Published in final edited form as:

J Neurosci Methods. 2017 November 01; 291: 61–68. doi:10.1016/j.jneumeth.2017.08.007.

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

Qingbao Yu¹, Yuhui Du^{1,2}, Jiayu Chen¹, Hao He^{1,3}, Jing Sui^{1,4}, Godfrey Pearlson^{5,6,7}, and Vince D. Calhoun^{1,3,6}

¹The Mind Research Network, Albuquerque, NM 87106, USA

²School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China

³Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM 87106, USA

⁴Brainnetome Center and National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Science, Beijing 100190, China

⁵Olin Neuropsychiatry Research Center, Hartford, CT 06106, USA

⁶Department of Psychiatry, Yale University, New Haven, CT 06520, USA

⁷Department of Neuroscience, Yale University, New Haven, CT 06520, USA

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.

Corresponding Author: Qingbao Yu, Ph.D. The Mind Research Network, 1101 Yale BLVD NE, Albuquerque, NM 87106, USA, qyu@mrn.org (Qingbao Yu), vcalhoun@unm.edu (Vince D. Calhoun).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

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 methods than ROI approaches in real fMRI data.

Keywords

brain graph; ROI; ICA; simulation; ground truth

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 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; Yu et al., 2013a; Yu et al., 2013b; 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, $i = 1, 2, 3, \dots M$, 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,

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

Where $\mathbf{Y}_i^{\text{nn}}$ 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, $\mathbf{b}_i$ 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 Figure 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. Figure 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 translation, rotation, and spread via normal deviates. Translation in the horizontal and vertical directions of each source has a standard deviation of 0.1 voxels (with a mean of 0). Rotation has a standard deviation of 1 degree (with a mean of 0). Experimental paradigms are block design with 2 conditions. Temporal similarity between SMs is controlled by modifying the amplitude of conditions to each SM. For all SMs, percent signal change is centered at 3 with a standard deviation of 0.25. Rician noise is added to the data of each subject to reach the desired contrast-to-noise ratio (CNR) level. Here, CNR is uniformly distributed over subjects from 0.65 to 2, reflecting the range of typical CNR values for BOLD contrast at conventional voxel sizes and field strengths. We perform the simulation in four scenarios. In scenario 1, twenty-nine SMs are with little overlap. In scenario 2, twenty-nine SMs are with little overlap plus three artificial SMs. For the positions of the three artifacts see Figure S3. In scenario 3, twenty-nine SMs are largely overlapped. In scenario 4, twenty-nine SMs are largely overlapped plus 3 artificial SMs as in scenario 2. Figure 1 shows locations of the SMs in all of the four Scenarios.

2.3. Group ICA

Group ICA is performed on the simulated data in each scenario with the GIFT toolbox (<http://icatb.sourceforge.net/>) which uses a temporal concatenation approach to combine data across subjects (Calhoun et al., 2001). Prior to the ICA decomposition, subject datasets are orthogonalized and whitened using principal component analysis (PCA) (Erhardt et al., 2011). Whitened subject datasets are temporally concatenated to form the aggregate data matrix and a group-level PCA reduced the dimension of the aggregate data to the number of components to be estimated with ICA. The number of the ICs is set as same as the number of spatial sources (29) plus artifacts (0 or 3) if any. Spatial ICA (Bell and Sejnowski, 1995; Himberg et al., 2004; McKeown et al., 1998a) is performed using the infomax algorithm with 10 repetitions in ICASSO (<http://research.ics.aalto.fi/ica/icasso>) to estimate the reliability of the decomposition. Individual subject maps and related TCs are then estimated by back-projecting the aggregate mixing matrix to the subject space based on the PCA reducing matrices. Detected spatial components are used as nodes when building the ICA graph in the following analysis.

2.4. Graph analysis

Four kinds of undirected weighted graphs are built in each scenario. The number of nodes (N) is same in all graphs (N = 29 in this study). In scenario 2 and 4, three artificial SMs associated nodes are excluded when building the four types of graphs.

1. Ground truth graph. In this graph, the nodes are SMs (excluding artifacts if any) in the simulation. Edges are the Pearson correlation between each pair of simulated time courses of the SMs.
2. ICA graph. In this graph, the nodes are spatial independent components obtained by group ICA performed on the simulated data. Edges are the Pearson correlation between each pair of ICA time courses of the components.
3. ROI graph. In this graph, the nodes are ROIs which are defined base on the SMs of the simulation using a threshold of 0.8. Before computing the correlation between time courses of ROIs, global mean of the simulated data is removed (Power et al., 2016). Time series of voxels within each ROI are averaged into one representative time courses. Edge weights of the graph are the Pearson correlation between each pair of representative time courses of the ROIs.
4. Modified ROI graph (MROI-graph). In fMRI data, ROIs are usually defined base on structural brain atlas in which not all voxels are functionally similar. To simulate the effects of defective functional ROIs, and compare with the ideal ROI scenario, we build and examine a MROI-graph. The only difference of this graph to ROI graph is that the nodes are modified ROIs by moving the location of each ROI to 15% of its area. To do this, we change the locations of all 29 SMs in simTB and identify the modified ROIs base on the modified SMs using the same threshold as in building ROI graph. See Figure S4 for the locations of modified SMs used for defining the modified ROIs comparing with original SMs.

When building the graph, to distinguish between positive and negative correlations, we do not use the absolute value of the correlation, but compute a similarity measure defined as (Mason et al., 2009; Mumford et al., 2010; Yu et al., 2015; Yu et al., 2011a)

$$s_{ij} = \frac{r_{ij} + 1}{2} \quad (2)$$

Global level metrics of the weighted graph S including connectivity strength [also called weighted degree or graph strength (Rubinov and Sporns, 2010)], clustering coefficient, and global efficiency of positive and negative graphs are then computed.

2.5. Equations for computing graph measures

We denote G as the set of all nodes in the weighted graph W ($W = S$ in this study), and N ($N = 29$ in this study) is the number of nodes. Connectivity strength (CS) of node i is defined as below:

$$CS_i = \sum_{j \in G} w_{ij} \quad (3)$$

The connectivity strength of whole graph (global level) is the average of the connectivity strength of all the nodes in the graph:

$$CS_{net} = \frac{1}{N} \sum_{i \in G} CS_i \quad (4)$$

Nodal level clustering coefficient (CC) is computed using the below equation:

$$CC_i = \frac{1}{CS_i(CS_i - 1)} \sum_{j, k \in G} (w_{ij} w_{ik} w_{jk})^{1/3} \quad (5)$$

The clustering coefficient of whole graph is the average of the clustering coefficient of all the nodes in the graph:

$$CC_{net} = \frac{1}{N} \sum_{i \in G} CC_i \quad (6)$$

Global efficiency (GE) of node i is defined as:

$$GE_i = \frac{\sum_{j \in N, j \neq i} (d_{ij})^{-1}}{N - 1} \quad (7)$$

In which

$$d_{ij} = \sum_{a_{uv} \in g_{i \leftrightarrow j}^w} f(w_{uv}) \quad (8)$$

Where f is a map (here is an inverse) from weight to length and $g_{i \leftrightarrow j}^w$ is the shortest weighted path between i and j . Global efficiency of the whole graph is the average of the global efficiency of all the nodes in the graph.

2.6. Statistical analysis

Paired t-tests are performed to examine if the graph metrics differ significantly between any pair of graphs with different nodes. Correlations of graph measures between ground truth graph and any other graph are computed to assess which node definition is better to represent the ground truth. For a flowchart of the methods of this work see Figure 2.

3. Results

3.1. Spatial components obtained by group ICA

Group ICA separates the simulated fMRI data into components which are similar to ground truth SMs. We computed the spatial correlation between ICA components and ground truth SMs. In little overlap case (scenario 1), the correlation values range from 0.43 to 0.99. In little overlap with artifacts case (scenario 2), correlation values range from 0.24 to 0.99. In large overlap case (scenario 3), correlation values range from 0.16 to 0.99. In large overlap with artifacts case (scenario 4), correlation values range from 0.0977 to 0.9862. As an example, we show the spatial maps obtained by group ICA in large overlap case (scenario 4) in supplementary Figure S5.

3.2. Graph measures

Structures of the ICA graph, ROI graph, and modified ROI graph are similar to the ground truth graph in all of the four scenarios (see Figures S6 – S9).

Paired t-tests show that the graph measures of ICA, ROI, MROI graphs are significantly ($P < 0.001$) different from ground truth in all cases. When comparing any other pair of graphs, connectivity strength between ROI graph and MROI graph is not showing significant difference in scenario 4. All graph metrics in all other cases are showing significant difference between any pair of graphs (paired t-tests, $p < 0.05$).

In all scenarios, values of all graph metrics for the ICA graph (compared to the ROI and MROI graphs) are closer to the ground truth graph. In the no artifacts cases (scenario 1 and

3), graph measures in modified ROI graphs are far away from ground truth compared to the ROI graph. However, in cases with artifacts (scenario 2 and 4), graph measures of the modified ROI graph are closer to ground truth than the ROI graph (see Figure 3).

For the correlation of graph metrics across subjects between the ground truth and any other graph, in little overlap case and little overlap with artifacts case (scenario 1, 2), the ROI graph is more strongly correlated with ground truth than ICA graph. However, modifying the location of ROI largely decrease the correlation. In large overlap and large overlap with artifacts cases (scenario 3, 4), the ICA graph is more strongly correlated with ground truth than ROI and MROI graphs. Modifying the location of ROIs also heavily decrease the correlation with ground truth (see Figure 4).

Discussion

In this simulation study, simulated fMRI data are generated using SimTB which implements a linear model (data are expressed as the product of time courses and spatial maps). When doing the simulations, spatial maps are created in four scenarios: 1. Twenty-nine SMs with little overlap; 2. Twenty-nine SMs with little overlap plus three artificial SMs; 3. Twenty-nine SMs with large overlap; 4. Twenty-nine SMs with large overlap plus three artificial SMs. Using the simulated data, four kinds of graphs are built. A. ground truth graph in which nodes are ground truth SMs and edges are the correlation between simulated time courses of the SMs. B. ICA graph in which nodes are ICA spatial components and edges are the correlation of ICA time courses of the components. C. ROI graph in which nodes are ROIs defined base on the simulation SMs by using a threshold of 0.8 and edges are the correlation between time courses of ROIs which are the averaged time course of all voxels within each ROI. D. MROI graph in which nodes are modified ROIs by moving the location of the SMs and edges are the correlation of time courses of the modified ROIs which are averaged time courses of all voxels within each modified ROI. By comparing graph measures between ground truth and any other graph, we find that graph measures are closer to ground truth for the ICA graph compare to the ROI and MROI graphs for all scenarios. In cases without artifacts, moving the location of the ROIs induces the graph metrics being far away from the ground truth. Whereas in cases with artifacts, moving the location of ROIs can make the graph metrics closer to ground truth, presumably because the modified locations are less impacted by artifacts. When looking at the correlation of the graph metrics between the ground truth and the various scenarios across subjects, we find that the ROI graph measures are more strongly correlated with the ground truth than ICA graph in the little overlap and the little overlap with artifacts cases (scenario 1, 2). Measures of ICA graph are more strongly correlated with ground truth than ROI and MROI graphs in the large overlap and large overlap with artifacts cases (scenario 3, 4). Moving the location of ROIs greatly decreases the correlation in all scenarios. By summarizing all of these findings, as predicted, the metrics of ICA graph represent the ground truth better than ROI and MROI graphs in most cases.

In studies with real fMRI data, it has been suggested that an ideal node definition for the brain graph should define functionally homogeneous nodes, represent functional heterogeneity across nodes, and account for spatial relationships (Fornito et al., 2013). There

are two types of approaches for defining the nodes which are data-driven methods and non-data-driven methods. ROI methods of seed-based or anatomical-atlas based techniques are not data-driven, requiring definition based on prior experience (Rajtmajer et al., 2015). A few studies have evaluated some graph metrics of brain networks with different nodes in real fMRI data. For example, a study examined the predominate schemes that are currently used for defining nodes in functional brain networks: voxel-wise, anatomical atlas-based, and functional activation meta-analytic. Results imply that voxel-wise networks are the most likely to result in new discoveries of unknown brain properties. The other methods are limited by a priori knowledge, and therefore, limit the ability to make new discoveries. Voxel-wise networks also allow for the discovery of key nodes that are not located within a putative functional unit. Overall, the data-driven method is better for building brain networks than non-data-driven methods (Stanley et al., 2013). Another study developed a data-driven method for generating an atlas by parcellating whole brain resting-state fMRI data into spatially coherent regions of homogeneous functional connectivity. By comparing the suitability of the data-driven parcellation atlas against four ROI atlases (Talairach and Tournoux, Harvard-Oxford, Eickoff-Zilles, and Automatic Anatomical Labeling) and a random parcellation approach, anatomical atlases exhibit poor ROI homogeneity and do not accurately reproduce functional connectivity patterns present at the voxel scale (Craddock et al., 2012). Consistent with these previous works with real fMRI data, overall findings of this simulation study imply that, the data-driven ICA method for defining the brain nodes performs better than ROI method.

To our knowledge, this is the first simulation study which aims to directly compare ICA and ROI methods for defining nodes to determine which is more appropriate when building a brain graph. The advantage of using simulated rather than real data is that graph metrics computed in graphs with different nodes can be compared to a ground truth. Results of this work may provide evidence indicating which method is more appropriate when performing graph analyses in real fMRI data. Results generally show that ICA graph represents ground truth better than ROI graph. Moving the location of ROIs largely decreases the correlation of graph measures between ground truth and ROI graphs in all scenarios. However, when comparing the values of graph metrics between graphs, moving the location of ROIs does not always make the graph measure far away from ground truth. When there are artifacts (scenario 2 and 4), the measures of MROI graph are closer to ground truth than ROI graph implying that the simulated artifacts influence the estimated graph measures. In addition, in two little overlap cases (scenario 1 and 2), graph metrics of ROI graph are more strongly correlated with ground truth than ICA graph. This makes sense and represented a near perfect scenario in which the ROI selection is known a priori. Such a scenario is unlikely in real fMRI data, since ROIs are often defined base on anatomical atlases which exhibit poor ROI homogeneity. In summary of all results of this work, we suggest that, the use of functionally inaccurate ROIs as graph nodes is extremely damaging to brain network estimation, and results derived from inappropriate ROI definition should be regarded with great caution, also in line with a previous study which compared different connectivity estimation approaches (Smith et al., 2011).

Methodological limitations of this study need to be discussed. First and foremost, the simulated data are generated using a linear model (the product of time courses and spatial

maps). Although ICA which is also a linear model has been successfully applied to both resting state and task fMRI data and revealed robust brain connectivity networks (Abou-Elseoud et al., 2010; Abou Elseoud et al., 2011; Allen et al., 2011; Arbabshirani and Calhoun, 2011; Calhoun et al., 2001; Calhoun et al., 2008; Kiviniemi et al., 2009; McKeown et al., 1998a; McKeown et al., 1998b; McKeown and Sejnowski, 1998), it is not clear if the linear model is a perfect representation of real fMRI data. In addition, only four scenarios are simulated, thus conclusions are limited to those scenarios. Also, when building the ROI graph, to make it comparable to ground truth graph, nodes are defined using a threshold base on SMs (created by simTB toolbox for generating simulated data). This is different from practice in which ROIs are usually defined base on a brain atlas. However, the advantage is the ROIs are based on the ground truth, which should be the best scenario for the ROI analysis.

When building the ROI graph, we define the ROIs using a threshold of 0.80. To ensure the results are not dependent on this specific threshold, we change the threshold from 0.50 to 0.95 in steps of 0.05 and compute the graph measures of the ROI graph built with these different thresholds. It shows that the graph measures are very close to each other and all are lower than ground truth in all scenarios (see Figures S10-S13). Thus the findings and conclusions of this study are robust to changes in the threshold for building the ROI graph.

In studies with real fMRI data, graphs built with ROI nodes usually have about 90 nodes (Liu et al., 2008) and graphs built with ICA nodes usually have about 48 – 57 nodes (Yu et al., 2015; Yu et al., 2013c). The simTB toolbox simulates just one slice (2D) data, thus the nodes number 29 in this work is much lower than those studies. To test if the findings are consistent when change the number of nodes, we create 11 more spatial sources (totally 40) in simTB and repeat the analysis by building graphs with 40 nodes (see Figure S14 for the locations of 40 SMs). Results show that the results of graph with 40 nodes are in line with the results of graph with 29 nodes (see Figures S15-S16).

Besides global level measures, nodal level graph metrics are also important to quantitatively describe the topological properties of graphs. Thus we assess nodal level graph measures as well. As shown in Figures S17-S74, results of the graph measure of most single nodes are consistent with global level measures.

Single voxels can also be used as graph nodes in fMRI studies. Thus we build a single voxel level graph in which each node is the voxel with the highest intensity value in each of the 29 SMs. Results show that global efficiency of single voxel level graph is even farther from ground truth graph compared to the ICA graph in all scenarios. The connectivity strength of the single voxel level graph is far from ground truth graph compared to the ICA graph in scenarios 1, 3, and 4. The clustering coefficient of single voxel level graph is closer to ground truth graph compared to the ICA graph (see Figure S75). However, the simulated data are generated using SMs and corresponding time courses in this work, thus ICA graph and ROI graph in which nodes are defined using a group of voxels may be more comparable to ground truth graph than single voxel level graph. Future work should explore additional simulation scenarios including a larger range of regions and single voxel cases.

Conclusions

In summary, the aim of this study is to examine which method (ROI vs ICA) for defining nodes is more appropriate when performing a brain graph analysis. By performing simulations, graph metrics of brain networks with nodes defined by ICA method or ROI method are compared with ground truth. Correlations of the graph measures between those graphs with ground truth are also evaluated. Overall results imply that graph with nodes defined by data-driven ICA method reflects ground truth better than ROI method. In addition, moving the location of ROIs would largely affect the graph metrics. Since in studies with real fMRI data, ROIs are typically defined using anatomical atlases which may not correspond well to real functional boundaries in the data, we strongly suggest researchers to use data-driven ICA method (rather than ROI method) to define graph nodes in brain network studies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work is supported by the National Institutes of Health (NIH) grants including a COBRE grant (P20GM103472), R01 grants (R01EB005846, 1R01EB006841, 1R01DA040487, REB020407, EB000840), and other grants (5P20RR021938, R37 MH43775 PI: Pearlson). This work is partly supported by the “100 Talents Plan” of Chinese Academy of Sciences, the state high-tech development plan of China (863) 2015AA020513 (PI: Sui J), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB02060005), Chinses NSF (81471367, PI: Sui J), and natural science foundation of Shanxi (2016021077, PI: YHD).

References

- Abou-Elseoud A, Starck T, Remes J, Nikkinen J, Tervonen O, Kiviniemi V. The effect of model order selection in group PICA. *Hum Brain Mapp.* 2010; 31:1207–1216. [PubMed: 20063361]
- Abou Elseoud A, Littow H, Remes J, Starck T, Nikkinen J, Nissila J, Timonen M, Tervonen O, Kiviniemi V. Group-ICA Model Order Highlights Patterns of Functional Brain Connectivity. *Front Syst Neurosci.* 2011; 5:37. [PubMed: 21687724]
- Allen EA, Erhardt EB, Damaraju E, Gruner W, Segall JM, Silva RF, Havlicek M, Rachakonda S, Fries J, Kalyanam R, Michael AM, Caprihan A, Turner JA, Eichele T, Adelsheim S, Bryan AD, Bustillo J, Clark VP, Feldstein Ewing SW, Filbey F, Ford CC, Hutchison K, Jung RE, Kiehl KA, Kodituwakku P, Komesu YM, Mayer AR, Pearlson GD, Phillips JP, Sadek JR, Stevens M, Teuscher U, Thoma RJ, Calhoun VD. A baseline for the multivariate comparison of resting-state networks. *Front Syst Neurosci.* 2011; 5:2. [PubMed: 21442040]
- Allen EA, Erhardt EB, Wei Y, Eichele T, Calhoun VD. Capturing inter-subject variability with group independent component analysis of fMRI data: a simulation study. *Neuroimage.* 2012; 59:4141–4159. [PubMed: 22019879]
- Arbabshirani, MR., Calhoun, VD. Functional network connectivity during rest and task: comparison of healthy controls and schizophrenic patients. Conference proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society Conference; 2011; 2011. p. 4418-4421.
- Bassett DS, Bullmore E. Small-world brain networks. *Neuroscientist.* 2006; 12:512–523. [PubMed: 17079517]
- Bassett DS, Gazzaniga MS. Understanding complexity in the human brain. *Trends Cogn Sci.* 2011; 15:200–209. [PubMed: 21497128]

- Bassett DS, Nelson BG, Mueller BA, Camchong J, Lim KO. Altered resting state complexity in schizophrenia. *Neuroimage*. 2012; 59:2196–2207. [PubMed: 22008374]
- Bassett DS, Wymbs NF, Porter MA, Mucha PJ, Carlson JM, Grafton ST. Dynamic reconfiguration of human brain networks during learning. *Proc Natl Acad Sci U S A*. 2011; 108:7641–7646. [PubMed: 21502525]
- Bell AJ, Sejnowski TJ. An information-maximization approach to blind separation and blind deconvolution. *Neural Comput*. 1995; 7:1129–1159. [PubMed: 7584893]
- Betzal RF, Bassett DS. Multi-scale brain networks. *Neuroimage*. 2016
- Butts CT. Revisiting the foundations of network analysis. *Science*. 2009; 325:414–416. [PubMed: 19628855]
- Buxton RB, Frank LR. A model for the coupling between cerebral blood flow and oxygen metabolism during neural stimulation. *Journal of Cerebral Blood Flow and Metabolism*. 1997; 17:64–72. [PubMed: 8978388]
- Buxton RB, Wong EC, Frank LR. Dynamics of blood flow and oxygenation changes during brain activation: The balloon model. *Magnetic Resonance in Medicine*. 1998; 39:855–864. [PubMed: 9621908]
- Calhoun VD, Adali T, Pearlson GD, Pekar JJ. A method for making group inferences from functional MRI data using independent component analysis. *Hum Brain Mapp*. 2001; 14:140–151. [PubMed: 11559959]
- Calhoun VD, Kiehl KA, Pearlson GD. Modulation of temporally coherent brain networks estimated using ICA at rest and during cognitive tasks. *Hum Brain Mapp*. 2008; 29:828–838. [PubMed: 18438867]
- Cao M, Wang Z, He Y. Connectomics in psychiatric research: advances and applications. *Neuropsychiatric disease and treatment*. 2015; 11:2801–2810. [PubMed: 26604764]
- Contreras JA, Goni J, Risacher SL, Sporns O, Saykin AJ. The Structural and Functional Connectome and Prediction of Risk for Cognitive Impairment in Older Adults. *Curr Behav Neurosci Rep*. 2015; 2:234–245. [PubMed: 27034914]
- Craddock RC, James GA, Holtzheimer PE 3rd, Hu XP, Mayberg HS. A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Hum Brain Mapp*. 2012; 33:1914–1928. [PubMed: 21769991]
- de Reus MA, van den Heuvel MP. The parcellation-based connectome: limitations and extensions. *Neuroimage*. 2013; 80:397–404. [PubMed: 23558097]
- Erhardt EB, Allen EA, Wei Y, Eichele T, Calhoun VD. SimTB, a simulation toolbox for fMRI data under a model of spatiotemporal separability. *Neuroimage*. 2012; 59:4160–4167. [PubMed: 22178299]
- Erhardt EB, Rachakonda S, Bedrick EJ, Allen EA, Adali T, Calhoun VD. Comparison of multi-subject ICA methods for analysis of fMRI data. *Hum Brain Mapp*. 2011; 32:2075–2095. [PubMed: 21162045]
- Fornito A, Bullmore ET. Connectomics: a new paradigm for understanding brain disease. *Eur Neuropsychopharmacol*. 2015; 25:733–748. [PubMed: 24726580]
- Fornito A, Zalesky A, Breakspear M. Graph analysis of the human connectome: promise, progress and pitfalls. *Neuroimage*. 2013; 80:426–444. [PubMed: 23643999]
- Fornito A, Zalesky A, Breakspear M. The connectomics of brain disorders. *Nat Rev Neurosci*. 2015; 16:159–172. [PubMed: 25697159]
- Fornito A, Zalesky A, Pantelis C, Bullmore ET. Schizophrenia, neuroimaging and connectomics. *Neuroimage*. 2012; 62:2296–2314. [PubMed: 22387165]
- Friston KJ. Functional and effective connectivity: a review. *Brain connectivity*. 2011; 1:13–36. [PubMed: 22432952]
- Friston KJ, Holmes AP, Worsley KJ, Poline JP, Frith CD, Frackowiak RSJ. Statistical parametric maps in functional imaging a general linear approach. *human Brain Mapping*. 1995;22.
- Friston KJ, Mechelli A, Turner R, Price CJ. Nonlinear responses in fMRI: The balloon model, volterra kernels, and other hemodynamics. *Neuroimage*. 2000; 12:466–477. [PubMed: 10988040]

- He H, Yu Q, Du Y, Vergara V, Victor TA, Drevets WC, Savitz JB, Jiang T, Sui J, Calhoun VD. Resting-state functional network connectivity in prefrontal regions differs between unmedicated patients with bipolar and major depressive disorders. *Journal of affective disorders*. 2016; 190:483–493. [PubMed: 26551408]
- Himberg J, Hyvarinen A, Esposito F. Validating the independent components of neuroimaging time series via clustering and visualization. *Neuroimage*. 2004; 22:1214–1222. [PubMed: 15219593]
- Kiviniemi V, Starck T, Remes J, Long X, Nikkinen J, Haaapea M, Veijola J, Moilanen I, Isohanni M, Zang YF, Tervonen O. Functional segmentation of the brain cortex using high model order group PICA. *Hum Brain Mapp*. 2009; 30:3865–3886. [PubMed: 19507160]
- Liu Y, Liang M, Zhou Y, He Y, Hao Y, Song M, Yu C, Liu H, Liu Z, Jiang T. Disrupted small-world networks in schizophrenia. *Brain*. 2008; 131:945–961. [PubMed: 18299296]
- Mandeville JB, Marota JJA, Ayata C, Zaharchuk G, Moskowitz MA, Rosen BR, Weisskoff RM. Evidence of a cerebrovascular postarteriole windkessel with delayed compliance. *Journal of Cerebral Blood Flow and Metabolism*. 1999; 19:679–689. [PubMed: 10366199]
- Mason MJ, Fan G, Plath K, Zhou Q, Horvath S. Signed weighted gene co-expression network analysis of transcriptional regulation in murine embryonic stem cells. *BMC Genomics*. 2009; 10:327. [PubMed: 19619308]
- McKeown MJ, Jung TP, Makeig S, Brown G, Kindermann SS, Lee TW, Sejnowski TJ. Spatially independent activity patterns in functional MRI data during the stroop color-naming task. *Proc Natl Acad Sci U S A*. 1998a; 95:803–810. [PubMed: 9448244]
- McKeown MJ, Makeig S, Brown GG, Jung TP, Kindermann SS, Bell AJ, Sejnowski TJ. Analysis of fMRI data by blind separation into independent spatial components. *Hum Brain Mapp*. 1998b; 6:160–188. [PubMed: 9673671]
- McKeown MJ, Sejnowski TJ. Independent component analysis of fMRI data: examining the assumptions. *Hum Brain Mapp*. 1998; 6:368–372. [PubMed: 9788074]
- Mumford JA, Horvath S, Oldham MC, Langfelder P, Geschwind DH, Poldrack RA. Detecting network modules in fMRI time series: a weighted network analysis approach. *Neuroimage*. 2010; 52:1465–1476. [PubMed: 20553896]
- Power JD, Plitt M, Laumann TO, Martin A. Sources and implications of whole-brain fMRI signals in humans. *Neuroimage*. 2016
- Rajtmajer SM, Roy A, Albert R, Molenaar PCM, Hillary FG. A voxelwise approach to determine consensus regions-of-interest for the study of brain network plasticity. *Frontiers in Neuroanatomy*. 2015; 9
- Ribeiro de Paula D, Ziegler E, Abeyasinghe PM, Das TK, Cavaliere C, Aiello M, Heine L, di Perri C, Demertzi A, Noirhomme Q, Charland-Verville V, Vanhaudenhuyse A, Stender J, Gomez F, Tshibanda JL, Laureys S, Owen AM, Soddu A. A method for independent component graph analysis of resting-state fMRI. *Brain Behav*. 2017; 7:e00626. [PubMed: 28293468]
- Rubinov M, Sporns O. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*. 2010; 52:1059–1069. [PubMed: 19819337]
- Shirer WR, Ryali S, Rykhlevskaia E, Menon V, Greicius MD. Decoding subject-driven cognitive states with whole-brain connectivity patterns. *Cereb Cortex*. 2012; 22:158–165. [PubMed: 21616982]
- Smith SM. The future of FMRI connectivity. *Neuroimage*. 2012; 62:1257–1266. [PubMed: 22248579]
- Smith SM, Miller KL, Salimi-Khorshidi G, Webster M, Beckmann CF, Nichols TE, Ramsey JD, Woolrich MW. Network modelling methods for FMRI. *Neuroimage*. 2011; 54:875–891. [PubMed: 20817103]
- Stam CJ, Reijneveld JC. Graph theoretical analysis of complex networks in the brain. *Nonlinear Biomed Phys*. 2007; 1:3. [PubMed: 17908336]
- Stanley ML, Moussa MN, Paolini BM, Lyday RG, Burdette JH, Laurienti PJ. Defining nodes in complex brain networks. *Frontiers in computational neuroscience*. 2013; 7:169. [PubMed: 24319426]
- Yu Q, Allen EA, Sui J, Arbabshirani MR, Pearlson G, Calhoun VD. Brain Connectivity Networks In Schizophrenia Underlying Resting State Functional Magnetic Resonance Imaging. *Current topics in medicinal chemistry*. 2012; 12:2415–2425. [PubMed: 23279180]

- Yu Q, Erhardt EB, Sui J, Du Y, He H, Hjelm D, Cetin MS, Rachakonda S, Miller RL, Pearlson G, Calhoun VD. Assessing dynamic brain graphs of time-varying connectivity in fMRI data: application to healthy controls and patients with schizophrenia. *Neuroimage*. 2015; 107:345–355. [PubMed: 25514514]
- Yu Q, Plis SM, Erhardt EB, Allen EA, Sui J, Kiehl KA, Pearlson G, Calhoun VD. Modular Organization of Functional Network Connectivity in Healthy Controls and Patients with Schizophrenia during the Resting State. *Frontiers in systems neuroscience*. 2011a; 5:103. [PubMed: 22275887]
- Yu Q, Sui J, Kiehl KA, Pearlson G, Calhoun VD. State-related functional integration and functional segregation brain networks in schizophrenia. *Schizophr Res*. 2013a; 150:450–458. [PubMed: 24094882]
- Yu Q, Sui J, Liu J, Plis SM, Kiehl KA, Pearlson G, Calhoun VD. Disrupted correlation between low frequency power and connectivity strength of resting state brain networks in schizophrenia. *Schizophr Res*. 2013b; 143:165–171. [PubMed: 23182443]
- Yu Q, Sui J, Liu J, Plis SM, Kiehl KA, Pearlson G, Calhoun VD. Disrupted correlation between low frequency power and connectivity strength of resting state brain networks in schizophrenia. *Schizophrenia Research*. 2013c; 143:165–171. [PubMed: 23182443]
- Yu Q, Sui J, Rachakonda S, He H, Gruner W, Pearlson G, Kiehl KA, Calhoun VD. Altered topological properties of functional network connectivity in schizophrenia during resting state: a small-world brain network study. *PLoS One*. 2011b; 6:e25423. [PubMed: 21980454]
- Yu Q, Wu L, Bridwell DA, Erhardt EB, Du Y, He H, Chen J, Liu P, Sui J, Pearlson G, Calhoun VD. Building an EEG-fMRI Multi-Modal Brain Graph: A Concurrent EEG-fMRI Study. *Front Hum Neurosci*. 2016; 10:476. [PubMed: 27733821]

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.

Figure 1.

Locations of spatial sources in the simulations. A, scenario 1, twenty-nine SMs with little overlap. B, scenario 2, twenty-nine SMs are with little overlap plus three artificial SMs. C, scenario 3, twenty-nine SMs are largely overlapped. D, scenario 4, twenty-nine SMs are largely overlapped plus 3 artificial SMs.

Figure 2.

An analysis pipeline of this study. Simulated fMRI data are generated in four scenarios using SimTB. Group ICA is performed on the simulated data. Graphs with different nodes (ground truth nodes, ICA nodes, ROI nodes, and modified ROI nodes) are then built. Graph metrics between the ground truth graph and any other graph are compared.

Figure 3.

Measures of graphs with different nodes (GG: ground truth graph; IG: ICA graph; RG: ROI graph; MG: modified ROI graph) in all four scenarios. CS between RG and MG is not showing significant difference in scenario 4. All graph metrics are showing significant difference between any pair of graphs in all other cases (CS: connectivity strength; CC: clustering coefficient; GE: global efficiency)

Figure 4.

Correlation of graph measures across subjects between other graphs and the ground truth in all four scenarios. (IG: between ICA graph and ground truth graph; RG: between ROI graph and ground truth graph; MG: between modified ROI graph and ground truth graph; CS-corr: correlation of connectivity strength; CC-corr: correlation of clustering coefficient; GE-corr: correlation of global efficiency)