NeuroImage 108 (2015) 251–264

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Improved statistical evaluation of group differences in connectomes by screening–filtering strategy with application to study maturation of brain connections between childhood and adolescence

Djalel-Eddine Meskaldji [a,b,*], Lana Vasung [c], David Romascano [d], Jean-Philippe Thiran [d,e], Patric Hagmann [e], Stephan Morgenthaler [f], Dimitri Van De Ville [a,b]

[a] Institute of Bioengineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
[b] Department of Radiology and Medical Informatics, University of Geneva, Geneva, Switzerland
[c] Division of Development and Growth, Department of Pediatrics, University of Geneva, Geneva, Switzerland
[d] LTS5, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
[e] Department of Radiology, University Hospital Center and University of Lausanne (CHUV-UNIL), Lausanne, Switzerland
[f] Applied Statistics, Institute of Mathematics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

ARTICLE INFO

Article history:
Accepted 29 November 2014
Available online 9 December 2014

Keywords:
Brain connectivity
Complex networks
Multiple testing
Family-wise error rate (FWER)
False discovery rate FDR
Fronto-limbic circuitry
Adolescence

ABSTRACT

Detecting local differences between groups of connectomes is a great challenge in neuroimaging, because the large number of tests that have to be performed and the impact on multiplicity correction. Any available information should be exploited to increase the power of detecting true between-group effects. We present an adaptive strategy that exploits the data structure and the prior information concerning positive dependence between nodes and connections, without relying on strong assumptions. As a first step, we decompose the brain network, i.e., the connectome, into subnetworks and we apply a screening at the subnetwork level. The subnetworks are defined either according to prior knowledge or by applying a data driven algorithm. Given the results of the screening step, a filtering is performed to seek real differences at the node/connection level. The proposed strategy could be used to strongly control either the family-wise error rate or the false discovery rate. We show by means of different simulations the benefit of the proposed strategy, and we present a real application of comparing connectomes of preschool children and adolescents.

Introduction

The study of brain connectivity has become an important aspect of neuroscience as it can help to understand brain organization and function (Fornito et al., 2013; Sporns, 2011). Moreover, the metrics of brain connectivity, assessed through neuroimaging methods, have been recognized as an important marker indicating the level of brain maturation or psychopathology. Through recent innovations in medical imaging and image analysis, the determination of interregional brain connectivity became feasible. Different types of connectivity can be obtained depending on the imaging modality and measure of connectivity, e.g., structural connectivity from diffusion-weighted MRI and fiber tracking (Cammoun et al., 2012; Hagmann et al., 2008), or functional connectivity from functional MRI and statistical dependence on time (Smith et al., 2013; Friston, 2011; van den Heuvel and Hulsoff-Pol, 2010; Achard et al., 2006).

Global brain connectivity can be modeled by a network (a weighted graph) called *connectome* (Sporns et al., 2005), where the N nodes stand for brain regions of interest (ROIs), and each edge weight characterizes a measure of connectivity between pairs of ROIs.

Investigating differences in connectivity between distinct populations based on connectivity matrices is attractive, but also comes with a certain number of problems (Fornito et al., 2013; Varoquaux and Craddock, 2013), among them, the high number of multiple comparisons.

Effectively, when the comparison between brain networks are studied at the level of nodes (vertices) ($\mathcal{O}(N)$) that represent brain ROIs, or connections (edges) ($\mathcal{O}(N^2)$) that link brain ROIs, a huge number of tests have to be performed on the same data, especially, in the case of testing at the level of connections, in which the number of tests basically grows quadratically with the number of nodes. If the multiplicity of tests is ignored, the risk of committing false discoveries increases. As a consequence, erroneous conclusions are frequently drawn (Meskaldji et al., 2013a). On the other hand, considering multiplicity could dramatically decrease the chance of detecting real between-group effects. This is a fact that is commonly reported by researchers especially when the conventional Bonferroni procedure is used for the multiplicity correction

* Corresponding author at: Department of Radiology and Medical Informatics, University of Geneva, Geneva, Switzerland.
E-mail address: djalel.meskaldji@epfl.ch (D.-E. Meskaldji).

http://dx.doi.org/10.1016/j.neuroimage.2014.11.059
1053-8119/

and the strong control on the number of false discoveries is exerted. For example, if $N = 100$ nodes, the Bonferroni threshold for significant p-values is of order 10^{-4} when testing at the level of nodes and 10^{-6} when testing at the level of connections.

Depending on the field of application and the nature of the data, many strategies have been adopted in order to face the multiplicity challenge in the presence of positively correlated test statistics. These strategies consist in exploiting the data structure and positive dependence that could be present between tests. This can have an important impact on the power of detecting true alternatives. For example, this concept was adopted in the widely recognized software package for analyzing fMRI data, the Statistical Parametric Mapping (SPM) (Friston et al., 1995; Frackowiak et al., 1997) and in its extensions such as wavelet extensions (Van De Ville et al., 2004, 2007), based on the idea that voxels of a neurological type belonging to a unique anatomical region will usually exhibit positively correlated behavior (Penny and Friston, 2003; Genovese et al., 1999). In this case, the data is supposed to be smooth and follow a multi-dimensional Gaussian distribution. For this reason, a smoothing has to be applied to the data (Nichols and Hayasaka, 2003). A permutation approach is performed to define active clusters.

The same concept has been followed to derive specific statistical methods in the brain connectivity context. Zalesky et al. (2010) proposed the network based statistic (NBS) as a method to correct for the *FWER* (the probability of having at least one false positive connection), in the framework of multiple testing applied to the brain network connections. The method relies on a first identification of connected components (in the graph theoretical sense), by thresholding the set of p-values at an arbitrary threshold. An iterative procedure based on permutation testing allows thereafter identifying connected components that carry a between-group effect. These methods have, however, some limitations. First, the inference is obtained at the level of connected components and only exerts a weak FWER control, that is, once a component is declared to be significantly different, nothing could be said at the level of individual nodes or connections belonging to the component. In other words, the type I error metric controlled at the level of nodes/connections is unknown. Second, the results strongly depend on the arbitrary choice of the threshold. The same can be said about the spatial pairwise clustering (SPC), proposed by the same authors, where the definition of components is based on geometrical distance in addition to the connectedness in the graph theoretical sense (Zalesky et al., 2012).

It is commonly admitted that most mental diseases or cognitive trait exhibit changes not in the entire brain uniformly, but rather specific in functional systems or brain regions and this to a different extent. Meskaldji et al. (2011a) proposed an adaptive strategy that exploits the network structure of the brain connectivity by considering brain subnetworks, which results in reducing the number of tests and a considerable improvement in power. The strategy was applied to detect differences in both structural and functional brain connectivities (Owen et al., 2013a, 2013b; Meskaldji et al., 2011a). However, besides the generality of this strategy in terms of summary statistics that could be used, and in terms of the diversity of the brain decomposition methods that could be applied, it suffers from the same drawback as the NBS and the SPC, that is, nothing could be said concerning the statistical evidence of nodes and connections that constitute the significant subnetworks. Nevertheless, the subnetworks could be chosen as small as possible to obtain statistical evidence at finer scales. We will give throughout this paper some highlights on the differences between these weak control methods.

The question that we will investigate in this paper is to go beyond the cluster/subnetwork level and investigate the differences at the single node/connection level. Inspired by Benjamini and Heller (2007), we propose a screening–filtering strategy that exploits the data structure and positive dependence that could exist between connections/nodes. The advantage of the proposed strategy is that it exerts a strong control of type I error rates under weak assumptions (i.e., weaker than assumptions needed by SPM, NBS and SPC). We study the performance of the screening–filtering approach on simulated networks and on structural brain connectivity matrices. In particular, we examine the influence of the network decomposition and the screening threshold on the statistical inference. We also discuss the conceptual differences between our proposed strategy and some of existing methods in the literature.

As far as we know, this method is the first adaptive strategy that guarantees the strong control of type I error rate at the level of nodes or connections. For this reason, the performances of the proposed strategy will only be compared to the standard node/connection-wise inference, that is, methods that exert a strong control, but do not consider neither data structure nor positive dependence between tests.

The paper is organized as follows. We first give the general processing pipeline and the mathematical formulation of the screening–filtering approach. Then, we show by simulations, the benefit of using the proposed strategy. Finally, we present a practical application on real data, from children and adolescents, which consists in comparing structural human brain connectomes between these populations.

Methods

We present in this section the different steps of local procedures that exert a strong false positives control. In particular, we outline two strategies: the standard methods and the screening and filtering methods.

Local network-based measures

Since the imaging measures of connectivity can be used to model the brain as a network, it is worth to locally compare populations not only cell by cell of the connectivity matrices, but also by estimating the network measures that characterize the topological properties of the brain network (Fornito et al., 2013; Meskaldji et al., 2013a; Bassett et al., 2008). The combination of the local and the global inferences gives a better understanding of the network organization (Meskaldji et al., 2013a). In this paper, we focus on the local measures. Sporns (2011); Rubinov and Sporns (2010) among others are good sources for a comprehensive list of important measures with their interpretations in the brain connectivity context.

For non-homogeneous populations, it is strongly recommended to correct for covariables such as the age or the gender of the subjects, by taking the residuals of a regression as the new observations (Meskaldji et al., 2013a).

This step ends up with a vector of local observations for each node/connection and for each subject.

Testing and p-value computation

Let us assume that the aim of a brain connectivity study is to compare different groups of connectomes. Comparing two populations at the level of nodes or connections or any local unit that we call *atom*, usually consists in performing a (univariate or multivariate) two-sample test for each node/connection (Meskaldji and Van De Ville, 2014). When more than two groups are compared, an analysis of variance (ANOVA) is performed with a predefined contrast. This ends up with M p-values, where M is of the order ($\mathcal{O}(N)$) when testing at the level of nodes, and M goes like ($\mathcal{O}(N^2)$) when testing at the level of connections.

Let p_j denotes the p-value of atom $j = 1, \ldots, M$. The standard method (SM) consists in performing a multiple testing procedure to the set of p-values to control a type I error metric. For example, one could apply the Bonferroni procedure to the p-values $\{p_1, \ldots, p_M\}$ by declaring