Bayesian network models in brain functional connectivity analysis

Jaime S. Ide a,b,∗, Sheng Zhang b, Chiang-shan R. Li b,c,d

a Department of Science and Technology, Universidade Federal de Sao Paulo, Sao Jose dos Campos, SP 12231, Brazil
b Department of Psychiatry, Yale University School of Medicine, New Haven, CT 06519, USA
c Interdepartmental Neuroscience Program, Yale University School of Medicine, New Haven, CT 06520, USA
d Department of Neurobiology, Yale University School of Medicine, New Haven, CT 06520, USA

ABSTRACT

Much effort has been made to better understand the complex integration of distinct parts of the human brain using functional magnetic resonance imaging (fMRI). Altered functional connectivity between brain regions is associated with many neurological and mental illnesses, such as Alzheimer and Parkinson diseases, addiction, and depression. In computational science, Bayesian networks (BN) have been used in a broad range of studies to model complex data set in the presence of uncertainty and when expert prior knowledge is needed. However, little is done to explore the use of BN in connectivity analysis of fMRI data. In this paper, we present an up-to-date literature review and methodological details of connectivity analyses using BN, while highlighting caveats in a real-world application. We present a BN model of fMRI dataset obtained from sixty healthy subjects performing the stop–signal task (SST), a paradigm widely used to investigate response inhibition. Connectivity results are validated with the extant literature including our previous studies. By exploring the link strength of the learned BNs and correlating them to behavioral performance measures, this novel use of BN in connectivity analysis provides new insights to the functional neural pathways underlying response inhibition.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

Biomedical imaging, including computed tomography, positron emission tomography, and magnetic resonance imaging has become important tools for medical diagnosis, surgical planning, and/or neuroscience research. In particular, brain imaging has provided important new information to the operations of the complex human brain, and improved our understanding of many psychiatric and neurological diseases, such as Alzheimer and Parkinson diseases, schizophrenia, addiction, and depression. In a recent review, Friston [1] referred to neuroimaging as the predominant technique in human behavioral and cognitive neuroscience, and highlighted the exponentially growing number of publications in the field. In particular, functional connectivity analysis, the study of how different parts of the brain are integrated during execution of cognitive tasks, is of growing importance in neuroscience research [1,2].

Several computational and statistical tools, from basic image processing to complex network analysis, have made possible these advances in neuroimaging. Approximate reasoning and Bayesian statistics are a de facto standard when dealing with complex and uncertain data, which often requires the use of expert and prior knowledge [3–6]. Although much progress has been made in the field, the interdisciplinary dialogue between statistician, computer scientist, physician and psychologist only began recently. For instance, although investigations have successfully applied Bayesian network (BN) and probabilistic graphical models in functional connectivity analysis [2,7], suggesting a large venue of opportunities to be explored, few...
functional connectivity studies systematically addressed the use of BN modeling in magnetic resonance imaging (fMRI) of a large number of subjects executing a cognitive task [8,9].

The goals of this paper are two-fold. First, we provide an up-to-date review of the current BN methodology in fMRI connectivity analysis, and highlight the practical issues involved in these applications. Second, and more importantly, we present a BN group modeling of fMRI dataset obtained from sixty healthy subjects performing the stop-signal task (SST) [10,11], to demonstrate the real-world application of this approach. By showing the link strength [12,13] of the learned BNs and their correlation to group behavioral measures, these new data provide novel insights to the brain networks of response inhibition. This application also highlights the caveats and potential pitfalls, and suggests plausible solutions to these issues.

The paper is organized as follows. In Section 2, we present a summary of BN applications in functional connectivity analyses, and provide a critical review of the caveats and methodological alternatives. In Section 3, we describe the details involved in a functional connectivity analysis of real fMRI data set using BN. In Sections 4 and 5, we demonstrate the results and conclude with a few additional remarks.

2. Bayesian networks (BN) in connectivity analysis of fMRI

In this section, we describe the methodology for applying BN in connectivity analysis of fMRI data. We review extant works of BN in Section 2.1, present the caveats and challenges in Section 2.7, and list available software tools in Section 2.8.

2.1. Literature review

The application of BN in connectivity analysis of fMRI data has a comparatively short history, and became a relatively established approach around 2000s [1]. The development of efficient fMRI data preprocessing methods and accessible software’s made possible a more reliable execution of group connectivity analyses. Pioneering studies applying BN to functional connectivity analysis of fMRI were published by Labatut and colleagues [14], followed by others [15,16,9,17,18]. Recently, BN and its methodological extensions were suggested to be useful in inferring causal relationships between activations by Glymour and colleagues [7]. The causal inference in fMRI is closely related to effective connectivity (influence that one neuronal system has on another [1]). On the other hand, Lindquist and Sobel [19] disagreed with Glymour and colleagues [20], and expressed their concern in using directed graphical models for revealing causal relations between regional activations. Pearl [21] entered in the discussion by clarifying the relation between the Structural Equation Modeling (SEM) [22] and causal graphical models in the analyses of brain connectivity. In particular, Lindquist and Sobel [23] demonstrated their concern about inferring causal relationships using graphical models in a data-driven fashion without prior substantive assumptions.

In the following, we revisit most of these works, by detailing the different approaches that are possible when applying BN in functional connectivity analysis.

2.2. fMRI data acquisition

FMRI provides an indirect measure of brain activity by means of a blood-oxygenation-level dependent contrast or BOLD signal in short, as discovered by Ogawa and colleagues [24]. The BOLD signal measures changes in the amount of deoxygenated hemoglobin, resulting from a large consumption of oxygen when neurons are activated, and increased blood flow and volume. Active neurons generate an increase in the amount of deoxygenated hemoglobin, leading to brighter MR images or an increase in BOLD signal. The term BOLD hemodynamic response (HR) is commonly used to refer to the shape of the BOLD signal following a neuronal event, and could be conceptualized as the impulse response of BOLD. The hemodynamic response shows its peak in 6s and an undershoot around 10s after the stimulus is presented, as a result of the changing dynamics of cerebral blood flow and volume. It must be noted that the hemodynamic response will vary depending on the duration and number of the environmental stimuli driving neuronal responses. Complex image acquisition and reconstruction processes are involved to turn BOLD signals into 3D images. Paul Lauterbur and Peter Mansfield, recipients of the 2003 Nobel Prize in Medicine, developed the echo-planar imaging (EPI) technique, which made rapid collection of BOLD signals possible. An excellent introduction and review of fMRI is available in Huettel et al. [25].

In terms of image acquisition, two main issues apply to BOLD fMRI:

• Spatial resolution: the ability to distinguish signal differences between nearby spatial locations. In fMRI, images are usually obtained in 64 × 64 or 256 × 256 matrix (referred as slice). For a field of view of 220 by 220 mm and a matrix of 64x64, the in-plane voxel size would be 3.44 by 3.44 mm. The third dimension is generally the same as or larger than the in-plane voxel size, and depends on how many slices one might want to obtain for one particular study. The term echo time (TE) is used to refer to the time interval between excitation pulse and data acquisition, and it is related to image reconstruction.

• Temporal resolution: the ability to distinguish signal differences between sequential observations. Temporal resolution is given by the repetition time (TR). TR is usually between 500ms and 3s, and is the time necessary for the scan to resample from the same voxel. One of the MR scanner parameters related to temporal resolution is the flip angle, which is the change in net longitudinal magnetization after excitation pulse. The shorter the TR, the smaller the flip angle has to
be. Thus, smaller TR means higher temporal resolution, but it also means decreased BOLD signal and reduced spatial coverage, because of the reduced flip angle.

In terms of the stimulus presented to the subjects during the scan, we can classify the fMRI experiments into three classes:

- **Block design**: the stimuli are presented in a multiple, consecutive way or constantly, and divided into different experimental conditions. For example, a visual stimulus can be presented in one block, and a constant “bip” sound can be presented in another condition, forming two distinct conditions. One can thus differentiate the BOLD signal level across conditions, and localize the brain activations specific to each of the two conditions. Most of the connectivity studies of BN used fMRI data from block-design tasks [7, 8, 14, 16–18, 26, 27].

- **Event-related design**: the stimuli are presented in short and discrete intervals, each constituting a single event or trial. As we will see in the experiment used in this work, each imaging session is formed by a sequence of stimuli, constituting different conditions depending on a particular set of stimuli. The duration between two consecutives events is known as inter-stimulus interval (ISI). Burge and colleagues [15, 28] used fMRI data of event-related design experiment in their BN modeling.

- **Resting-state experiment**: there are no stimuli and the aim is to collect the brain’s BOLD signal during a resting state, in order to investigate the brain’s underlying activity when no external stimulus is presented. Resting state BOLD data have been instrumental in defining clusters of brain regions that are functionally connected including the default-mode network [29]. More recently, investigators started to apply BN modeling to resting-state fMRI data sets [2, 30].

Because resting-state fMRI is straightforward to run and easy to compare across different investigations, it has become one of the most popular approaches in functional brain connectivity analysis [31]. On the other hand, block and event-related designs are required to investigate functional connectivity during cognitive challenges.

Different from other static MR image modalities – such as T1, T2, and diffusion weighted images – fMRI data comprise a sequence of 3D images, where in each voxel we have a time-series of BOLD signals sampled according to the TR.

### 2.3. fMRI data preprocessing

After fMRI acquisition and reconstruction, there are a series of computational procedures to correct for unwanted variability (noise) and artifacts, and to prepare the images for statistical analysis. These procedures are known as preprocessing in the fMRI literature and similar regardless of the experimental design and the purpose of statistical analyses. The most common preprocessing steps in fMRI are:

- **Realignment**: corrects for misalignment of the images across the slices and scan sessions, originating basically from the head movement. Assuming the head as a solid, rigid-body transformations (rotation and translation) are applied, based on some image similarity measures.

- **Slice-timing correction**: when associating each slice (in-plane image) with a time point, it must be noted that there is a time delay from the start to the end of scanning of each single slice. This phase delay must be corrected, using some interpolation technique.

- **Unwarping**: corrects for distortions in the images, generated by inhomogeneities in the magnetic field. In addition to the machine’s imprecision, head size and location are sufficient to create inhomogeneity or bias field.

- **Spatial normalization**: registers different subject’s brains to a common stereotaxic space, such as Talairach space and Montreal Neurological Institute (MNI) space. Normalization is necessary when comparing subject’s BOLD signals in a group. Normalization also allows the localization of a particular brain structure of interest through an anatomical atlas.

- **Smoothing**: a spatial filtering is performed by means of a Gaussian kernel to reduce noise and enhance the statistical power for group comparisons. This also has to do with the imprecise nature of the spatial normalization process.

These preprocessing stages are not directly related to the functional connectivity analysis, but as we will discuss further, they may have important impacts on the functional studies.

### 2.4. Selection of regions of interest and functional connectivity

In fMRI, it is very common to select regions of interests (ROIs) in order to study the functional integration (connectivity) between these ROIs. In terms of how ROIs are defined, it is possible to classify functional connectivity methods into three main categories:

- **Data-driven approach**: regions are selected in a completely data-driven fashion. Multiple time-series are classified into groups or clusters according to some criterion, such as independent component analysis (ICA) [32], in which clusters are formed based on the spatiotemporal characteristics of the BOLD signal of every single voxel. ICA have been successfully applied to fMRI connectivity analyses and might be used in conjunction with BN modeling [17].

Please cite this article in press as: J.S. Ide et al., Bayesian network models in brain functional connectivity analysis, Int. J. Approx. Reason (2013), http://dx.doi.org/10.1016/j.ijar.2013.03.013
• **Seed-based methods**: a single ROI is selected, an average or representative time-series is extracted for this ROI, and a functional connectivity measure is computed for every pair ROI-to-voxel across the whole-brain. Thus, these are bivariate approaches, and sometimes referred to as *whole-brain* methods. In this category, we have the correlation analysis of resting-state data [33], bivariate Granger causality [34] and *psychophysiological interaction* or PPI [35].

• **ROI-based methods**: multiple ROIs are selected from a previous functional localization analysis of the fMRI data (*activation maps*); i.e., parts of the brain that are active during a condition of interest; or from an anatomical atlas of the brain. A connectivity hypothesis is tested based on the network obtained from the time-series modeling of these ROIs. In this category, we have techniques based on *structural equation modeling* or SEM [22], multivariate Granger causality [36], and *dynamic causal modeling* or DCM [37]. Most of the previous BN applications to fMRI connectivity analyses have focused on this approach, because of the multivariate modeling offered by the BN learning methods.

Although we presented these approaches separately, they are complementary and usually are employed in conjunction. For instance, ICA can be used to select networks of interest, which in turn are used as the ROIs of the multivariate Granger causality [38]. PPI analysis is used to localize regions connected to a particular ROI, followed by a multivariate Granger causality analysis [39].

2.5. Modeling fMRI data with BN

As we discussed earlier, the ROI-based approach is a natural way for BN modeling of functional connectivity, where each ROI constitutes a *node* and the functional connection between the ROIs is represented by an *arc* of the graph. Furthermore, the conditional distribution can be used to compute the strength of the connection, as we will explore in this work. In Figure 1, we summarize the steps described so far, in order to apply BN in connectivity analyses.

There are two alternatives when entering the BOLD signal time-series into BN modeling:

• **Linear Gaussian interactions**: it assumes that the neural interaction between regions can be modeled by a linear relationship with Gaussian noise [27]. This allows implementation of efficient BN learning methods but runs the risk of missing important nonlinear interactions, as shown by [40]. Nonetheless, it seems to be a valid approach in light of several successful applications of Gaussian BN in connectivity analysis [8,26,27].

• **Discrete nodes**: BOLD time-series are *discretized* or *quantized* into a finite number of categories [15,18]. This approach is very attractive since it allows modeling of nonlinear relationships. However, it may require a larger number of samples to estimate the model parameters, and has the drawback of losing information. This approach has been successfully applied by [17,18,28].

In parallel, we can formulate the BN modeling in two ways:

• **Static Bayesian networks**: the functional network is considered *static* across time. That is, the BN represents a single snapshot of the functional connectivity during the entire experiment [27]. Because BN is represented by a *directed acyclic graph* (DAG), it is not possible to model cyclic or bidirectional connections with a single BN. Despite these limitations, as an example of this general approach, *graphical causal models* have been successfully applied to connectivity analysis in fMRI [2,7].

• **Dynamic Bayesian networks (DBN)**: the functional network is considered *dynamic* across the experiment, which is a more realistic assumption of how the brain works. This is a natural way of using BN to model temporal characteristics of the BOLD time-series, and has been employed in several functional connectivity investigations [16,18,28]. However, to implement DBN in practice, one has to assume that activations of brain regions are *stationary* and follow a *Markovian* condition, which are strong assumptions, because these activations are constantly influenced by external and internal stimuli.

Please cite this article in press as: J.S. Ide et al., Bayesian network models in brain functional connectivity analysis, Int. J. Approx. Reason (2013), http://dx.doi.org/10.1016/j.ijar.2013.03.013
2.6. Group connectivity analysis

In fMRI studies, it is common to have a large number of participating subjects to delineate a statistically significant brain network of the group, which can be of healthy control or patients. Group studies are common if the rule in medical research, because of individual differences and the need to extract a representative brain for the group. McKeown and colleagues [8] have systematically investigated and compared different methods to combine individual fMRI data to make group network inferences, and offer the following approaches:

- **Virtual-typical-subject (VTS) approach**: it assumes that every subject within the group has the same brain network, with the same functional connectivity. The practical consequence of VTS approach is that BOLD time-series from each individual are concatenated and treated as if they are sampled from a single virtual subject. This approach has a theoretical drawback, because it can result in statistical dependencies in the combined data that do not exist in any of the individual subject [7], and a random effects analysis (RFX) is needed for modeling multiple subjects [41]. Nonetheless, most of the previous studies have obtained representative and consistent brain networks with this approach [8, 18, 28, 42], likely because VTS allows the construction of a large data set with an enhanced signal-to-noise ratio [8].

- **Individual-structure (IS) approach**: it learns individual networks for each subject separately, and performs group-analysis on these individual networks [9]. Thus, for each subject we learn a BN structure with a score. This approach is more appropriate when there is large individual network variability across the group, which can occur in patients. On the other hand, it is not a trivial task to obtain a group statistically significant network, when individual network patterns are diverse [8]. IS approach is a natural way to perform a random effects group analysis.

- **Common-structure (CS) approach**: it considers a common brain network within the group as well as different patterns of connectivity for each individual subject [8]. In practice, within a single common BN structure, each individual’s data provide the parameter estimates and their respective scores. For group analysis, individual scores are summed forming a group score for comparison between competing models. CS is an intermediate approach between the two extremes, VTS and IS, and can be understood as a fixed effects analysis (FFX). For instance, CS approach is the one employed by [7] to compare causal graphical models.

2.7. Caveats and challenges

In this subsection we summarize some of the concerns and pitfalls that previous studies have raised when performing functional connectivity analysis on fMRI data [2, 7, 23], highlight those issues that are especially relevant to BN modeling, and point out possible directions to resolve these issues. We can organize these concerns as follows:

- **Imaging artifacts**: BOLD signals suffer from several artifacts: distortion, partial volume effects, large vessel effects [25]. Thus, when applying BN in fMRI, one must pay attention to the quality of preprocessed output images.

- **Imperfect normalization**: Despite the availability of very accurate normalization algorithms, no alignment between subject’s brains can be perfect because of the inter-subject variability in brain anatomy. One should be aware of this inaccuracy when extracting the average BOLD time-series from a selected ROI. This might have direct impact on the functional connectivity results, because the latter is affected by functionally inaccurate ROIs, as observed by [2]. We suggest that VTS approach, in contrast to the IS group analysis, is more robust to this artifact, since the concatenated data set would likely diminish the intersubject variability. In other words, the results of network modeling based on a single subject’s data are more susceptible to ROI selection, which is impure in most cases.

- **Hemodynamic response (HR) variability**: It has long been known that different regions in the brain have different HR [43], and this poses a critical concern to the lag-based connectivity methods, since they take the temporal precedence as granted to infer connectivity. In particular, recent investigations have shown poor performance of these lag-based methods in connectivity modeling [2]. We think that HR variability is a potential issue when applying DBN (in which temporal precedence of the nodes is assumed) to fMRI data, which requires future investigations. For instance, simulations showed that the variability of HRF significantly affected the accuracy of the learned structures [18].

- **Group analysis**: There does not seem to be a simple and single solution for connectivity group analysis of estimated BNs, as pointed out by [8]. We propose that the experimental design and the group specificity must always be considered for the selection of the best group analysis approach. The IS approach is more favorable with a large but heterogeneous sample while the VTS approach is more favorable with a relatively small but homogeneous sample.

- **Network structure learning**: As we increase the number of regions of interest, learning the best BN structure becomes challenging, since we have an exponential increase of the model search space. Existing Monte Carlo and local search methods can potentially fail, and exact structure learning methods are only feasible under prior structural assumptions [44]. Causal structure learning methods are yet possible controlling for false discovery rate [45].

2.8. Software tools

Several software tools are currently available for analyzing fMRI data. Within the Matlab framework, the most popular one is **Statistical Parametric Mapping (SPM)**, freely available at http://www.fil.ion.ucl.ac.uk/spm/. SPM provides a
complete set of tools to process fMRI data, from preprocessing to group including functional connectivity analyses: PPI and DCM. In Functional Connectivity Toolbox, obtainable at https://sites.google.com/site/functionalconnectivitytoolbox/home, several methods are available. For resting-state fMRI, an interesting one is the Resting-State Toolkit (REST), available at http://www.restfmri.net/forum/REST_V1.4. Other popular tools to analyze fMRI data and perform functional connectivity analyses are the FMRIB Software Library (FSL), freely available at http://www.fmrib.ox.ac.uk/fsl/, and BrainVoyager, a commercial software popular for medical use.

3. Experimental procedures: modeling a response inhibition brain network using BN

In this section, we present a case study to illustrate the application of BN in functional connectivity analysis of fMRI, as described in Section 2. We present an application using BN to model the functional connectivity of regions involved in an important aspect of cognitive control: response inhibition. Deficits in response inhibition characterize many clinical conditions including attention deficit hyperactivity disorder (ADHD) [46] and drug addiction [47]. Details of the experiment and the significance of functional connectivity analysis of the inhibitory network are described in our previous works [48,49]. Here, we offer a brief summary of the behavioral task and previous findings, which are necessary for readers to follow the discussion on BN modeling.

3.1. Subjects and behavioral task

Sixty subjects (30 men, 22–45 years of age, all right-handed) were paid to participate in the study. All subjects signed a written consent after details of the study were explained, in accordance to institute guidelines and procedures approved by the Yale Human Investigation Committee.

We employed a simple reaction time task in this stop-signal paradigm [10,11,50–52], illustrated in Figure 2. There were two trial types: “Go” (~75%) and “Stop” (~25%), randomly intermixed, with an inter-trial interval of 2s. A small fixation dot appeared on the screen to engage attention at the beginning of the trials. After a randomized time interval (fore-period or FP) between 1 and 5 s, the dot turned into a circle (the “go” signal), prompting the subjects to quickly press a button. The circle disappeared at a button press or after 1 s had elapsed, and the trial terminated. The time between go signal and the button press is known as reaction time (RT). A premature button press prior to the appearance of the circle also terminated the trial. Both premature responses and go trials in which no responses were made constituted a go error (GE). On a Stop trial, an additional “X,” the “stop” signal, replaced the go signal and instructed subjects to withhold their response (button press). Similar to Go trials, a Stop trial terminated at button press or 1 s after the appearance of the stop signal. Failure to withhold the go response for the 1 s constituted a stop error (SE), and stop success (SS) otherwise. The stop signal delay (SSD) – the time interval between go and stop signals – started at 200 ms and was adjusted according to a staircase procedure: if the subject succeeded in withholding the go response, SSD increased by 67 ms; conversely, if he failed, SSD decreased by 67 ms [53]. The staircase procedure ensured that subjects succeeded in withholding their response in approximately half of the stop trials. The stop signal reaction time (SSRT) is an index of motor response inhibition performance, and can be estimated by subtracting the critical SSD from the median of Go trials RT. The critical SSD was computed by taking the mean of all mid-run SSDs [54].

Subjects were instructed to respond to the go signal quickly while keeping in mind that a stop signal could come up in a small number of trials. Prior to the fMRI study, each subject had a practice session outside the scanner. In the scanner, each subject completed four 10-min runs of the task. Depending on the actual stimulus timing (that is, FP duration) and speed of response, the total number of trials varied slightly across subjects.

3.2. Imaging protocol

Conventional functional, blood oxygenation level dependent (BOLD) signals were acquired with a single-shot gradient echo echoplanar imaging (EPI) sequence. Thirty-two axial slices parallel to the AC–PC line covering the whole brain were
acquired with TR = 2,000 ms, TE = 25 ms, bandwidth = 2004 Hz/pixel, flip angle = 85°, field of view = 220 x 220 mm, matrix = 64 x 64, 32 slices with slice thickness = 4 mm and no gap. Three hundred images were acquired in each run (session), for a total of 4 runs.

3.3. Preprocessing

Data were analyzed with Statistical Parametric Mapping version 5 (SPM5, Wellcome Department of Imaging Neuroscience, University College London, U.K.). Images from the first five TRs at the beginning of each trial were discarded to enable the signal to achieve steady-state equilibrium between RF pulsing and relaxation. Images of each individual subject were first corrected for slice timing, realigned (motion-corrected) and unwarped (Andersson et al. 2001; Hutton et al., 2002). A mean functional image volume was constructed for each subject for each run from the realigned image volumes. These mean images were normalized to an MNI (Montreal Neurological Institute) EPI template with affine registration followed by nonlinear transformation (Ashburner and Friston, 1999; Friston et al., 1995a). The normalization parameters determined for the mean functional volume were then applied to the corresponding functional image volumes for each subject. Finally, images were smoothed with a Gaussian kernel of 10 mm at Full Width at Half Maximum. The data were high-pass filtered (128 sec cutoff) to remove low-frequency signal drifts.

These preprocessing procedures were aimed to reduce the imaging artifacts as described before. Also, we should be aware of the inexact nature of the normalized functional images. To assure the preprocessing quality, as well as the data set integrity, individual images were manually examined. One subject was excluded from the analysis because of excessive head movement.

3.4. Selection of regions of interest (ROI) and hypothesis

We selected the brain regions previously implicated in response inhibition [48,55,56]: right inferior frontal cortex (IFC), left subthalamic nucleus (STN), pre-supplementary motor area (pSMA), and left primary motor cortex (PMC), and employed the same masks as used in our previous studies [48]. IFC, pSMA, and PMC masks were obtained from functional localization study of response inhibition [10], and STN was obtained from an atlas [57]. IFC has been associated with attention and inhibition functions and is hypothesized to project to STN – the so-called hyperdirect pathway – for motor inhibitory control [56]. Along with several other investigations [10,58–61] that have shown the importance of pSMA in response inhibition, we showed evidence supporting a specific role of attentional monitoring for IFC and motor inhibitory control for the pSMA. This latter model calls for an IFC connection to STN through the pSMA [48]. In Figure 6, we depict the selected ROIs and the obtained functional connections according to the current analysis.

3.5. Data quantization

For each subject, the spatially preprocessed BOLD time-series were averaged across all voxels inside the selected ROIs. The average time-series were concatenated across four sessions (each containing 295 time points) after linear de-trending and normalization (subtraction of temporal mean and division by standard deviation) [62], to eliminate potential linear trends due to imaging artifacts, and scanning differences across distinct sessions. Thus, for each subject and ROI, we constructed a time-series with 1,180 points.

Following previous studies of BN with categorical variables to model non-linear functional relationships between ROIs [17,18,28], we quantized every BOLD time-series into two to five categories, equally distributed. For example, we quantized the BOLD signals into: very low (1), low (2), medium (3), high (4), very high (5), with each of the five categories containing 20% of the data points. In Figure 3, we show part of the original time-series in the IFC (solid line) and its quantized series (dashed line) for a subject.

3.6. Discrete BN model construction

According to the SST experimental conditions, the quantized BOLD signals were split into three conditions: GS, SS, and SE (GE was ignored because they are very few, <3%), and for each condition we constructed a BN. We employed the BNT

![Fig. 3. Sample of BOLD signal in the IFC region. Original time-series (black line) and quantized time-series (blue line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)](http://dx.doi.org/10.1016/j.ijar.2013.03.013)
toolbox for Matlab, freely available at http://code.google.com/p/bnt/. The division into conditions was executed as follows. To account for the HR delay, the condition (stimulus) onsets were convolved by a balloon model hemodynamic response function [63] or HRF (Fig. 4). In Fig. 5a, we depict a sequence of experimental conditions during a session: “1” for GS, “2” for GE, “3” for SS, and “4” for SE. GS is the most frequent condition, while GE is very rare. Each of the condition onsets was convolved by HRF separately, setting “1” whenever the stimulus occurred. In Fig. 5b, we show the overlay of the convolved condition onsets. In Fig. 5c, we plot the original BOLD signal in IFC and PMC regions, in the same selected time window. It is important to note that the BOLD signal is synchronized not with the condition onset (Fig. 5a), but with the convolved onset (Fig. 5b), illustrating the existence of the HR delay. For instance, the condition SS happens at slices 927 and 955, and as expected an increase of the BOLD signal is observed in IFC, because the latter is implicated in attentional monitoring and response inhibition.

3.7. Best group model search leaving-one-out

For each condition, we selected the best representative connectivity model, employing the following group analysis, as described previously: VTS, IS, and CS. The Bayesian information criterion (BIC) was employed for model comparisons, and the model evidences were computed to evaluate the Bayesian factor (BF) for pair-wise model comparisons [64]. Since we selected four ROIs or nodes (i.e., a total of 543 possible BN structures), an exhaustive search was employed. That is, BIC scores were computed for every single BN structure and the best model was obtained. The exhaustive search allowed us to assure the global maximum model and focus on functional modeling and group analysis per se, rather than the particularities of the model search algorithm. We considered all the BNs within the same Markov equivalent class [65] as a single model, with a focus on the conditional independence properties of the network but not on the directionality of the connections. As discussed in the beginning, the directionality of connectivity is an important yet debated topic [23] that requires future investigations.

For the VTS approach, we employed cross-validation by concatenating the 58 subject data sets, and leaving one subject out, generating a total of 59 VTS group BN best models. The rationale is that the group best model should not be influenced by a single subject data set.

3.8. Follow-up link strength analysis of learned BN using cross-validation

The utility of BN in functional connectivity analysis derives from using the learned parameters to estimate the strength of the connectivity in biologically plausible models. Earlier studies have computed connection strengths of learned parameters using Gaussian DBNs [8,26]; however, as far as we know, none have explored the strength of connection in discrete BNs and examined for association with behavioral measures. In this work, we used the link strength of an arc [12,13], as it reflects the strength of connection solely among that arc, and correlated them with the SSRT, a measure of motor inhibition performance. Given the random variables $X, Y$, the link strength of $X \rightarrow Y$ was defined as the $\text{MI}(X, Y|Z) = U(Y|Z) - U(Y|X, Z)$, where $\text{MI}$ is the mutual information, $U$ is entropy and $Z$ is the other parents of variable $Y$. 

![Fig. 4. Hemodynamic response function (HRF).](image-url)
4. Results and discussions

In this work, we will focus on the results obtained from BOLD signals quantized into five categories and group analysis based on VTS (Virtual-typical-subject) approach, since it provided the most meaningful and interpretable results. In our dataset, group analyses using IS and CS approaches did not provide useful results since the selected group best models were all disconnected networks.

4.1. Group best BN model

For each condition, we obtained a sorted list of BN models according to their BIC scores. For the top-ten best scored models, we computed the log of the Bayesian Factor (BF) for all individual pairs of models, and confirmed for each condition that the best model was significantly better than the second best model (BF values where \( > 100 \)). As we are interested in the inhibition network, we focus on the two stop conditions: Stop Success (SS) and Stop Error (SE). The best BN model family is illustrated at Figure 6. Here, a family of selected best BN models represented the same conditional independence, having the same BIC scores. We treated this family of equivalent classes as a single model, with no assertion about the direction of influence. Importantly, the best BN model provided evidence of pSMA mediating the projection from IFC to STN, in support of our previous results [48].

4.2. Model stability

To test the robustness of the group network result, a cross-validation approach was used to check whether the best BN model was biased by some of the subject dataset. In Table 1, we show the number of times that each of the top-ten models was ranked in that position. A lower number indicated how much that model was influenced by subject selection. Clearly, for SS and SE conditions, models ranked after second and fourth position, respectively, are dependent on the group
of subjects that were selected; thus, none of them is a robust model. This approach is particularly useful when dealing with large networks, when an exhaustive search is impossible and a meta-heuristic is needed.

4.3. Link strength significantly modulated by experimental conditions

According to our hypothesis, the expected connectivity pattern would be IFC → pSMA → STN → PMC. Thus, we fixed this network for link strength (LS) analysis, where we compared the magnitude of LS during different conditions. In Table 2, we show the average LS of each connection, across different conditions, and identify the connectivity patterns that are modulated by conditions: pSMA → STN and STN → PMC. These results showing trend and significant modulation each of pSMA → STN and STN → PMC, but not IFC → pSMA, connection strengths, by experimental conditions (SS and SE, GS and SE) are consistent with our previous studies that show a direct role of pSMA, and not IFC, in response inhibition [48].

4.4. Link strength correlated with behavioral performance measure

Most interestingly, the link strengths of different connections are significantly correlated with the SSRT (stop signal reaction time), during different conditions (Table 3, and Fig. 7). In particular, the connection strength between the STN and PMC during stop success trials is negatively correlated with the SSRT. That is, higher connection strength between the STN and PMC during stop success trials is associated with shorter SSRT and better inhibitory control.
Fig. 7. Significant correlations between link strength of inhibition nodes and an inhibition performance measured by SSRT.

Taken together, the findings that the connection IFC→pSMA is correlated with SSRT in both conditions SS (successful stop) and SE (no stop), but is not modulated by SS and SE per se, suggests the attentional rule of the IFC in detecting stop signals, and an executive role of the pSMA to inhibit the prepotent go response [48].

4.5. Limitations and future works

In this work, we focused on a small size BN with four variables to allow an interpretable validation of the learned functional connectivity pattern, and to perform an exhaustive search of the group best model, without entering in the discussion of network structure learning algorithms. Interesting extensions are to increase the number of ROIs, and use exact structure learning methods with some priors (partially fixed structures) [44], and also employ causal structure learning algorithms [7,45].

5. Concluding remarks

In this work, we presented a literature review of BN modeling in fMRI research. We also presented novel results on the response inhibition network by applying BN modeling to fMRI data collected from a stop-signal task. The patterns of connectivity identified by BN is not only biological plausible but also consistent with our previous findings obtained with other connectivity analysis methodologies. These results support the utility of BN modeling in connectivity analysis of fMRI data and testing competing hypotheses for patterns of functional connectivity.

Acknowledgements

This study was supported by Young Researcher Award (Sao Paulo Research Foundation FAPESP 2011/08573-4) to J.S.I, and NIH grants R01DA023248, K02DA026990, and R21AA018004 to C.-S.L.