Journal of Neuroscience Methods 291 (2017) 61–68

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

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

Research Paper

Comparing brain graphs in which nodes are regions of interest or independent components: A simulation study

Qingbao Yu[a,*], Yuhui Du[a,b], Jiayu Chen[a], Hao He[a,c], Jing Sui[a,d,e], Godfrey Pearlson[f,g,h], Vince D. Calhoun[a,c,g,*]

[a] The Mind Research Network, Albuquerque, NM, 87106, USA
[b] School of Computer and Information Technology, Shanxi University, Taiyuan, 030006, China
[c] Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM, 87106, USA
[d] Brainnetome Center and National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Science, Beijing, 100190, China
[e] CAS Center for Excellence in Brain Science and Intelligence Technology, University of Chinese Academy of Sciences in Beijing, 100049, China
[f] Olin Neuropsychiatry Research Center, Hartford, CT, 06106, USA
[g] Department of Psychiatry, Yale University, New Haven, CT, 06520, USA
[h] Department of Neuroscience, Yale University, New Haven, CT, 06520, USA

HIGHLIGHTS

- Graphs with different nodes are compared with ground truth in simulated fMRI data.
- Graphs with ICA nodes more accurately represent the ground truth.
- It is more appropriate to define nodes using ICA rather than ROI in fMRI data.

ARTICLE INFO

Article history:
Received 4 April 2017
Received in revised form 2 July 2017
Accepted 8 August 2017
Available online 12 August 2017

Keywords:
Brain graph
ROI
ICA
Simulation
Ground truth

ABSTRACT

Background: A key challenge in building a brain graph using fMRI data is how to define the nodes. Spatial brain components estimated by independent components analysis (ICA) and regions of interest (ROIs) determined by brain atlas are two popular methods to define nodes in brain graphs. It is difficult to evaluate which method is better in real fMRI data.
New method: Here we perform a simulation study and evaluate the accuracies of a few graph metrics in graphs with nodes of ICA components, ROIs, or modified ROIs in four simulation scenarios.
Results: Graph measures with ICA nodes are more accurate than graphs with ROI nodes in all cases. Graph measures with modified ROI nodes are modulated by artifacts. The correlations of graph metrics across subjects between graphs with ICA nodes and ground truth are higher than the correlations between graphs with ROI nodes and ground truth in scenarios with large overlapped spatial sources. Moreover, moving the location of ROIs would largely decrease the correlations in all scenarios.
Comparison with existing method (s): Evaluating graphs with different nodes is promising in simulated data rather than real data because different scenarios can be simulated and measures of different graphs can be compared with a known ground truth.
Conclusion: Since ROIs defined using brain atlas may not correspond well to real functional boundaries, overall findings of this work suggest that it is more appropriate to define nodes using data-driven ICA than ROI approaches in real fMRI data.

1. Introduction

The activity of different human brain areas is correlated rather than independent (Friston, 2011). The brain performs like a complex, interconnected network even during the resting state (Bassett and Bullmore, 2006; Bassett and Gazzaniga, 2011). Functional Magnetic Resonance Imaging (fMRI) is a powerful tool for assessing

* Corresponding authors at: The Mind Research Network, Albuquerque, NM, 87106, USA.
E-mail addresses: qyu@mrn.org (Q. Yu), vcalhoun@unm.edu (V.D. Calhoun).

http://dx.doi.org/10.1016/j.jneumeth.2017.08.007
0165-0270/

functional connectivity of brain networks. Graph theory based analysis provides a mechanism for quantitatively characterize the architecture of these brain networks and is a popular technique to explore human brain fMRI data in health, disease, development, and aging (Bassett et al., 2012; Bassett et al., 2011; Betzel and Bassett, 2016; Cao et al., 2015; Contreras et al., 2015; Fornito and Bullmore, 2015; Fornito et al., 2015; Fornito et al., 2012; Stam and Reijneveld, 2007; Yu et al., 2012).

When performing graph theory based analyses in fMRI data, the first step is typically to identify a set of functional entities that are represented as a vertex set. Each element of this set is called a node. The connections or "edges" between these nodes are then estimated usually by computing the correlation between time courses of each pair of defined nodes (Butts, 2009). Despite exciting advances in studying functional brain connectivity using graph theory based analysis, it remains a challenge to define the nodes when building a brain network in fMRI data. An ideal node definition for building an fMRI brain graph should define functionally homogeneous nodes, represent functional heterogeneity across nodes, and account for spatial relationships (Fornito et al., 2013). The method of defining a brain node varies considerably in the literature (Ribeiro de Paula et al., 2017; Stanley et al., 2013). In fMRI studies, nodes are often defined as spatial regions of interest (ROIs) in which anatomical approaches utilize atlases (e.g. automated anatomical labeling, AAL template) to define the nodes based on brain structure. Alternatively, independent component analysis (ICA) can be run to detect independent components (ICs, spatial brain maps), which can be considered as graph nodes (He et al., 2016; Smith, 2012; Smith et al., 2011; Yu et al., 2015; Yu et al., 2011a,b ; Yu et al., 2013a,b; Yu et al., 2011b; Yu et al., 2016). While the "correct" method for defining the brain nodes remains an open question that deserves further extensive research (Stanley et al., 2013).

Previous work has shown that different approaches of node definition may significantly modulate the quantitative measures of graph metrics in the brain network. Although a few studies have reviewed or compared the graph measures of brain networks in which nodes are defined using different methods (de Reus and van den Heuvel, 2013; Fornito et al., 2013; Rajtmajer et al., 2015; Shirer et al., 2012; Smith et al., 2011; Stanley et al., 2013), no study directly compared the brain graphs constructed by ROI versus ICA methods. It is difficult to evaluate which method is better in real fMRI data sets. A promising approach to evaluate which method is "correct" for defining the nodes when building a brain graph is to use simulated data in which different scenarios can be estimated and captured measures of graphs with different nodes can be compared to a known ground truth. To this end, we perform a simulation analysis in this work. Simulated fMRI data are generated by SimTB (http://mialab.mrn.org/software/simtb/index.html) (Allen et al., 2012; Erhardt et al., 2012). Graph metrics of graphs with ROI or ICA nodes are compared to the ground truth. The aim of this study is to examine which graph is more accurately represent the ground truth. Since ICA is a data driven method and previous studies have shown the advantages of data driven method for defining nodes in brain networks, we predict ICA method would perform better than ROI method.

2. Materials and methods

2.1. SimTB

Simulated data are generated with the MATLAB (https://www.mathworks.com/) toolbox, SimTB (Erhardt et al., 2012), which is developed by our group. The SimTB implements a data generation model consistent with spatiotemporal separability, that is, data can be expressed as the product of time courses (TCs) and spatial maps (SMs). Specifically, for each subject, i = 1, 2, 3, ... M, it is assumed there are up to C sources or components, each consisting of a SM, activation TC, and an amplitude. The no-noise data are a linear combination of amplitude-scaled and baseline-shifted TC and SM sources,

$$\mathbf{Y}_i^{nn} = \left[\mathbf{R}_i diag\left(\mathbf{g}_i\right)\mathbf{S}_i + \mathbf{J}_T^V\right] \odot \mathrm{b}_i\mathbf{J}_T^1\mathbf{u}^{\mathrm{T}} \quad (1)$$

Where $\mathbf{Y}_i^{nn}$ is the time-by-voxel (T-by-V) no-noise data for subject i, $\mathbf{R}_i$ is a matrix of C column vectors of TCs, $\mathbf{S}_i$ is a matrix of C row vectors of SMs, $\mathbf{g}_i$ is a vector of C source amplitudes defined as a percent signal change of the baseline, b_i is a baseline intensity scalar, $\mathbf{u}$ is a vector of voxel tissue type baseline modifiers, $\mathbf{J}_T^V$ is a T-by-V matrix of ones, and $\odot$ denotes the Hadamard (element-wise) matrix product.

A template of 30 default SMs in SimTB is shown in Fig. S1 on a square image of $V = \sqrt{V}X\sqrt{V}$ voxels, where side length $\sqrt{V}$ is specified by the user. SimTB has no requirements regarding the shape of the SMs (or TCs), and users can specify SMs using any 2-D function defined on x, y $\in$ [-1, 1]. Default SMs are modeled after sources commonly seen in axial slices of real fMRI data and most are created by combinations of simple Gaussian distributions.

The length of each spatial source TC is T time points, where the user specifies the repetition time (TR) in seconds per sample. TCs are constructed under the assumption that source activations are induced from underlying neural events as well as noise. Neural events can follow block or event-related experimental designs, or can represent unexplained, random deviations from baseline. We refer to an underlying event time series as TS to distinguish it from the subsequent TC that is created with a hemodynamic model.

Experimental paradigms are designed with task blocks and task events which can be assigned to several sources and can be identical across subjects, while unique events refer to unexplained deviations that are unique to each source and subject. Each task block is described by a block length and an inter-stimulus interval. For a given source, the TS is created by adding together amplitude-scaled task blocks, task events, and unique events. Amplitudes for task inputs can be negative or positive (indicating suppression or activation with the task); or can be zero (indicating that source activation does not follow the task).

Generating the fMRI blood oxygen level-dependent (BOLD) – like TCs from the event TS may be done in several ways, including linear convolution with a canonical hemodynamic response function (HRF) (difference of two gamma functions) (Friston et al., 1995) and the Windkessel balloon model (Buxton and Frank, 1997; Buxton et al., 1998; Friston et al., 2000; Mandeville et al., 1999). Users may vary hemodynamic parameters between sources and subjects, and define their own TC source models. After creation of the TCs, each source TC is scaled to have a peak-to-peak range of one. As with the SMs, Gaussian noise distributed as $N(0, 2.5 \times 10^{-5})$ is added to ensure non-zero TCs.

A baseline intensity, b_i, is specified for each subject. Fig. S2 displays the default baseline intensity map where four tissues are defined: sinus signal dropout, cerebrospinal fluid (CSF), white matter, and gray matter. By default, b = 800, thus the intensity map ranges from $0.3 \times 800 = 240$ in areas with signal dropout to $1.5 \times 800 = 1200$ in CSF.

2.2. Simulation parameters

In this work, we use 29 (C = 29) of 30 SMs (excluding the whole-brain spatial source) in the SimTB. The size of each SM is set to be 70% (with a standard deviation of 1% across subjects) of the default setting in SimTB. We simulate M = 100 subjects and T = 300 time points in length with a repetition time (TR) of 2 s/sample. To mimic between-subject variability, the SMs are given a small amount of