



OPEN ACCESS

EDITED BY
Meryem Ayse Yücel,
Boston University, United States

REVIEWED BY
Rickson C. Mesquita,
University of Campinas, Brazil
Yuanyuan Gao,
Boston University, United States

*CORRESPONDENCE
Daoqiang Zhang
dqzhang@nuaa.edu.cn

SPECIALTY SECTION
This article was submitted to
Brain Imaging Methods,
a section of the journal
Frontiers in Neuroscience

RECEIVED 22 July 2022
ACCEPTED 22 November 2022
PUBLISHED 07 December 2022

CITATION
Wen X, Yang M, Hsu L and Zhang D
(2022) Test-retest reliability
of modular-relevant analysis in brain
functional network.
Front. Neurosci. 16:1000863.
doi: 10.3389/fnins.2022.1000863

COPYRIGHT
© 2022 Wen, Yang, Hsu and Zhang.
This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Test-retest reliability of modular-relevant analysis in brain functional network

Xuyun Wen^{1,2}, Mengting Yang^{1,2}, Liming Hsu³ and
Daoqiang Zhang^{1,2*}

¹College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu, China, ²MIIT Key Laboratory of Pattern Analysis and Machine Intelligence, Nanjing, Jiangsu, China, ³Center for Animal MRI, University of North Carolina, Chapel Hill, Chapel Hill, NC, United States

Introduction: The human brain could be modeled as a complex network *via* functional magnetic resonance imaging (fMRI), and the architecture of these brain functional networks can be studied from multiple spatial scales with different graph theory tools. Detecting modules is an important mesoscale network measuring approach that has provided crucial insights for uncovering how brain organizes itself among different functional subsystems. Despite its successful application in a wide range of brain network studies, the lack of comprehensive reliability assessment prevents its potential extension to clinical trials.

Methods: To fill this gap, this paper, using resting-state test-retest fMRI data, systematically explored the reliabilities of five popular network metrics derived from modular structure. Considering the repeatability of network partition depends heavily on network size and module detection algorithm, we constructed three types of brain functional networks for each subject by using a set of coarse-to-fine brain atlases and adopted four methods for single-subject module detection and twelve methods for group-level module detection.

Results: The results reported moderate-to-good reliability in modularity, intra- and inter-modular functional connectivities, within-modular degree and participation coefficient at both individual and group levels, indicating modular-relevant network metrics can provide robust evaluation results. Further analysis identified the significant influence of module detection algorithm and node definition approach on reliabilities of network partitions and its derived network analysis results.

Discussion: This paper provides important guidance for choosing reliable modular-relevant network metrics and analysis strategies in future studies.

KEYWORDS

test-retest reliability, brain functional network, module detection, modular structure, network metric

Introduction

The human brain could be modeled as a complex network *via* functional magnetic resonance imaging (fMRI) consisting of brain regions as nodes linked by edges, estimated with functional connectivity (FC) (Friston, 2011; Park and Friston, 2013). As with other real-world connected systems, investigating the topology of brain functional interactions has profound implications for understanding its complex principles (Heuvel and Pol, 2010), e.g., underlying neural-mechanisms of human complex behaviors and cognitive capacities (Bullmore and Sporns, 2009) and pathogenesis of brain diseases (Sang et al., 2015; Dai et al., 2018), etc. The brain functional network can be studied from different spatial scales with graph theory tools, such as global network efficiency at macroscale (van den Heuvel et al., 2009), functional sub-systems' coordination at mesoscale (Chan et al., 2014), and regional properties at microscale (Cheng et al., 2015). In recent decades, network metrics at each scale have provided crucial insights for understanding the organization of human brain in health and disease.

With the extension of the use of FCs coupled to graph theory analysis to clinical trials becoming possible, increasing attention has been given to the reliability of these studies (Braun et al., 2012; Liang et al., 2012; Liao et al., 2013; Cao et al., 2014; Andellini et al., 2015; Du et al., 2015; Wang et al., 2017; Jin et al., 2018; Zeng et al., 2019). Test-retest reliability (TRT), which quantifies within-subject stability of a measure under multiple occasions in a group of subjects (Lavrakas, 2008), has been widely used for reliability assessment of fMRI data analysis. Based on this approach, the robustness of various graph-theory analyses at different spatial scales in functional brain networks has been extensively explored over the past decade. Macroscale studies found that most global network metrics, including global efficiency, clustering coefficient, characteristic path length, small-worldness, modularity, and hierarchy, are reproducible with reliability ranging from moderate to good (Braun et al., 2012). For nodal topological properties at microscale, their reliabilities displayed a spatially heterogeneous distribution, wherein regions in association areas exhibit moderate reliability (Liang et al., 2012), and nodal degree and nodal efficiency are more reliable than nodal betweenness (Liao et al., 2013). Compared to macroscale and microscale network metrics with their TRT reliability systematically assessed, network properties at mesoscale still lack a comprehensive reliability evaluation.

In recent years, module detection is becoming a popular approach for investigating mesoscale properties of brain functional networks (Gallos et al., 2012; Park and Friston, 2013; Sporns and Betzel, 2016; Baum et al., 2017). It decomposes the human brain into several functional modules, each of which represents groups of brain regions that perform specialized neurophysiological function (Power et al., 2011; Cole et al., 2014). Following suits, a variety of module-related network metrics were proposed, including modularity (Sporns and

Betzel, 2016), intra- and inter-modular FCs (Ma et al., 2017), within-modular degree, and participation coefficient (Guimera and Amaral, 2005). More importantly, these measurements have been used as biomarkers for brain disease diagnosis. For example, previous studies reported abnormal changes of modularity, intra- and inter-modular FCs in neurological ailments like Alzheimer's disease (Haan et al., 2012), Parkinson's disease (Cai et al., 2018), and depression (Andric and Hasson, 2015). Furthermore, the detected modular structure also influences the classification of nodes into different functional roles through within-modular degree (reflecting intra-modular coordination) and participation coefficient (reflecting inter-modular coordination) (Meunier et al., 2009a; Bertolero et al., 2015; Wen et al., 2019). These hub regions have been found to have increased susceptibility in case of brain disorders and cause major functional disruption in case of brain injury (Power et al., 2013). All abovementioned studies suggest the promising future of modular-relevant network analysis, and timely assessment of its reliability is crucial for its broader applications because only with adequate reliability, we can then expect the detected modular-relevant biomarkers to be reproducible.

The reliability of modular-relevant analysis in brain functional networks depends on two factors, i.e., the reliability of edges (i.e., FCs among brain regions) and the repeatability of network partition. TRT reliability regarding to FCs have been well assessed, which can be affected by many factors, including imaging acquisition parameters (Shirer et al., 2015), imaging preprocessing steps (Parkes et al., 2017), and FC estimation methods (Liang et al., 2012; Fiecas et al., 2013; Mejia et al., 2018), etc. A recent review literature (Noble et al., 2019) concludes that fMRI studies with high reliability tend to have the following characteristics: (1) eyes open and awake recordings of subjects during imaging scanning, (2) as much frames of fMRI data for each subject as possible, (3) test and retest measurements collected at shorter inter-scan intervals, and (4) connectivity estimated *via* full correlation. This study provides an important guidance for choosing the effective approach to improve the reliability of FCs. In contrast to systematic studies of FC reliability, how to guarantee the repeatability of network partition in brain functional networks remains unclear. It is well known that, the non-deterministic nature of module detection is an NP-hard problem (Xu et al., 2010; Gong et al., 2014), the reliability and quality of its resultant network partition is potentially affected *module searching strategy* and *network size*. Therefore, in reliability evaluation of modular-relevant analysis, the influence of module detection algorithm and node definition approach should be highlighted.

In this work, we systematically explored the reliability of various modular-relevant network measurements, including modular structure and five modular-relevant network properties. We aimed to investigate the following three questions: (1) *Are individual- and group-level modular-relevant network analysis methods reliable and reproducible?* (2) *Does the*

module detection algorithm affect reliability of modular-relevant measurements? (3) Does the node definition approach affect reliability of modular-relevant measurements? To this end, we used a test-retest fMRI dataset from 45 healthy subjects in Human Connectome Project (HCP) database and adopted three coarse-to-fine brain atlases to construct individual brain functional networks. For each kind of networks, we used four individual-level module detection methods and 12 group-level module detection methods to capture the corresponding network partitions and then calculated the related network measurements. TRT reliability was computed with intra-class correlation coefficient (ICC).

Materials and methods

A summary of the processing pipelines in this paper was given in **Figure 1**. Specifically, based on a publicly available HCP test-retest database, we first conducted module detection on each single subject and the population group to obtain individual- and group-level modular structures, and then calculated five popular modular-relevant network properties. After that, we assessed the reproducibility of all above estimations to give a comprehensive evaluation of TRT reliability of modular-relevant analysis methods in brain functional networks.

Participants

This study was carried out using test-retest dataset from the publicly available HCP database S1200 release¹ (Van Essen et al., 2013). It includes 45 healthy subjects (13 males, aged 22~35 years old) with full 3T imaging scans. All participants were free of current psychiatric or neurologic illness. Extensive descriptions of the data, please refer to Marcus et al. (2013).

Each subject has two resting-state fMRI (rs-fMRI) sessions, each of which consists of two runs scanned with the same multi-band sequence (Moeller et al., 2010) but the different readout directions (i.e., one with LR phase coding direction and the other with RL direction). The mean interval between two sessions of subjects is ~140 days. The first session is used as test data and the second session is used as retest data. All rs-fMRI data were collected with eye open and relaxed fixation on a projected bright cross-hair on a dark background. The other acquisition parameters for rs-fMRI were: TR = 720 ms, TE = 33.1 ms, flip angle = 52°, voxel size = 2 mm³ (isotropic), 72 slices, and total volumes = 1,200 (15 min). For more detailed information of rs-fMRI in HCP, please refer to the previous work (Smith et al., 2013).

¹ <https://www.humanconnectome.org/study/hcp-young-adult>

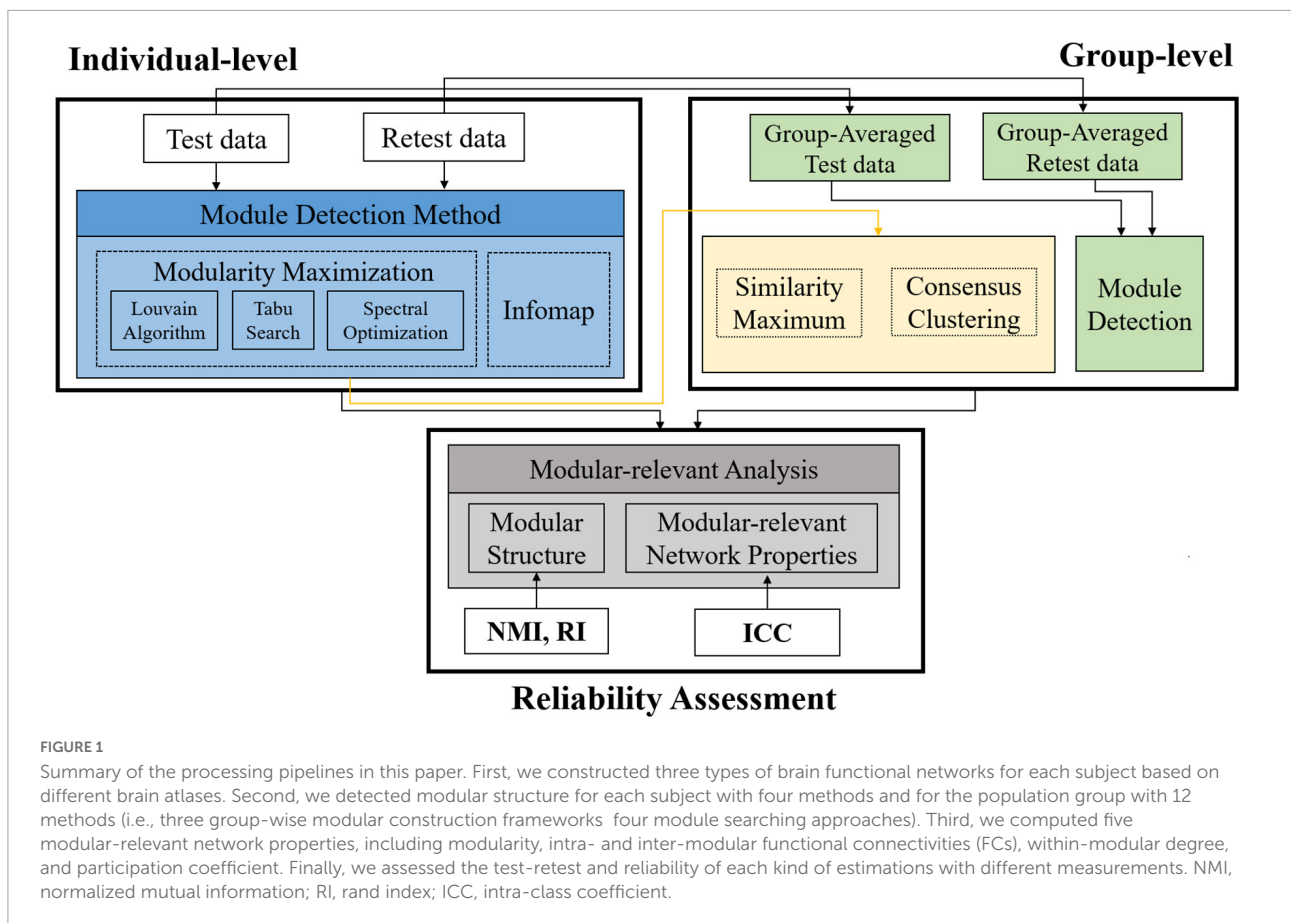
Data preprocessing

All rs-fMRI data were first preprocessed by “HCP fMRIVolume” minimal preprocessing pipeline (Glasser et al., 2013), including the following procedures: (1) gradient distortion correction, (2) head motion correction, (3) EPI distortion correction, (4) registration to the Montreal Neurological Institute (MNI) space, (5) intensity normalization to a global mean, and (6) masking out non-brain voxels. After that, we further adopted independent component analysis (ICA) based FIX Xnoiseifier to remove artifacts from fMRI data (Griffanti et al., 2014; Salimi-Khorshidi et al., 2014). During this cleanup, 24 head motion parameters (including 6 rigid-body motion parameters, their backward temporal derivatives, and squares of those 12 time series) and “bad components” estimated from ICA were regressed from blood oxygen level dependent (BOLD) signals of each scan.

Brain functional network construction

Considering the node definition approach may affect repeatability of module detection methods, this paper adopted three different types of brain atlases to define nodes in brain functional networks. These three atlases were defined with different methods (or brain features) and parcellate the human brain from a coarse to fine scale. They are (1) *automated anatomical labeling (AAL)* atlas derived from anatomical landmarks and parcellate the human brain into 90 brain regions (Tzourio-Mazoyer et al., 2002); (2) *Shen’s FC-based atlas* generated from rs-fMRI data and includes 217 brain regions (Shen et al., 2013); (3) *Zalesky’s random atlas* defined with the random parcellation method and consists of 1,024 brain regions (Zalesky et al., 2010). Notably, we excluded cerebellar regions of AAL atlas and Shen atlas due to the difficulty of cerebellum registration.

After obtaining parcels with brain atlas, we constructed the corresponding functional brain network for each scan using the following steps. Specifically, we first extracted rs-fMRI time series for each brain region (i.e., brain parcel) by averaging blood oxygen level dependent (BOLD) signals of all voxels within it. We then computed the Pearson’s correlation between each pair of time series and used them as edges of brain functional networks. After that, we conducted *r*-to-*z* value conversion on each FC matrix with Fisher-*z* transformation to improve the data normality. Since each session has two runs, each subject has two FC matrices, one of which was generated from rs-fMRI data with LR readout direction and the other from RL direction. We followed the suggestion in Smith et al. (2013) to construct the representative FC network for each subject by averaging LR and RL FC matrices. Finally, for each subject, we set all the negative values of FC network of each subject to zeros for their unclear biological



meanings (Garrison et al., 2015). In summary, each subject in each dataset would have a 90×90 FC matrix, a 217×217 FC matrix, a $1,024 \times 1,024$ FC matrix. It should be noted that the weighted brain networks were used in the following experiments.

Module detection algorithms

The module detection in brain networks is broadly divided into two categories, one of which is to detect modular structure for one single subject (i.e., individual-level module detection) and the other is to generate a common network partition for a cohort of subjects (i.e., group-level module detection). For a comprehensive evaluation, we assessed TRT reliability of modular-relevant analysis at both individual and group levels. For each category, we selected a variety of widely used module detection methods to generate different network partition results to evaluate the influence of module detection algorithm on reliability. To ensure the quality of network partitions, we only kept top 10% strongest positive connections to ensure the sparsity of FC networks (Wen et al., 2019; Zhang et al., 2021). After obtaining individual-

and group-level modular structures, we projected them on brain surface for further comparison between different methods and different brain atlases. The brain parcellation template generated on a large population by using the clustering method was adopted as ground truth (Yeo et al., 2014), whose spatial distribution is displayed in [Supplementary Figure 1](#).

Individual-level module detection algorithms

Among various module detection algorithms, “modularity maximization” is the most widely adopted method in brain module detection in FC-related studies (Sporns and Betzel, 2016). This approach divides the network into modules by maximizing modularity quality function, during which the optimization strategy has an effect on the quality of the solution. Considering this situation, this study employed three different approaches to search for optimal modularity, including Louvain algorithm (Blondel et al., 2008), Tabu search (Arenas et al., 2008), and spectral optimization (Newman, 2006). For each approach, we repeated module detection 50 times on each network, and used consensus clustering to construct consistent modules to avoid their uncertainty (Bassett et al., 2013). In addition to the above three methods, we also used

Infomap algorithm for module detection due to its good performance in other domains (Rosvall and Bergstrom, 2008). Unlike the first three methods, Infomap detects modules by minimizing the description length of random walks in the network. In summary, for each single subject, we employed four module detection methods to capture its modular structure.

Group-level module detection algorithms

Currently, there are three popular group-wise frameworks for determining modular structure on the population group, which can be named as “Ave,” “SimMax,” and “Consensus,” respectively. “Ave” based methods first construct a representative brain network by averaging FC matrixes across all subjects, and then apply module detection on this averaged network for generating group-level modular structure (Meunier et al., 2009b). “SimMax” based methods conduct module detections on all subjects and choose the partition that is most similar to others as the group-level result (Meunier et al., 2009b). “Consensus” based methods first perform module detection for each subject to construct the association matrix, where each element represents the probability of assigning a pair of nodes to the same module across subjects, and then employ the second clustering on association matrix to generate group-level network partition (Bassett et al., 2013). By combining three group-wise modular construction schemes with four individual-level module detection methods, there are 12 group-level module detection methods, named as Louvain + Ave, Tabu + Ave, Spectral + Ave, Infomap + Ave, Louvain + SimMax, Tabu + SimMax, Spectral + SimMax, Infomap + SimMax, Louvain + Consensus, Tabu + Consensus, Spectral + Consensus, Infomap + Consensus.

Modular-relevant network properties

Based on the detected module structure, we computed five popular network properties, including modularity, intra-modular FC, inter-modular FC, within-modular degree, and participation coefficient. Among these metrics, the first three assessed the information coordination capacity among functional modules of the whole brain network (including separation and integration), and the latter two evaluated the topological role of nodes within module and between different modules, which are regional level measurements. There are differences for network property calculation between individual and group levels. Specifically, for a given network, individual-level network metrics were computed based on modular structure detected on itself, while group-level network metrics were measured by warping group-wise representative on the network. All positive connections in FC network were used for computing network properties.

Modularity

It is defined as the difference between the fraction of edges within the given modules and the expected fraction if edges are distributed at random (Newman, 2006), which is formulated as

$$Q = \frac{1}{2m} \sum_{ij} \left(a_{ij} - \frac{k_i k_j}{2m} \right) \delta(\sigma_i, \sigma_j) \quad (1)$$

where a_{ij} represents the link strength between nodes i and j , and k_i indicates the degree of node i . m is the total link strength of the whole network. The Kronecker delta function, $\delta(\sigma_i, \sigma_j)$, is equal to 1 when nodes i and j belong to the same module; otherwise, it is set to 0. Modularity is a network-level measurement for evaluating network segregation.

Intra-modular functional connectivity

It measures the total link strength of FC within the modules, which is defined as the averaged FC strength across all the within-modular connections (Chan et al., 2014). This index is generally used for evaluate network segregation.

Inter-modular functional connectivity

It measures the total link strength of FC between different modules, which is defined as the averaged FC strength across all the edges that linked different modules (Chan et al., 2014). This index can be used as an indicator for network integration.

Within-modular degree

It characterizes the importance of a region within its own module (Guimera and Amaral, 2005). Given a modular partition $\mathcal{C} = \{c_1, c_2, \dots, c_A\}$, the within-module degree WD of a node n in module c_i is formulated as

$$WD_n = \frac{K_{c_i}^n - \overline{K_{c_i}}}{\delta_{K_{c_i}}} \quad (2)$$

where $K_{c_i}^n$ is the total FC strength between node n and the other nodes in module c_i , and $\overline{K_{c_i}}$ and $\delta_{K_{c_i}}$ represent the average and standard deviation of $K_{c_i}^n$ across all nodes in module c_i , respectively.

Participation coefficient

It evaluates the region's importance in connecting with different modules (Guimera and Amaral, 2005). Given a modular partition $\mathcal{C} = \{c_1, c_2, \dots, c_A\}$, the participation coefficient of a node n in module c_i is defined as

$$PC_n = 1 - \sum_{i=1}^A \left(\frac{K_{c_i}^n}{k_n} \right)^2 \quad (3)$$

where $K_{c_i}^n$ is the total FC strength between node n and the other nodes in module c_i and k_n is the sum of FC strength connecting node n .

Reliability assessment

To systematically assess the reliability of modular-relevant analysis in brain functional networks, we computed the reproducibility of three network analysis methods by using different evaluation indexes, including modular structure and modular-relevant network metrics.

Modular structure

The repeatability of modular structure between test and retest data was evaluated with normalized mutual information (NMI, Ana and Jain, 2003) and rand index (RI, Rand, 1971). Suppose there are two partitions $\mathcal{A} = \{c_1, c_2, \dots, c_A\}$ and $\mathcal{B} = \{c_1, c_2, \dots, c_B\}$. NMI is an information-theory-based index, which is calculated as follows:

$$NMI(\mathcal{A}, \mathcal{B}) = \frac{-2 \sum_{i=1}^A \sum_{j=1}^B N_{ij} \log\left(\frac{N_{ij}N}{N_i N_j}\right)}{\sum_{i=1}^A N_i \log\left(\frac{N_i}{N}\right) + \sum_{j=1}^B N_j \log\left(\frac{N_j}{N}\right)} \quad (4)$$

In Eq. 4, N is confusion matrix, where each element N_{ij} denotes the number of nodes in module c_i of partition \mathcal{A} that appears in module c_j of partition \mathcal{B} . N_i is the sum over row i of matrix N and N_j is the sum over column j . N is the sum over all elements in N . Different with NMI, RI calculates the fraction of correctly classified elements to all elements, which can be formatted as

$$RI(\mathcal{A}, \mathcal{B}) = \frac{2(n_{11} + n_{00})}{n(n-1)} \quad (5)$$

where n_{11} is the number of node pairs that are in the same mode under \mathcal{A} and \mathcal{B} , and n_{00} is the number of node pairs that in different modules under \mathcal{A} and \mathcal{B} . n is the number of nodes in FC networks. Both NMI and RI scores range from 0 to 1, where 1 corresponds to identical network partitions between \mathcal{A} and \mathcal{B} , and 0 indicates two entirely dissimilar network partitions.

Modular-relevant network properties

The TRT reliability of five modular-relevant network properties were evaluated using two-way mixed single-measures intra-class correlation coefficient (ICC) (Shrout and Fleiss, 1979), which is defined as

$$ICC = \frac{BMS - EMS}{BMS + (s-1)EMS} \quad (6)$$

where BMS is the between-subjects mean square, EMS denotes the error mean square and s is the number of repeated measurement (here, $s = 2$). ICC coefficient conceptually ranges from 0 (not reliable at all) to 1 (perfectly consistent between repeated measurements), but its estimation can be negative in a few cases. In this paper, we followed previous literatures and set the negative ICC values to be zeros (Zhang et al., 2011, 2017). Based on the values of ICC, reliability is usually categorized

as poor (0~0.2), fair (0.2~0.4), moderate (0.4~0.6), good (0.6~0.8), and excellent (>0.8) (Koch, 1977; Chen et al., 2015).

Results

Reliability of individual-level modular-relevant analysis

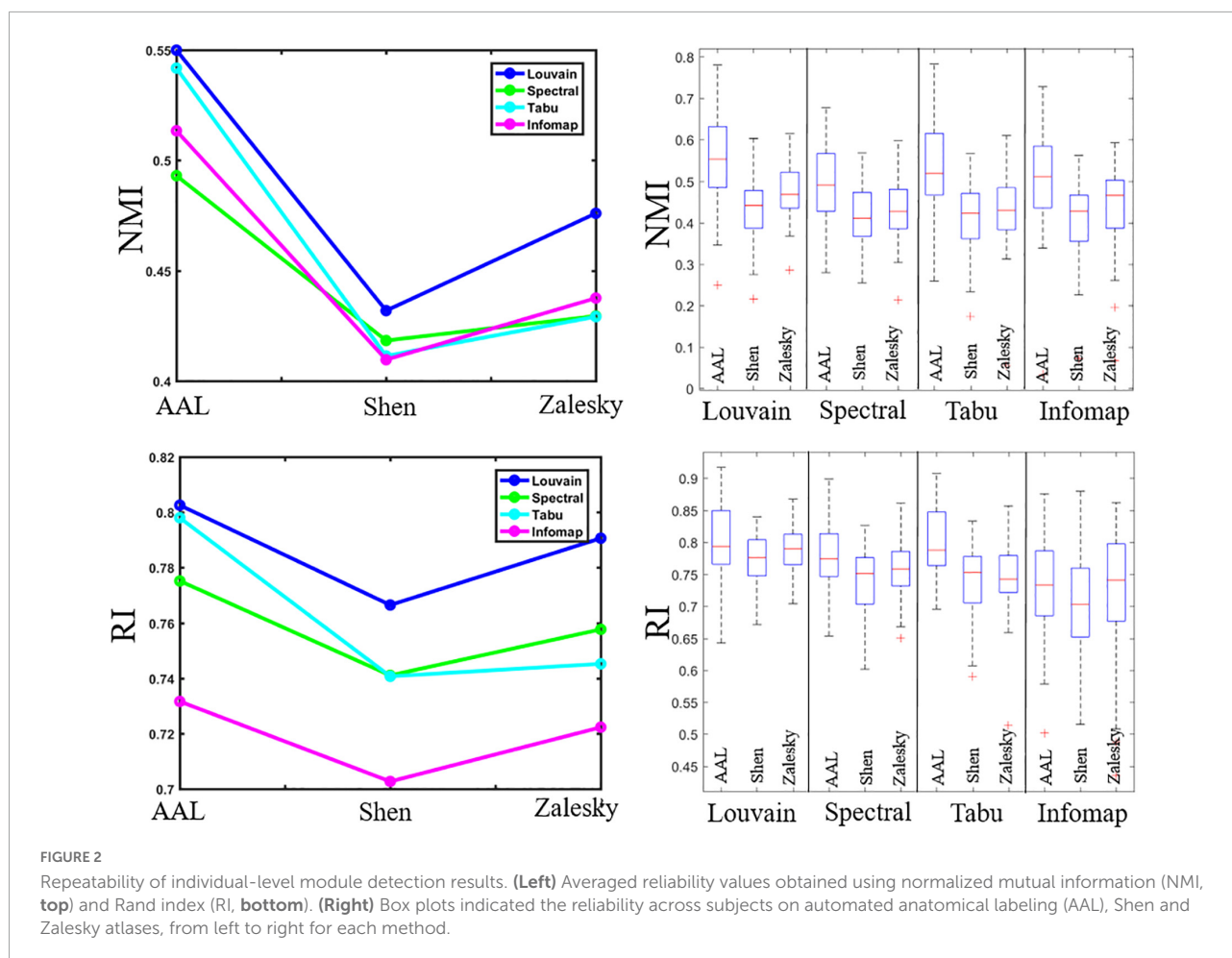
In this section, we reported TRT reliability results of modular-relevant analysis on individual level, including modular structure and five modular-relevant network properties. Modular structure results were presented in Figures 2, 3. Modular-relevant network metric results, including modularity, intra- and inter-modular FCs, within-module degree and participation coefficient, were displayed in Figure 4 and Table 1.

Modular structure

Figure 2 displayed the repeatability of individual-level module detection results computed by NMI and RI. For ease of comparison, we reported averaged evaluation measures in the form of line graphs for all computed results, and also showed box plots alongside line graphs to represent the variability across subjects.

In Figure 2, we observed that the repeatability of network partition results depends heavily on module detection algorithms. We could draw two conclusion. *First*, modularity maximization based methods yielded higher reliability than Infomap algorithm in almost all experiments, suggesting it is more suitable for module detection in brain functional network. *Second*, of three estimated modularity optimization strategies, Louvain algorithm displayed the highest reliability in all experiments, and its advantage became more and more prominent as the network size increases. Specifically, on the brain network with 90 nodes (i.e., AAL atlas), the Louvain algorithm performed slightly better than Tabu method, but with the number of nodes increases to 217 (i.e., Shen atlas), their gap became larger. When the number of nodes increases to 1,024 (i.e., Zalesky atlas), the performance of Louvain algorithm was significantly better than the other methods. This result implies that the choice of module detection algorithm is important to ensure the reliability of module detection results, especially for the large-scale network.

Regarding the brain atlas, we found that AAL atlas yielded the highest reliability in all experiments. This result is due to its much smaller number of nodes than the other two atlases reduces the difficulty of module detection. A surprising finding is that, in most of cases, Shen atlas with fewer nodes generated lower NMIs and RIs than Zalesky atlas with more nodes. This finding suggests that the brain atlas construction approach



can also affect the repeatability of module detection on brain functional networks.

Furthermore, to verify the availability of module detection results by different methods on different brain atlases, we randomly selected an exemplar subject (subject ID: 114823) and visualized his/her module detection results on brain surface in **Figure 3**. We observed that (1) most of module detection results captured four available functional subsystems, including visual network (VN), sensorimotor network (SMN), default mode network (DMN), and frontal-parietal network (FPN), and (2) brain regions in each module derived from AAL atlas were more spatially contiguous than those from Shen and Zalesky atlases. Compared with Yeo's atlas (Yeo et al., 2014), the modular structures detected on Shen and Zalesky atlases were closer to the true spatial distribution of functional systems than the AAL atlas. As red line labeled, AAL atlas incorrectly assigned a large brain region in frontal areas in SMN to DMN.

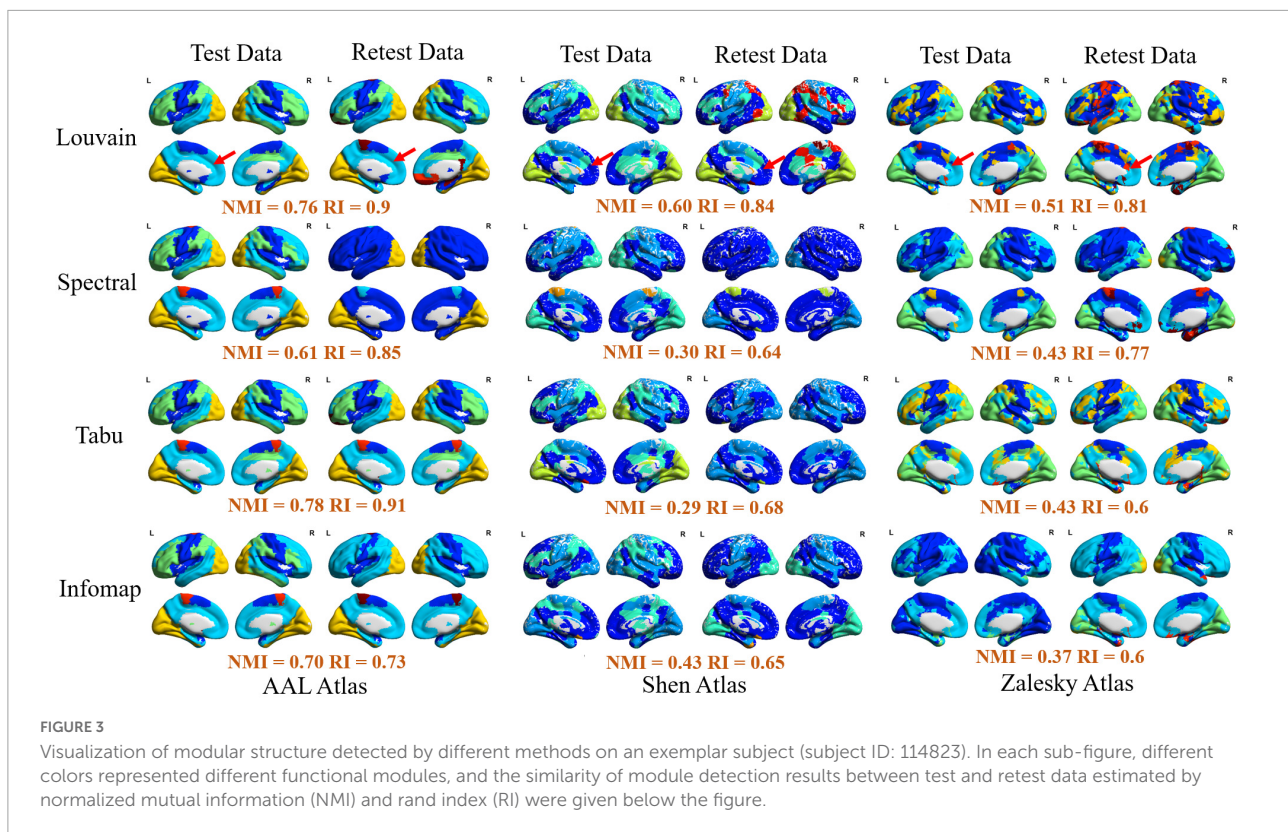
Modular-relevant network properties

Test-retest reliability on five network properties derived from individual-level module detection were given in **Figure 4**, where the results of modularity, intra- and inter-modular FCs

were visualized as bar figures and the results of within-modular degree and participation coefficient were displayed as box plots. For ease of comparison, for each network metric estimated on each brain atlas, we ranked four individual-level module detection methods based on their ICC values as summarized in **Table 1**.

As shown in **Figure 4** and **Table 1**, among five estimated network properties, modularity yielded the highest ICC values with the reliability above good level on all three brain atlases, and the reliability of the other four properties ranged from moderate to good. But overall, these network metrics derived from individual-level module detection in brain functional networks are reliable and reproducible.

We also observed a significant impact of module detection algorithm on the reliability of modular-relevant network metrics. Consistent with the repeatability of network partitions, Louvain algorithm performed best in reliability assessment experiments, and Infomap algorithm performed the worst. Specifically, as shown in **Table 1**, Louvain algorithm ranked first in 11/15 cases, whereas Infomap algorithm ranked last in 12/15 cases. For the other two methods, Tabu method yielded higher



ICC values than spectral method on smaller networks (i.e., AAL atlas and Shen atlas), but lower ICC values on larger networks (i.e., Zalesky atlas), indicating it cannot handle the large-scale brain network.

Regarding to the influence of brain atlas on network metrics, a surprising finding is that, the reliability of modular-relevant network metrics was not significantly changed with the increase of network size, except within-modular degree. Take results from Louvian algorithm as the example for illustration. As shown in **Figure 4**, although ICC values of modularity, intra- and inter-modular FCs, and participation coefficient displayed variation across different brain atlases, these changes were not very significant, and their reliabilities were still in the same level. As for within-modular degree, with the number of nodes changing from 90 to 1,024, its reliability gradually increased from moderate to excellent.

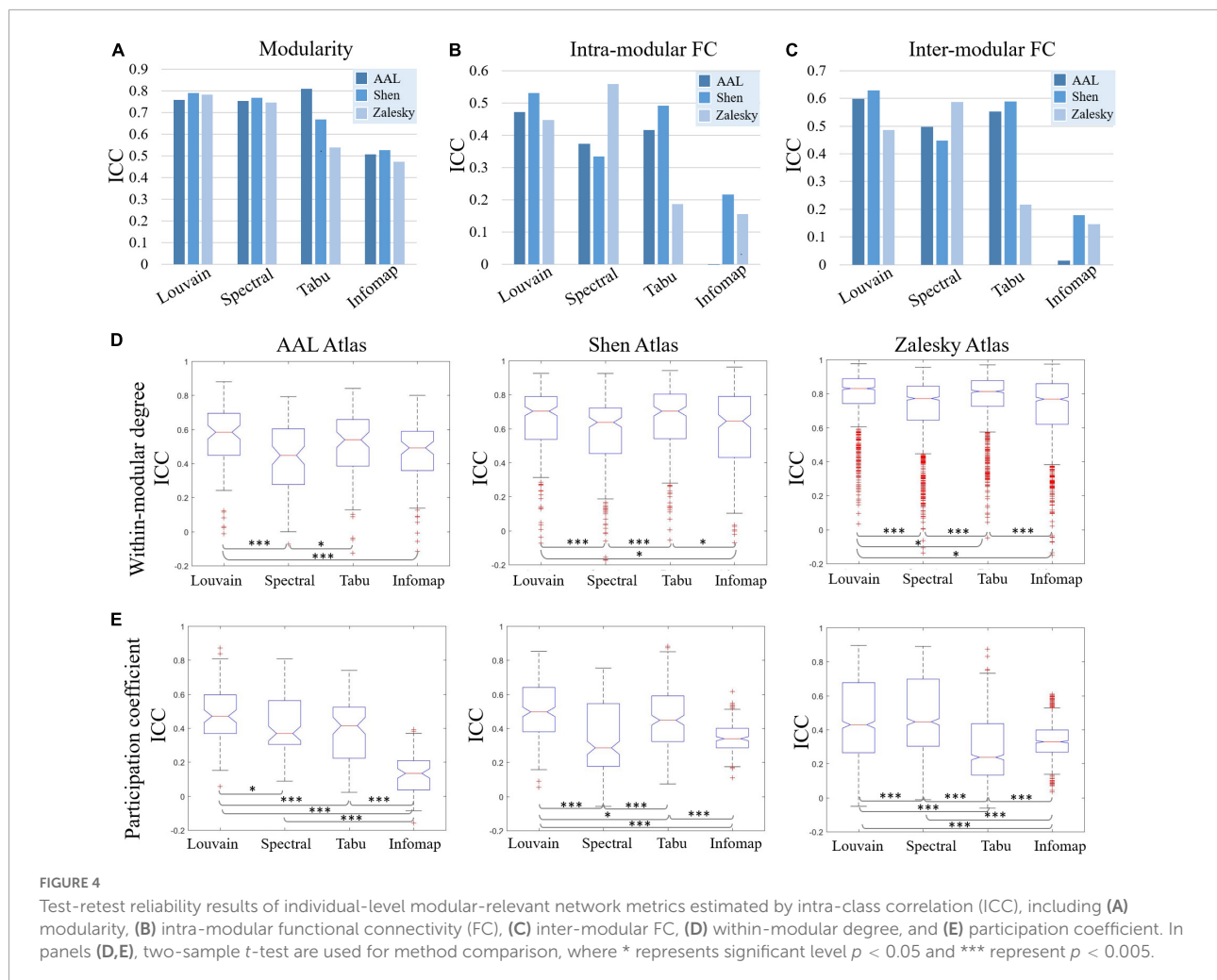
Reliability of group-level modular-relevant analysis

In this section, we systematically evaluated TRT reliability of group-level modular-relevant analysis, including modular structure, modularity, intra- and inter-modular FCs, within-module degree and participation coefficient. Modular structure results were displayed in **Figures 5, 6**. Five network metric results were given in **Figure 7** and **Table 2**.

Modular structure

Figure 5 displayed the similarity of group-level modular structure between test and retest data obtained with 12 different module detection methods, including four average based methods, four similarity maximum based methods, and four consensus clustering based methods. In **Figure 5**, an obvious finding is that, both NMI and RI values derived from similarity maximum based methods were much lower than those from average based and consensus clustering based methods. It indicates that, in group-level module detection, the methods considering all subjects can yield more robust modular structure than those by selecting a representative subject. Further comparisons among four average based methods, we found the inconsistent result with individual-level analysis, i.e., none of the module searching approach displayed better performance than others in terms of repeatability of modular structure. The similar finding was also found among four consensus clustering based methods. All these results suggest that the repeatability of group-level network partition were sensitive to group-wise modular structure construction framework rather than module searching strategy.

Additionally, the effect of brain atlas was not exactly same as that found in individual level. In subject level, the modular structure detected on Shen atlas displayed the lowest repeatability among three brain atlases in most of experiments (as shown in **Figure 2**). However, in group-level analysis, the



similar results were only found on three average based methods, i.e., Spectral + Ave, Tabu + Ave, and Infomap + Ave. For the other two frameworks, the repeatability of the resultant network partitions was decreased with the increase of network size.

Finally, for each atlas, we also projected 12 group-level network partition results obtained from different module detection algorithms on brain surface for the further comparison. The visualizations on AAL, Shen and Zalesky atlases were, respectively, given in [Figure 6](#) and [Supplementary Figures 2, 3](#). Similar to individual-level results, four functional subsystems could be also successfully captured by most of group-level module detection methods, i.e., VN, SMN, DMN, and FPN. Taking network partitions from Louvain algorithm as example, we further compared differences of spatial distribution of functional subsystems between different group-wise module construction schemes and different brain atlases. The network partition map generated by [Yeo et al. \(2014\)](#) were adopted as ground truth. In [Figure 6](#), we observed two results. *First*, average and consensus clustering based methods yield the similar network partition results. *Second*,

functional modules generated on Shen atlas were closer to the real spatial distributions than the other two atlases. For example, the results on AAL and Zalesky atlases assigned superior frontal areas to DMN, but they were not constituent brain areas of this system.

Modular-relevant network properties

We reported TRT reliability results of modularity, intra- and inter-modular FCs, within-modular degree, and participation coefficient estimated from group-level module detection in [Figure 7](#). Besides, in all 15 experiments (5 network metrics \times 3 brain atlases), we further sorted 12 group-level module detection methods based on ICC values and visualized the top four in [Table 2](#) for the further comparisons.

As shown in [Figure 7](#) and [Table 2](#), the reliabilities of five group-level modular-relevant network metrics were ranged from moderate to good with modularity yielded the highest ICC values. These results suggest that, the estimation of network properties from group-level module detection in brain functional networks were reliable.

TABLE 1 Rankings of four individual-level module detection algorithms on five network metrics estimated on three brain atlases.

Metrics	Atlas	ICC rank
Modularity	AAL	Tabu (excellent) > Louvain (good) > Spectral (good) > Infomap (moderate)
	Shen	Louvain (good) > Spectral (good) > Tabu (good) > Infomap (moderate)
	Zalesky	Louvain (good) > Spectral (good) > Tabu (moderate) > Infomap (moderate)
Intra-modular FC	AAL	Louvain (moderate) > Tabu (moderate) > Spectral (fair) > Infomap (poor)
	Shen	Louvain (moderate) > Tabu (moderate) > Spectral (fair) > Infomap (poor)
	Zalesky	Spectral (moderate) > Louvain (moderate) > Tabu (fair) > Infomap (poor)
Inter-modular FC	AAL	Louvain (moderate) > Tabu (moderate) > Spectral (moderate) > Infomap (poor)
	Shen	Louvain (good) > Tabu (moderate) > Spectral (moderate) > Infomap (poor)
	Zalesky	Spectral (moderate) > Louvain (moderate) > Tabu (poor) > Infomap (poor)
Within-module degree	AAL	Louvain (moderate) > Tabu (moderate) > Infomap (moderate) > Spectral (moderate)
	Shen	Louvain (good) > Tabu (good) > Spectral (good) > Infomap (good)
	Zalesky	Louvain (excellent) > Tabu (excellent) > Spectral (good) > Infomap (good)
Participation coefficient	AAL	Louvain (moderate) > Tabu (moderate) > Spectral (fair) > Infomap (poor)
	Shen	Louvain (moderate) > Tabu (moderate) > Infomap (fair) > Spectral (moderate)
	Zalesky	Spectral (moderate) > Louvain (moderate) > Infomap (fair) > Tabu (fair)

Poor reliability ($0 < ICC < 0.2$), fair reliability ($0.2 < ICC < 0.4$), moderate reliability ($0.4 < ICC < 0.6$), good ($0.6 < ICC < 0.8$), and excellent reliability ($ICC > 0.8$).

We further compared ICC values of 12 group-level module detection algorithms in each experiment, and found that reliability was relevant to group-wise module construction framework rather than module searching approach. Specifically, as shown in Table 2, among all listed module detection algorithms, there are 31 consensus clustering based methods, 24 average based methods, and 5 similarity maximum based methods, indicating the superiority of consensus clustering framework in terms of reliability assessment. However, the further comparison between four module searching approaches within each group-wise scheme did not find the significant difference between each other. Taking the consensus clustering framework as example, the proportions for Louvain algorithm, spectral method, Tabu method and Infomap algorithm were 7/31, 8/31, 11/31, and 5/31, respectively, and none of the method displayed significantly better performance than the others.

Regarding to the brain atlas, the similar results were found with individual-level analysis, that is, the reliability of most network properties derived from group-level modular detection was not significantly changed with the increase of network size. The only network metric with dramatic change was inter-modular FC, whose ICC values were slightly decreased with the increase of network size.

Discussion

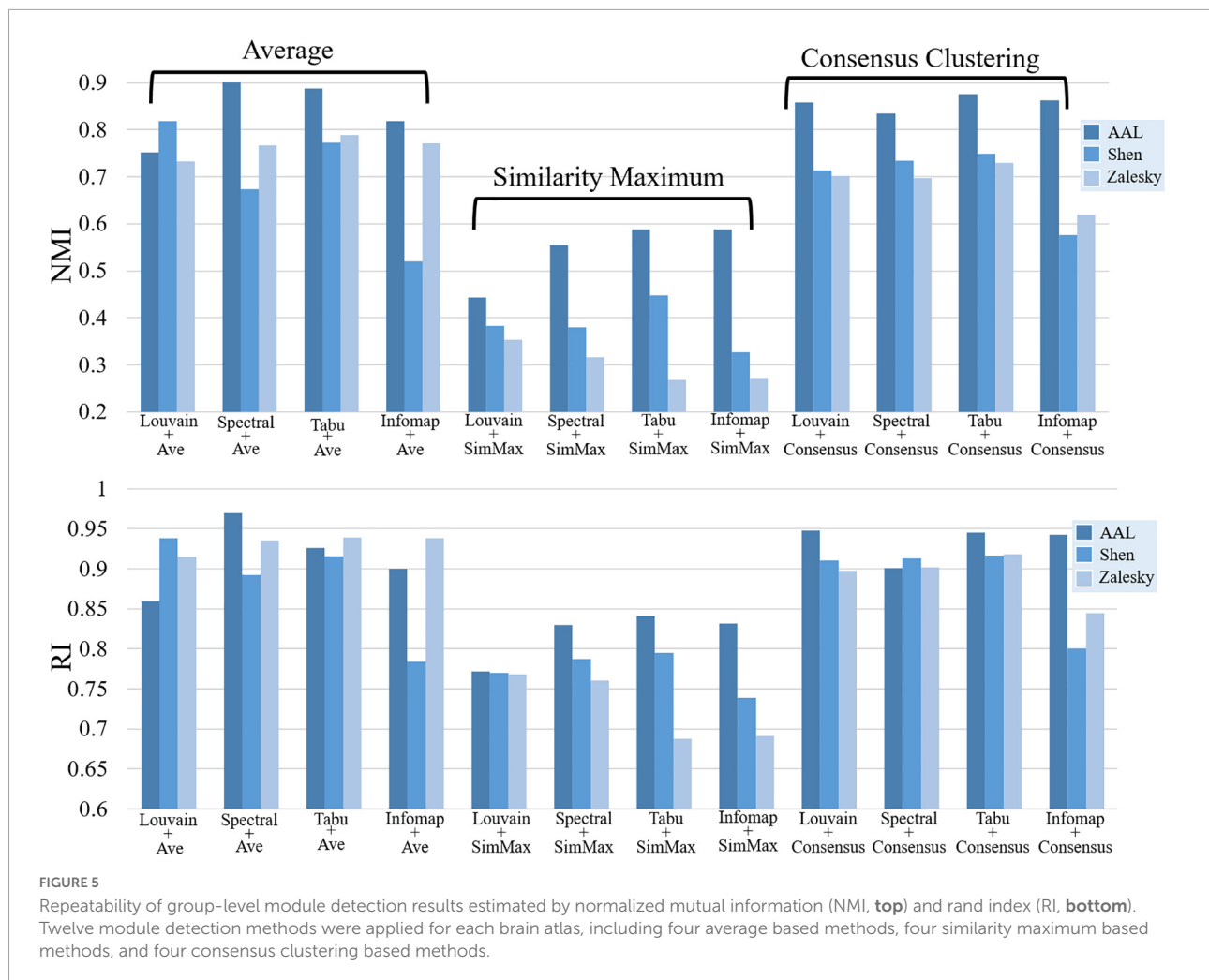
The main purpose of this study was to systematically investigate the reliability of modular-relevant analysis in brain functional network from rs-fMRI data, including modular structure and popular module-derived network metrics. We were also interested in evaluating how module detection algorithm and node definition approach affect the reliability.

Through a series of experiments, we drew the following conclusion. *First*, five modular-relevant network metrics were reliable and reproducible with TRT reliability ranging from moderate to good. *Second*, the reliability of network metrics could be affected by module detection algorithm but there was difference between individual- and group-level analyses. Individual-level reliabilities were sensitive to module searching strategy, whereas group-level reliabilities depended heavily on group-wise modular structure construction framework. *Third*, node definition approach was another potential affecting factor for reliability of modular-relevant analysis, and both the generation approach and parcellation granularity of brain atlas could impact ICC values.

To our best knowledge, there is no such comprehensive reliability study before specific to mesoscale module-related network measuring approaches from rs-fMRI data. Although many studies have investigated reliabilities of FCs and network properties at macroscale and microscale (Shehzad et al., 2009; Anderson et al., 2011; Shou et al., 2013; Mueller et al., 2015; Shirer et al., 2015; Laumann et al., 2017; Tomasi et al., 2017), they cannot guarantee the repeatability of module detection results and their derived network analysis approaches. This study filled this gap that focused on this important mesoscale graph theory tool to provide a potential guidance for the usage of module detection in brain functional networks in future studies.

Reliability of modular-relevant network analysis

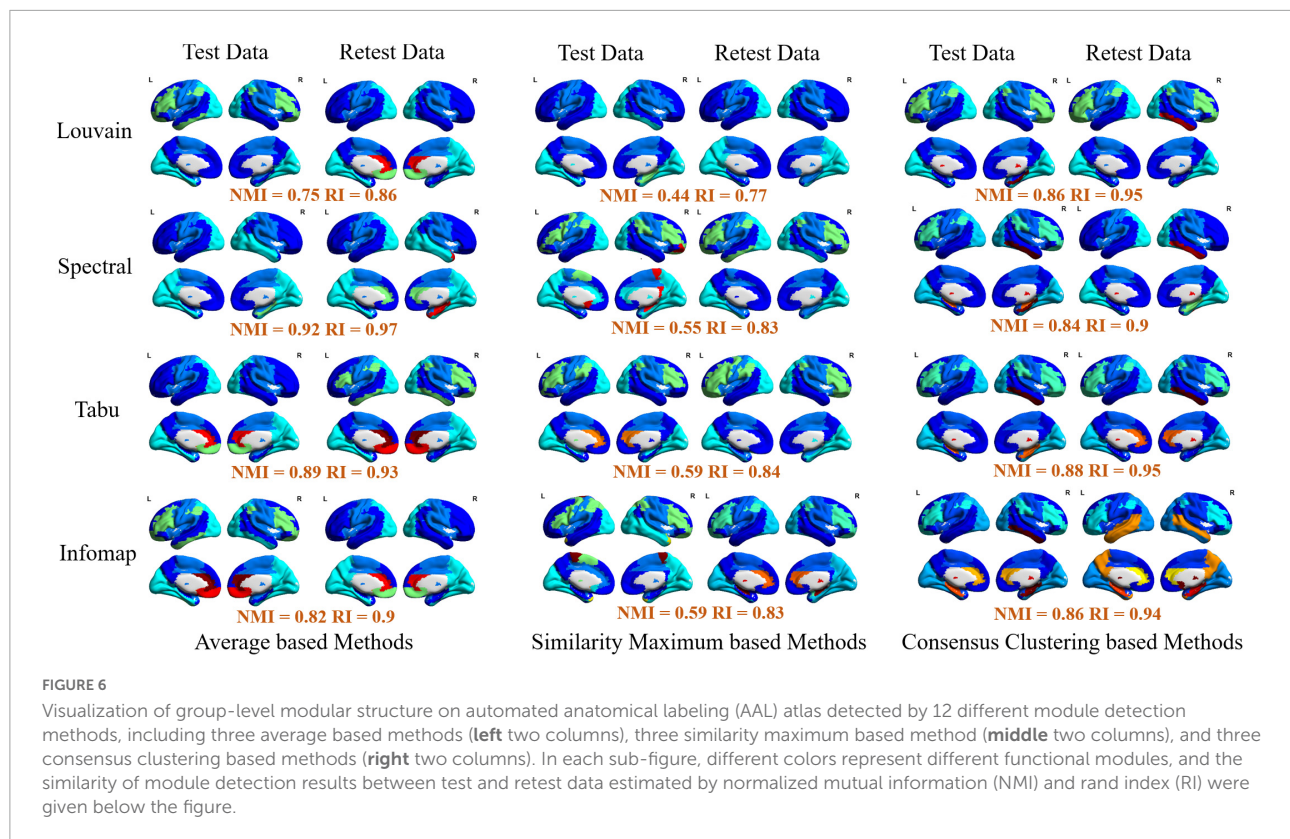
The results from a series of reliability evaluation experiments indicate five popular network measurements from functional modules are reliable and reproducible. This



finding not only fills the gap of mesoscale reliability assessment of brain functional networks, but also provides an evidence for the potential application of module detection in future clinical trials. Of the five network metrics, we noted that modularity yields the highest ICC values. This is because it is calculated by directly comparing the modular structure of the given network and stochastic model (Sporns and Betzel, 2016), thus avoiding the secondary introduction of unreliable FCs to reduce reliability values. In addition, the present work reported good-to-excellent reliability in modularity, which were much higher than the previous studies (Braun et al., 2012; Du et al., 2015). This discrepancy may due to different imaging acquisition parameters and network analysis approaches. Specifically, compared with Braun et al. (2012), the higher reliability in our work to some extent is benefited from better rs-fMRI data acquisition parameters, including subject's eye open and much more frames per subject, etc. A recent FC reliability literature review in support of this view. It using the meta-analysis method concluded that studies with rs-fMRI collected in subject awake, eye open and more within-subject

data tend to generate more reliable FCs (Noble et al., 2019), and thus more reproducible network properties. As for Du et al. (2015), the main reason for its poorer reliability is due to the large voxel-wise brain functional network intensifies the non-determinacy of module detection algorithm, making it difficult to obtain a stable and accurate modular partition. This gives us a hint of the care should be taken of voxel-wise analysis when using graph theory tools of module detection in brain functional networks.

Regarding regional-level module-derived network metrics, we found both within-modular degree and participation coefficient from brain functional networks were less reliable than those from structural networks. Comparing to moderate-to-good reliability from rs-fMRI in present work, a recent reliability assessment in structural brain network from diffusion tensor imaging (DTI) using the similar approach reported good TRT reliability in within-modular degree and excellent reliability in participation coefficient on AAL atlas (Dimitriadis et al., 2021). This finding proves again that structural connectivity (SC) from DTI is more robust than FC



from rs-fMRI as previous studies announced good reliability in SC of averaged ICC = 0.62 (Buchanan et al., 2014) versus poor reliability in FC of averaged ICC = 0.29 (Noble et al., 2019).

Effects of module detection algorithm

We found that module detection algorithm has a significant influence on the reliabilities of modular-relevant network measurements. This is not surprising because the repeatability of network partition results on test and retest data heavily determines the reliability of network attributes derived from them. An interesting finding is that module detection method affects individual- and group-level analyses through different manners. Individual-level results reported Louvain algorithm performed better than the other methods in almost all experiments (as shown in Figure 4), suggesting module searching strategy has a great impact on reliability of network metrics. However, group-level experiments did not find the superiority of any module discovery method, but the group-wise representative module determination framework significantly affect the ICC values (e.g., average and consensus clustering based frameworks perform much better than the similarity maximum as shown in Figure 7). This difference is due to the fact that individual-level module detection algorithm needs to ensure the partition quality of each tested brain functional

network; while group-level detection aims to obtain consistent partition results across the population group, which has a relatively low requirement for the partition accuracy of each single network. The above findings can provide a guidance for the selection of module detection algorithms in the subsequent brain functional network research.

Of four individual-level module detection algorithms, Louvain algorithm obtained the best results in almost all brain networks ranging from small to large networks. This is due to its hierarchical module searching strategy, which not only makes it easier to approach the optimal solution in NP-hard problems, but also suitable for large-scale networks (Blondel et al., 2008). In contrary, Tabu method can only work well on small brain network limited by its exhaustive searching approach, and spectral method is not prominent in all networks. A more interesting finding is that, in terms of reliability, Infomap method is not suitable for module detection in brain functional networks. We speculate the main reason is that the random walk used in this method is highly dependent on the reliability of edges in network, but FCs are usually unstable due to imaging noise (Birn, 2012; Murphy et al., 2013). One possible solution is to improve the stability of the function connection or use a binary network. In group-level analysis, the average and consensus clustering based module detection algorithms yielded comparable performance in reliability assessment, which

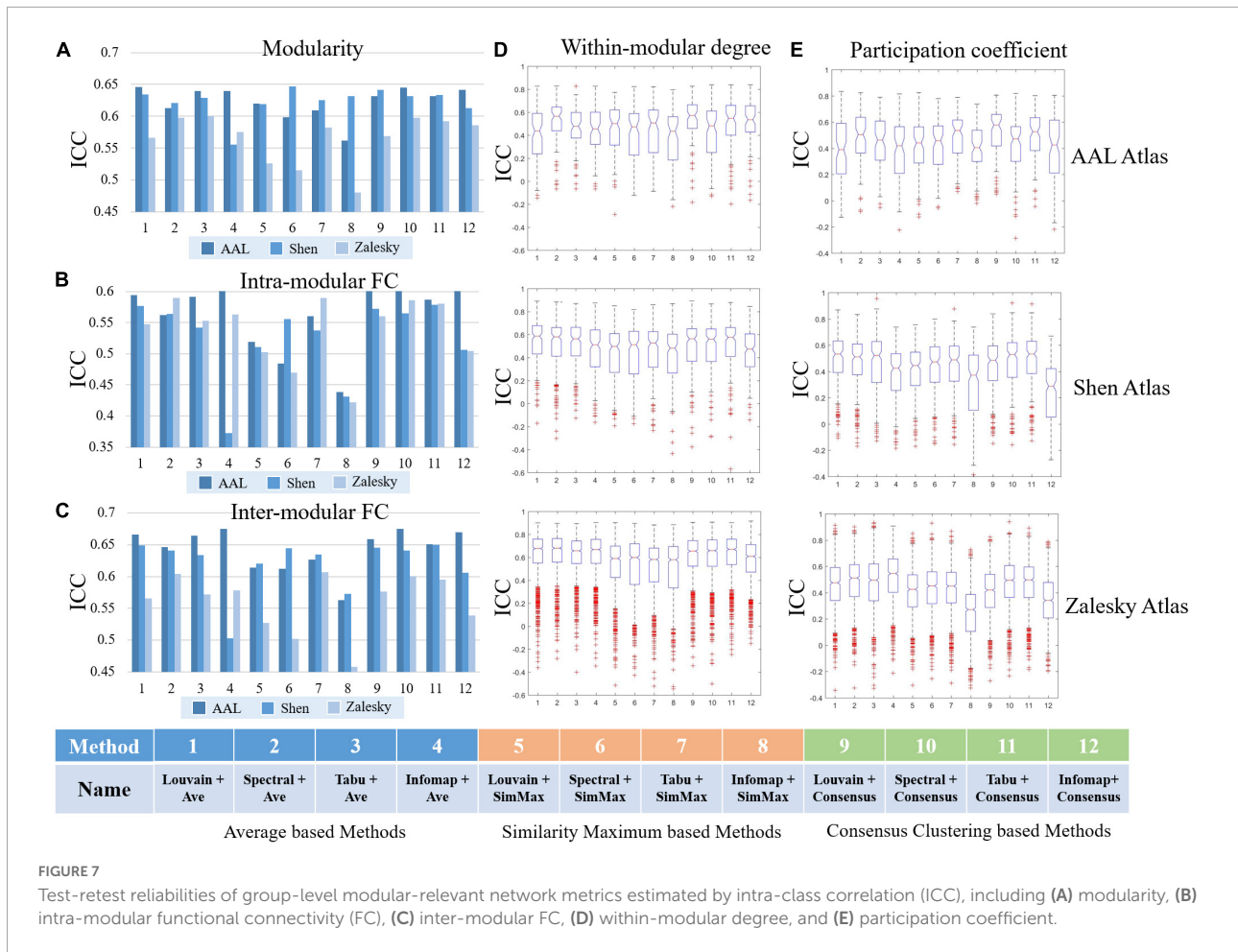


TABLE 2 List of top four group-level module detection methods with the highest intra-class correlation coefficient (ICC) values on five network metrics estimated on three brain atlases.

Property	Atlas	ICC rank
Modularity	AAL	Louvain + Ave (good) > Spectral + Consensus (good) > Infomap + Consensus (good) > Tabu + Ave (good)
	Shen	Spectral + SimMax (good) > Louvain + Consensus (good) > Louvain + Ave (good) > Tabu + Consensus (good)
	Zalesky	Tabu + Ave (good) > Spectral + Ave (moderate) > Spectral + Consensus (moderate) > Tabu + Consensus (moderate)
Intra-modular FC	AAL	Infomap + Consensus (good) > Spectral + Consensus (good) > Infomap + Ave (good) > Louvain + Consensus (good)
	Shen	Tabu + Consensus (moderate) > Louvain + Ave (moderate) > Louvain + Consensus (moderate) > Infomap + Consensus (moderate)
	Zalesky	Tabu + SimMax (moderate) > Spectral + Ave (moderate) > Spectral + Consensus (moderate) > Tabu + Consensus (moderate)
Inter-modular FC	AAL	Infomap + Ave (good) > Spectral + Consensus (good) > Infomap + Consensus (good) > Tabu + Ave (good)
	Shen	Tabu + Consensus (good) > Louvain + Ave (good) > Louvain + Consensus (good) > Spectral + SimMax (good)
	Zalesky	Tabu + SimMax (good) > Spectral + Ave (good) > Spectral + Consensus (good) > Tabu + Consensus (moderate)
Within-module degree	AAL	Louvain + Consensus (moderate) > Spectral + Ave (moderate) > Tabu + Consensus (moderate) > Infomap + Consensus (moderate)
	Shen	Louvain + Ave (moderate) > Spectral + Ave (moderate) > Tabu + Consensus (moderate) > Louvain + Consensus (moderate)
	Zalesky	Louvain + Ave (good) > Spectral + Tabu (good) > Tabu + Consensus (good) > Infomap + Ave (good)
Participation coefficient	AAL	Louvain + Consensus (good) > Tabu + SimMax (moderate) > Tabu + Consensus (moderate) > Spectral + Ave (moderate)
	Shen	Louvain + Ave (moderate) > Spectral + Consensus (moderate) > Tabu + Consensus (moderate) > Tabu + Ave (moderate)
	Zalesky	Infomap + Ave (moderate) > Spectral + Ave (moderate) > Spectral + Consensus (moderate) > Tabu + Ave (moderate)

Poor reliability (0 < ICC < 0.2), fair reliability (0.2 < ICC < 0.4), moderate reliability (0.4 < ICC < 0.6), good (0.6 < ICC < 0.8), and excellent reliability (ICC > 0.8).

is much better than those using similarity maximum as group-wise module determination framework. This conclusion is in line with our expectation because the similarity maximum framework is to select network partition from one subject to represent group-level modular structure, which is difficult to characterize the module characteristics of all the tested brain functional networks.

Effects of node definition approach

While the impact of node definition approach on the topological organization of brain functional networks has been widely acknowledged (Wang et al., 2009; Fornito et al., 2010; Zalesky et al., 2010; Power et al., 2011), the question of how parcellation scheme influences the reliability is still unclear. Most studies assessed the reliability of network metrics using common AAL atlas (Schwarz and McGonigle, 2011; Braun et al., 2012; Chou et al., 2012; Guo et al., 2012; Liang et al., 2012; Fiecas et al., 2013; Jin et al., 2018), and only few studies have focused on this feature (Wang et al., 2011; Cao et al., 2014). These studies compared the performance between structural [AAL atlas and Harvard-Oxford atlas (Kennedy et al., 1998)] and functional atlas [F-DOS (Dosenbach et al., 2010) and funROI (Power et al., 2011)], and concluded that the reliability of network metrics is certainly influenced by the chosen brain atlas, but it is difficult to define which parcellation scheme yields the highest reliability. Similar conclusion were reached in our work that the robustness of modular-relevant analysis was modulated by strategies of network node definition, and none of brain atlas outperformed the others in all cases. For example, in the individual level, Shen atlas displayed the lowest repeatability in network partition results (shown in Figure 2), whereas the group-level module detection found Zalesky atlas performed worst. The discrepancy between individual and group levels may due to that the reliability of modular-relevant analysis can be influenced by both parcellation granularity of the brain atlas and its generation approach. On one hand, it is well known that increasing network size would worsen the uncertainty of module detection results, thus the brain atlas with finer parcellation generally display lower repeatability than coarser parcellation. Therefore, in group-level modular analysis, the Zalesky atlas with the largest number of nodes exhibited the lowest reliability. On the other hand, in terms of atlas generation, ROIs from AAL, Shen and Zalesky atlases were, respectively, obtained based on anatomical features of sulcal pattern (Tzourio-Mazoyer et al., 2002), brain region's functional connectivity, and geometric information. A previous study found that the geometric parcellation could yield more reliable FCs than the data-driven FC-based parcellation scheme (Zeng et al., 2019). This may be the reason of the poorer performance of Shen atlas than Zalesky atlas in individual-level analysis. Recent studies have also attempted to determine the best brain

parcellation in terms of quality and reliability but failed to figure out a clear winner (Arslan et al., 2018), which needs to be further investigated in future studies.

Limitations and technique considerations

Our study has a few of limitations. *Firstly*, we only explored the influence of module detection algorithm and node definition approach on reliability of modular-relevant network metrics, without going through imaging acquisition and preprocessing procedures. Previous studies have identified differences in acquisition and processing procedures also impact test-retest reliability (Zuo et al., 2013; Aurich et al., 2015; Varikuti et al., 2017), e.g., temporal resolution of the fMRI data (Birn et al., 2014; Shah et al., 2014; Zuo and Xing, 2014), global signal regression (Shirer et al., 2015; Varikuti et al., 2017), and motion correction (Zuo and Xing, 2014), etc. Hence, the reliability results in this work are specific to the acquisition protocol of HCP dataset and specific imaging preprocessing. *Secondly*, the reproducibility of brain functional networks can be affected by FC estimation methods (Liang et al., 2012) and network sparsity strategies (Cao et al., 2014). These variables should also be considered in functional brain networks following our data-driven analysis. However, our work only examined two factors (i.e., module detection algorithm and node definition approach) which can impact the repeatability of network partitions directly. *Lastly*, using rs-fMRI, we examined the reliability of modular-relevant analysis in functional brain networks. Previous studies have performed similar analyses of structural brain networks using DTI data (Dimitriadis et al., 2021). A systematic reliability evaluation using multimodal data from the same population is warranted to gain a deeper understanding of the structural and functional architecture of the human brain.

Conclusion

In this study, we systematically investigated test-retest reliability of diverse modular-relevant analyses in brain functional networks as well as how they are affected by module detection algorithm and node definition approach. The results showed that five popular modular-derived network metrics could provide robust results with modularity achieving the highest reliability. The extensive analysis identified different module detection methods and brain atlases yield different reliability in network metrics. Our analysis indicate that much attention should be paid to the choice of module detection algorithms and node definition approaches in order to extract reliable results at the mesoscale of brain networks.

Data availability statement

The original contributions presented in this study are included in the article/**Supplementary material**, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by Human Connectome Project, WU-Minn Consortium. The patients/participants provided their written informed consent to participate in this study.

Author contributions

XW and MY developed, implemented, and evaluated the algorithms, wrote, and revised the manuscript. LH and DZ reviewed and supervised the research. All authors contributed to the article and approved the submitted version.

Funding

This study was supported in part by National Natural Science Foundation of China (Grant Nos. 62001222 and

62136004), China Postdoctoral Science Foundation funded project (Grant Nos. 2021TQ0150 and 2021M701699), and 2021 Jiangsu Shuangchuang Talent Program.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnins.2022.1000863/full#supplementary-material>

References

- Ana, L. F., and Jain, A. K. (2003). "Robust data clustering," in *Proceedings of the 2003 IEEE computer society conference on computer vision and pattern recognition* (Madison, WI: IEEE).
- Andellini, M., Cannata, V., Gazzellini, S., Bernardi, B., and Napolitano, A. (2015). Test-retest reliability of graph metrics of resting state MRI functional brain networks: A review. *J. Neurosci. Methods* 253, 183–192. doi: 10.1016/j.jneumeth.2015.05.020
- Anderson, J. S., Ferguson, M. A., Lopez-Larson, M., and Yurgelun-Todd, D. (2011). Reproducibility of single-subject functional connectivity measurements. *Am. J. Neuroradiol.* 32, 548–555.
- Andric, M., and Hasson, U. (2015). Global features of functional brain networks change with contextual disorder. *Neuroimage* 117, 103–113. doi: 10.1016/j.neuroimage.2015.05.025
- Arenas, A., Fernandez, A., and Gómez, S. (2008). Analysis of the structure of complex networks at different resolution levels. *N. J. Phys.* 10:53039.
- Arslan, S., Ktena, S. I., Makropoulos, A., Robinson, E. C., Rueckert, D., and Parisot, S. (2018). Human brain mapping: A systematic comparison of parcellation methods for the human cerebral cortex. *Neuroimage* 170, 5–30.
- Aurich, N. K., Filho, J. O., Silva, A. M., and Franco, A. R. (2015). Evaluating the reliability of different preprocessing steps to estimate graph theoretical measures in resting state fMRI data. *Front. Neurosci.* 9:48. doi: 10.3389/fnins.2015.00048
- Bassett, D. S., Porter, M. A., Wymbs, N. F., Grafton, S. T., Carlson, J. M., and Mucha, P. J. (2013). Robust detection of dynamic community structure in networks. *Chaos Interdiscip. J. Nonlinear Sci.* 23:13142.
- Baum, G. L., Ćirić, R., Roalf, D. R., Betzel, R. F., Moore, T. M., Shinohara, R. T., et al. (2017). Modular segregation of structural brain networks supports the development of executive function in youth. *Curr. Biol.* 27, 1561–1572. doi: 10.1016/j.cub.2017.04.051
- Bertolero, M. A., Yeo, B. T., and D'Esposito, M. (2015). The modular and integrative functional architecture of the human brain. *Proc. Natl. Acad. Sci. U.S.A.* 112, E6798–E6807.
- Birn, R. M. (2012). The role of physiological noise in resting-state functional connectivity. *Neuroimage* 62, 864–870.
- Birn, R. M., Cornejo, M. D., Molloy, E. K., Patriat, R., Meier, T. B., Kirk, G. R., et al. (2014). The influence of physiological noise correction on test-retest reliability of resting-state functional connectivity. *Brain Connect.* 4, 511–522.
- Blondel, V. D., Guillaume, J. L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *J. Stat. Mech.* 2008:10008. doi: 10.1103/PhysRevE.83.036103
- Braun, U., Plichta, M. M., Esslinger, C., Sauer, C., Haddad, L., Grimm, O., et al. (2012). Test-retest reliability of resting-state connectivity network characteristics using fMRI and graph theoretical measures. *Neuroimage* 59, 1404–1412.
- Buchanan, C. R., Pernet, C. R., Gorgolewski, K. J., Storkey, A. J., and Bastin, M. E. (2014). Test-retest reliability of structural brain networks from diffusion MRI. *Neuroimage* 86, 231–243.
- Bullmore, E., and Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* 10, 186–198.
- Cai, J., Liu, A., Mi, T., Garg, S., Trappe, W., McKeown, M. J., et al. (2018). Dynamic graph theoretical analysis of functional connectivity in Parkinson's disease: The importance of Fiedler value. *IEEE J. Biomed. Health Inform.* 23, 1720–1729. doi: 10.1109/JBHI.2018.2875456
- Cao, H., Plichta, M. M., Schäfer, A., Haddad, L., Grimm, O., Schneider, M., et al. (2014). Test-retest reliability of fMRI-based graph theoretical properties during

- working memory, emotion processing, and resting state. *Neuroimage* 84, 888–900. doi: 10.1016/j.neuroimage.2013.09.013
- Chan, M. Y., Park, D. C., Savalia, N. K., Petersen, S. E., and Wig, G. S. (2014). Decreased segregation of brain systems across the healthy adult lifespan. *Proc. Natl. Acad. Sci. U.S.A.* 111, E4997–E5006.
- Chen, B., Xu, T., Zhou, C., Wang, L., Yang, N., Wang, Z., et al. (2015). Individual variability and test-retest reliability revealed by ten repeated resting-state brain scans over one month. *PLoS One* 10:e144963. doi: 10.1371/journal.pone.0144963
- Cheng, H., Newman, S., Goñi, J., Kent, J. S., Howell, J., Bolbecker, A., et al. (2015). Nodal centrality of functional network in the differentiation of schizophrenia. *Schizophr. Res.* 168, 345–352.
- Chou, Y., Panych, L. P., Dickey, C. C., Petrella, J. R., and Chen, N. (2012). Investigation of long-term reproducibility of intrinsic connectivity network mapping: A resting-state fMRI study. *Am. J. Neuroradiol.* 33, 833–838.
- Cole, M. W., Bassett, D. S., Power, J. D., Braver, T. S., and Petersen, S. E. (2014). Intrinsic and task-evoked network architectures of the human brain. *Neuron* 83, 238–251.
- Dai, Z., Lin, Q., Li, T., Wang, X., Yuan, H., Yu, X., et al. (2018). Disrupted structural and functional brain networks in Alzheimer's disease. *Neurobiol. Aging* 45, 71–82.
- Dimitriadis, S. I., Messaritaki, E., and Jones, D. K. (2021). The impact of graph construction scheme and community detection algorithm on the reliability of community and hub identification in structural brain networks. *Hum. Brain Mapp.* 42, 4261–4280. doi: 10.1002/hbm.25545
- Dosenbach, N. U., Nardos, B., Cohen, A. L., Fair, D. A., Power, J. D., Church, J. A., et al. (2010). Prediction of individual brain maturity using fMRI. *Science* 329, 1358–1361.
- Du, H., Liao, X., Lin, Q., Li, G., Chi, Y., Liu, X., et al. (2015). Test-retest reliability of graph metrics in high-resolution functional connectomics: A resting-state functional MRI study. *CNS Neurosci. Ther.* 21, 802–816. doi: 10.1111/cns.12431
- Fiecas, M., Ombao, H., Lunen, D. V., Baumgartner, R., Coimbra, A., and Feng, D. (2013). Quantifying temporal correlations: A test-retest evaluation of functional connectivity in resting-state fMRI. *Neuroimage* 65, 231–241. doi: 10.1016/j.neuroimage.2012.09.052
- Fornito, A., Zalesky, A., and Bullmore, E. T. (2010). Network scaling effects in graph analytic studies of human resting-state FMRI data. *Front. Syst. Neurosci.* 4:22. doi: 10.3389/fnsys.2010.00022
- Friston, K. J. (2011). Functional and effective connectivity: A review. *Brain Connect.* 1:13.
- Gallos, L. K., Makse, H. A., and Sigman, M. (2012). A small world of weak ties provides optimal global integration of self-similar modules in functional brain networks. *Proc. Natl. Acad. Sci. U.S.A.* 109, 2825–2830. doi: 10.1073/pnas.1106612109
- Garrison, K. A., Scheinost, D., Finn, E. S., Shen, X., and Constable, R. T. (2015). The (in) stability of functional brain network measures across thresholds. *Neuroimage* 118, 651–661.
- Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., et al. (2013). The minimal preprocessing pipelines for the human connectome project. *Neuroimage* 80, 105–124.
- Gong, M., Cai, Q., Chen, X., and Ma, L. (2014). Complex network clustering by multiobjective discrete particle swarm optimization based on decomposition. *IEEE Trans. Evol. Comput.* 18, 82–97.
- Griffanti, L., Salimi-Khorshidi, G., Beckmann, C. F., Auerbach, E. J., Douaud, G., Sexton, C. E., et al. (2014). ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. *Neuroimage* 95, 232–247. doi: 10.1016/j.neuroimage.2014.03.034
- Guimera, R., and Amaral, L. A. N. (2005). Functional cartography of complex metabolic networks. *Nature* 433:895.
- Guo, C. C., Kurth, F., Zhou, J., Mayer, E. A., Eickhoff, S. B., Kramer, J. H., et al. (2012). One-year test-retest reliability of intrinsic connectivity network fMRI in older adults. *Neuroimage* 61, 1471–1483. doi: 10.1016/j.neuroimage.2012.03.027
- Haan, W. d., Flier, W. M., Koene, T., Smits, L. L., Scheltens, P., and Stam, C. J. (2012). Disrupted modular brain dynamics reflect cognitive dysfunction in Alzheimer's disease. *Neuroimage* 59, 3085–3093. doi: 10.1016/j.neuroimage.2011.11.055
- Heuvel, M. D., and Pol, H. (2010). Exploring the brain network: A review on resting-state fMRI functional connectivity. *Eur. Neuropsychopharmacol.* 20, 519–534.
- Jin, D., Xu, K., Liu, B., Jiang, T., and Liu, Y. (2018). “Test-retest reliability of functional connectivity and graph metrics in the resting brain network,” in *Proceedings of the 2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC)* (Honolulu, HI: IEEE).
- Kennedy, D. N., Lange, N., Makris, N., Bates, J., Meyer, J., and Jr, V. S. (1998). Gyri of the human neocortex: An MRI-based analysis of volume and variance. *Cereb. Cortex* 8, 372–384. doi: 10.1093/cercor/8.4.372
- Koch, L. G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174.
- Laumann, T. O., Snyder, A. Z., Mitra, A., Gordon, E. M., Gratton, C., Adeyemo, B., et al. (2017). On the stability of BOLD fMRI correlations. *Cereb. Cortex* 27, 4719–4732.
- Lavrakas, P. J. (2008). “Test retest reliability,” in *Encyclopedia of survey research methods*, ed. P. J. Lavrakas (Thousand Oaks, CA: SAGE). doi: 10.4135/9781412963947.n581
- Liang, X., Wang, J., Yan, C., Shu, N., Xu, K., Gong, G., et al. (2012). Effects of different correlation metrics and preprocessing factors on small-world brain functional networks: A resting-state functional MRI study. *PLoS One* 7:e32766. doi: 10.1371/journal.pone.0032766
- Liao, X., Xia, M., Xu, T., Dai, Z., Cao, X., Niu, H., et al. (2013). Functional brain hubs and their test-retest reliability: A multiband resting-state functional MRI study. *Neuroimage* 83, 969–982.
- Ma, Q., Huang, B., Wang, J., Seger, C., Yang, W., Li, C., et al. (2017). Altered modular organization of intrinsic brain functional networks in patients with Parkinson's disease. *Brain Imaging Behav.* 11, 430–443. doi: 10.1007/s11682-016-9524-7
- Marcus, D. S., Harms, M. P., Snyder, A. Z., Jenkinson, M., Wilson, J. A., Glasser, M. F., et al. (2013). Human connectome project informatics: Quality control, database services, and data visualization. *Neuroimage* 80, 202–219. doi: 10.1016/j.neuroimage.2013.05.077
- Mejia, A. F., Nebel, M. B., Barber, A. D., Choe, A. S., Pekar, J. J., Caffo, B. S., et al. (2018). Improved estimation of subject-level functional connectivity using full and partial correlation with empirical Bayes shrinkage. *Neuroimage* 172:478. doi: 10.1016/j.neuroimage.2018.01.029
- Meunier, D., Achard, S., Morcom, A., and Bullmore, E. (2009a). Age-related changes in modular organization of human brain functional networks. *Neuroimage* 44, 715–723.
- Meunier, D., Lambiotte, R., Fornito, A., Ersche, K. D., and Bullmore, E. T. (2009b). Hierarchical modularity in human brain functional networks. *Front. Neuroinform.* 3:37. doi: 10.3389/neuro.11.037.2009
- Moeller, S., Yacoub, E., Olman, C. A., Auerbach, E., Strupp, J., Harel, N., et al. (2010). Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. *Magn. Reson. Med.* 63, 1144–1153. doi: 10.1002/mrm.22361
- Mueller, S., Wang, D., Fox, M. D., Pan, R., Lu, J., Li, K., et al. (2015). Reliability correction for functional connectivity: Theory and implementation. *Hum. Brain Mapp.* 36, 4664–4680. doi: 10.1002/hbm.22947
- Murphy, K., Birn, R. M., and Bandettini, P. A. (2013). Resting-state fMRI confounds and cleanup. *Neuroimage* 80, 349–359.
- Newman, M. E. (2006). Modularity and community structure in networks. *Proc. Natl. Acad. Sci. U.S.A.* 103, 8577–8582.
- Noble, S., Scheinost, D., and Constable, R. T. (2019). A decade of test-retest reliability of functional connectivity: A systematic review and meta-analysis. *Neuroimage* 203, 116–157. doi: 10.1016/j.neuroimage.2019.116157
- Park, H., and Friston, K. (2013). Structural and functional brain networks: From connections to cognition. *Science* 342:1238411.
- Parkes, L., Fulcher, B., Yücel, M., and Fornito, A. (2017). An evaluation of the efficacy, reliability, and sensitivity of motion correction strategies for resting-state functional MRI. *Neuroimage* 171, 415–436. doi: 10.1016/j.neuroimage.2017.12.073
- Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., et al. (2011). Functional network organization of the human brain. *Neuron* 72, 665–678.
- Power, J. D., Schlaggar, B. L., Lessov-Schlaggar, C. N., and Petersen, S. E. (2013). Evidence for hubs in human functional brain networks. *Neuron* 79, 798–813.
- Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *J. Am. Stat. Assoc.* 66, 846–850.
- Rosvall, M., and Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure. *Proc. Natl. Acad. Sci. U.S.A.* 105:1118. doi: 10.1073/pnas.0706851105
- Salimi-Khorshidi, G., Douaud, G., Beckmann, C. F., Glasser, M. F., Griffanti, L., and Smith, S. M. (2014). Automatic denoising of functional MRI data: Combining

- independent component analysis and hierarchical fusion of classifiers. *Neuroimage* 90, 449–468. doi: 10.1016/j.neuroimage.2013.11.046
- Sang, L., Zhang, J., Wang, L., Zhang, J., Zhang, Y., Li, P., et al. (2015). Alteration of Brain functional networks in early-stage Parkinson's disease: A resting-state fmri study. *PLoS One* 10:e0141815. doi: 10.1371/journal.pone.0141815
- Schwarz, A. J., and McGonigle, J. (2011). Negative edges and soft thresholding in complex network analysis of resting state functional connectivity data. *Neuroimage* 55, 1132–1146. doi: 10.1016/j.neuroimage.2010.12.047
- Shah, N., Dakin, S. C., Whitaker, H. L., and Anderson, R. S. (2014). Effect of scoring and termination rules on test–retest variability of a novel high-pass letter acuity chart. *Invest. Ophthalmol. Vis. Sci.* 55, 1386–1392. doi: 10.1167/iov.13-13340
- Shehzad, Z., Kelly, A. M., Reiss, P. T., Gee, D. G., Gotimer, K., Uddin, L. Q., et al. (2009). The resting brain: Unconstrained yet reliable. *Cereb. Cortex* 19, 2209–2229.
- Shen, X., Tokoglu, F., Papademetris, X., and Constable, R. T. (2013). Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. *Neuroimage* 82, 403–415. doi: 10.1016/j.neuroimage.2013.05.081
- Shirer, W. R., Jiang, H., Price, C. M., Ng, B., and Greicius, M. D. (2015). Optimization of rs-fMRI pre-processing for enhanced signal-noise separation, test-retest reliability, and group discrimination. *Neuroimage* 117, 67–79. doi: 10.1016/j.neuroimage.2015.05.015
- Shou, H., Eloyan, A., Lee, S., Zippunnikov, V., Crainiceanu, A. N., Nebel, N. B., et al. (2013). Quantifying the reliability of image replication studies: The image intraclass correlation coefficient (I2C2). *Cogn. Affect. Behav. Neurosci.* 13, 714–724. doi: 10.3758/s13415-013-0196-0
- Shrout, P. E., and Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychol. Bull.* 86:420.
- Smith, S. M., Beckmann, C. F., Andersson, J., Auerbach, E. J., Bijsterbosch, J., Douaud, G., et al. (2013). Resting-state fMRI in the human connectome project. *Neuroimage* 80, 144–168.
- Sporns, O., and Betzel, R. F. (2016). Modular brain networks. *Annu. Rev. Psychol.* 67, 613–640.
- Tomasi, D. G., Shokri-Kojori, E., and Volkow, N. D. (2017). Temporal evolution of brain functional connectivity metrics: Could 7 min of rest be enough? *Cereb. Cortex* 27, 4153–4165. doi: 10.1093/cercor/bhw227
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15, 273–289.
- van den Heuvel, M. P., Stam, C. J., Kahn, R. S., and Pol, H. E. (2009). Efficiency of functional brain networks and intellectual performance. *J. Neurosci.* 29, 7619–7624.
- Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., Ugurbil, K., et al. (2013). The WU-Minn human connectome project: An overview. *Neuroimage* 80, 62–79.
- Varikuti, D. P., Hoffstaedter, F., Genon, S., Schwender, H., Reid, A. T., and Eickhoff, S. B. (2017). Resting-state test–retest reliability of a priori defined canonical networks over different preprocessing steps. *Brain Struct. Funct.* 222, 1447–1468. doi: 10.1007/s00429-016-1286-x
- Wang, J., Ren, Y., Hu, X., Nguyen, V. T., Guo, L., Han, J., et al. (2017). Test–retest reliability of functional connectivity networks during naturalistic fMRI paradigms. *Hum. Brain Mapp.* 38, 2226–2241.
- Wang, J., Wang, L., Zang, Y., Yang, H., Tang, H., Gong, Q., et al. (2009). Parcellation-dependent small-world brain functional networks: A resting-state fMRI study. *Hum. Brain Mapp.* 30, 1511–1523. doi: 10.1002/hbm.20623
- Wang, J., Zuo, X., Gohel, S., Milham, M. P., Biswal, B. B., and He, Y. (2011). Graph theoretical analysis of functional brain networks: Test-retest evaluation on short-and long-term resting-state functional MRI data. *PLoS One* 6:e21976. doi: 10.1371/journal.pone.0021976
- Wen, X., Zhang, H., Li, G., Liu, M., Yin, W., Lin, W., et al. (2019). First-year development of modules and hubs in infant brain functional networks. *Neuroimage* 185, 222–235. doi: 10.1016/j.neuroimage.2018.10.019
- Xu, G., Bennett, L., Papageorgiou, L. G., and Tsoka, S. (2010). Module detection in complex networks using integer optimisation. *Algorithms Mol. Biol.* 5:36. doi: 10.1186/1748-7188-5-36
- Yeo, B. T., Krienen, F. M., Chee, M. W., and Buckner, R. L. (2014). Estimates of segregation and overlap of functional connectivity networks in the human cerebral cortex. *Neuroimage* 88, 212–227.
- Zalesky, A., Fornito, A., Harding, I. H., Cocchi, L., Yücel, M., Pantelis, C., et al. (2010). Whole-brain anatomical networks: Does the choice of nodes matter? *Neuroimage* 50, 970–983. doi: 10.1016/j.neuroimage.2009.12.027
- Zeng, J., Dang, A. T., and Atluri, G. (2019). “Test-retest reliability of functional networks for evaluation of data-driven parcellation,” in *Connectomics in neuroimaging. CNI 2019. Lecture notes in computer science*, Vol. 11848, eds M. Schirmer, A. Venkataraman, I. Rekik, M. Kim, and A. Chung (Cham: Springer). doi: 10.1007/978-3-030-32391-2_10
- Zhang, H., Chen, X., Zhang, Y., and Shen, D. (2017). Test-retest reliability of “high-order” functional connectivity in young healthy adults. *Front. Neurosci.* 11:439. doi: 10.3389/fnins.2017.00439
- Zhang, H., Duan, L., Zhang, Y., Lu, C., Liu, H., and Zhu, C. (2011). Test-retest assessment of independent component analysis-derived resting-state functional connectivity based on functional near-infrared spectroscopy. *Neuroimage* 55, 607–615. doi: 10.1016/j.neuroimage.2010.12.007
- Zhang, Y., Wang, Y., Chen, N., Guo, M., Wang, X., Chen, G., et al. (2021). Age-associated differences of modules and hubs in brain functional networks. *Front. Aging Neurosci.* 12:607445. doi: 10.3389/fnagi.2020.607445
- Zuo, X., and Xing, X. (2014). Test-retest reliabilities of resting-state FMRI measurements in human brain functional connectomics: A systems neuroscience perspective. *Neurosci. Biobehav. Rev.* 45, 100–118. doi: 10.1016/j.neubiorev.2014.05.009
- Zuo, X., Xu, T., Jiang, L., Yang, Z., Cao, X., He, Y., et al. (2013). Toward reliable characterization of functional homogeneity in the human brain: Preprocessing, scan duration, imaging resolution and computational space. *Neuroimage* 65, 374–386. doi: 10.1016/j.neuroimage.2012.10.017