of networks because the definitions of adjusted p-values provided in eq. (3.7) have been borrowed to the CTP. Truncation can be separately applied for adjusting p-values associated with null hypotheses involving intra-subnetworks or inter-subnetworks, resulting in two separate sets of adjusted p-values: one of size m for investigating differences in the distributions of each intra-subnetwork induced by the partition elements and another one of size $m(m-1)/2$ for investigating differences in the distributions of each inter-subnetwork induced by any pair of partition elements. We coin this procedure the *α-truncated multiscale testing procedure* (α-TMTP). It is summarized by Algorithm 2 in Appendix B. In the following, we will denote by $\widetilde{p}$ any adjusted p-value provided by the α-TMTP in contrast with the notation p for adjusted p-values coming from the CMTP.

Remark 3.6. *The CMTP and the α-TMTP both identify the same set of significantly different sub-networks when using α as significance level. More specif-*

ically, if a sub-network is statistically different between the two samples, the p-values found with the CMTP and with the α-TMTP are exactly the same, i.e., $\widetilde{p}_{V_i} = p_{V_i}$ and $\widetilde{p}_{V_i \cup V_j} = p_{V_i \cup V_j}$. Conversely, if a sub-network is not statistically different between the two samples, the p-value found with the CMTP is always larger or equal than the one found with the α-TMTP, i.e., $\widetilde{p}_{V_i} \leq p_{V_i}$ and $\widetilde{p}_{V_i \cup V_j} \leq p_{V_i \cup V_j}$.

More generally, once the α-TMTP has been run, the corresponding adjusted p-values that are smaller or equal to α coincide with those obtained with the CMTP. Hence, it is possible to rely on adjusted p-values from the α-TMTP also in cases where we are interested in significant sub-networks at a level $\alpha_<$ smaller than α. The inverse is however not true. In effect, adjusted p-values obtained with the α-TMTP that are greater than α are underestimated w.r.t. those obtained with the CMTP. It is thus not recommended to use them for discussing significance at levels $\alpha_>$ larger than α.

Lemma 3.2. *The truncated multiscale testing procedure run at a level α that matches the FWER guarantees strong control of the FWER over the entire family of tests performed on the union set of networks $\mathcal{G}^{\text{full}} \cup \mathcal{G}^{\text{intra}} \cup \mathcal{G}^{\text{inter}}$, where $\mathcal{G}^{\text{full}}$, $\mathcal{G}^{\text{intra}}$ and $\mathcal{G}^{\text{inter}}$ are defined in eqs. (3.4) to (3.6). This means that, if $G \in \mathcal{G}^{\text{full}} \cup \mathcal{G}^{\text{intra}} \cup \mathcal{G}^{\text{inter}}$ is the largest subnetwork where H_0 is true, then we have that:*

$$\mathbb{P}\left[\left(\exists G_{V_i}^{\text{intra}} \subseteq G : \widetilde{p}_{V_i} \leq \alpha\right) \vee \left(\exists G_{V_i \cup V_j}^{\text{inter}} \subseteq G : \widetilde{p}_{V_i \cup V_j} \leq \alpha\right)\right] \leq \alpha.$$

Proof. From Remark 3.6, it follows that, if α is the level at which the α-TMTP has been run, then $\widetilde{p}_{V_i} \leq \alpha$ if and only if $p_{V_i} \leq \alpha$ and $\widetilde{p}_{V_i \cup V_j} \leq \alpha$ if and only if $p_{V_i \cup V_j} \leq \alpha$. Hence, if α is equal to the family-wise error rate, the proof is a trivial consequence of Lemma 3.1. □

3.4 Computation times

In this section, we compare the computational costs of the CMTP and α-TMTP in terms of number of tests required to achieve the p-value adjustment. These costs depends on two parameters: (i) the number m of elements in the chosen vertex partition and (ii) the number k of (intra or inter) sub-networks induced by partition elements that exhibit significant differences between the two samples. We derive analytic expression of the costs separately for the intra– and inter–sub-networks.

3.4.1 Cost of intra–hypothesis tests

In the α-TMTP, we progressively test on sub-networks induced by elements of $\sigma(\mathcal{V})$ of decreasing dimension from m to 1. All tests are performed for elements of dimension m to $m - k$ (number of non-significant partition elements). On the contrary, for elements of dimension i in between $m - k - 1$ to 1, some tests

become unnecessary. Specifically, all tests on sub-networks induced by elements of dimension i composed exclusively of non-significant partition elements can be discarded which spares a total of $\binom{m-k}{i}$ tests. Provided that $m > k$, the total of spared tests thus amounts to

$$\sum_{i=1}^{m-k-1} \binom{m-k}{i} = 2^{m-k} - 2.$$

The CMTP performs instead $2^m - 1$ tests. The relative computational saving of the α-TMTP w.r.t. the CMTP is thus:

$$R_{\text{intra}} = \begin{cases} \frac{2^{m-k}-2}{2^m-1} & \text{if } m > k \\ 0 & \text{otherwise.} \end{cases}$$

3.4.2 Cost of inter–hypothesis tests

The argumentation is the same for inter–subnetworks except that elements of $\sigma(\mathcal{V})$ of dimension 1 are not considered at all. Hence, provided that $m > k + 1$, the total number of spared tests amounts to

$$\sum_{i=2}^{m-k-1} \binom{m-k}{i} = 2^{m-k} - m + k - 2,$$

whereas the CMTP performs $2^m - m - 1$ tests. The relative computational saving of the α-TMTP w.r.t. the CMTP is thus:

$$R_{\text{inter}} = \begin{cases} \frac{2^{m-k}-m+k-2}{2^m-m-1} & \text{if } m > k+1 \\ 0 & \text{otherwise.} \end{cases}$$

Figure 9 in Appendix B illustrates the computation savings for a small value of $m = 5$ and a large value of $m = 10$. Observe that both relative computational savings, in the large m regime, are independent from m and tend to 2^{-k} (the red curve). Actually, already for $m \geq 10$, there is practically no difference between the actual saving and the 2^{-k} approximation. In general, we can observe that we save the most in cases where there are few partition elements that induce subnetworks exhibiting differences between the two samples. Specifically, looking at the $m = 10$ case, we read, for large m, that the saving is of 100% if $k = 0$, 50% if $k = 1$ and 25% for $k = 3$.

4 Simulation studies

The aim of this section is to explore the potential of our methodology on simulated data sets where different levels and kinds of differences are present. In detail, the goal of this section is twofold. First of all we focus on the identification

of local differences, showing that our procedure is able to detect in which regions of interest (RoI) inside the network there are difference between the two populations, while controlling the family wise error rate; we also show that a naive approach that does not control for multiple testing fail in controlling the family wise error rate. Second, we generate a particular data set in order to highlight how important it is to consider the entire network approach also in the case of local inference instead of other summary objects that somehow summarize the entire structure of the network and we show how this latter approach loses power. Inspired by atlases commonly used in the clinical practise, we generate samples of networks with 68 vertices. We use a partition with four elements and the sample sizes are $n_1 = n_2 = 10$. We used a total of 1000 replicates and an α level equal to 0.05.

4.1 Identification of local differences

4.1.1 Simulated scenarios

In this first simulation study, we simulate four different scenarios, all characterized by the presence of specific subnetworks with different edge strength distributions; what distinguishes the scenarios is which subnetworks are different between the two populations. We rely on the stochastic block model [Holland et al., 1983], often abbreviated SBM, for generating the samples. This is a popular and useful model that allows to choose the probability of existence of an edge within and between pre-specified RoIs inside the network. The parameters of an SBM are the partition of the vertex set into disjoint subsets $C_1, \ldots, C_m$ and an edge probability matrix P with dimension $m \times m$ whose element p_{ij} stores the probability of existence of an edge between vertices belonging to RoI C_i and vertices belonging to RoI C_j. Therefore, the SBM allows to specify the probabilities both of edges connecting vertices within blocks ($i = j$) and of edges connecting vertices between blocks ($i \neq j$). Selecting $C_1, \ldots, C_m = V_1, \ldots, V_m$, we are able, through this model, to generate samples that have *intra*–differences or *inter*–differences in different RoIs. Table 1 summarizes the 4 scenarios as 4×4 matrices. The diagonal informs about *intra*–differences while off-diagonal elements inform about *inter*–differences. A cross ($\times$) stands for significant difference (null hypothesis is false) while a checkmark ($\checkmark$) stands for no difference (null hypothesis is true).

In details, we start with the case where differences are present in all 4 *intra*–subnetworks and all 6 *inter*–subnetworks defined by the partition (see Table 1a). The 2nd and 3rd scenarios explore separately *intra*– and *inter*–differences. The 2nd scenario (see Table 1b) includes all the *intra*–differences only while the 3rd scenario (see Table 1c) includes all the *inter*–differences only. Finally, we explore a more realistic scenario in Table 1d, where the differences between the two populations are located in some *intra*– and some *inter*–subnetworks. The four elements of the vertex partition contain 17 vertices each. Edge probability

matrices for each specific scenario are reported in appendix A.1.

Table 1: Summary tables of *intra*– and *inter*–differences in the first simulation study.

	RoI 1	RoI 2	RoI 3	RoI 4
RoI 1	×	×	×	×
RoI 2		×	×	×
RoI 3			×	×
RoI 4				×

(a) First scenario

	RoI 1	RoI 2	RoI 3	RoI 4
RoI 1	×	✓	✓	✓
RoI 2		×	✓	✓
RoI 3			×	✓
RoI 4				×

(b) Second scenario

	RoI 1	RoI 2	RoI 3	RoI 4
RoI 1	✓	×	×	×
RoI 2		✓	×	×
RoI 3			✓	×
RoI 4				✓

(c) Third scenario

	RoI 1	RoI 2	RoI 3	RoI 4
RoI 1	✓	✓	×	✓
RoI 2		✓	×	✓
RoI 3			×	×
RoI 4				✓

(d) Fourth scenario

4.1.2 Estimation of probability of rejection

Table 2 reports the estimated probability of rejection of the test on the four simulated scenarios described in the previous section. In correspondence to table 1, for each simulation, we report a 4×4 table with the estimated probability of rejection for each tested hypothesis. The entry (i, j) refers to the comparison of the subnetwork identified by the RoIs i and j. When $i = j$, the resulting subnetwork is seen as an *intra*–subnetwork, while when $i \neq j$, the resulting subnetwork is seen as an *inter*–subnetwork. The results show that in all the generated scenarios the α-TMTP is sensitive to the violation of the null hypothesis, independently from the type of difference (*intra* or *inter*).

We also compared the two samples in these same four scenarios by means of a naive approach that simply tests all the null hypotheses separately without applying any correction for multiple comparisons. In this case, a total of 10 null hypotheses is tested (4 *intra*–hypotheses and 6 *inter*–hypotheses). Table 3 reports the estimated power for each tested hypothesis. We observe that *intra*– and *inter*–differences are all correctly detected and that, whenever the null hypothesis is true, the power of the individual tests is at the nominal level as expected.

What dramatically changes between the naive approach and the α-TMTP is the control of the FWER. Table 4 summarizes the estimated FWER obtained from both approaches for the scenarios 2, 3 and 4. Scenario 1 is not included because the null hypothesis is false everywhere, preventing us from getting an estimate of the FWER in that case. The results show that the α-TMTP conservatively controls the FWER at level $\alpha = 5\%$ while the naive approach does not

Table 2: Estimated probabilities of rejection obtained with the α-TMTP applied to the generated data sets of the first simulation.

	RoI 1	RoI 2	RoI 3	RoI 4
RoI 1	1.000	1.000	1.000	1.000
RoI 2		1.000	1.000	1.000
RoI 3			1.000	1.000
RoI 4				1.000

(a) First scenario.
Global power: 1.000 (intra), 1.000 (inter)

	RoI 1	RoI 2	RoI 3	RoI 4
RoI 1	1.000	0.004	0.005	0.002
RoI 2		1.000	0.000	0.001
RoI 3			1.000	0.001
RoI 4				1.000

(b) Second scenario.
Global power: 1.000 (intra), 0.050 (inter)

	RoI 1	RoI 2	RoI 3	RoI 4
RoI 1	0.002	1.000	1.000	1.000
RoI 2		0.002	1.000	1.000
RoI 3			0.006	1.000
RoI 4				0.003

(c) Third scenario.
Global power: 0.069 (intra), 1.000 (inter)

	RoI 1	RoI 2	RoI 3	RoI 4
RoI 1	0.000	0.000	1.000	0.002
RoI 2		0.000	1.000	0.002
RoI 3			1.000	1.000
RoI 4				0.001

(d) Fourth scenario.
Global power: 1.000 (intra), 1.000 (inter)

Table 3: Estimated probabilities of rejection obtained with the naive approach applied to the generated data sets of the first simulation.

	RoI 1	RoI 2	RoI 3	RoI 4
RoI 1	1.000	1.000	1.000	1.000
RoI 2		1.000	1.000	1.000
RoI 3			1.000	1.000
RoI 4				1.000

(a) First scenario.
Global power: NA (intra), NA (inter)

	RoI 1	RoI 2	RoI 3	RoI 4
RoI 1	1.000	0.042	0.048	0.065
RoI 2		1.000	0.049	0.054
RoI 3			1.000	0.049
RoI 4				1.000

(b) Second scenario.
Global power: NA (intra), NA (inter)

	RoI 1	RoI 2	RoI 3	RoI 4
RoI 1	0.045	1.000	1.000	1.000
RoI 2		0.045	1.000	1.000
RoI 3			0.042	1.000
RoI 4				0.058

(c) Third scenario.
Global power: NA (intra), NA (inter)

	RoI 1	RoI 2	RoI 3	RoI 4
RoI 1	0.058	0.039	1.000	0.051
RoI 2		0.050	1.000	0.055
RoI 3			1.000	1.000
RoI 4				0.044

(d) Fourth scenario.
Global power: NA (intra), NA (inter)

and exhibits estimated FWER around four times the pre-specified significance level α.

Table 4: Estimated FWER for scenarios 2, 3 and 4 of the first simulation study.

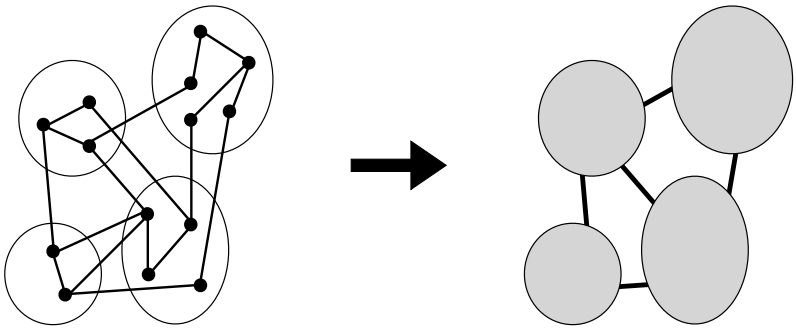
	Scenario 2	Scenario 3	Scenario 4
α-TMTP	0.012	0.013	0.005
Naive approach	0.255	0.180	0.260

4.2 Usefulness of exploiting the whole network structure

4.2.1 Simulated scenarios

Most of the time, when one defines a partition of the vertex set of a network, the interest is put on *inter*–differences, i.e. differences between the two samples on the *inter*–subnetwork induced by a pair of partition elements. A classic approach pertains to considering the *aggregated network*, in which vertices in a same partition element are collapsed into a single one and edge weights between partition elements (now single vertices) are aggregated most often by mean or sum of the weights in the original network. Figure 3 illustrates the principle of aggregation.

Figure 3: Schematic illustration of the process of network aggregation.



This second simulation aims at comparing the resulting inference of the α-TMTP applied to the original complex network and to the aggregated network in the search for differences in the subnetworks induced between pairs of partition elements. For this purpose, we generated two samples of networks with the same vertex set of size 68. We chose a partition of four elements (with 10, 20, 17 and 21 vertices in each element). We generated fully connected random networks by sampling the edge weights from a Poisson distribution with parameter 8. We introduced an *inter*–difference between the first two elements of the partition. Specifically, we introduced the difference in 24 edge weights out of the 100 existing ones. Table 5 summarizes where the difference lies. In the first

sample, we modified 12 edge weights over 24 by using a Poisson distribution with parameter 5 and the other 12 edge weights by using a Poisson distribution with parameter 11. In the second sample, we modified the exact same edge weights but we switched the two Poisson distributions. We report the details on edge weight modifications in appendix A.2. We finally ran the α-TMTP with $\alpha = 5\%$ on the two samples obtained from both the original and aggregated forms.

Table 5: Summary tables of *intra*– and *inter*–differences in the second simulation study.

	RoI 1	RoI 2	RoI 3	RoI 4
RoI 1	✓	×	✓	✓
RoI 2		✓	✓	✓
RoI 3			✓	✓
RoI 4				✓

4.2.2 Estimated probabilities of rejection

Table 6 reports the Monte-Carlo estimates of the probabilities of the rejection obtained from the original networks (Table 6a) and from the aggregated ones (Table 6b). There are two important aspects to discuss. First, the aggregation of the edge weights between partition elements makes the testing procedure insensitive to the kind of *inter*–difference we introduced, while the testing procedure ran on the original complex networks nicely captures that difference. This translates into the near-to-zero statistical power of the test between RoI 1 and RoI 2 when testing samples of aggregated networks and into a global statistical power reaching the significance level α. Second, one can observe that there are no values on the diagonal in table 6b. This is because the principle of aggregation completely discards the information on the structure of the *intra*–subnetworks, preventing the discovery of any possible *intra*–difference.

Table 6: Estimated probabilities of rejection obtained from whole (a) and aggregated (b) networks.

	RoI 1	RoI 2	RoI 3	RoI 4
RoI 1	0	1	0.001	0.004
RoI 2		0.001	0.004	0.005
RoI 3			0	0.002
RoI 4				0

(a) Using whole networks
Global power: 0.049 (intra), 1.000 (inter)

	RoI 1	RoI 2	RoI 3	RoI 4
RoI 1		0.001	0.008	0.004
RoI 2			0	0
RoI 3				0.002
RoI 4				

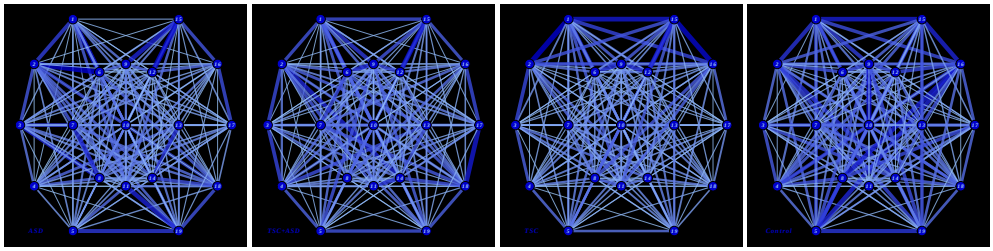
(b) Using aggregated networks
Global power: NA (intra), 0.042 (inter)

5 Application to the study of Autism Spectrum Disorder

5.1 Description of the data set

We applied our methodology to brain functional networks of electroencephalographic (EEG) connectivity data of children with tuberous sclerosis complex (TSC) and autism spectrum disorder (ASD) as previously studied in Peters et al. [2013]. TSC is a multi-system, autosomal dominant disorder affecting children and adults and it results from mutations in one of two genes, TSC1 or TSC2 [Crino et al., 2006]. Approximately 40% of patients affected by TSC develop ASD [Jeste et al., 2008]. However, it remains unclear whether the symptoms usually associated with autism are caused by the same neurobiological mechanism in patients with TSC or without TSC (the latter is often called non-syndromic ASD). The data has been collected from patients with TSC without ASD ($n = 29$), patients with TSC and ASD ($n = 14$), patients with non-syndromic ASD ($n = 16$) and healthy controls ($n = 13$). See Peters et al. [2013] for the details on the process of identification and diagnosis of the patients included in the study. Each network has 19 vertices identified by the electrode locations from the international 10-20 system of electrode placement while the edges are given by the coherence measure (see Peters et al. [2013] for the detail on this measure and the validity of this approach) between the EEG signals of two electrodes. Coherence has a value between 0 and 1 and networks are therefore weighted with weight between 0 and 1.

Figure 4: Brain network representations for a patient with non-syndromic ASD (a), a patient with syndromic ASD (b), a patient with only TSC (c) and a healthy subject (d).



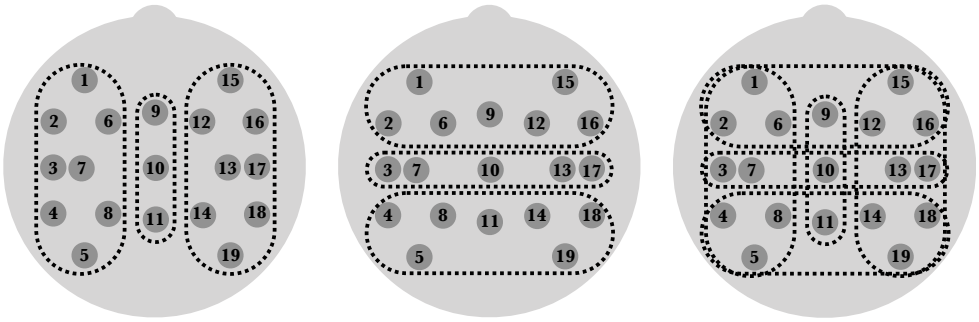
(a) ASD (b) TSC+ASD (c) TSC (d) Healthy

5.2 Choice of the partitions

We explored 3 partitions to investigate whether local differences are present between children with ASD and controls and between syndromic and non-syndromic ASD. The 1st partition, indicated with $\mathcal{P}_{LR}$, separates the two hemispheres and it practically translates into a partition of 3 elements (see fig. 5a). The 2nd partition, indicated with $\mathcal{P}_{FT}$, is composed of 3 elements: frontal/anterior area,

intermediate area and posterior area (see fig. 5b). For the last partition, we considered the union of the first two. This is not an actual partition but, as stated in Remark 3.2, we can still construct the σ–algebra generated by it. It boils down to considering a partition $\mathcal{P}_{LR-FT}$ of 9 elements (see fig. 5c). For the analysis, we used the α-TMTP with $\alpha = 1\%$. We carried out the tests using the adjacency matrix representation and the Frobenius distance.

Figure 5: The 3 partitions of the vertex set considered for the EEG data set. Partition $\mathcal{P}_{LR}$ of the brain in left hemisphere, intermediate area and right hemisphere (a); partition $\mathcal{P}_{FT}$ of the brain in frontal lobe, intermediate area and posterior lobe (b); Partition $\mathcal{P}_{LR-FT}$ induced by the σ-algebra from partitions $\mathcal{P}_{LR}$ and $\mathcal{P}_{FT}$ (c).



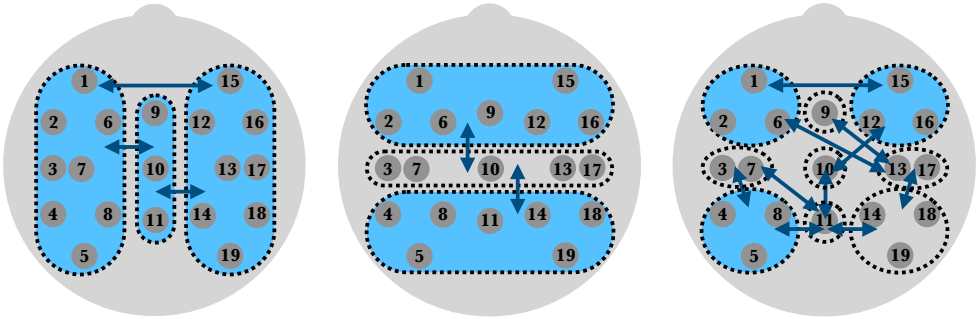
(a) $\mathcal{P}_{LR}$ (b) $\mathcal{P}_{FT}$ (c) $\mathcal{P}_{LR-FT}$

5.3 Results

When comparing syndromic against non-syndromic ASD (see fig. 6), local differences are found to be widespread when using partitions $\mathcal{P}_{LR}$ and $\mathcal{P}_{FT}$ with the exception of fronto-posterior and intermediate-intermediate connections. When using the more granular partition $\mathcal{P}_{LR-FT}$, more specificity emerges. Differences in frontal area are mostly confined to fronto-lateral regions (i.e. within F.L. and F.R. and between F.L. and F.R. but not those involving F.C.). In the intermediate and posterior areas, connections between homologous regions (i.e. between I.L. and I.R. and between P.L. and P.R.) are spared while connections within the same hemispheres (i.e. within P.L., between I.L. and P.L. and between P.R. and I.R.) are affected. Taken together, this findings suggest that the neural mechanism underpinning autism is locally different between syndromic and non-syndromic ASD with differences in connectivity involving the fronto-lateral regions (both inter- and intra-hemispheric) and the intermedio-posterior regions (confined to single hemispheres).

When comparing non-syndromic ASD and controls (see fig. 7, a remarkably similar distribution of differences is observed. In the frontal area, differences are mostly noticeable within and between fronto-leteral regions (i.e. within F.L. and F.R. and between F.L. and F.R.), while differences within the same hemispheres of the intermedio-posterior regions are also identified (i.e. within P.L., within

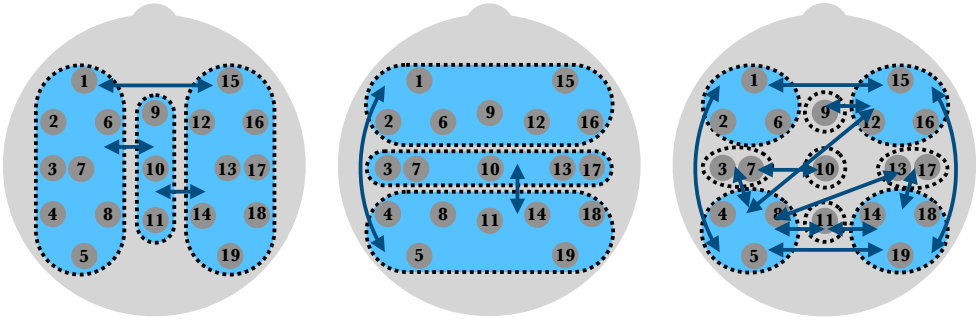
Figure 6: Locations of the differences between patients with syndromic or non-syndromic ASD. Light blue areas locate *intra*–differences, while dark blue arrows locate *inter*–differences.



(a) With $\mathcal{P}_{LR}$ (b) With $\mathcal{P}_{FT}$ (c) With $\mathcal{P}_{LR-FT}$

P.R., between I.L. and P.L. and between P.R. and I.R.). The overlap of sub-networks that are significantly different in the comparisons between ASD and controls and between syndromic and non-syndromic ASD suggests that what differs between syndromic and non-syndromic ASD is the degree to which sub-networks are affected (rather than their nature). This parallels the findings based on the comparison of specific white matter tracts between TSC with and without ASD [Lewis et al., 2012, Peters et al., 2012]. There are however additional differences observed when comparing ASD and controls which are not observed in the comparison between syndromic and non-syndromic ASD. Anatomically speaking, the most striking ones are significant differences between F.L. and P.L. and between F.R. and both P.L. and P.R. This suggests that such long-distance connections are a common basis in the mechanism of autism regarless of the presence or absence of TSC. Such long-distance connections between the frontal and posterior areas have been repeatedly demonstrated to be altered in ASD [Kana et al., 2009, Tan et al., 2010, Barttfeld et al., 2011].

Figure 7: Locations of the differences between patients with non-syndromic ASD and controls. Light blue areas locate *intra*–differences, while dark blue arrows locate *inter*–differences.



(a) With $\mathcal{P}_{LR}$ (b) With $\mathcal{P}_{FT}$ (c) With $\mathcal{P}_{LR-FT}$

Tables 7 and 8, Appendix C detail the p-values for these two comparisons.

6 Discussion

In this paper, we addressed the important problem of discriminating different form of autism at an early age. Specifically, we looked at the brain and modeled it as a network. We subsequently proposed an appropriate methodology for properly localizing significant differences on the brain network structure while providing a finite-sample strong control of the family-wise error rate. Thanks to this methodology, we managed to be very specific in the location of the differences between syndromic ASD (i.e. autism combined with tuberous sclerosis complex) and non-syndromic ASD (i.e. autism without TSC), which shed lights onto the underlying biomechanisms underpinning each form of autism.

The finite-sample aspect is key in this application since it is difficult to gather a large number of children diagnosed with autism and, at the same time, compliant enough for an EEG. Our proposed framework is fully non-parametric and relies on the theory of permutation tests. As such, very few assumptions are required on the data (only exchangeability under the null hypothesis and stochastic dominance of the test statistic under the alternative w.r.t. the null). We thus believe that it can easily be extended for studying object data in general, be it networks or other complex data. Although not yet available on CRAN, the `R` package providing the full implementation of the proposed method is available on Github under the name nevada (in a branch named *local*).

acknowledgements

We thank Boston Children's Hospital for making the EEG data available for this analysis. At this point in time, we cannot however publish the data because it would violate patients' privacy rights.

References

Pablo Barttfeld, Bruno Wicker, Sebastián Cukier, Silvana Navarta, Sergio Lew, and Mariano Sigman. A big-world network in asd: dynamical connectivity analysis reflects a deficit in long-range connections and an excess of short-range connections. *Neuropsychologia*, 49(2):254–263, 2011.

Danielle S Bassett and Michael S Gazzaniga. Understanding complexity in the human brain. *Trends in cognitive sciences*, 15(5):200–209, 2011.

Danielle Smith Bassett and ED Bullmore. Small-world brain networks. *The neuroscientist*, 12(6):512–523, 2006.

Ed Bullmore and Olaf Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10 (3):186, 2009.

Ed Bullmore and Olaf Sporns. The economy of brain network organization. *Nature Reviews Neuroscience*, 13(5):336, 2012.

P. B. Crino, K. L. Nathanson, and E. P. Henske. The tuberous sclerosis complex. *N Engl J Med*, 356, 2006.

Reinhard Diestel. *Graph theory*. Springer Publishing Company, Incorporated, 2018.

D Durante and D B Dunson. Bayesian inference and testing of group differences in brain networks. *Bayesian Anal.*, 13(1):29–58, 2018.

Daniele Durante, David B. Dunson, and Joshua T. Vogelstein. Nonparametric bayes modeling of populations of networks. *J. Am. Statist. Ass.*, 112(520): 1516–1530, 2017. doi: 10.1080/01621459.2016.1219260. URL https://doi.org/10.1080/01621459.2016.1219260.

Michael D Fox, Abraham Z Snyder, Justin L Vincent, Maurizio Corbetta, David C Van Essen, and Marcus E Raichle. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences*, 102(27):9673–9678, 2005.

C E Ginestet, J Li, P Balanchandran, S Rosenberg, and E D Kolaczyk. Hypothesis testing for network data in functional neuroimaging. *The Annals of Applied Statistics*, 11(2):725–750, 2017.

Michael D Greicius, Ben Krasnow, Allan L Reiss, and Vinod Menon. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences*, 100(1):253–258, 2003.

P W Holland, K B Laskey, and S Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5:109–137, 1983.

Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, pages 65–70, 1979.

Gerhard Hommel. A stagewise rejective multiple test procedure based on a modified bonferroni test. *Biometrika*, 75(2):383–386, 1988.

Shafali S. Jeste, Mustafa Sahin, Patrick Bolton, George B. Ploubidis, and Ayla Humphrey. Characterization of autism in young children with tuberous sclerosis complex. *Journal of Child Neurology*, 23(5):520–525, 2008. doi: 10.1177/0883073807309788. URL https://doi.org/10.1177/0883073807309788. PMID: 18160549.

Rajesh K Kana, Timothy A Keller, Vladimir L Cherkassky, Nancy J Minshew, and Marcel Adam Just. Atypical frontal-posterior synchronization of theory of

mind regions in autism during mental state attribution. *Social neuroscience*, 4(2):135–152, 2009.

G Keppel and TD Wickens. Simultaneous comparisons and the control of type i errors. *Design and analysis: A researchers handbook. 4th ed. Upper Saddle River (NJ): Pearson Prentice Hall. p*, pages 111–130, 2004.

William W Lewis, Mustafa Sahin, Benoit Scherrer, Jurriaan M Peters, Ralph O Suarez, Vanessa K Vogel-Farley, Shafali S Jeste, Matthew C Gregas, Sanjay P Prabhu, Charles A Nelson III, et al. Impaired language pathways in tuberous sclerosis complex patients with autism spectrum disorders. *Cerebral cortex*, 23(7):1526–1532, 2012.

I. Lovato, A. Pini, A. Stamm, and S. Vantini. Model-free two-sample test for network-valued data. *MOX - Tech. Report, Dep. of Mathematics, Politecnico di Milano*, 2018.

Mary-Ellen Lynall, Danielle S Bassett, Robert Kerwin, Peter J McKenna, Manfred Kitzbichler, Ulrich Muller, and Ed Bullmore. Functional connectivity and brain networks in schizophrenia. *Journal of Neuroscience*, 30(28):9477–9487, 2010.

Dante Mantini, Mauro G Perrucci, Cosimo Del Gratta, Gian L Romani, and Maurizio Corbetta. Electrophysiological signatures of resting state networks in the human brain. *Proceedings of the National Academy of Sciences*, 104 (32):13170–13175, 2007.

Ruth Marcus, Eric Peritz, and K. R. Gabriel. On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*, 63(3):655–660, 1976. ISSN 00063444. URL http://www.jstor.org/stable/2335748.

M. E. J. Newman. The structure and function of complex networks. *SIAM Rev.*, 45(2):167–256, 2003.

F. Pesarin and L. Salmaso. *Permutation Tests for Complex Data*. Wiley, 2010.

Jurriaan M Peters, Mustafa Sahin, Vanessa K Vogel-Farley, Shafali S Jeste, Charles A Nelson III, Matthew C Gregas, Sanjay P Prabhu, Benoit Scherrer, and Simon K Warfield. Loss of white matter microstructural integrity is associated with adverse neurological outcome in tuberous sclerosis complex. *Academic radiology*, 19(1):17–25, 2012.

Jurriaan M Peters, Maxime Taquet, Clemente Vega, Shafali S Jeste, Ivn Snchez Fernndez, Jacqueline Tan, Charles A Nelson III, Mustafa Sahin, and Simon K Warfield. Brain functional networks in syndromic and non-syndromic autism: a graph theoretical study of EEG connectivity. *BMC Medicine*, 54(11), 2013.

Mary L Phillips and Holly A Swartz. A critical appraisal of neuroimaging studies of bipolar disorder: toward a new conceptualization of underlying neural circuitry and a road map for future research. *American Journal of Psychiatry*, 171(8):829–843, 2014.

Belinda Phipson and Gordon K. Smyth. Permutation P-values Should Never Be Zero: Calculating Exact P-values When Permutations Are Randomly Drawn. *Stat Appl Genet Mol Biol.*, 9(1):1–12, 2010. ISSN 1544-6115. doi: 10.2202/1544-6115.1585.

Maria Giulia Preti, Thomas AW Bolton, and Dimitri Van De Ville. The dynamic functional connectome: state-of-the-art and perspectives. *Neuroimage*, 160: 41–54, 2017.

Th Royen. Generalized maximum range tests for pairwise comparisons of several populations. *Biometrical Journal*, 31(8):905–929, 1989.

Mikail Rubinov and Olaf Sporns. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, 52(3):1059–1069, 2010.

Aswin Sekar, Allison R Bialas, Heather de Rivera, Avery Davis, Timothy R Hammond, Nolan Kamitaki, Katherine Tooley, Jessy Presumey, Matthew Baum, Vanessa Van Doren, et al. Schizophrenia risk from complex variation of complement component 4. *Nature*, 530(7589):177, 2016.

Juliet Popper Shaffer. Modified sequentially rejective multiple test procedures. *Journal of the American Statistical Association*, 81(395):826–831, 1986.

Stephen M Smith, Diego Vidaurre, Christian F Beckmann, Matthew F Glasser, Mark Jenkinson, Karla L Miller, Thomas E Nichols, Emma C Robinson, Gholamreza Salimi-Khorshidi, Mark W Woolrich, et al. Functional connectomics from resting-state fmri. *Trends in cognitive sciences*, 17(12):666–682, 2013.

Olaf Sporns, Giulio Tononi, and Rolf Kötter. The human connectome: a structural description of the human brain. *PLoS computational biology*, 1(4):e42, 2005.

Geoffrey CY Tan, Thomas F Doke, John Ashburner, Nicholas W Wood, and Richard SJ Frackowiak. Normal variation in fronto-occipital circuitry and cerebellar structure with an autism-associated polymorphism of cntnap2. *Neuroimage*, 53(3):1030–1042, 2010.

Martijn P Van Den Heuvel and Hilleke E Hulshoff Pol. Exploring the brain network: a review on resting-state fmri functional connectivity. *European neuropsychopharmacology*, 20(8):519–534, 2010.

Peter H Westfall and Randall D Tobias. Multiple testing of general contrasts: Truncated closure and the extended shaffer–royen method. *Journal of the American Statistical Association*, 102(478):487–494, 2007.

Andrew Zalesky, Alex Fornito, and Edward T Bullmore. Network-based statistic: identifying differences in brain networks. *Neuroimage*, 53(4):1197–1207, 2010.

A Appendix: Data generation for the simulations

A.1 Details about the first simulated data set

We report here the edge probability matrices used to generate the simulated scenarios in the first simulation study in Section 4.1.

Scenario	p_1	p_2
1	$\begin{pmatrix} 0.4 & 0.1 & 0.4 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.4 \\ 0.4 & 0.1 & 0.1 & 0.4 \\ 0.1 & 0.4 & 0.4 & 0.1 \end{pmatrix}$	$\begin{pmatrix} 0.1 & 0.4 & 0.1 & 0.4 \\ 0.4 & 0.4 & 0.4 & 0.1 \\ 0.1 & 0.4 & 0.4 & 0.1 \\ 0.4 & 0.1 & 0.1 & 0.4 \end{pmatrix}$
2	$\begin{pmatrix} 0.1 & 0.4 & 0.4 & 0.1 \\ 0.4 & 0.4 & 0.1 & 0.1 \\ 0.4 & 0.1 & 0.4 & 0.4 \\ 0.1 & 0.1 & 0.4 & 0.1 \end{pmatrix}$	$\begin{pmatrix} 0.4 & 0.4 & 0.4 & 0.1 \\ 0.4 & 0.1 & 0.1 & 0.1 \\ 0.4 & 0.1 & 0.1 & 0.4 \\ 0.1 & 0.1 & 0.4 & 0.4 \end{pmatrix}$
3	$\begin{pmatrix} 0.1 & 0.4 & 0.1 & 0.1 \\ 0.4 & 0.1 & 0.4 & 0.4 \\ 0.1 & 0.4 & 0.4 & 0.1 \\ 0.1 & 0.4 & 0.1 & 0.4 \end{pmatrix}$	$\begin{pmatrix} 0.1 & 0.1 & 0.4 & 0.4 \\ 0.1 & 0.1 & 0.1 & 0.1 \\ 0.4 & 0.1 & 0.4 & 0.4 \\ 0.4 & 0.1 & 0.4 & 0.4 \end{pmatrix}$
4	$\begin{pmatrix} 0.1 & 0.4 & 0.1 & 0.4 \\ 0.4 & 0.4 & 0.1 & 0.4 \\ 0.1 & 0.1 & 0.4 & 0.4 \\ 0.4 & 0.4 & 0.4 & 0.4 \end{pmatrix}$	$\begin{pmatrix} 0.1 & 0.4 & 0.4 & 0.4 \\ 0.4 & 0.4 & 0.4 & 0.4 \\ 0.4 & 0.4 & 0.1 & 0.1 \\ 0.4 & 0.4 & 0.1 & 0.4 \end{pmatrix}$

A.2 Details about the second simulated data set

Here, we provide details about the edge weight modification process used for introducing a difference between the two samples in the second simulation study. Recall that, first, all edge weights were drawn from a Poisson distribution with parameter 8 – $\mathcal{P}(8)$ – for both samples. Then, half of the edges that connect vertices within the first two elements of the partition were modified by drawing their weight from $\mathcal{P}(5)$ in sample 1 and $\mathcal{P}(11)$ in sample 2 and the other half had their weight drawn from $\mathcal{P}(11)$ for sample 1 and $\mathcal{P}(5)$ for sample 2. The following table details precisely which weights were drawn from which distribution.

New distribution	Sample 1	Sample 2
$\mathcal{P}(5)$	1-11, 1-13, 2-12, 2-17, 2-18, 3-14, 3-19, 4-12, 4-16, 5-11, 5-15, 5-17	6-20, 7-20, 7-23, 7-27, 8-24, 8-26, 8-28, 9-21, 9-22, 9-29, 10-24, 10-25
$\mathcal{P}(11)$	6-20, 7-20, 7-23, 7-27, 8-24, 8-26, 8-28, 9-21, 9-22, 9-29, 10-24, 10-25	1-11, 1-13, 2-12, 2-17, 2-18, 3-14, 3-19, 4-12, 4-16, 5-11, 5-15, 5-17

B Appendix: Details on the closed testing procedures

B.1 Toy Example

In this section, we report a toy example that shows the differences between the complete multiscale testing procedure (CMTP) and the α-truncated multiscale testing procedure (α-TMTP). For that purpose, we used a partition of 4 elements and report in the following figure the output of each procedure for controlling the FWER on on the set of tests performed on *intra*-subnetworks.

B.2 Algorithms

In this section, we provide the detailed algorithms for running the two proposed multiscale testing procedures as they are implemented in our `R` package nevada.
First, Algorithm 1 provides the implementation details for the CMTP:
Next, Algorithm 2 provides the implementation details for the α-TMTP:

B.3 Computational burden

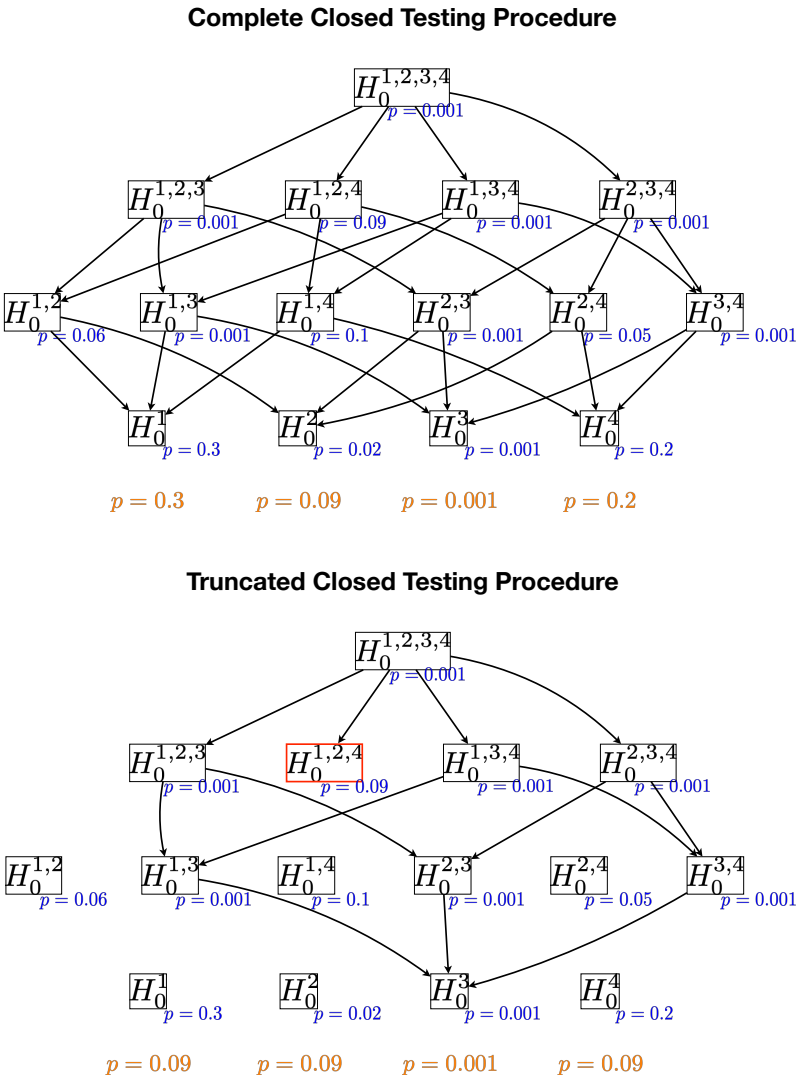
In this section, we compare the computational times required to perform the CMTP and the α-TMTP. Specifically here, we plot the computation times for both procedures based on the equations established in Section 3.4. We use two partition sizes: $m = 5$ which can be considered as a small size and $m = 10$ which is a large size.

C Appendix: P-values for the analysis of the EEG data

The following table reports the adjusted p-values obtained from the α-TMTP with $\alpha = 10\%$ for the comparison between non-syndromic and syndromic ASD:

The following table reports the adjusted p-values obtained from the α-TMTP with $\alpha = 10\%$ for the comparison between non-syndromic ASD and controls:

Figure 8: An example of closed testing procedure (top) and its truncated version (bottom). Hypotheses of interest are at the bottom of the pyramid.



Algorithm 1 Complete Multiscale Testing Procedure

procedure CMTP($\mathcal{G}_1, \mathcal{G}_2, \mathcal{V}$)
  # Initializations
  $N \leftarrow \text{length}(\mathcal{V})$;
  for $(i = 1; i \leq N; {+}{+}i)$ **do**
    $p_{V_i}^{\text{intra}} \leftarrow 0$;
    for $(j = i + 1; j \leq N; {+}{+}j)$ **do**
      $p_{V_i \cup V_j}^{\text{inter}} \leftarrow 0$;
  # Main loop
  for $A \in \sigma(\mathcal{V})$ **do**
    # P-value associated with full hypothesis
    $p_A^{\text{full}} \leftarrow$ permutation p-value investigating $\mathcal{G}_1$ vs $\mathcal{G}_2$ on G_A^{full} as in eq. (3.1);
    # P-value associated with intra hypothesis
    $p_A^{\text{intra}} \leftarrow$ permutation p-value investigating $\mathcal{G}_1$ vs $\mathcal{G}_2$ on G_A^{intra} as in eq. (3.2);
    # P-value associated with inter hypothesis
    $p_A^{\text{inter}} \leftarrow$ permutation p-value investigating $\mathcal{G}_1$ vs $\mathcal{G}_2$ on G_A^{inter} as in eq. (3.3);
    # P-value adjustment
    for $(i = 1; i \leq N; {+}{+}i)$ **do**
      if $V_i \subseteq A$ **then**
        $p_{V_i}^{\text{intra}} = \max(p_{V_i}^{\text{intra}}, p_A^{\text{full}}, p_A^{\text{intra}})$
        for $(j = i + 1; j \leq N; {+}{+}j)$ **do**
          if $V_i \cup V_j \subseteq A$ **then**
            $p_{V_i \cup V_j}^{\text{inter}} = \max(p_{V_i \cup V_j}^{\text{inter}}, p_A^{\text{full}}, p_A^{\text{inter}})$;
  return $\{p_{V_i}^{\text{intra}} : i = 1, \ldots, N\}$ and $\{p_{V_i \cup V_j}^{\text{inter}} : i, j = 1, \ldots, N \text{ s.t. } i < j\}$.

Algorithm 2 Truncated Multiscale Testing Procedure

procedure TMTP($\mathcal{G}_1, \mathcal{G}_2, \mathcal{V}, \alpha$)
  # Initializations
  $N \leftarrow \text{length}(\mathcal{V})$;
  skipIntraTests $\leftarrow$ skipInterTests $\leftarrow \emptyset$;
  for $(i = 1; i \leq N; {+}{+}i)$ **do**
    $p^{\text{intra}}_{V_i} \leftarrow 0$;
    for $(j = i + 1; j \leq N; {+}{+}j)$ **do**
      $p^{\text{inter}}_{V_i \cup V_j} \leftarrow 0$;
  # Main loop
  for $(k = N; k \geq 1; {+}{+}k)$ **do**
    for $A \in \sigma(\mathcal{V}) : \dim(A) = k$ **do**
      if $(A \in \text{skipIntraTests}) \cap (A \in \text{skipInterTests})$ **then**
        continue;
      # P-value associated with full hypothesis
      $p^{\text{full}}_A \leftarrow$ permutation p-value investigating $\mathcal{G}_1$ vs $\mathcal{G}_2$ on G^{full}_A as in eq. (3.1);
      for $(i = 1; i \leq N; {+}{+}i)$ **do**
        if $V_i \subseteq A$ **then**
          $p^{\text{intra}}_{V_i} = \max(p^{\text{intra}}_{V_i}, p^{\text{full}}_A)$;
          for $(j = i + 1; j \leq N; {+}{+}j)$ **do**
            if $V_i \cup V_j \subseteq A$ **then**
              $p^{\text{inter}}_{V_i \cup V_j} = \max(p^{\text{inter}}_{V_i \cup V_j}, p^{\text{full}}_A)$;
      if $p^{\text{full}}_A > \alpha$ **then**
        for $B \in \sigma(\mathcal{V}) : B \subseteq A$ **do**
          skipIntraTests $\leftarrow$ skipIntraTests $\cup B$;
          skipInterTests $\leftarrow$ skipInterTests $\cup B$;
      if $A \notin$ skipIntraTests **then**
        # P-value associated with intra hypothesis
        $p^{\text{intra}}_A \leftarrow$ permutation p-value investigating $\mathcal{G}_1$ vs $\mathcal{G}_2$ on G^{inra}_A as in eq. (3.2);
        for $(i = 1; i \leq N; {+}{+}i)$ **do**
          if $V_i \subseteq A$ **then**
            $p^{\text{intra}}_{V_i} = \max(p^{\text{intra}}_{V_i}, p^{\text{intra}}_A)$;
        if $p^{\text{intra}}_A > \alpha$ **then**
          for $B \in \sigma(\mathcal{V}) : B \subseteq A$ **do**
            skipIntraTests $\leftarrow$ skipIntraTests $\cup B$;
      if $A \notin$ skipInterTests **then**
        # P-value associated with inter hypothesis
        $p^{\text{inter}}_A \leftarrow$ permutation p-value investigating $\mathcal{G}_1$ vs $\mathcal{G}_2$ on G^{inter}_A as in eq. (3.3);
        for $(i = 1; i \leq N; {+}{+}i)$ **do**
          if $V_i \subseteq A$ **then**
            for $(j = i + 1; j \leq N; {+}{+}j)$ **do**
              if $V_i \cup V_j \subseteq A$ **then**
                $p^{\text{inter}}_{V_i \cup V_j} = \max(p^{\text{inter}}_{V_i \cup V_j}, p^{\text{inter}}_A)$;
        if $p^{\text{inter}}_A > \alpha$ **then**
          for $B \in \sigma(\mathcal{V}) : B \subseteq A$ **do**
            skipInterTests $\leftarrow$ skipInterTests $\cup B$;
  return $\{p^{\text{intra}}_{V_i} : i = 1, \ldots, N\}$ and $\{p^{\text{inter}}_{V_i \cup V_j} : i, j = 1, \ldots, N \text{ s.t. } i < j\}$.

Figure 9: **Relative computational savings of the α-TMTP w.r.t. the CMTP.** Black solid line for the intra–subetwork case, black dashed line for the inter–subnetwork case and red solid line for the 2^{-k} approximation.

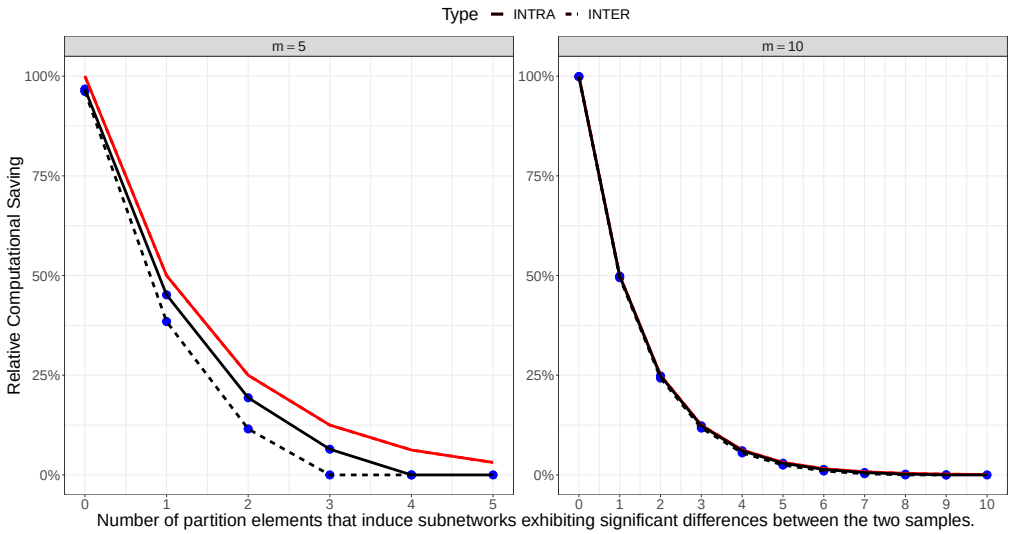


Table 7: Non-Syndromic vs Syndromic ASD. Intra- and inter- p-values adjusted with the 10%-TMTP.

	Left	Central	Right
Left	$<$ **0.001**	**0.002**	**0.002**
Central		**0.016**	**0.002**
Right			**0.004**

Left and right hemisphere.
$p_{\text{intra}} = \mathbf{0.0002} - p_{\text{inter}} = \mathbf{0.0006}$

	Front.	Interm.	Temp.
Front.	**0.002**	$<$ **0.001**	$\geq$ 0.231
Interm.		$\geq$ 0.169	$<$ **0.001**
Temp.			**0.002**

Frontal and posterior lobe.
$p_{\text{intra}} = \mathbf{0.0002} - p_{\text{inter}} = \mathbf{0.0004}$

	F.L.	F.C.	F.R.	I.L.	I.C.	I.R.	T.L.	T.C.	T.R.
F.L.	**0.031**	$\geq$0.269	**0.023**	$\geq$0.181	$\geq$0.269	**0.093**	$\geq$0.197	$\geq$0.193	$\geq$0.269
F.C.			$\geq$0.296	$\geq$0.181	$\geq$0.287	**0.093**	$\geq$0.296	$\geq$0.208	$\geq$0.287
F.R.			**0.094**	$\geq$0.112	**0.067**	$\geq$0.152	$\geq$0.206	$\geq$0.208	$\geq$0.112
I.L.				$\geq$0.309	$\geq$0.169	$\geq$0.169	**0.050**	**0.095**	$\geq$0.181
I.C.						$\geq$0.169	$\geq$0.140	**0.022**	$\geq$0.287
I.R.						$\geq$0.510	$\geq$0.152	$\geq$0.510	**0.060**
T.L.							**0.037**	**0.072**	$\geq$0.187
T.C.									**0.060**
T.R.									$\geq$0.309

Right and left hemisphere, frontal and posterior lobe.
$p_{\text{intra}} = \mathbf{0.0004} - p_{\text{inter}} = \mathbf{0.0004}$

Table 8: Non-Syndromic ASD vs Controls. Intra- and inter- p-values adjusted with the 10%-TMTP.

	Left	Central	Right
Left	$<$ **0.001**	**0.003**	**0.001**
Central		**0.026**	**0.002**
Right			$<$ **0.001**

	Front.	Interm.	Temp.
Front.	$<$ **0.001**	$\geq$ 0.146	**0.021**
Interm.		**0.010**	**0.003**
Temp.			$<$ **0.001**

Left and right hemisphere.
$p_{\text{intra}} = \mathbf{0.0002}$ – $p_{\text{inter}} = \mathbf{0.0010}$

Frontal and posterior lobe.
$p_{\text{intra}} = \mathbf{0.0002}$ – $p_{\text{inter}} = \mathbf{0.0034}$

	F.L.	F.C.	F.R.	I.L.	I.C.	I.R.	T.L.	T.C.	T.R.
F.L.	**0.003**	$\geq$0.172	**0.005**	$\geq$0.163	$\geq$0.172	$\geq$0.172	**0.033**	$\geq$0.110	$\geq$0.163
F.C.			**0.038**	$\geq$0.163	$\geq$0.172	$\geq$0.172	$\geq$0.151	$\geq$0.110	$\geq$0.163
F.R.			**0.001**	$\geq$0.209	$\geq$0.171	$\geq$0.209	**0.032**	$\geq$0.209	**0.095**
I.L.				$\geq$0.155	**0.077**	$\geq$0.209	**0.006**	$\geq$0.209	$\geq$0.163
I.C.						$\geq$0.172	$\geq$0.639	$\geq$0.171	$\geq$0.151
I.R.						$\geq$0.187	**0.032**	$\geq$0.219	**0.085**
T.L.							**0.002**	**0.008**	**0.029**
T.C.									**0.038**
T.R.									**0.006**

Right and left hemisphere, frontal and posterior lobe.
$p_{\text{intra}} = \mathbf{0.0002}$ – $p_{\text{inter}} = \mathbf{0.0006}$

MOX Technical Reports, last issues

Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

38/2019 Massi, M.C.; Ieva, F.; Gasperoni, F.; Paganoni, A.M.
Minority Class Feature Selection through Semi-Supervised Deep Sparse Autoencoders

36/2019 Salvador, M.; Dede', L.; Quarteroni, A.
An intergrid transfer operator using radial basis functions with application to cardiac electromechanics

37/2019 Menafoglio, A.; Secchi, P.
O2S2: a new venue for computational geostatistics

34/2019 Antonietti, P. F.; Mazzieri, I.; Melas, L.; Paolucci, R.; Quarteroni, A.; Smerzini, C.; Stupazzini,
Three-dimensional physics-based earthquake ground motion simulations for seismic risk assessment in densely populated urban areas

35/2019 Zancanaro, M.; Ballarin, F.; Perotto, S.; Rozza, G.
Hierarchical model reduction techniques for flow modeling in a parametrized setting

33/2019 Regazzoni, F.; Dede', L.; Quarteroni, A.
Machine learning of multiscale active force generation models for the efficient simulation of cardiac electromechanics

32/2019 Fedele, M.
Polygonal surface processing and mesh generation tools for numerical simulations of the complete cardiac function.

31/2019 Pagani, S.; Vitulano, G.; De Blasi, G.; Frontera, A.
High density characterization of the atrial electrical substrate during sinus rhythm in patients with atrial fibrillation

30/2019 Pagani, S.; Manzoni, A.; Quarteroni, A.
Forward uncertainty quantification in cardiac electrophysiology by reduced order modeling and machine learning

29/2019 Dal Santo, N.; Manzoni, A.; Pagani, S.; Quarteroni, A.
Reduced order modeling for applications to the cardiovascular system