

HYPOTHESIS TESTING FOR NETWORK DATA IN FUNCTIONAL NEUROIMAGING

BY CEDRIC E. GINESTET^{*,1}, JUN LI^{†,2}, PRAKASH BALACHANDRAN[†],
STEVEN ROSENBERG^{†,2} AND ERIC D. KOLACZYK^{†,1,2}

King's College London and Boston University†*

In recent years, it has become common practice in neuroscience to use networks to summarize relational information in a set of measurements, typically assumed to be reflective of either functional or structural relationships between regions of interest in the brain. One of the most basic tasks of interest in the analysis of such data is the testing of hypotheses, in answer to questions such as “Is there a difference between the networks of these two groups of subjects?” In the classical setting, where the unit of interest is a scalar or a vector, such questions are answered through the use of familiar two-sample testing strategies. Networks, however, are not Euclidean objects, and hence classical methods do not directly apply. We address this challenge by drawing on concepts and techniques from geometry and high-dimensional statistical inference. Our work is based on a precise geometric characterization of the space of graph Laplacian matrices and a nonparametric notion of averaging due to Fréchet. We motivate and illustrate our resulting methodologies for testing in the context of networks derived from functional neuroimaging data on human subjects from the 1000 Functional Connectomes Project. In particular, we show that this global test is more statistically powerful than a mass-univariate approach. In addition, we have also provided a method for visualizing the individual contribution of each edge to the overall test statistic.

1. Introduction. Functional neuroimaging data has been central to the advancement of our understanding of the human brain. Neuroimaging data sets are increasingly approached from a graph-theoretical perspective using the tools of modern network science [Bullmore and Sporns (2009)]. This has elicited the interest of statisticians working in that area. At the level of basic measurements, neuroimaging data can be said to consist typically of a set of signals (usually time series) at each of a collection of pixels (in two dimensions) or voxels (in three dimensions). Building from such data, various forms of higher-level data representations are employed in neuroimaging. Traditionally, two- and three-dimensional

Received August 2014; revised November 2016.

¹Supported in part by a grant from the Air Force Office for Scientific Research (AFOSR) whose grant number is FA9550-12-1-0102.

²This material is also based upon work supported by, or in part by, the U.S. Army Research Laboratory and the U.S. Army Research Office under contract/grant number W911NF1510440.

Key words and phrases. Fréchet mean, fMRI, graph Laplacian, hypothesis testing, matrix manifold, network data, object data.

images have, naturally, been the norm, but increasingly in recent years there has emerged a substantial interest in network-based representations.

1.1. *Motivation.* Let $G = (V, E)$ denote a graph based on $d = |V|$ vertices. In this setting, the vertices $v \in V$ correspond to regions of interest (ROIs) in the brain, often predefined through considerations of the underlying neurobiology (e.g., the putamen or the cuneus). Edges $\{u, v\} \in E$ between vertices u and v are used to denote a measure of association between the corresponding ROIs. Depending on the imaging modality used, the notion of “association” may vary. For example, in diffusion tensor imaging (DTI), associations are taken to be representative of structural connectivity between brain regions. On the other hand, in functional magnetic resonance imaging (fMRI), associations are instead thought to represent functional connectivity, in the sense that the two regions of the brain participate together in the achievement of some higher-order function, often in the context of performing some task (e.g., counting from 1 to 10).

With neuroimaging now a standard tool in clinical neuroscience, and with the advent of several major neuroscience research initiatives—perhaps most prominent being the recently announced Brain Research Accelerated by Innovative Neurotechnologies (BRAIN) initiative—we are quickly moving toward a time in which we will have available databases composed of large collections of secondary data in the form of network-based data objects. Faced with databases in which networks are a fundamental unit of data, it will be necessary to have in place the statistical tools to answer such questions as, “What is the ‘average’ of a collection of networks?” and “Do these networks differ, on average, from a given nominal network?”, as well as “Do two collections of networks differ on average?” and “What factors (e.g., age, gender, etc.) appear to contribute to differences in networks?”, or finally, say, “Has there been a change in the networks for a given subpopulation from yesterday to today?” In order to answer these and similar questions, we require network-based analogues of classical tools for statistical estimation and hypothesis testing.

While these classical tools are among the most fundamental and ubiquitous in use in practice, their extension to network-based data sets, however, is not immediate and, in fact, can be expected to be highly nontrivial. The main challenge in such an extension is due to the simple fact that networks are not Euclidean objects (for which classical methods were developed)—rather, they are combinatorial objects, defined simply through their sets of vertices and edges. Nevertheless, our work here in this paper demonstrates that networks can be associated with certain natural subsets of Euclidean space, and furthermore demonstrates that through a combination of tools from geometry, probability on manifolds and high-dimensional statistical analysis it is possible to develop a principled and practical framework in analogy to classical tools. In particular, we focus on the development of an asymptotic framework for one- and two-sample hypothesis testing.

Key to our approach is the correspondence between an undirected graph G and its Laplacian, where the latter is defined as the matrix $L = D - W$, with W denoting the $d \times d$ adjacency matrix of G and D a diagonal matrix with the vertex degrees along the diagonal. When G has no self-loops and no multi-edges, the correspondence between graphs G and Laplacians L is one-to-one. Our work takes place in the space of graph Laplacians. Importantly, this requires working not in standard Euclidean space \mathbb{R}^n , but rather on certain subsets of Euclidean space which are either submanifolds of \mathbb{R}^n or submanifolds of \mathbb{R}^n with corners. While these subsets of Euclidean space have the potential to be complicated in nature, we show that, in the absence of any nontrivial structural constraints on the graphs G , the geometry of these subsets is sufficiently “nice” to allow for a straightforward definition of distance between networks to emerge.

Our goal in this work is the development of one- and two-sample tests for network data objects that rely on a certain sense of “average.” We adopt the concept of Fréchet means in defining what average signifies in our context. Recall that, for a metric space, (\mathcal{X}, ρ) , and a probability measure, Q , on its Borel σ -field, under appropriate conditions, the Fréchet mean of Q is defined as the (possibly nonunique) minimizer

$$(1) \quad \mu := \operatorname{argmin}_{x \in \mathcal{X}} \int_{\mathcal{X}} \rho^2(x, y) Q(dy).$$

Similarly, for any sample of realizations from Q on \mathcal{X} , denoted $Y := \{Y_1, \dots, Y_n\}$, the corresponding sample Fréchet mean is defined as

$$(2) \quad \hat{\mu}_n(Y) := \operatorname{argmin}_{x \in \mathcal{X}} \frac{1}{n} \sum_{i=1}^n \rho^2(x, Y_i).$$

Thus, the distance ρ that emerges from our study of the geometry of the space of networks implicitly defines a corresponding notion of how to “average” networks.

Drawing on results from nonparametric statistical inference on manifolds, we are then able to establish a central limit theory for such averages and, in turn, construct the asymptotic distributions of natural analogues of one- and two-sample z -tests. These tests require knowledge of the covariance among the edges of our networks, which can be expected to be unavailable in practice. Nevertheless, we show how recent advances in the estimation of large, structured covariance matrices can be fruitfully brought to bear in our context, and provide researchers with greater statistical power than a mass-univariate approach, which is the standard approach in this field.

1.2. The 1000 Functional Connectomes Project. Our approach is motivated by and illustrated with data from the 1000 Functional Connectomes Project (FCP). This major MRI data-sharing initiative was launched in 2010 [Biswal et al. (2010)]. The impetus for the 1000 FCP was given by a need to make widely accessible neuroimaging data, which are costly and time-consuming to collect [Biswal et al.

(2010)]. This was conducted within the so-called “discovery science” paradigm, paralleling similar initiatives in systems biology. The 1000 FCP constituted the largest data set of its kind at the time of its release. As for the use of such large data sets in genetics, it is believed that facilitating access to high-throughput data generates economies of scale that are likely to lead to more numerous and more substantive research findings.

The 1000 FCP describes functional neuroimaging data from 1093 subjects, located in 24 community-based centers. The mean age of the participants is 29 years, and all subjects were 18 years old or older. Each individual scan lasted between 2.2 and 20 minutes. The strength of the MRI scanner varied across centers, with $n = 970$ scans at 3T and $n = 123$ at 1.5T. Voxel-size was 1.5–5 mm within the plane; and slice thickness was 3–8 mm. The ethics committee in each contributing data center approved the project; and the institutional review boards of the NYU Langone Medical Center and of the New Jersey Medical School approved the dissemination of the data. This freely available data set has been extensively used in the neuroimaging literature [Tomasi and Volkow (2010), Yan et al. (2013), Zuo et al. (2012)].

The individual fMRI scans were parcellated into a set of 50 cortical and subcortical regions using the Automated Anatomical Labeling (AAL) template [Tzourio-Mazoyer et al. (2002)]. Note that the resulting connectivity networks are sensitive to our particular choice of parcellation, and that the results in this paper need not generalize to other templates [see Wang et al. (2009) for a review]. The voxel-specific time series in each of these regions were aggregated to form mean regional time series, as commonly done in the study of the human connectome [see, e.g., Achard et al. (2006)]. The resulting regional time series were then compared using two different measures of association. We here considered the correlation coefficient since this measure has proved to be popular in the neuroimaging literature [Ginestet and Simmons (2011), Micheloyannis et al. (2009), Pachou et al. (2008)].

Subjects in the 1000 FCP data can be subdivided with respect to sex. Several groups of researchers have previously considered the impact of sex differences on resting-state connectivity [Biswal et al. (2010), Tomasi and Volkow (2011)]. It is hypothesized that sexual dimorphism in human genomic expression is likely to affect a wide range of physiological variables [Ellegren and Parsch (2007)]. In particular, differences in hormonal profiles (e.g., estrogen) during brain development are known to be related to region-specific effects [McEwen (1999)]. Thus, it is of interest to compare the subject-specific networks of males and females in the 1000 FCP data set (see Figure 1). Observe that previous research in this field has established *local* sex differences in connectivity by considering individual edge weights [Biswal et al. (2010), Tomasi and Volkow (2011)]. By contrast, we are here investigating the effect of sex differences on *entire* networks.

It is here useful to distinguish between these two types of network data analysis in neuroimaging. While local analysis focuses on edge-specific statistics, global analysis instead considers network topological properties such as the shortest-path

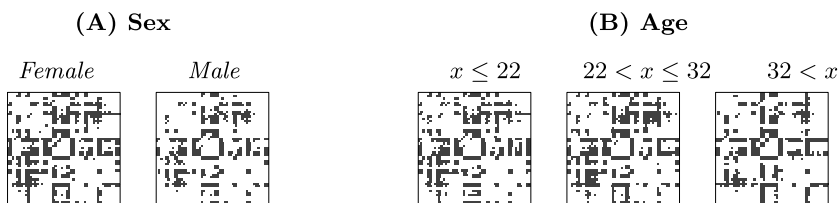


FIG. 1. Descriptive statistics for the 1000 FCP data set. In panel (A), the group-specific mean Laplacians for networks over 50 AAL vertices in females ($n_f = 555$) and males ($n_m = 462$). Similarly, in panel (B), the age variable has been divided into three groups ($n_1 = 386$, $n_2 = 297$ and $n_3 = 334$), and the respective means are reported for each age group. The Laplacians have been binarized with respect to the 75th percentile of the overall distribution of the entries in the full 1000 FCP database. (Black indicates entries greater or equal than that percentile.)

length. In this paper, we are extending the latter by providing a framework for identifying the mean network and characterizing the space of all possible such networks.

The organization of this paper is as follows. In Section 2, we describe the statistical and mathematical background of this type of research question. In Section 3, we provide a geometrical characterization of the space of networks under scrutiny. In Section 4, we describe how certain central limit theorems can be adapted to this space in order to construct a statistical inferential framework for network data. A simulation study exploring the relationship between statistical power and various aspects of neuroimaging data is reported in Section 5. In Section 6, we apply this framework to the analysis of a subset of the data from the 1000 FCP. These results and the potential extensions of the proposed statistical tests are then discussed in Section 7.

2. Related work. At the heart of the class of statistical problems we wish to address is a desire to summarize and compare groups of network data objects in a statistically principled manner. There are, of course, already a variety of numerical devices available for carrying out certain descriptive summaries and comparisons. Basic set-theoretic operations (e.g., union, intersection, symmetric difference) are all well defined for graphs. More broadly, various metrics, such as the Hamming distance, have been borrowed from other fields and applied to graphs. Currently, the mainstay in the analysis of network data in neuroimaging is the mass-univariate approach in which independent tests are conducted for every edge, adjusting for multiple testing; see [Ginestet, Fournel and Simmons \(2014\)](#) for a survey of such methods in the context of functional neuroimaging.

Such mass-univariate approaches, however, fail to draw inference about networks as a whole. In particular, it is unclear whether multiple local differences necessarily lead to globally significant differences. One may tackle this problem by treating network data objects as data points. What is lacking to achieve this, however, is the necessary mathematical foundation—establishing a formal “space”

of graphs, equipped with a formal metric with understood geometric and topological properties so that a formal notion of probability and measure can be defined, all underlying the desired theory and methods for the hypothesis testing problems of interest here.

Networks are not the only data type for which standard Euclidean-based methods are insufficient. Statistical inference on manifolds—in particular on spheres and shapes spaces—has a fairly long history. There is a substantial literature on statistics on spheres, or so-called directional statistics, going back to a seminal paper by R. A. Fisher in 1953 [Fisher (1953)], and works by Watson (1983), Mardia and Jupp (2000), and Fisher, Lewis and Embleton (1987), among others. Statistical analysis on shapes that are landmark-based was pioneered by Kendall (1977, 1984) and Bookstein (1978). Inference in these settings takes various forms. Nonparametric forms of inference typically employ a notion of averaging due to Fréchet (1948), as we do in this paper. Nevertheless, little work has been pursued with manifolds given as some general metric space—such as the spaces of networks that are our main interest. The most related work seems to be due to Billera, Holmes and Vogtmann (2001) and Barden, Le and Owen (2013), who study the metric geometry of the space of phylogenetic trees and derive a central limit theorem for the Fréchet mean in such spaces. Also, see the related work of Marron and colleagues in the context of so-called object-oriented data analysis with trees [Aydin et al. (2009), Wang and Marron (2007)].

In order to establish a formal characterization of a well-defined “space” of networks, it is natural to associate a network with a matrix. And while there are several such matrices that might be used, we have found that the (combinatorial) graph Laplacian is particularly appropriate. The Laplacian falls in the cone of symmetric positive (semi)definite (PSD) matrices. A substantial amount of effort has been expended on uncovering the mathematical properties of the PSD cone [Bhatia (1997), Moakher and Zéraï (2011)]. In addition, there has in recent years been quite a lot of work exploring the various notions of “average” induced upon this manifold by the underlying choices of geometry [Arsigny et al. (2007), Bonnabel and Sepulchre (2009), Moakher (2005)]. Finally, depending on the choice of average adopted, there are results establishing the probabilistic and statistical properties of averages through CLTs [Bhattacharya and Patrangenaru (2003, 2005), Bhattacharya and Bhattacharya (2012), Kendall and Le (2011)]. Much of this research has been motivated by shape analysis [Le (2001), Le and Kume (2000)], but many of these results have been developed in other areas of applications where matrices play a key role such as in DTI [Dryden, Koloydenko and Zhou (2009)].

However, the space of graph Laplacians forms a *subset* of the PSD cone and, furthermore, by definition this subset intersects in a nontrivial fashion with the boundary of this cone. Therefore, results for PSD matrices do not carry over immediately to the space of graph Laplacians—the latter must necessarily be studied in its own right. At present, while graph Laplacians as individual objects are well

studied [see Chung (1997), who discusses discrete eigenvalue and isoperimetric estimates analogous to Riemannian estimates; see also Chavel (1984), Xia (2013)], there appears to be no formal body of results to date establishing the properties of the *space* of graph Laplacians—and certainly none that reflects the impact of what have become established canonical properties of complex networks (e.g., sparseness, small-world, etc.). The closest work of which we are aware is, for example, recent work in the signal processing literature, characterizing subspaces of the PSD cone corresponding to subsets of covariance matrices sharing certain simple structural properties such as rank or trace constraints [Krishnamachari and Varanasi (2013)].

A certain notion of embedding is crucial to the mathematical and probabilistic theory underlying our approach. There are, in fact, different uses of the term “embedding.” Our work involves averaging or comparing different networks/graphs via the distance between network Laplacians computed by first embedding (i.e., smoothly injecting) the set of Laplacian matrices into a Euclidean space; here “embedding” is defined as in the differentiable topology literature [see Chapter 7 in Lee (2006)]. This seems to have advantages over comparing networks via, for example, isometric embeddings of the graph itself into \mathbb{R}^3 , for which computation of the types of distance functions that have been useful (e.g., Gromov–Hausdorff distance) is impractical.

In addition, there is also the large literature on graph embedding, which maps a graph onto a typically low-dimensional Euclidean space using eigenvector/eigenvalue information of the adjacency matrix or associated Laplacian [Fu and Ma (2013), Linal (2002), Linal, London and Rabinovich (1995), Yan et al. (2007)]. Graph embedding methods are very different from differentiable topology techniques. In particular, the image of a graph embedding is often used as a dimension-reduction tool. This map in general has some distortion, and so is not an isometry. This change in the geometry from the domain space to the range space implies that the precise inference framework for manifolds that we employ here, as described below, cannot be applied to graph embeddings. Thus, there is no natural notion of average and projection onto the image under a graph embedding, and in fact such a projection may not exist. On the other hand, our notion of embedding, which considers the spaces of Laplacians as a manifold, does not reduce dimension, preserves all the raw information in a specific graph, and allows analysis of averages and projections by geometric methods.

3. Characterization of spaces of networks. In this section, we establish the necessary mathematical properties associated with a certain notion of a “space” of networks from which a natural notion of “averaging” emerges. In fact, we offer several variations of a space of networks and, in doing so, illustrate how even relatively simple constraints on network topology affect the geometry of these spaces. The geometry is important when seeking to develop the corresponding probabilistic behavior of averages of networks, as we do in Section 4, which also informs the sampling distributions of the one- and two-sample test statistics that we develop.

3.1. *Main results.* Let $G = (V, E, W)$ be a *weighted* undirected graph for weights $w_{ij} = w_{ji} \geq 0$, where equality with zero holds if and only if $\{i, j\} \notin E$. Assume G to be simple (i.e., no self-loops or multi-edges). We associate uniquely with each graph G its graph Laplacian $L = D(W) - W$, where D is a diagonal matrix of weighted degrees (also called vertex strengths), that is, $D_{jj} = d_j(W) = \sum_{i \neq j} w_{ij}$. We further assume in most of what follows that G is connected, in which case L has one (and only one) zero eigenvalue and all the others are positive (and hence L is positive semi-definite).

Under the assumption that G is simple, there is a one-to-one correspondence between a graph G and its Laplacian matrix L . We therefore define our space of networks using the corresponding space of Laplacians. In the following theorem, we show that an initial notion of the space of graph Laplacians over d nodes admits a relatively simple topology, which can be described as a convex subset of an affine space in \mathbb{R}^{d^2} .

THEOREM 1. *The set \mathcal{L}_d of $d \times d$ matrices A , satisfying*

- (1) $\text{Rank}(A) = d - 1$,
- (2) *symmetry*, $A' = A$,
- (3) *positive semi-definiteness*, $A \geq 0$,
- (4) *the entries in each row sum to 0*,
- (5) *the off-diagonal entries are negative*, $a_{ij} < 0$,

forms a submanifold of \mathbb{R}^{d^2} of dimension $d(d - 1)/2$. In fact, \mathcal{L}_d is a convex subset of an affine space in \mathbb{R}^{d^2} of dimension $d(d - 1)/2$.

A proof of this theorem is in the supplemental article [Ginestet et al. (2017)]. The practical importance of this result is that \mathcal{L}_d admits several Riemannian metrics [Arsigny et al. (2007)], which give rise to a restricted class of distance functions. For example, any one of these metrics turns \mathcal{L}_d into a length space in the sense of Gromov (2007), that is, the distance between any two points $A, B \in \mathcal{L}_d$ is the length of some path from A to B . Also, all the usual notions of curvature, and its influence on variations of geodesics, come into play.

However, we note that the definition of \mathcal{L}_d requires that *every* potential edge in G be present, with edges only distinguishable in terms of the relative magnitude of their weights. Consider the description of the 1000 FCP data in Section 1.2. For the case where our network is defined to be, say, the matrix W of empirical correlations of signals between pairs of ROIs, the space \mathcal{L}_d is appropriate. On the other hand, if we chose instead to work with a thresholded version of such matrices, then it is important that we allow for *both* the presence/absence of edges by allowing weights to be zero. The result of Theorem 1 can be extended to include such networks, as described in the following corollary. This leads to a manifold that possesses corners. A good introduction to manifolds with corners can be found in

standard texts on smooth manifolds [see Chapter 14 in Lee (2006)]. Moreover, this manifold is also a convex subset of Euclidean space.

COROLLARY 1. *In Theorem 1, if condition (5) is replaced by*

(5') the off-diagonal entries are nonpositive, $a_{ij} \leq 0$,

then the corresponding matrix space \mathcal{L}'_d is a manifold with corners of dimension $d(d-1)/2$. Furthermore, \mathcal{L}'_d is a convex subset of an affine space in \mathbb{R}^{d^2} of dimension $d(d-1)/2$.

A proof of this corollary is also provided in the supplemental article [Ginestet et al. (2017)]. Importantly, the above theorem and its corollary indicate that the Euclidean metric (i.e., the Frobenius distance on the space of $d \times d$ matrices with real-valued entries) is a natural choice of distance function on our spaces of Laplacians. The metric space of interest is therefore composed of, for example, (\mathcal{L}'_d, ρ_F) , where ρ_F is the Frobenius distance

$$\rho_F(X, Y) := \|X - Y\|_F^2 = \sum_{i,j} (x_{ij} - y_{ij})^2$$

for any pair of matrices $X, Y \in \mathcal{L}'_d$. As we shall see momentarily below, in Section 4, the concept of a Fréchet mean and its sample-based analogue, as detailed in equations (1) and (2), may now be brought to bear, yielding a well-defined sense of an average of networks.

3.2. Extensions: Implications of constraints on network topology. In ending this section, we note that our definition of a “space of networks” is intentionally minimal in lacking constraints on the topology of the networks. However, one of the most fundamental results that has emerged from the past 20 years of complex network research is the understanding that real-world networks typically (although not exclusively) tend to possess a handful of quite marked structural characteristics. Examples include sparseness (i.e., number of edges scaling like the number of vertices), heavy-tailed degree distributions and the presence of cohesive subgraphs (aka communities); see Chapter 8 in Newman (2010), for example, for details and a more comprehensive summary. In the context of neuroimaging, it can be expected that the networks of interest will be sparse due to a trade-off between wiring cost and topological complexity [Bullmore and Sporns (2012)]. Importantly, this fact suggests that the appropriate differential or metric measure geometry of the “space of all networks”—or, more formally, the space of Laplacians corresponding to such networks—depends on the constraints imposed on these networks/Laplacians.

While a detailed study of these implications are beyond the scope of this paper, we illustrate them through the following theorem, which extends the previous

results to the more general case of graphs composed of different numbers of connected components. In particular, we can generalize Theorem 1 to spaces of Laplacians representing graphs with a fixed number of components, ℓ . (Recall that the rank of a Laplacian is equal to d minus the number of communities in that graph.)

THEOREM 2. *The set \mathcal{L}_ℓ of $d \times d$ matrices E , satisfying*

- (1 $_\ell$) $\text{Rank}(E) = \ell$,
- (2) E is symmetric,
- (3) E is positive semidefinite,
- (4) the sum of the entries of each column is zero,
- (5) each off-diagonal entry is negative,

forms a submanifold of \mathbb{R}^{d^2} of dimension $d\ell - \ell(\ell + 1)/2$.

A proof of this theorem is in the supplemental article [Ginestet et al. (2017)]. Intuitively, this result is stating that the number of connected components of the average of two graphs can be smaller than the number of components of each graph, but it cannot be larger; that is, the average of two graphs may decrease the number of communities, but it cannot increase that number. Indeed, when taking the Euclidean average of several graphs with non-negative edge weights, we can only maintain existing edges or create new edges.

4. Statistical inference on samples of networks. Having characterized a space of networks, it becomes possible to construct an inferential framework for comparing one or more samples of networks. We here describe some analogues of the classical one- and two-sample t -statistics in this setting. These are obtained by first selecting a notion of averaging and deriving a central limit theorem for sequences of network averages, next appealing to Wald-like constructions of test statistics, and, finally, utilizing recent results on high-dimensional covariance estimation.

4.1. A central limit theorem. Let G_1, \dots, G_n denote n graphs, each simple and assumed to have the same number of vertices d ; and let L_1, \dots, L_n be the corresponding combinatorial Laplacians. The L_i 's are assumed to be independent and identically distributed according to a distribution Q . In the context of neuroimaging, for example, these might be the correlation networks from resting-state fMRI images obtained from a group of human subjects matched for various demographic characteristics (e.g., age, gender) and health status (e.g., clinical manifestation of a given neurodegenerative disease).

The results of the previous section tell us that an appropriate sense of distance between pairs of networks is given by the Euclidean distance between their corresponding Laplacians. Combining these results with the definition of average in

equations (1) and (2), indicates that a principled way in which to define the average of n networks is through elementwise averaging of the entries of their Laplacians (and hence their adjacency matrices). Such an average is, of course, easily computed. However, this is not always the case when computing averages on manifolds; see, for instance, Chapter 6 in Bhatia (2007) for an illustration of the difficulties that may arise, when computing the matrix mean in the cone of positive definite symmetric matrices with respect to the geodesic distance on that manifold.

In the context of the 1000 FCP database, we wish to compare networks with respect to the sex of the subjects and over different age groups. It is thus necessary to compute the means in each subgroup of networks. This was done, for example, in Figure 1, by constructing the Euclidean mean of the Laplacians for each group of subjects in different age groups. Such group-specific mean Laplacians can then be interpreted as the mean functional connectivity in each group.

The sample Fréchet mean \widehat{L}_n is a natural statistic upon which to build our hypothesis tests about the average of networks or groups of networks. In order to do so, we require an understanding of the behavior of \widehat{L}_n as a random variable. Under broad regularity conditions, $\widehat{L}_n \rightarrow \Lambda$ almost surely; that is, the sample Fréchet mean, \widehat{L}_n , is a consistent estimator of the true mean Λ [see Ziezold (1977)]. In addition, under further assumptions, we can also derive a central limit theorem for the sample Fréchet mean of Laplacians, with respect to the half-vectorization map, ϕ .

THEOREM 3. *If the expectation $\Lambda := \mathbb{E}[L]$ does not lie on the boundary of \mathcal{L}'_d , and $\mathbb{P}[U] > 0$, where U is an open subset of \mathcal{L}'_d with $\Lambda \in U$ and \mathcal{L}'_d defined as Corollary 1, and under some further regularity conditions [see supplemental article Ginestet et al. (2017)], we obtain the following convergence in distribution:*

$$n^{1/2}(\phi(\widehat{L}_n) - \phi(\Lambda)) \longrightarrow N(0, \Sigma),$$

where $\Sigma := \text{Cov}[\phi(L)]$ and $\phi(L)$ denotes the half-vectorization of L .

Theorem 3 assumes that the true Fréchet mean does not lie on the boundary of the parameter space, which requires that all of its off-diagonal entries are nonzero. This potentially conflicts with the fact that neuroimaging networks are often hypothesized to be sparse [Bullmore and Sporns (2012)]. Note, however, that we are only requiring such entries to be nonzero in expectation. Thus, any positive value would suffice to ensure that the true Fréchet mean is away from the boundary, although structural zeros (i.e., zeros in the true Fréchet mean) would be problematic.

A proof of this theorem and the full set of assumptions are provided in the supplemental article [Ginestet et al. (2017)]. The argument is a specialization of a general result due to Bhattacharya and Lin (2017). The result stated in the theorem has fundamental significance regarding our goal of developing analogues of classical testing strategies for the analysis of network data objects. It is an asymptotic result stating that, given a sufficient number of samples from a population of

networks, an appropriately defined notion of sample average behaves in a classical manner: It possesses a statistical distribution that is approximately multivariate normal, centered on the population mean μ and with covariance Σ . Note that this population covariance is assumed to be positive definite, even though its sample estimate needs not be positive definite, as we will see in our examples. In such cases, we will estimate that covariance matrix using a method due to Schäfer and Strimmer (2005).

Theorem 3 can be straightforwardly extended in order to select specific entries in the Laplacians under scrutiny. Such submatrices would correspond to certain subgraphs of interest. In particular, given an orthogonal projection matrix, Γ , and using the fact that such central limit results are preserved under linearity, we obtain the following generalization of Theorem 3:

$$(3) \quad n^{1/2}(\Gamma\phi(\widehat{L}_n) - \Gamma\phi(\Lambda)) \longrightarrow N(0, \Gamma\Sigma\Gamma').$$

In the context of neuroimaging, this may allow to identify subnetworks of specific interest, such as the so-called default mode network, for instance [Greicius et al. (2003)]. We will study the properties of this neuroanatomical network in the sequel.

4.2. *One-sample, two-sample and k-sample tests.* As an immediate consequence of this central limit theorem, we can define natural analogues of classical one- and k -sample hypothesis tests. Consider, for example, the null hypothesis that the expectation $\Lambda = \mathbb{E}[L]$ is equal to some prespecified value, that is, $H_0 : \Lambda = \Lambda_0$. In the context of neuroimaging, the choice of Λ_0 might correspond to a reference connectivity pattern, derived from a large study, such as the 1000 FCP, for instance. In addition to the conditions stated in Theorem 3, let us now assume that the true covariance matrix, Σ , is nonsingular. Moreover, it is also assumed that the target Laplacian, Λ_0 , is known.

COROLLARY 2. *Under the assumptions of Theorem 3, and under the null hypothesis $H_0 : \mathbb{E}[L] = \Lambda_0$, the test statistic,*

$$T_1 := n(\phi(\widehat{L}) - \phi(\Lambda_0))' \widehat{\Sigma}^{-1}(\phi(\widehat{L}) - \phi(\Lambda_0)),$$

converges to a χ^2 -distribution with $p := \binom{d}{2}$ degrees of freedom, and where $\widehat{\Sigma} := 1/(n - 1) \sum_{i=1}^n (\phi(L_i) - \phi(\widehat{L}))(\phi(L_i) - \phi(\widehat{L}))'$ denotes the sample covariance.

See Theorem 5.2.3 of Anderson (2003) for a proof. Similarly, one can also construct a statistical test for two independent samples using the same framework. Assume that we have two independent sets of Laplacians of dimension $d \times d$, and consider the problem of testing whether these sets have in fact been drawn from the same population. Each sample of Laplacians has the form, L_{in_j} , where $i = 1, \dots, n_j$ for every $j = 1, 2$. The population means are denoted Λ_j , while the sample means of these sets of Laplacians are denoted by \widehat{L}_j . Then, as a direct corollary to Theorem 3, we also have the following asymptotic result.

COROLLARY 3. Assume that every Λ_j does not lie on the boundary of \mathcal{L}'_d , and that $\mathbb{P}[U] > 0$, where U is an open subset of \mathcal{L}'_d , such that $L_j \in U$ for each $j = 1, 2$. Moreover, also assume that $n_j/n \rightarrow p_j$ for every sample, with $n := \sum_{j=1}^2 n_j$ and $0 < p_j < 1$. Then, under the null, $H_0 : \Lambda_1 = \Lambda_2$, we have

$$T_2 := \frac{n_1 n_2}{n_1 + n_2} (\phi(\widehat{L}_1) - \phi(\widehat{L}_2))' \widehat{\Sigma}^{-1} (\phi(\widehat{L}_1) - \phi(\widehat{L}_2)) \rightarrow \chi_p^2,$$

where \widehat{L}_j denotes the j th sample mean, and $\widehat{\Sigma} := (\sum_{j=1}^2 n_j \widehat{\Sigma}_j) / (n_1 + n_2 - 2)$ is the pooled covariance estimate, with the $\widehat{\Sigma}_j$'s denoting the individual covariance matrices of each subsample.

Finally, we can also derive a test statistic T_k to test that $H_0 : \Lambda_1 = \dots = \Lambda_k$ is true. This can be performed using an analogue of Wilks's statistic, by partitioning the variance of interest, using the asymptotic result in Theorem 3.

4.3. *Covariance estimation.* We note that, in order to use any of the above results in a practical setting, we must have knowledge of the covariance matrix $\Sigma = \text{Cov}[\phi(L)]$. It can be expected that we must use a sample-based estimate. However, because the dimension of this matrix is $O(d^2) \times O(d^2)$, and the sample size n is potentially much smaller than $O(d^2)$, the traditional sample covariance $\widehat{\Sigma}$ is likely to be numerically unstable, and is not guaranteed to be positive definite.

Fortunately, the development of estimators of Σ in such low-sample/high-dimension contexts has been an active area of statistical research over the past few years. Typically, borrowing regularization strategies from the field of nonparametric function estimation, optimization of a cost function combining the Frobenius norm or penalized maximum likelihood with a regularization term yields a convex optimization problem that can be solved efficiently. Generally, the choice of a regularization term is linked to the assumed structure of the covariance matrix, for example, assumptions of banding [Bickel and Levina (2008b)] or sparseness [Bickel and Levina (2008a), Cai and Liu (2011), El Karoui (2008)]. There is also a substantial recent literature on the closely related problem of estimating the inverse covariance matrix Σ^{-1} ; see Cai, Liu and Luo (2011) for a recent example and associated citations.

In our context, there is little understanding of how the covariance matrices of the off-diagonal entries of graph Laplacians should behave. Accordingly, as an alternative to the sample covariance, we have adopted a shrinkage estimator due to Schäfer and Strimmer (2005) which is particularly well suited to large data sets. The method of Schäfer and Strimmer (2005) is a popular generic method for high-dimensional covariance estimation, which works through shrinkage toward substructures. This was deemed sufficiently flexible for our purpose.

Moreover, since in finite samples the estimator $\widehat{\Sigma}$ may not necessarily be a positive definite matrix, we have therefore adopted an algorithm due to Higham

(2002) in order to locate the nearest positive definite matrix in the Frobenius norm [see also Cheng and Higham (1998)]. The resulting matrix, say $\tilde{\Sigma}_{PD}$, is then used in place of $\hat{\Sigma}$ in the test statistics described in the previous section.

4.4. *Visualization of differences.* The contribution of each node to the test statistics in the one- and two-sample tests can be visualized through a linear decomposition of the above test statistics. Consider, for example, the one-sample statistic, T_1 . By taking the square root of the inverted covariance matrix, this particular quantity admits the following decomposition: $T_1/n = \hat{\lambda}'\hat{\lambda}$, in which $\hat{\lambda} := \Sigma^{-1/2}(\phi(\hat{L}) - \phi(\Lambda_0))$. Consequently, we can reformulate this one-sample test as a sum of squares, $T_1 = n \sum_{i=1}^p \hat{\lambda}_i^2$, with $p := \binom{d}{2}$. The square of the i th entry of $\hat{\lambda}$ can then serve as an indicator of the contribution of the i th edge to the value of T_1 .

Similarly, this visualization strategy can be extended to the comparison of two groups; that is, the statistic T_2 in Corollary 3 can be represented as $(n_1 + n_2)/(n_1 n_2) T_2 = \sum_{j=1}^2 \hat{\lambda}'_j \hat{\lambda}_j$, where the two p -dimensional vectors, $\hat{\lambda}_1$ and $\hat{\lambda}_2$, are defined as $\hat{\lambda}_j := \hat{\Sigma}^{-1/2}(\phi(\hat{L}_j) - \phi(\hat{L}))$ for $j = 1, 2$. Consequently, we can decompose the weighted version of T_2 as a sum of p terms of the form

$$\frac{n_1 + n_2}{n_1 n_2} T_2 = \sum_{j=1}^2 \hat{\lambda}'_j \hat{\lambda}_j = \sum_{j=1}^2 \sum_{i=1}^p \hat{\lambda}_{ij}^2 = \sum_{i=1}^p \left(\sum_{j=1}^2 \hat{\lambda}_{ij}^2 \right) =: \sum_{i=1}^p \hat{\kappa}_i,$$

in which $\hat{\lambda}_{ij}$ stands for the i th element of $\hat{\lambda}_j$, and $\hat{\kappa}_i := \sum_{j=1}^2 \hat{\lambda}_{ij}^2$. Therefore, as for T_1 , each $\hat{\kappa}_i$ can be treated as the specific contribution of the i th edge to the value of T_2 .

In the sequel, we will plot the $\hat{\kappa}_i$'s in order to provide a fine-grained visualization of the differences between the families of networks under scrutiny. The empirical distribution of the edgewise contributions, $\hat{\kappa}_i$'s, to the estimate of the statistic T_2 will be computed; and we will report all values above a certain threshold. In Figure 4, for instance, we have plotted the edgewise contributions above the 85th and 95th percentiles of the distribution of the κ_i 's. Since each $\hat{\kappa}_i$ is positive, our strategy consists in partitioning T_2 into a sum of positive values, and the plotted $\hat{\kappa}_i$'s in Figure 4 represent the largest such values.

5. Simulation studies. In this empirical study, we evaluate the statistical power of the two-sample test T_2 for Laplacians under different choices of number of vertices and for increasing sample sizes. We simulate network-based data for n subjects in each group, and focus our attention on two-sample experimental designs. Motivated by the neuroimaging application underlying the methodological development just described, the data-generating process relies on (i) the selection of a network topology and the construction of an associated covariance matrix, (ii) the generation of multivariate time series for each network model, and (iii) the construction of subject-specific Laplacians based on the covariance matrices.

5.1. *Network topologies.* In these simulations, we consider two types of network topology, specified through binary matrices, A_1 and A_2 , of order $d \times d$. First, we consider a block-diagonal structure A_1 , which represents the grouping of several vertices into two homogeneous communities,

$$A_1 := \begin{pmatrix} X & R \\ R & Y \end{pmatrix},$$

where X and Y are square matrices of dimensions $\lceil d/2 \rceil$ and $\lfloor d/2 \rfloor$, respectively. The elements of X and Y are given a value of 1 according to independent Bernoulli variates with proportion $p_1 := 4/d$, whereas the elements of R take a value of 1 with a probability of $p_2 := 1/(2d)$. These choices of p_1 and p_2 ensure that the corresponding block models are *sparse* in the sense that their number of edges are proportional to their number of vertices as d grows.

Second, we specify a small-world network structure, A_2 , by constructing a regular network with a ring topology, whose number of edges is taken to be proportional to d , which again enforces sparsity. The edges of this network are then randomly rewired [Watts and Strogatz (1998)]. The choice of N_e —the number of edges—is here motivated by a desire to maintain some level of comparison between the block-diagonal model and the small-world topology. Using such N_e 's, we ensure that both types of networks have approximately the same number of edges. These two families of network topologies are illustrated in Figure 2 for simulated networks of size $d = 50$.

For both of these models, we generated mean covariance matrices, S_{gm} 's, where $g = 1, 2$ denoting the group of subjects, and $m = 1, 2$ denoting the block model and small-world model, respectively. These were constructed using a mixture model based on the binary matrices, A_m 's. The S_{2m} 's were expressed as a function of the S_{1m} 's. For the diagonal elements of the S_{1m} 's,

$$S_{aa,1m} \stackrel{\text{i.i.d.}}{\sim} \exp(\lambda), \quad a = 1, \dots, d,$$

whereas the off-diagonal elements of the S_{1m} 's are constrained by the corresponding off-diagonal elements in the adjacency matrices, A_m 's, as follows:

$$S_{ab,1m} | A_{ab,m} \stackrel{\text{ind.}}{\sim} |A_{ab,m} N(\mu_1, \sigma^2) + (1 - A_{ab,m}) N(\mu_2, \sigma^2)|$$

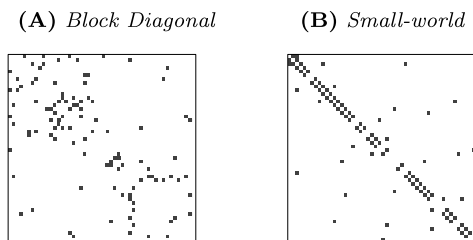


FIG. 2. Simulated matrices over $d = 50$ vertices. In panels (A) and (B), matrices with a block-diagonal structure and a small-world topology are respectively represented.

for every $a \neq b$, and where the parameters of the mixture model are given the following values, $\lambda := 4$, $\mu_1 = 1$, $\mu_2 = 0$ and $\sigma^2 = 0.2$ for all simulation scenarios, thereby producing a high signal-to-noise ratio, permitting to distinguish between the different types of entries in the matrices, S_{1m} 's. Note that none of the simulation scenarios guarantees that the resulting S_{1m} 's are positive definite. Consequently, we projected the resulting matrices to the nearest positive definite matrices in the Frobenius norm, using the method described in Section 4.3. Once the S_{1m} 's were obtained, they were fixed for each scenario, and used to generate the covariance matrix in the second group as follows: $S_{2m} := C(\eta - 1)S_{1m}$, where η controlled the distance between the two population means, which was interpreted as the effect size, and the constant C was set to a small value, $C := 0.03$, throughout the simulations.

5.2. Noise models. Resting-state or default-mode brain networks have been investigated by a large number of researchers in neuroimaging [Beckmann et al. (2005), Thirion et al. (2006)]. The main difficulty in simulating these networks stems from the absence of a prior to produce such resting-state patterns of activities [Kang et al. (2012), Leon et al. (2013)]. For each subject, we here constructed a set of d sequences of T realizations, where d represents the number of ROIs, and T denotes the total number of time points. These sequences of realizations were drawn from a multivariate Gaussian, such that, for every subject, $i = 1, \dots, n$, the random vectors, $X_{itgm} \in \mathbb{R}^d$, were given by

$$X_{itgm} \stackrel{\text{i.i.d.}}{\sim} N_d(0, S_{gm}) \quad \forall t = 1, \dots, T,$$

where $g = 1, 2$ denotes group affiliation, and $m = 1, 2$ denotes the choice of the underlying adjacency matrix: block-diagonal model and small-world model.

5.3. Simulation results. Four main factors were made to vary in this set of simulations. In line with the subsequent real-data analysis, we considered sample sizes of $n = 100, 200, 300$ and 400 per group. This was deemed representative of the number of subjects found in most neuroimaging studies. Second, we varied network sizes, with d taking values $10, 20, 30$ and 40 . This range of network sizes allowed us to identify the effect of network size on the statistical power of our test. Larger dimensions were expected to decrease power.

In each of these scenarios, we computed the statistical power of the two-sample tests using different effect sizes. Here, the effect size was defined with respect to the value of the parameter η . Recall that η controlled the distance between the two population means, such that $S_{2m} := C(\eta - 1)S_{1m}$. For each set of conditions, the simulations were repeated 100 times in order to obtain an empirical estimate of the theoretical power of the two-sample test statistic for Laplacians under these conditions.

The results of these simulations are reported in Figure 3. The power of the two-sample test for Laplacians was found to be empirically well behaved for all the

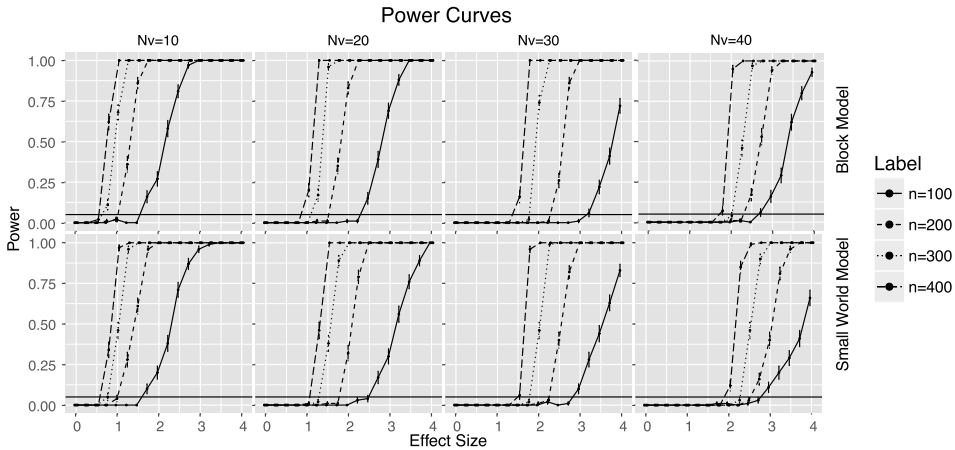


FIG. 3. Power curves for the simulated two-sample tests using the covariance estimation procedure, under a multivariate Gaussian model, with error bars based on one standard error from the mean. The y-axis indicates the probability of rejecting the null hypothesis when it is false, whereas the x-axis is a proxy measure of effect size (see Section 5.3). These results are presented for networks on $d = 10, 20, 30$ and 40 vertices, with different sample sizes and over $T = 50$ time points, and based on 100 iterations per condition with respect to the block (top row) and small-world (bottom row) topologies. A horizontal line has been added to indicate a power of 0.05.

scenarios considered. In particular, this was true for both the block-diagonal and small-world topologies, as illustrated in the first and second row in Figure 3. As expected, the power of the test tended to increase with larger sample sizes, albeit that increase was mitigated by the size of the underlying networks.

6. Analysis of the 1000 FCP data set. Different aspects of the 1000 FCP data set were considered. First, we used a one-sample test for comparing the Laplacian mean to a subsample of the data. We then tested for sex and age differences using the two- and k -sample tests for Laplacians. Finally, we analyzed differences in subnetworks, including the default-mode network (DMN). After excluding subjects for which demographics data were incomplete, we obtained a sample size of $n = 1017$.

6.1. *Inference on full data set.* As described in Section 1.2, the 1000 FCP data provides a unique opportunity for neuroscientists to extract a reference template of human connectivity. We tested the reliability of that template using a one-sample Laplacian test for some random subsample of the data. We computed the reference mean Laplacian over the full FCP sample, which is here treated as a *population parameter*, Λ_0 . This was compared with a large random subsample of 917 subjects, that is, after removing 100 subjects from the original FCP data. We then tested for the null hypothesis that the sample mean, \widehat{L}_1 , was equal to the reference mean Λ_0 . As expected, the test failed to reject the null hypothesis

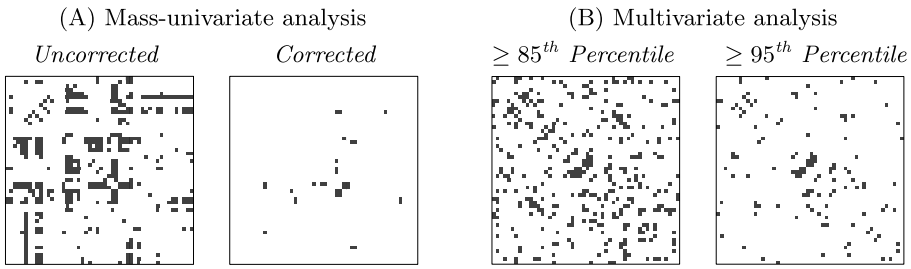


FIG. 4. Comparison of edgewise differences using either a standard mass-univariate analysis or our proposed multivariate method. In panel (A), mass-univariate analyses were conducted to test for local differences in connectivity due to sex in the full FCP data set ($n = 1017$). In each case, $\binom{50}{2}$ tests were performed independently for each of the off-diagonal entries in the Laplacians. The first matrix denotes the entries that were found to be significantly different between the groups at $\alpha = 0.05$, whereas the second matrix represents the significant entries after Bonferroni correction. Black denotes significant entries. In panel (B), the same edgewise effects are reported for our multivariate approach using the visualization method described in Section 4.4. The edges whose contribution to the T_2 -statistic is greater than the 85th and 95th percentiles (of the distribution of such individual contributions) have been plotted in black in the third and fourth panels, respectively.

$[T_1 = 926, df = \binom{50}{2}, p \approx 1]$, since the sample and reference means were drawn from the same population.

The partitioning of the 1000 FCP data set by sex is provided in Figure 1(A). This consists of $n_f = 555$ female and $n_m = 462$ male subjects. We tested whether such sex differences were significant using the two-sample test for Laplacians. The null hypothesis of no group differences was rejected with high probability [$T_2 = 1689.5, df = \binom{50}{2}, p < 0.001$]. These results should be compared with the use of a mass-univariate approach, in which a single hypothesis test is run for each voxel. The significant voxel-level differences detected using a mass-univariate approach for sex is reported in Figure 4.

Subjects in the 1000 FCP database can also be grouped according to age. In Figure 1(B), we have divided the FCP sample into three subgroups of approximately equal sizes, with 386, 297 and 334 subjects, for subjects younger than 22, between 22 and 32, and older than 32, respectively. The k -sample Laplacian test (or Wilks's Lambda) was performed to evaluate the hypothesis stating that these $k = 3$ groups were drawn from the same population. The null hypothesis was also rejected with high probability in this case [$\Lambda = 0.106, n = 1017, df = \binom{40}{2}, p < 0.001$]. (For computational convenience, we here restricted our attention to networks with 40 nodes, which yielded invertible sample covariance matrices for the Wilks's test.)

6.2. Inference on partial data set. The results of the previous section were compared with another analysis based on a small subset of connectomes. The 1000 FCP data set is indeed exceptionally large for the field of neuroimaging. By contrast, most papers using MRI data tend to report results based on smaller data sets,

commonly containing between 20 and 100 subjects. Here, we have replicated the various statistical tests described in the last section for such small sample sizes in order to produce an analysis more reflective of what might be performed by, say, a single lab.

The conclusions of the network-level tests for the different hypotheses of interest were found to be robust to a large decrease in sample size. As for the full data set, sex differences remain close to significance [$T_2 = 836.76$, $df = \binom{40}{2}$, $p = 0.077$], when solely considering 100 female and 100 male subjects. Note, however, that our proposed global test failed to reject the null hypothesis when considering smaller data sets. Indeed, we restricted our attention to smaller subsets of subjects, composed of 20 cases in each group, and such a test did not reject the null hypothesis [$T_2 = 514.96$, $df = \binom{40}{2}$, $p \approx 1$].

These results should be contrasted with the use of a mass-univariate approach. We compared the conclusions of a network-level Laplacian test for sex with the ones of a mass-univariate approach based on 100 female and 100 male subjects. No local differences were here found after correcting for multiple comparisons, and solely one edge out of $\binom{40}{2}$ was found to significantly differ between groups at a threshold of 0.001. This highlights one of the important advantages of using a global test in this context. While the mass-univariate approach fails to detect any sex differences at the local level, our proposed global test, by contrast, had sufficient power to reject the null hypothesis at a global level.

6.3. Default-Mode Network. The Default-Mode Network (DMN) is a widely studied portion of the functional network characterizing brain activity in both humans and animals [Buckner, Andrews-Hanna and Schacter (2008), Greicius et al. (2003)]. This network tends to be active when an individual is not engaged in a cognitive task. The DMN is composed of a set of hubs that include the precuneus, posterior cingulate, medial prefrontal cortex and angular gyri, as well as prefrontal cortices, temporo-parietal junctions, the hippocampi and the parahippocampi. In the parcellation template used in this paper, these regions corresponded to $d = 24$ AAL areas.

We tested for the effect of sex in the full FCP sample by applying the projection method described in equation (3). The hypothesis of no difference between males and females was not rejected for the DMN network [$T_2 = 1128$, $df = \binom{24}{2}$, $p \approx 1$]. (The mean Laplacians for these subnetworks are reported in Figure 5.) This demonstrates that such multivariate methods also tend to lose power when restricted to subnetworks.

7. Discussion. In this paper, we have analyzed a large neuroimaging data set using a novel framework for network-based statistical testing. The development of this framework is grounded in a formal asymptotic theory for network averages, developed within the context of a well-defined notion of the space of graph Laplacians. Importantly, we have showed that using the global tests that result from our

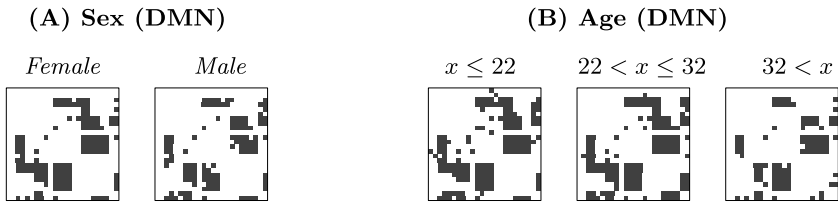


FIG. 5. Descriptive statistics for Default-Mode Network (DMN) in the 1000 FCP data set. In panels (A) and (B), we have provided the mean Laplacians after dividing the sample according to sex and age, respectively. As before, the Laplacians have been binarized with respect to the 75th percentile in the distribution of the FCP database.

framework may provide the researcher with decidedly more statistical power than when using a mass-univariate approach, which is the standard approach in the field.

To the best of our knowledge, we are the first to ascribe a notion of a “space” to the collection of graph Laplacians and to describe the geometrical properties of this space. While we have found it convenient for the purposes of exposition simply to summarize these results in the main body of the paper, and to collect details in the appendices, it is important to note that this initial step is crucial in allowing us to bring to bear recent probabilistic developments in the field of shape analysis to produce our key central limit theorem, upon which the distribution theory for our tests lies. We note too that the framework we offer is quite general and should, therefore, as a result be quite broadly applicable. Nevertheless, this initial work also has various limitations, and furthermore sets the stage for numerous directions for extensions, which we describe briefly below.

7.1. Limitations. It can be expected that there be a trade-off in the performance of our tests between sample size n and the dimension d of the networks in the sample. This expectation is confirmed in our simulations, where one can observe that, for a given sample size n , the rate of type I error increases beyond the nominal rate as d increases. Since our test can be seen to be equivalent to a Hotelling T^2 on the off-diagonal elements of the Laplacians, it follows that sample sizes of order $O(d^2)$ would be required to control for this increase in type I error rate. For the analysis of the full FCP data set, this condition was approximately satisfied, since this data set contains more than 1000 subjects, and we were here comparing networks with 50 vertices. In their current forms, such global statistical tests may therefore be most applicable to very large data sets or to relatively small networks. However, our analysis of the smaller subsets of the FCP data (i.e., mimicking analysis at the level of a single lab) suggests that even at low sample sizes the test is well powered against the alternative of differences in network group averages.

Computationally, the method employed in this paper was also challenging since the application of the Laplacian test required the inversion of a large covariance

matrix. We have here resorted to different methods to facilitate this process, including the use of modern shrinkage estimation techniques [Schäfer and Strimmer (2005)], as well as the modification of the resulting sample covariance matrix estimates in order to force positive definiteness [Cheng and Higham (1998), Higham (2002)]. Practically, however, such methods remain computationally expensive, and may therefore limit the size of the networks that one may wish to consider when using such Laplacian tests.

Finally, observe that the networks of interest in this paper have been constructed using fMRI data. This preliminary step necessitated the estimation of covariance matrices for each subject, and such estimation has not been directly taken into account in the final analysis. Further research may therefore need to adopt a global modeling strategy in which the uncertainty at the first level of the analysis is propagated to the second level, in which we compare groups of subject-specific networks.

7.2. Extensions. In our work here (specifically, as described in Section 3) we show that the “space” of networks—*without any structural constraints*—behaves “nicely” from the mathematical perspective, and therefore we are able to develop a corresponding probability theory and statistical methods for one- and two-sample assessment of network data objects. However, one of the most fundamental results that has emerged from the past 20 years of complex network research is the understanding that real-world networks typically (although not exclusively) in fact tend to possess a handful of quite marked structural characteristics. For example, most networks are relatively sparse, in the sense that the number of edges is on the same order of magnitude as the number of vertices. Other common key properties include heterogeneous degree distributions, cohesive subgraphs (aka communities) and small-world behavior [see Newman (2010), Chapter 8].

The ubiquity of such characteristics in real-world networks has been well established. Importantly, this fact suggests that the appropriate (differential or metric measure) geometry of the “space of all networks”—or, more formally, the space of Laplacians corresponding to such networks—depends both on the constraints imposed on these networks/Laplacians and the geometry chosen for the larger space \mathcal{P} of PSD matrices. In our case, it is natural to choose a Euclidean geometry rather than geometries associated to \mathcal{P} as a homogeneous space. In particular, other choices of network constraints can lead to metric geometry problems embedded inside Riemannian geometry problems. For examples, imposing sparseness on a network or allowing for directed edges lead to nontrivial geometries. The Euclidean average of two sparse networks/matrices need not be sparse, and apart from simple scalings, one expects the set \mathcal{L} of sparse matrices, properly defined, to be a discrete subset of the manifold of positive semi-definite matrices (PSD), and hence far from convex. Thus, it is natural to define the average of two sparse matrices to be the sparse matrix closest to the Euclidean average, but this may be computationally unappealing. Moreover, the Riemannian measure on PSD

does not determine a measure on \mathcal{L} , and so computing Fréchet means becomes problematic. Of course, one can impose a uniform distribution on \mathcal{L} , but this risks losing all geometric relations between \mathcal{L} and PSD. Hence, there are a variety of open problems to be studied examining the implications of network structural constraints on the space \mathcal{L} .

Furthermore, since the asymptotic theory we exploit from shape analysis relies heavily on the topological and geometrical properties of the space within which they are brought to bear, we can expect that different network constraints will require different levels of effort in producing central limit theorems. More precisely, while a general asymptotic distribution theory for Fréchet means in metric spaces has recently been derived by [Bhattacharya and Lin \(2017\)](#), this theory requires that a number of conditions be satisfied, the verification of which can be expected to become increasingly difficult as the geometry of the space becomes complicated. Thus, accompanying the various extensions in geometry described above are likely to be corresponding challenges in probability theory and shape analysis. Some progress in this direction has been spearheaded by [Bhattacharya et al. \(2011\)](#) and [Hotz et al. \(2013\)](#), who have considered stratified spaces and sticky CLTs for open books, respectively. Moreover, similar data object analyses have been conducted using phylogenetic trees [[Skwerer et al. \(2014\)](#)]. In object data analysis, the approach adopted in this paper would be regarded as extrinsic, in the sense that it embeds the manifold of interest in an ambient space. Further research may also investigate intrinsic approaches to study the set of graph Laplacians.

Finally, while the 1000 FCP data set is unique in its magnitude and richness, which in turn has allowed us to pose and answer a good number of questions relevant to neuroscience in the analyses using our proposed testing framework, there remains much additional empirical work to be done applying our methods in order to more fully establish both their capabilities and their limitations. We would anticipate that with the recently started BRAIN initiative, and other endeavors like it, that within five years there will be a plethora of databases of network-based objects in neuroscience, providing more than ample motivation not only for the further testing of methods like the ones we have proposed here, but also for extending other tools from classical statistics to network data.

Acknowledgments. The data from the 1000 Functional Connectome Project was accessed through the International Neuroimaging Data-sharing Initiative (INDI), which was designed for unrestricted data-sharing via the Neuroimaging Informatics Tool and Resources Clearinghouse (NITRC). We are indebted to Sean Markan, Lizhen Lin, Emily Stephen and Heather Shappell for useful suggestions and discussion. We are also very grateful for the comments that we have received from three anonymous referees, one associate editor and one of the AOAS area editors.

SUPPLEMENTARY MATERIAL

Proofs of theorems (DOI: [10.1214/16-AOAS1015SUPP](https://doi.org/10.1214/16-AOAS1015SUPP); .pdf). Therein we here provide detailed proofs of the main results in this paper.

REFERENCES

- ACHARD, S., SALVADOR, R., WHITCHER, B., SUCKLING, J. and BULLMORE, E. (2006). A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *J. Neurosci.* **26** 63–72.
- ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. Wiley, Hoboken, NJ. [MR1990662](#)
- ARSIGNY, V., FILLARD, P., PENNEC, X. and AYACHE, N. (2007). Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM J. Matrix Anal. Appl.* **29** 328–347. [MR2288028](#)
- AYDIN, B., PATAKI, G., WANG, H., BULLITT, E. and MARRON, J. S. (2009). A principal component analysis for trees. *Ann. Appl. Stat.* **3** 1597–1615. [MR2752149](#)
- BARDEN, D., LE, H. and OWEN, M. (2013). Central limit theorems for Fréchet means in the space of phylogenetic trees. *Electron. J. Probab.* **18** 1–25. [MR3035753](#)
- BECKMANN, C. F., DELUCA, M., DEVLIN, J. T. and SMITH, S. M. (2005). Investigations into resting-state connectivity using independent component analysis. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.* **360** 1001–1013.
- BHATIA, R. (1997). *Matrix Analysis. Graduate Texts in Mathematics* **169**. Springer, New York. [MR1477662](#)
- BHATIA, R. (2007). *Positive Definite Matrices*. Princeton Univ. Press, Princeton, NJ. [MR2284176](#)
- BHATTACHARYA, A. and BHATTACHARYA, R. (2012). *Nonparametric Inference on Manifolds with Applications to Shape Spaces*. Cambridge Univ. Press, New York.
- BHATTACHARYA, R. and LIN, L. (2017). Omnibus CLTs for Fréchet means and nonparametric inference on non-Euclidean spaces. *Proc. Amer. Math. Soc.* **145** 413–428. [MR3565392](#)
- BHATTACHARYA, R. and PATRANGENARU, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. I. *Ann. Statist.* **31** 1–29. [MR1962498](#)
- BHATTACHARYA, R. and PATRANGENARU, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds. II. *Ann. Statist.* **33** 1225–1259. [MR2195634](#)
- BHATTACHARYA, R., BUIBAS, M., DRYDEN, I., ELLINGSON, L., GROISSER, D., HENDRIKS, H., HUCKEMANN, S., LE, H., LIU, X. and MARRON, J. (2011). Extrinsic data analysis on sample spaces with a manifold stratification. In *Advances in Mathematics, Invited Contributions at the Seventh Congress of Romanian Mathematicians, Brasov* 148–156.
- BICKEL, P. J. and LEVINA, E. (2008a). Covariance regularization by thresholding. *Ann. Statist.* **36** 2577–2604. [MR2485008](#)
- BICKEL, P. J. and LEVINA, E. (2008b). Regularized estimation of large covariance matrices. *Ann. Statist.* **36** 199–227. [MR2387969](#)
- BILLERA, L. J., HOLMES, S. P. and VOGTMANN, K. (2001). Geometry of the space of phylogenetic trees. *Adv. in Appl. Math.* **27** 733–767. [MR1867931](#)
- BISWAL, B. B., MENNES, M., ZUO, X.-N., GOHEL, S. and KELLY, C. et al. (2010). Toward discovery science of human brain function. *Proc. Natl. Acad. Sci. USA* **107** 4734–4739.
- BONNABEL, S. and SEPULCHRE, R. (2009). Riemannian metric and geometric mean for positive semidefinite matrices of fixed rank. *SIAM J. Matrix Anal. Appl.* **31** 1055–1070.
- BOOKSTEIN, F. (1978). *The Measurement of Biological Shape and Shape Change*. Springer, London.
- BUCKNER, R. L., ANDREWS-HANNA, J. R. and SCHACTER, D. L. (2008). The brain's default network: Anatomy, function and relevance to disease. *Ann. N.Y. Acad. Sci.* **1124** 1–38.

- BULLMORE, E. and SPORNS, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nat. Rev., Neurosci.* **10** 186–198.
- BULLMORE, E. and SPORNS, O. (2012). The economy of brain network organization. *Nat. Rev., Neurosci.* **13** 336–349.
- CAI, T. and LIU, W. (2011). Adaptive thresholding for sparse covariance matrix estimation. *J. Amer. Statist. Assoc.* **106** 672–684.
- CAI, T., LIU, W. and LUO, X. (2011). A constrained L_1 minimization approach to sparse precision matrix estimation. *J. Amer. Statist. Assoc.* **106** 594–607. [MR2847973](#)
- CHAVEL, I. (1984). *Eigenvalues in Riemannian Geometry. Pure and Applied Mathematics* **115**. Academic Press, Inc., Orlando, FL. Including a chapter by Burton Randol. With an appendix by Jozef Dodziuk. [MR0768584](#)
- CHENG, S. H. and HIGHAM, N. J. (1998). A modified Cholesky algorithm based on a symmetric indefinite factorization. *SIAM J. Matrix Anal. Appl.* **19** 1097–1110. [MR1636528](#)
- CHUNG, F. R. K. (1997). *Spectral Graph Theory. CBMS Regional Conference Series in Mathematics* **92**. Amer. Math. Soc., Providence, RI. [MR1421568](#)
- DRYDEN, I. L., KOLOYDENKO, A. and ZHOU, D. (2009). Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *Ann. Appl. Stat.* **3** 1102–1123.
- EL KAROUI, N. (2008). Operator norm consistent estimation of large-dimensional sparse covariance matrices. *Ann. Statist.* **36** 2717–2756. [MR2485011](#)
- ELLEGREN, H. and PARSCH, J. (2007). The evolution of sex-biased genes and sex-biased gene expression. *Nat. Rev., Genet.* **8** 689–698.
- FISHER, R. (1953). Dispersion on a sphere. *Proc. R. Soc. Lond. Ser. A* **217** 295–305. [MR0056866](#)
- FISHER, N. I., LEWIS, T. and EMBLETON, B. J. J. (1987). *Statistical Analysis of Spherical Data*. Cambridge Univ. Press, Cambridge. [MR0899958](#)
- FRÉCHET, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. Henri Poincaré* **10** 215–310. [MR0027464](#)
- FU, Y. and MA, Y. (2013). *Graph Embedding for Pattern Analysis*. Springer, New York.
- GINESTET, C. E., FOURNEL, A. P. and SIMMONS, A. (2014). Statistical network analysis for functional MRI: Summary networks and group comparisons. *Front. Comput. Neurosci.* **8** Art. ID 51.
- GINESTET, C. E. and SIMMONS, A. (2011). Statistical parametric network analysis of functional connectivity dynamics during a working memory task. *NeuroImage* **5** 688–704.
- GINESTET, C. E., LI, J., BALACHANDRAN, P., ROSENBERG, P. and KOLACZYK, E. D. (2017). Supplement to “Hypothesis testing for network data in functional neuroimaging.” DOI:[10.1214/16-AOAS1015SUPP](#).
- GREICIUS, M. D., KRASNOW, B., REISS, A. L. and MENON, V. (2003). Functional connectivity in the resting brain: A network analysis of the default mode hypothesis. *Proc. Natl. Acad. Sci. USA* **100** 253–258.
- GROMOV, M. (2007). *Metric Structures for Riemannian and Non-Riemannian Spaces*, English ed. Birkhäuser, Inc., Boston, MA. [MR2307192](#)
- HIGHAM, N. J. (2002). Computing the nearest correlation matrix: A problem from finance. *IMA J. Numer. Anal.* **22** 329–343.
- HOTZ, T., HUCKEMANN, S., LE, H., MARRON, J. S., MATTINGLY, J. C., MILLER, E., NOLEN, J., OWEN, M., PATRANGENARU, V. and SKWERER, S. (2013). Sticky central limit theorems on open books. *Ann. Appl. Probab.* **23** 2238–2258. [MR3127934](#)
- KANG, H., OMBAO, H., LINKLETTER, C., LONG, N. and BADRE, D. (2012). Spatio-spectral mixed-effects model for functional magnetic resonance imaging data. *J. Amer. Statist. Assoc.* **107** 568–577. [MR2980068](#)
- KENDALL, D. G. (1977). The diffusion of shape. *Adv. in Appl. Probab.* **9** 428–430.
- KENDALL, D. G. (1984). Shape manifolds, Procrustean metrics, and complex projective spaces. *Bull. Lond. Math. Soc.* **16** 81–121. [MR0737237](#)

- KENDALL, W. S. and LE, H. (2011). Limit theorems for empirical Fréchet means of independent and non-identically distributed manifold-valued random variables. *Braz. J. Probab. Stat.* **25** 323–352.
- KRISHNAMACHARI, R. T. and VARANASI, M. K. (2013). On the geometry and quantization of manifolds of positive semi-definite matrices. *IEEE Trans. Signal Process.* **61** 4587–4599. [MR3096701](#)
- LE, H. (2001). Locating Fréchet means with application to shape spaces. *Adv. in Appl. Probab.* **33** 324–338. [MR1842295](#)
- LE, H. and KUME, A. (2000). The Fréchet mean shape and the shape of the means. *Adv. in Appl. Probab.* **32** 101–113.
- LEE, J. (2006). *Introduction to Smooth Manifolds*. Springer, London.
- LEON, P. S., KNOCK, S. A., WOODMAN, M. M., DOMIDE, L., MERSMANN, J., MCINTOSH, A. R. and JIRSA, V. (2013). The Virtual Brain: A simulator of primate brain network dynamics. *Front. Neuroinform.* **7** Art. ID 10.
- LINIAL, N. (2002). Finite metric spaces: Combinatorics, geometry and algorithms. In *Proceedings of the Eighteenth Annual Symposium on Computational Geometry* 63.
- LINIAL, N., LONDON, E. and RABINOVICH, Y. (1995). The geometry of graphs and some of its algorithmic applications. *Combinatorica* **15** 215–245.
- MARDIA, K. V. and JUPP, P. E. (2000). *Directional Statistics*. Wiley, Cichester.
- MCEWEN, B. S. (1999). Permanence of brain sex differences and structural plasticity of the adult brain. *Proc. Natl. Acad. Sci. USA* **96** 7128–7130.
- MICHELOYANNIS, S., VOURKAS, M., TSIRKA, V., KARAKONSTANTAKI, E., KANATSOULI, K. and STAM, C. J. (2009). The influence of ageing on complex brain networks: A graph theoretical analysis. *Hum. Brain Mapp.* **30** 200–208.
- MOAKHER, M. (2005). A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM J. Matrix Anal. Appl.* **26** 735–747. [MR2137480](#)
- MOAKHER, M. and ZÉRAÏ, M. (2011). The Riemannian geometry of the space of positive-definite matrices and its application to the regularization of positive-definite matrix-valued data. *J. Math. Imaging Vision* **40** 171–187. [MR2782125](#)
- NEWMAN, M. E. J. (2010). *Networks: An Introduction*. Oxford Univ. Press, Oxford. [MR2676073](#)
- PACHOU, E., VOURKAS, M., SIMOS, P., SMIT, D., STAM, C., TSIRKA, V. and MICHELOYANNIS, S. (2008). Working memory in schizophrenia: An EEG study using power spectrum and coherence analysis to estimate cortical activation and network behavior. *Brain Topogr.* **21** 128–137.
- SCHÄFER, J. and STRIMMER, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Stat. Appl. Genet. Mol. Biol.* **4** Art. ID 32.
- SKWERER, S., BULLITT, E., HUCKEMANN, S., MILLER, E., OGUZ, I., OWEN, M., PATRANGENARU, V., PROVAN, S. and MARRON, J. (2014). Tree-oriented analysis of brain artery structure. *J. Math. Imaging Vision* **50** 126–143.
- THIRION, B., FLANDIN, G., PINEL, P., ROCHE, A., CIUCIU, P. and POLINE, J.-B. (2006). Dealing with the shortcomings of spatial normalization: Multi-subject parcellation of fMRI datasets. *Hum. Brain Mapp.* **27** 678–693.
- TOMASI, D. and VOLKOW, N. D. (2010). Functional connectivity density mapping. *Proc. Natl. Acad. Sci. USA* **107** 9885–9890.
- TOMASI, D. and VOLKOW, N. D. (2011). Gender differences in brain functional connectivity density. *Hum. Brain Mapp.* **33** 849–860.
- TZOURIO-MAZOYER, N., LANDEAU, B., PAPATHANASSIOU, D., CRIVELLO, F., ETARD, O., DELCROIX, N., MAZOYER, B. and JOLIOT, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MINI MRI single-subject brain. *NeuroImage* **15** 273–289.
- WANG, H. and MARRON, J. (2007). Object oriented data analysis: Sets of trees. *Ann. Statist.* **35** 1849–1873. [MR2363955](#)

- WANG, J., WANG, L., ZANG, Y., YANG, H., TANG, H., GONG, Q., CHEN, Z., ZHU, C. and HE, Y. (2009). Parcellation-dependent small-world brain functional networks: A resting-state fMRI study. *Hum. Brain Mapp.* **30** 1511–1523.
- WATSON, G. S. (1983). *Statistics on Spheres*. University of Arkansas Lecture Notes in the Mathematical Sciences **6**. Wiley, New York. MR0709262
- WATTS, D. J. and STROGATZ, S. H. (1998). Collective dynamics of ‘small-world’ networks. *Nature* **393** 440–442.
- XIA, C. (2013). *Eigenvalues in Riemannian Geometry*. IMPA, Rio de Janeiro.
- YAN, S., XU, D., ZHANG, B., ZHANG, H.-J., YANG, Q. and LIN, S. (2007). Graph embedding and extensions: A general framework for dimensionality reduction. *IEEE Trans. Pattern Anal. Mach. Intell.* **29** 40–51.
- YAN, C.-G., CRADDOCK, R. C., ZUO, X.-N., ZANG, Y.-F. and MILHAM, M. P. (2013). Standardizing the intrinsic brain: Towards robust measurement of inter-individual variation in 1000 functional connectomes. *NeuroImage* **80** 246–262.
- ZIEZOLD, H. (1977). On expected figures and a strong law of large numbers for random elements in quasi-metric spaces. In *Transactions of the Seventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes and of the 1974 European Meeting of Statisticians*.
- ZUO, X.-N., EHMKE, R., MENNES, M., IMPERATI, D., CASTELLANOS, F. X., SPORNS, O. and MILHAM, M. P. (2012). Network centrality in the human functional connectome. *Cereb. Cortex* **22** 1862–1875.

C. E. GINESTET
 DEPARTMENT OF BIostatISTICS
 AND HEALTH INFORMATICS
 INSTITUTE OF PSYCHIATRY,
 PSYCHOLOGY AND NEUROSCIENCE
 KING’S COLLEGE LONDON
 LONDON SE5 8AF
 UNITED KINGDOM
 E-MAIL: cedric.ginestet@kcl.ac.uk

J. LI
 P. BALACHANDRAN
 S. ROSENBERG
 E. D. KOLACZYK
 DEPARTMENT OF MATHEMATICS
 AND STATISTICS
 BOSTON UNIVERSITY
 BOSTON, MASSACHUSETTS 02215
 USA
 E-MAIL: junli14@bu.edu
prakashb@bu.edu
sr@bu.edu
kolaczyk@bu.edu