

Article

Test-Retest Reliability of Resting Brain Small-World Network Properties across Different Data Processing and Modeling Strategies

Qianying Wu^{1,2,3,†} , Hui Lei^{2,4,†}, Tianxin Mao^{1,2}, Yao Deng^{1,2}, Xiaocui Zhang^{2,5,6,7}, Yali Jiang^{2,5}, Xue Zhong^{2,5}, John A. Detre² , Jianghong Liu^{8,*}  and Hengyi Rao^{1,2,*}

- ¹ Key Laboratory of Brain-Machine Intelligence for Information Behavior (Ministry of Education and Shanghai), School of Business and Management, Shanghai International Studies University, Shanghai 201613, China
- ² Department of Neurology, University of Pennsylvania, Philadelphia, PA 19104, USA
- ³ School of Life Sciences, University of Science and Technology of China, Hefei 230026, China
- ⁴ College of Education, Hunan Agricultural University, Changsha 410127, China
- ⁵ Medical Psychological Center, The Second Xiangya Hospital, Central South University, Changsha 410017, China
- ⁶ Medical Psychological Institute, Central South University, Changsha 410017, China
- ⁷ National Clinical Research Center for Mental Disorders, Changsha 410011, China
- ⁸ Department of Family and Community Health, School of Nursing, University of Pennsylvania, Philadelphia, PA 19104, USA
- * Correspondence: jhliu@nursing.upenn.edu (J.L.); hengyi@penmedicine.upenn.edu (H.R.)
- † These authors contributed equally to this work.

Abstract: Resting-state functional magnetic resonance imaging (fMRI) with graph theoretical modeling has been increasingly applied for assessing whole brain network topological organization, yet its reproducibility remains controversial. In this study, we acquired three repeated resting-state fMRI scans from 16 healthy controls during a strictly controlled in-laboratory study and examined the test-retest reliability of seven global and three nodal brain network metrics using different data processing and modeling strategies. Among the global network metrics, the characteristic path length exhibited the highest reliability, whereas the network small-worldness performed the poorest. Nodal efficiency was the most reliable nodal metric, whereas betweenness centrality showed the lowest reliability. Weighted global network metrics provided better reliability than binary metrics, and reliability from the AAL90 atlas outweighed those from the Power264 parcellation. Although global signal regression had no consistent effects on the reliability of global network metrics, it slightly impaired the reliability of nodal metrics. These findings provide important implications for the future utility of graph theoretical modeling in brain network analyses.

Keywords: resting-state fMRI; test-retest reliability; graph theoretical modeling; small-world network (SWN); intra-class correlation coefficient (ICC)



Citation: Wu, Q.; Lei, H.; Mao, T.; Deng, Y.; Zhang, X.; Jiang, Y.; Zhong, X.; Detre, J.A.; Liu, J.; Rao, H. Test-Retest Reliability of Resting Brain Small-World Network Properties across Different Data Processing and Modeling Strategies. *Brain Sci.* **2023**, *13*, 825. <https://doi.org/10.3390/brainsci13050825>

Academic Editors: Michele Ferrara and Aurora D'Atri

Received: 4 March 2023

Revised: 2 May 2023

Accepted: 12 May 2023

Published: 19 May 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Resting-state functional magnetic resonance imaging (rs-fMRI) has proven to be a powerful tool for examining spontaneous fluctuations in brain activities. The spontaneous activity of the resting brain, often referred to as the intrinsic baseline brain function, likely represents a ‘physiologic, functionally significant state of the brain’. In the resting state, task-evoked energy consumption appears to be less than 5% of that during basal metabolism [1,2]. Functional connectivity has been widely applied to characterize resting-state brain activities. It is defined as ‘the temporal correlation between neurophysiological measurements made in different brain areas’ [3] (p. 6) and reflects the level of information processing and transportation between anatomically separated brain regions [4].

With the concept of functional connectivity, the brain can be viewed as a highly connected network that consistently maintains a balance between the wiring cost and processing efficiency [5]. Such functional connectivity networks have been further modeled using graph theory. In graph theory, a graph consists of two fundamental elements: node and edge. Mapping onto the rs-fMRI data, nodes are typically defined by regions of interest (ROI) or voxels, and edges are represented by functional connectivity strengths.

Small-world architecture is an important graph-theoretical property of the brain network [6]. Consider two extreme situations: a regular network and a random network. The former is highly ordered, in which every node is connected to, and only connected to, its nearest neighbors (i.e., no probabilistic edge). The latter is entirely random, in which every pair of nodes can be connected independently with equal probabilities. A small-world network is in the middle and the connections are probabilistic; however, the probability follows some rules. For example, two nodes far away from each other are usually not directly connected by a single edge, whereas one can easily reach the other through a small number of steps (i.e., intermediate nodes and edges) [7]. Such properties of a small-world network are quantified by several metrics, such as the clustering coefficient and characteristic path length [8]. The brain network has been demonstrated to have a high clustering coefficient and short characteristic path length, which is crucial for maintaining efficient information segregation and integration [9]. Given such great importance of the brain's small-world properties, they have been extensively studied and applied as biomarkers to identify various psychopathologies [10–15].

Despite the explosive application of the resting-state BOLD fMRI and graph theory, some of the results may suffer from low test-retest reliability, as measured by intra-class correlation (ICC). A review of 15 fMRI studies found that the mean voxel- or ROI-based ICC across all the studies ranged from 0.16 to 0.88, with an average of 0.50 [16]. Similarly, a recent meta-analysis reported that the overall reliability of functional connectivity is poor (average ICC = 0.29) [17]. The ICC of small-world metrics was unsatisfying as well (e.g., clustering coefficient, characteristic path length, small-worldness, etc.), and many of them hardly exceeded 0.6 [18–21].

Such poor test-retest reliabilities of the graph-theoretical measures in the rs-fMRI data hamper their application in clinical settings; therefore, it is essential to explore the optimal strategy for network analysis. An important source of inconsistency may derive from the methods of data processing and network construction. When defining the nodes, different ROI parcellations can lead to different interpretations. For example, some atlases are generated by structural separations of the brain areas (e.g., Automated-Anatomical Labeling (AAL), Harvard-Oxford atlas (HOA), Brainnetome atlas (BN)), while many others are based on functional homogeneities (i.e., co-activations) (e.g., Power 264, funROI, DOS) [22–26]. Previous studies comparing ICC across atlases suggested that the HOA is generally more reliable than the AAL and DOS atlas [20]; and that finer parcellations, with more regions, could produce more reliable clustering coefficients and local efficiency [27]. It is also important to consider the trade-off between greater functional homogeneity (more ROIs) and better anatomical interpretability (fewer ROIs) [28].

There is further controversy regarding the definition of edge. The primary issue is whether to apply the global signal regression (GSR) during fMRI preprocessing. Global signal regression refers to the process of removing the average time series across voxels (i.e., the global signal) from each voxel signal to improve the signal-to-noise ratio [29]. However, GSR can produce spurious negative correlations, and global signals may correlate with the experimental manipulation [30]. Therefore, it is not appropriate to interpret them entirely as nuisance variables [31,32]. It is so far not clear whether GSR has a positive or negative effect on TRT reliability. Andellini et al., (2015) examined the ICC of five micro-level graph theoretical metrics (degree, clustering coefficient, local efficiency, global efficiency, and assortativity) and found that global signal regression would decrease the reliability of all the metrics [18]. On the contrary, Braun et al. (2012) reported that the effect of global signal

regression depends on the exact metric and network density; however, in most cases, the ICC improved after GSR [19].

Another issue regarding the choice of the edge is the use of binary versus weighted edges. Most studies prefer the binary network because it is straightforward to model and interpret. Despite the difficulty in modeling and interpretation, the weighted network contains more information and presents a more detailed picture of the brain [33]. Previous studies have compared the TRT reliability between these two modeling strategies. For example, Andellini et al., (2015) reported that the ICC of graph theoretical metrics was not significantly improved (although slightly improved) in the weighted network modeling compared to the binary network modeling [18]. However, Xiang et al., (2019) found that weighted modeling significantly benefits the TRT reliability clustering coefficient, shortest path length, and local and global efficiency [21].

In addition to the choice of data processing and network construction methods, variations and confounding factors during subject recruitment and data collection also impair TRT reliability. For example, subjects with sleep disorders or substance abuse need to be screened and excluded, since the organization of functional brain networks of those patients significantly differs from that of healthy controls [34–37]. Moreover, resting-state data collected with eyes closed were less reliable than those with eyes open [38,39], likely caused by drowsiness when eyes were closed [17]. In addition, the consistency of scanners and sites is a relevant factor: inter-site and inter-scanner differences decrease the TRT reliability of the temporal signal-to-noise ratio (tSNR) and functional connectivity [40,41]. Finally, the brain status may vary if the scans take place at different times of the day or different days of the week; the volume of the brain, glucose metabolism, regional cerebral blood flow, and rs-fMRI-based functional connectivity all fluctuate, possibly due to the effects of circadian rhythm on the brain [42–44].

Previous investigations on resting-state test-retest reliability were primarily based on fMRI data from existing datasets (e.g., [20,21,27,45,46]). However, the participants' health conditions were not well monitored, and many data collection details were not available. Moreover, there is no current consensus regarding data processing and network construction methods, and few studies have been able to consider multiple aspects of the analysis simultaneously. To address these important issues, we conducted a very strictly controlled experiment in which participants remained in the laboratory for five consecutive days (four nights). They were continuously monitored by research and hospital staff and were allowed to sleep for 8–9 h. All the scan sessions took place using the same scanner at the same time of the day. To further rule out potential confounders, caffeine, alcohol, tobacco, and medications were not permitted. After ruling out potential confounders in the data collection, we aimed to examine the optimal combination of strategies that would yield the highest test-retest reliability in the brain network's small-world properties. Two nodal (degree and betweenness centrality) and five global metrics (clustering coefficient, characteristic path length, small-worldness, global efficiency, and local efficiency) were applied to characterize the overall topology of the network. Brain networks were constructed using eight different processing strategies, and the ICC was calculated to reflect the TRT reliability. We sought to answer the following three questions: (1) Which graph-theoretical measure of the brain network is the most reliable? (2) Is it necessary to apply global signal regression and use a weighted network during modeling? (3) Is AAL-90 a reliable parcellation scheme for brain network modeling?

2. Materials and Methods

2.1. Subjects

Sixteen healthy adults (8 females, mean age = 35.4 ± 9.5 yrs) were recruited as the control subjects in a very strictly controlled in-laboratory sleep study [47,48].

Upon recruitment, participants reported habitual sleep duration between 6.5 h–8.5 h, bedtime between 22:00–00:00, and awakenings between 06:00–09:00. Prior to the in-laboratory study, their reports were confirmed using approximately one week of wrist

actigraphy. Participants assessed by questionnaire who reported habitual napping, sleep disturbances, and extreme morningness or eveningness chronotypes were excluded from the study. Screenings for acute or chronic medical and psychological conditions, as well as drug and alcohol intake, were conducted using questionnaires, physical examinations, and blood and urine tests. All participants were nonsmokers and did not participate in shift work, transmeridian travel, or irregular sleep-wake routines 60 days prior to the study. Starting one week before the end of the laboratory session, participants were not permitted to use caffeine, alcohol, tobacco, and medications (except oral contraceptives), as verified by urine screenings.

The study was approved by the Institutional Review Board (IRB) of the University of Pennsylvania (IRB ID# 811678). Informed consent was obtained before enrollment, and the subjects were compensated for their participation.

2.2. Experimental Design

To ensure adherence to the protocol, participants remained in the laboratory at the Clinical Translational Research Center at the Hospital of the University of Pennsylvania for 5 consecutive days (4 consecutive nights). They were behaviorally monitored by trained staff, allowed to watch television, read, play video or board games, and perform other sedentary activities, but they were not allowed to exercise or leave the laboratory.

Participants received 9 h time in bed (21:30–06:20) on day 1 to adjust to the laboratory environment and 8 h of sleep (22:30–06:30) on days 2–5. MRI scan sessions were on the morning of days 2, 3 and 5, from 7.00 a.m. to 10.00 a.m.

2.3. Imaging Data Acquisition and Preprocessing

Magnetic resonance imaging was conducted using a Siemens 3.0 Tesla Trio whole-body scanner (Siemens AG, Erlangen, Germany) and a standard array coil. Resting-state BOLD fMRI data were collected using the standard EPI sequence: TR = 2 s, TE = 24 ms, FOV = 220×220 mm², matrix = $64 \times 64 \times 36$, slice thickness = 4 mm, and inter-slice gap = 4 mm. A total of 210 images were acquired for each subject. Subjects were instructed to keep their eyes open and look at a cross fixation in the scanner. T1-weighted structural images were obtained using a standard 3D MPRAGE sequence: TR = 1.62 s, TE = 3.09 ms, FOV = 187×250 mm², matrix size = 192×256 , slice thickness = 5 mm, and inter-slice gap = 1 mm.

Rs-fMRI data were pre-processed and analyzed using the Data Processing Assistant for Resting-State fMRI (DPARSF V2.3_20130615; <http://rfmri.org/DPARSF>, accessed on 3 March 2023), which is based on Statistical Parametric Mapping software (SPM8, Wellcome Department of Cognitive Neurology, London, UK) and the REST_V1.8_130615 toolbox http://www.restfmri.net/forum/REST_V1.8, accessed on 3 March 2023) implemented in Matlab14 (MathWorks, Natick, MA, USA). The pipeline consisted of head motion correction, co-registration, smoothing with an 8 mm full-width at half-maximum (FWHM) isotropic Gaussian kernel, normalization to the standard Montreal Neurological Institute (MNI) space, and the removal of linear trends. All functional volumes were band-pass filtered ($0.01 \text{ Hz} < f < 0.08 \text{ Hz}$) in order to reduce low-frequency drift and physiological high-frequency respiratory and cardiac noise. Nuisance covariates including six head motion parameters, white matter signal, and CSF signal were regressed out.

2.4. Brain Network Construction

A summary of all eight methods is in Table 1. Four of the methods were developed using a combination of two network types (binary/weighted), and with/without global signal regression, primarily using the AAL-90 atlas because AAL-90 was used in most of the previous studies [11,19,49]. Next, a comparison with the Power264 atlas, another widely used parcellation, was added in the binary/weighted networks with/without GSR. Seven global measurements were analyzed using thresholds from 0.15 to 0.35 (i.e., the proportion of strongest functional connectivity to preserve) with a 0.05 step, while three

nodal measurements were analyzed based on the area under the curve (AUC) value from 0.05 to 0.50 densities.

Table 1. Summary of eight data processing methods.

	BG90	BNG90	WG90	WNG90	BG264	BNG264	WG264	WNG264
Network type	Binary	Binary	Weighted	Weighted	Binary	Binary	Weighted	Weighted
Global signal regression	Yes	No	Yes	No	Yes	No	Yes	No
Parcellation	AAL90	AAL90	AAL90	AAL90	Power264	Power264	Power264	Power264

Note: BG90: binary + global signal regression + AAL90, BNG90: binary + no global signal regression + AAL90, WG90: weighted + global signal regression + AAL90, BG264: binary + global signal regression + Power264, BNG264: binary + no global signal regression + Power264, WNG90: weighted + no global signal regression + AAL90, WG264: weighted + global signal regression + Power264, WNG264: weighted + no global signal regression + Power264.

2.5. Graph Theoretical Metrics

Global network metrics included mean clustering coefficient (C_p) and its normalized version, gamma (γ), characteristic path length (L_p) and its normalized version, lambda (λ), small-worldness (σ), global efficiency (E_g), and local efficiency (E_{loc}). Nodal metrics included degree centrality (D_c), betweenness centrality (B_c), and nodal efficiency (N_e). All the network metrics were calculated using the GRETNA toolbox [50].

Clustering coefficient (C_p): The clustering coefficient describes the level of closeness to form a completely connected subgraph [51]. In this study, we used the global clustering coefficient, which is equal to the average clustering coefficient of all the nodes. Gamma is the normalized C_p by random networks.

$$C_p = \frac{1}{n} \sum_{i \in N} C_i = \frac{1}{n} \sum_{i \in N} \frac{2t_i}{k_i(k_i - 1)}$$

$$\gamma = C_p / C_p^{rand}$$

Characteristic path length (L_p): The characteristic path length is the mean shortest path length over all possible pairs of nodes. It helps to quantify the functional integration level [33]. Lambda is the normalized L_p by random networks.

Nodal efficiency (N_e): Nodal efficiency is defined as the average inverse shortest path length between a given node and every other node in the network [33].

Global efficiency (E_g): Global efficiency is the average nodal efficiency across all the nodes in the network. Compared to the shortest path length, it is more immediately related to parallel information transmission [8].

Local efficiency (E_{loc}): The local efficiency is proportional to the clustering coefficient and is seen as the global efficiency computed on the neighborhood of the node, also called fault tolerance [8].

Small-worldness (Sigma): Compared with random networks, small-world networks can be quantified with a larger clustering coefficient and a comparable characteristic path length, leading to a sigma σ (i.e., small-worldness) larger than one.

$$\sigma = \gamma / \lambda$$

Degree centrality (D_c): Degree centrality, also called degree, is the simplest measure of centrality. For binary networks, degree is the number of edges incident to the node; for weighted networks, it is the sum of weights of all the edges of the node.

Betweenness centrality (B_c): Betweenness centrality measures the importance of a node in information communication. It is defined as the number of times the shortest path between any other node passes through a particular node.

2.6. Test-Retest Reliability

In order to measure the test-retest reliability of each graph-theoretical metric among three sessions, the intra-class correlation coefficient (ICC) was introduced. Specifically, as is defined and recommended in previous studies [52–54], we used ICC(A,1), a two-way random model, which assessed absolute agreement between measurements and considered session effects.

$$\text{ICC}(A,1) = \frac{BMS - EMS}{BMS + (k - 1)EMS + (JMS - EMS)\frac{k}{n}}$$

In this formula, *BMS* is the between-subject mean square, *JMS* is the between-session mean square, *EMS* is the mean square error, *k* is the number of sessions and *n* is the number of subjects. In this study, *k* = 3 and *n* = 16. According to Winer (1971), ICC < 0.25 is poor, 0.25–0.4 is low, 0.4–0.6 is fair, 0.6–0.75 is good, and 0.75–1.0 is excellent [55], which is what we assumed. ICCs were calculated based on SPSS (SPSS Inc. Released 2007; SPSS for Windows, Version 16.0; SPSS Inc., Chicago, IL, USA) and MATLAB 9.2 (MathWorks, Natick, MA, USA).

In addition, we calculated the within-subject coefficient of variation (CV) to account for the relative uncertainty. If the standard deviation is denoted by *S*, and the mean is denoted by *M*, then the coefficient of variation is calculated as:

$$\text{CV} = \frac{S}{M} \quad (1)$$

2.7. Statistical Analyses

We performed the following statistical analyses to systematically compare the TRT reliabilities of several metrics under different data processing and modeling strategies.

For global metrics, we first used a one-way repeated ANOVA to compare the effect of network thresholding on the ICC for all the metrics under all the methods. Because there were no significant differences in any metric across the thresholds, in the subsequent analysis, we focused on the average ICC across different thresholds. Second, we used paired sample *t*-tests to compare the ICC (pooled over all metrics) between methods, including contrasts between weighted and binary networks, the AAL90 atlas and Power264 atlas, and with and without GSR. Finally, we investigated the effect of the inter-scan interval on the TRT reliability of the graph-theoretical metrics. Because each subject was scanned three times within a week, we calculated pair-wise ICC between all pairs of visits: visit 1 and visit 2 (v1v2), visit 1 and visit 3 (v1v3), and visit 2 and visit 3 (v2v3). We then conducted a two-way repeated ANOVA on all the ICC, including a main effect of inter-scan intervals (three levels: v1v2, v1v3, v2v3), a main effect of modeling method (four levels: BG, binary network with global signal regression; BNG, binary network without global signal regression; WG, weighted network with global signal regression; WNG, weighted network without global signal regression), and an interaction effect between them. We also conducted a one-way repeated ANOVA using interval as the within-subject main effect (including post hoc *t*-tests) separately for each method. Additionally, we tested how individual characteristics affect TRT reliability by splitting the subjects based on biological sex (8 males and 8 females) and age (8 in the younger group and 8 in the older group, split by a median of 34.5). We calculated the ICC for each of the groups for all global metrics and across all methods and conducted paired sample *t*-tests between males and females, and younger and older subjects, respectively (each pair of samples had the same metric and method).

For nodal metrics, we performed a one-way ANOVA (plus post hoc pair-wise comparisons) on each metric (betweenness centrality, degree centrality, and nodal efficiency) with the main effect of the modeling method. We included all eight methods in the analysis. We also performed a paired *t*-test between all ICCs (pooled over all metrics) calculated with and without GSR.

3. Results

3.1. Test-Retest Reliability of Seven Global Metrics

Figure 1 displays the ICC values for all seven-network metrics across different data processing and modeling strategies. In general, most global network metrics exhibited poor to fair TRT reliability (ICC: 0.32 ± 0.15 , CV: $8.1\% \pm 6.0\%$). The least stable metric was small-worldness (sigma, ICC: 0.19 ± 0.07 , CV: $9.9\% \pm 4.8\%$), and the most reliable metric was the normalized characteristic path length (Lambda, ICC: 0.39 ± 0.16 , CV: $3.0\% \pm 1.6\%$). In addition, the ICC of the normalized characteristic path length reached its highest level (ICC: 0.71 ± 0.002 , CV: $3.5\% \pm 1.3\%$) using the AAL90 atlas, weighted network modeling, and GSR.

Because the main effect of the threshold (one-way repeated ANOVA) within the range of 0.15–0.35 was not significant ($F(1.70, 69.5) = 1.66$, $p = 0.20$), in the following statistical analysis we adopted average ICC over thresholds as independent samples within the groups.

To investigate the necessity of adopting GSR, weighted network, or Power264 atlas, we compared the mean ICC resulting from methods containing these elements. Paired sample *t*-tests (Figure 2) revealed that the reliability of weighted network metrics was significantly higher than that of binary network metrics ($T(27) = 2.41$, $p = 0.022$), and the use of the AAL90 atlas for brain parcellation provided higher ICC values than the use of the Power264 atlas ($T(27) = 3.95$, $p = 0.001$). However, GSR did not have a significant effect on the ICC values ($T(27) = 1.15$, $p = 0.26$).

We further assessed whether inherent individual characteristics affect the TRT reliability of the brain network metrics. We split our subjects into two types of groups based on their biological sex and age. Paired sample *t*-test (Figure 3) showed that sex did not have a significant impact on the overall ICC among brain metrics (ICC of male: 0.27 ± 0.21 , ICC of female: 0.30 ± 0.18 , $T(55) = -0.68$, $p = 0.50$), whereas age did have an impact, in which the older group had a significantly higher ICC than the younger group (younger group: age: 27.25 ± 3.95 , ICC: 0.19 ± 0.17 ; older group: age: 43.63 ± 5.13 , ICC: 0.40 ± 0.23 , $T(55) = -5.18$, $p < 0.0001$).

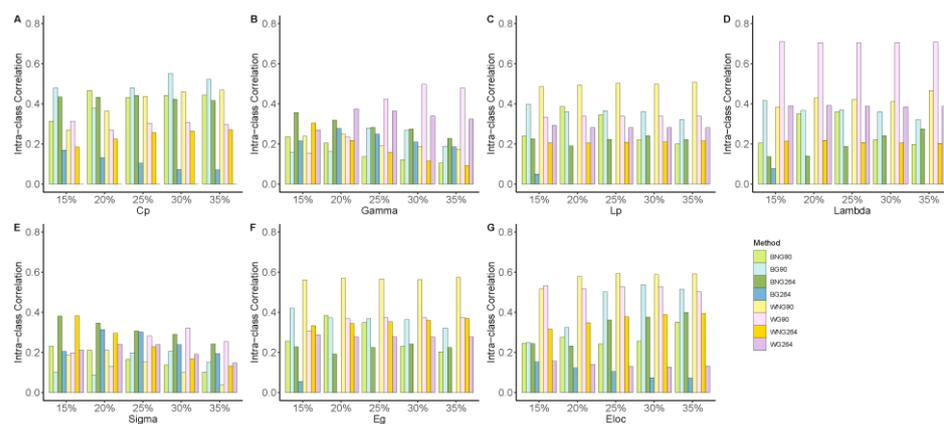


Figure 1. The ICC values for all seven-network metrics across different data processing and modeling strategies. Subfigures show the decomposition of ICC values across network density levels 15–35% for (A) Cp, (B) Gamma, (C) Lp, (D) Lambda, (E) Sigma, (F) Eg, and (G) Eloc, respectively. BG90: binary + global signal regression + AAL90, BNG90: binary + no global signal regression + AAL90, WG90: weighted + global signal regression + AAL90, WNG90: weighted + no global signal regression + AAL90, WG264: weighted + global signal regression + Power264, WNG264: weighted + no global signal regression + Power264; Cp: Clustering coefficient; Lp: Characteristic path length; Eg: Global efficiency; El: Local efficiency.

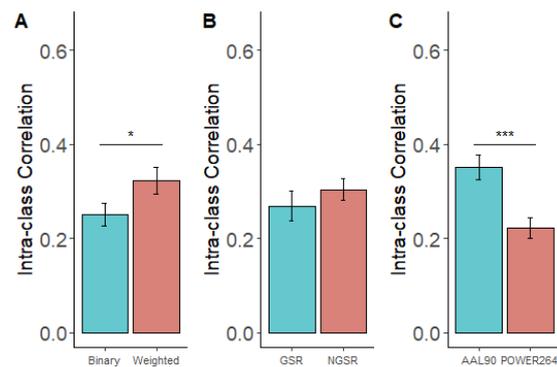


Figure 2. Comparisons of TRT reliabilities between three groups of methods. **(A)** Comparison of ICC values between the use of binary networks and weighted networks. **(B)** Comparison of ICC values between the use of GSR or not (NGSR). **(C)** Comparison of ICC values between the use of AAL90 and Power264 atlases. *: $p < 0.05$, ***: $p < 0.001$.

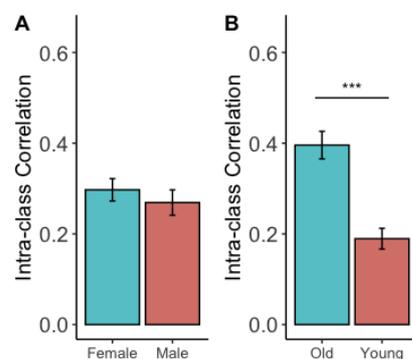


Figure 3. TRT reliabilities across different individual characteristics. **(A)** Comparison of the overall ICC values between males and females. **(B)** Comparison of the overall ICC values between the younger group and the older group. ***: $p < 0.001$.

3.2. Reliability of Two Visits

Using a two-way repeated ANOVA, we compared the TRT reliability between pairs of visits (i.e., different inter-scan intervals) across four methods (BG, BNG, WG, and WNG). Our results showed that there was not a significant main effect of inter-scan interval ($F(1.15, 6.88) = 0.354, p = 0.60$) or method ($F(1.88, 11.29) = 3.32, p = 0.076$), but a significant interaction effect ($F(6, 36) = 4.66, p = 0.001$) was observed between them. Figure 4 shows the relationship between TRT reliability and interval under four methods. WG yielded the highest value (ICC: 0.413 ± 0.059), while BNG yielded the lowest average ICC (ICC: 0.264 ± 0.032).

To test the effect of interval on each method, we conducted a one-way repeated ANOVA using interval as the within-subject factor. The main effects of the interval were not significant for BG ($F(1.1, 6.4) = 0.26, p = 0.77$), BNG ($F(1.1, 6.3) = 2.68, p = 0.15$), and WNG ($F(2, 12) = 0.174, p = 0.84$), but they were significant for WG ($F(2, 12) = 21.5, p < 0.001$). For WG, post hoc *t*-tests (Bonferroni) revealed significantly lower ICC between visit1 and visit3 than that between visit1 and visit2 ($p = 0.009$) or between visit2 and visit3 ($p < 0.001$).

Table 2 shows the ICC of the seven metrics between three pairs of visits averaged over the threshold of 0.15–0.35. Given a certain method, the ICC of different metrics changed along the interval in different patterns. For example, the reliabilities of the normalized clustering coefficient (γ) and small-worldness (σ) decreased with an increase in the interval, regardless of the network construction methods. In addition, under the WG method, the ICC of all metrics, except the normalized characteristic path length (Λ), decreased with a longer interval.

Table 2. ICC between pairs of visits.

	BG			BNG			WG			WNG		
	v1v2	v2v3	v1v3									
Cp	0.519	0.521	0.427	0.416	0.317	0.557	0.463	0.302	0.176	0.324	0.332	0.543
Gamma	0.477	0.262	0.000	0.209	0.247	0.066	0.516	0.389	0.202	0.217	0.299	0.127
Lambda	0.253	0.317	0.473	0.159	0.120	0.474	0.662	0.788	0.670	0.548	0.449	0.336
Lp	0.250	0.312	0.465	0.213	0.127	0.455	0.480	0.401	0.158	0.476	0.408	0.589
Sigma	0.435	0.198	0.000	0.208	0.226	0.096	0.295	0.308	0.132	0.201	0.220	0.057
Eg	0.258	0.321	0.471	0.231	0.130	0.460	0.487	0.449	0.188	0.492	0.606	0.593
Elocal	0.434	0.485	0.373	0.231	0.274	0.330	0.586	0.589	0.426	0.449	0.584	0.668

Note: v1: The first visit on day 2; v2: the second visit on day 3; v3: the third visit on day 5; BG: binary network with global signal regression; BNG: binary network without global signal regression; WG: weighted network with global signal regression; WNG: weighted network without global signal regression.

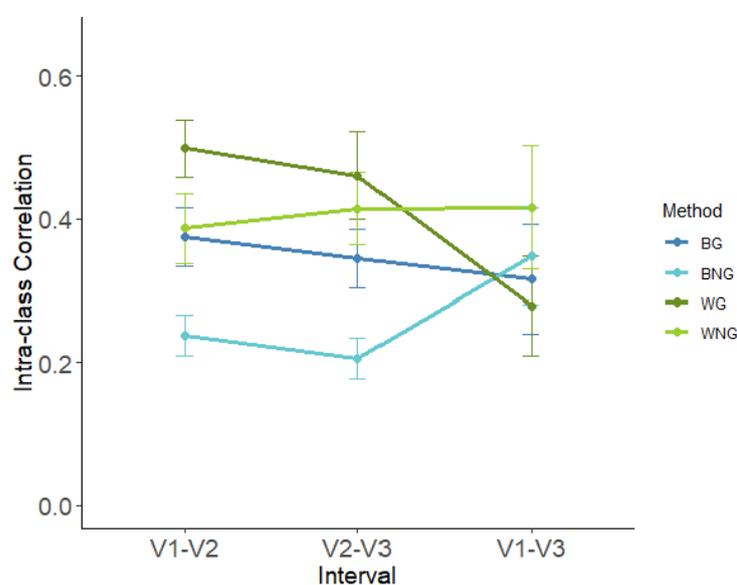


Figure 4. Average of test-retest reliability of global metrics as the function of interval. BG: binary + global signal regression, BNG: binary + no global signal regression, WG: weighted + global signal regression, WNG: weighted + no global signal regression. V1-V2: visit 1 (day 2) and visit 2 (day 3), V2-V3: visit 2 and visit 3 (day 5), and V1-V3: visit 1 and visit 3.

3.3. Reliability of Three Nodal Metrics

The results of the reliability analysis for nodal metrics are shown in Figure 5. Generally, nodal efficiency has the highest TRT reliability (ICC: 0.33 ± 0.07 , CV: $13.6\% \pm 5.5\%$), betweenness centrality has the least reliable nodal metric (ICC: 0.18 ± 0.05 , CV: $106.5\% \pm 31.1\%$), and the reliability of degree centrality is in the middle (ICC: 0.31 ± 0.06 , CV = $36.9\% \pm 12.3\%$).

Among all the modeling strategies, three of them resulted in higher reliabilities for each nodal metric compared with the rest of the strategies: BNG264, WNG90, and WNG264. In particular, the ICC of nodal efficiency under the WNG264 method was the highest (ICC: 0.45 ± 0.16 , CV: $20.7\% \pm 5.3\%$). A one-way ANOVA showed significant differences in the TRT reliability across the eight modeling strategies for betweenness centrality ($F(7, 1408) = 19.92$, $p < 0.001$), degree centrality ($F(7, 1408) = 22.98$, $p < 0.001$), and nodal efficiency ($F(7, 1408) = 36.07$, $p < 0.001$). The results of the post hoc analysis (Bonferroni corrected) are shown in Supplementary Tables S1–S3. WNG264 yielded a significantly higher ICC than BG90, WG90, and WG264 for all three metrics, and its reliability was even significantly better than WNG90 for nodal efficiency. Considering that all the top three methods did not use GSR, we further combined the average nodal ICCs of all the metrics to test the influence of GSR per sec by a paired *t*-test, which proved that regressing out the global signal significantly decreases the nodal TRT reliability (Δ ICC: 0.11 ± 0.03 , $p < 0.001$).

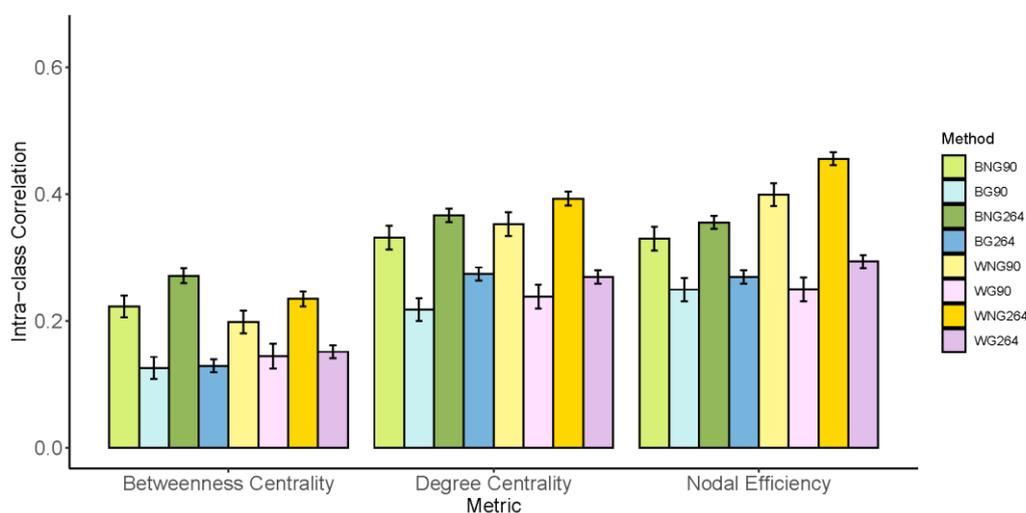


Figure 5. Test-retest reliability of nodal metrics across different data processing and modeling strategies. BG90: binary + global signal regression + AAL90, BNG90: binary + no global signal regression + AAL90, WG90: weighted + global signal regression + AAL90, WNG90: weighted + no global signal regression + AAL90, WG264: weighted + global signal regression + Power264, WNG264: weighted + no global signal regression + Power264.

4. Discussion

According to previous studies, the major threats to the TRT reliability of rs-fMRI include scan conditions [56], physiological noise [57,58], data preprocessing, and network construction strategies [18–20,58]. In this study, we systematically investigated the test-retest reliability of the brain network topology based on strictly controlled rs-fMRI data across different preprocessing and modeling strategies. The overall reliability among the global network metrics was poor to moderate (0.000~0.592), except for the normalized characteristic path length (λ). The ICC of λ reached a good level (0.705~0.711) when we applied weighted connections and global signal removal. Similar results were reported by Wang et al., (2011), who found moderate reliability in λ , despite poor to low reliabilities in all other global metrics [20]. The TRT reliability of individual nodes had a large nodal variation (0.000~0.811), and the nodal efficiency (0.456) had the highest average ICC among the three metrics when we constructed weighted networks using the Power 264 atlas without global signal removal.

One possible reason for the commonly found low ICC is that the ICC reflects the ratio of between-subject to within-subject variability. Since the functional connectivity of rs-fMRI across subjects could be highly homogeneous (i.e., small between-subject variability), the variation within the subjects was not small enough to yield a high ICC [59]. In support of this argument, higher reliability was found during the task state than the resting state due to the higher stability of event-related co-activations [60]. However, the reliability pattern depends on the content of the task; some tasks could improve the global ICC, whereas others could impair the global ICC [52]. In addition, despite a low ICC, the test-retest reliability could still be moderate to high [61]. Thus, it is important to also examine other criteria for TRT reliability. In this study, we also calculated the coefficient of variation. In contrast to the conclusion from ICC, the overall CV of global metrics was good ($8.1\% \pm 6.0\%$), indicating small within-subject variations. Nevertheless, the CV of some nodal metrics, especially the betweenness centrality, was high ($106.5\% \pm 31.1\%$), which was consistent with its poor ICC.

4.1. Factor Affecting the TRT Reliability of Global Metrics

For global metrics, comparisons between the methods revealed that the weighted network was generally more reliable than the binary network. One crucial aspect of the

weighted network is that it preserves more detailed information on connectivity strength. As a result, the weighted network can detect subtle changes in connectivity, and this complexity leads to high resistance to external disturbances. Many other studies have also focused on this issue, and one review [18] reported a slight advantage of weighted methods by analyzing data from many previous studies [19,20,62–65]. Nevertheless, most studies today still prefer binary networks, partly due to their simplicity of interpretation.

Removing the global signal showed a slight but non-significant disadvantage on the global metrics. Similar results have been reported by Andellini et al., (2015). Even though GSR had a significant negative effect on the reliability of the clustering coefficient, the overall effect across metrics was not significant [18]. In another study, the reliability of Lambda even increased after GSR [19]. Future research needs to investigate this issue with a broader sample size and should potentially address the effect of GSR while varying other steps during the network construction.

In addition to the major findings regarding data processing and modeling strategies, we also examined whether individual demographic characteristics (sex and age) affected the TRT reliability of network global metrics. While we did not find significant differences in ICC between males and females, we found that subjects in the older group (35–50 years) had higher ICC than those in the younger group (22–34 years). Previous studies comparing the ICC of BOLD fMRI across age groups suggested an inverse U-shaped relationship as follows: the ICC is lower during infancy and childhood, peaks in adulthood with the maturation of the brain, and decreases in older adults [66–68]. However, such evidence is scarce, and it is not clear which age range has the highest ICC. Our results, although preliminary and limited in sample size ($N = 8$ for each group), can provide a finer characterization of the relationship.

4.2. The Effect of Inter-Scan Interval on TRT Reliability

For TRT reliability between pairs of visits, we only found subtle but not significant differences. In other words, the overall TRT reliability of global metrics remained stable during the entire study and was independent of the passage of time. However, such robustness to inter-scan intervals relied on specific data processing strategies. For instance, the reliability of metrics under WG decreased when the inter-scan intervals were longer. The source of variability in longer intervals may be the fact that participants kept adjusting their lifestyles and biological clocks in the new environment. Among all metrics, longer intervals mostly affected the reliability of Gamma and Sigma.

4.3. TRT Reliability of Nodal Metrics

All three nodal metrics exhibited poor to low reliabilities on average, yet the degree centrality and nodal efficiency were more reliable than the betweenness centrality, consistent with the findings of Du et al., (2015) [45]. These results could be explained by the definitions of these metrics; the degree centrality and nodal efficiency of a node only depend on direct connections with it, whereas betweenness centrality is calculated by the connections of adjacent nodes. As a result, connectivity changes in a remote node will have an impact on the betweenness centrality of the current node, but not on the degree centrality or nodal efficiency. Therefore, the reliability of betweenness centrality is in general low.

What is evident from the comparison of methods is that applying GSR significantly lowers the TRT reliability for both binary and weighted networks and with both parcelations. This is inconsistent with Du et al., (2015), who detected slight but not significant disadvantages of the GSR [45]. Such a difference may be due to the different network sizes in the two studies; in Du et al., (2015), the networks were based on 25,218 voxels that were much larger than those in the current study [45]. Therefore, the benefit of GSR in improving sensitivity offsets the loss of reliability in much denser networks.

4.4. Opposite Effects of Parcellations on Global and Nodal Metrics

It is worth noting that the effect of parcellations on the global metrics was opposite to that on the nodal metrics: the AAL-90 atlas generated higher reliabilities for global metrics than the Power264 atlas, whereas the optimal processing and modeling strategy for nodal network metrics was the one that applied the Power264 atlas (i.e., WNG264). There are many differences between these two atlases. The AAL-90 defines 90 brain regions based on anatomical features, and each ROI encompasses a wide range of brain tissues, whereas the Power264 atlas defines 264 spherical ROIs based on functional co-activations, and each ROI is a spherical region containing a fixed and limited number of voxels. Therefore, on a global scale, ROIs from the AAL-90 atlas always come from the same inherent anatomical organizations, independent of the functional state of the brain, and produce a more robust pattern of connectivity in general. On the other hand, on a nodal scale, the Power264 atlas has higher spatial resolutions, and its use of smaller ROIs reduces the regional inhomogeneity of the ROI [69]. For a given local property, the reliability of a node can be less affected by the signals of its surrounding areas.

4.5. TRT Reliability of BOLD fMRI Compared to Other Modalities

In addition, the reliability of BOLD fMRI may become a disadvantage compared with other modalities. For example, cerebral blood flow (CBF) quantified by arterial spin-labeled (ASL) perfusion MRI couples with regional brain activity, perfusion, and metabolism [70], which also serve as biomarkers in clinical settings. While the common ICC of resting-state BOLD fMRI is poor to moderate (0.2–0.6) [20,27,45,46,71], that of CBF is usually greater than 0.6, falling in the good to excellent range [72–74]. In our previous work, we evaluated the TRT of resting-state and task-based absolute CBF, as well as task-induced relative CBF, and found that ICC values ranged from good to excellent (ICC > 0.6) for absolute CBF and poor for relative CBF (ICC < 0.4) [75]. On average, absolute CBF, rather than relative CBF, has a better TRT than the small world network properties based on rs-fMRI.

4.6. Limitations

Our study has several limitations. The first limitation is the choice of parcellation schemes. In the current comparison, we only included two representative parcellations: AAL-90 and Power264. To better understand the impact of parcellation, future studies should include more atlases in the comparison with different numbers of subdivisions and different parcellation algorithms. A fine-grained strategy, such as the voxel-wise analysis, should also be considered. For example, previous studies have reported overall good to high ICCs with voxel-wise network construction [76].

Another limitation is the limited range of the inter-scan interval. An eligible biomarker should remain stable in the long term in the absence of a disease. In our study, the longest interval was three days; thus, it was difficult to test such eligibility directly. Instead, we only examined the performance of graph theoretical metrics after a short-term manipulation and tried to predict the long-term effects.

Finally, in this initial study, we were only able to test a small sample of 16 control subjects in a very strictly controlled in-laboratory 5-day and 4-night study. Although the use of strict controls added to the methodological strength of the study to help rule out potential confounders during the data collection, a potential drawback is that it is unclear whether our findings can be generalized to most other studies that do not have such a stringent design. Future replications with larger sample sizes and different data collection protocols are needed to test the generalizability of the current findings.

5. Conclusions

In summary, in the existing literature, as of now we are the first to comprehensively investigate the influence of data processing and modeling strategies on the TRT reliability of both global- and nodal-level graph-theoretical metrics while strictly controlling the subjects'

behaviors. By strictly monitoring the daily activities of our subjects in the laboratory for five days, we attenuated the impact of external factors on brain activities.

Several important suggestions can be derived from our findings for the implications of the future utility of graph theoretical modeling in brain network analyses. First, when using a global network metric as a clinical biomarker, the normalized characteristic path length is highly recommended. Based on our results, it has the highest ICC among all the global metrics, especially when calculated using weighted networks and global signal regression. In terms of general methods, researchers should consider using weighted networks instead of binary networks and using AAL-90 instead of the Power264 atlas, as they both showed significantly higher ICC than their counterparts. Researchers should be cautious when applying global signal removal. Similarly to previous studies, we did not find a clear advantage of including or excluding GSR in the resulting ICC. When using a nodal network metric, we recommend degree centrality and nodal efficiency, as both have consistently better TRT reliability than betweenness centrality. Finally, the reliability of brain network metrics may decline longitudinally, which is a threat to experiments with long scan intervals.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/brainsci13050825/s1>, Table S1: Pair-wise differences of the betweenness centrality ICC among eight strategies. Table S2: Pair-wise differences of the degree centrality ICC among eight strategies. Table S3: Pair-wise differences of the nodal efficiency ICC among eight strategies.

Author Contributions: H.R. conceived the overall project and edited the manuscript. J.A.D. developed the protocol for data collection and edited the methodology. Q.W., H.L., X.Z. (Xue Zhong) and J.L. drafted the article and analyzed the data; T.M., Y.D., Y.J. and X.Z. (Xiaocui Zhang) collected and interpreted the data. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported in part by the National Natural Science Foundation of China (71942003), National Institutes of Health grants (R01-HL102119, R21-AG051981), and Shanghai International Studies University Research Projects (20171140020). The funders had no role in the study design, data collection and analysis, data interpretation, writing of the manuscript, or the decision to submit the article for publication.

Institutional Review Board Statement: The institutional review committees of the University of Pennsylvania approved the survey.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The original contributions presented in the study are included in the article. Further inquiries can be directed to the corresponding authors.

Acknowledgments: We thank Joy Rao for her help in the writing and editing of the manuscript. We are also grateful for the generosity of time and effort by all the participants and all the researchers who made this project possible.

Conflicts 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.

References

1. Fox, M.D.; Raichle, M.E. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nat. Rev. Neurosci.* **2007**, *8*, 700–711. [[CrossRef](#)] [[PubMed](#)]
2. Raichle, M.E.; Mintun, M.A. Brain work and brain imaging. *Annu. Rev. Neurosci.* **2006**, *29*, 449–476. [[CrossRef](#)] [[PubMed](#)]
3. Friston, K.J.; Frith, C.D.; Liddle, P.F.; Frackowiak, R.S. Functional connectivity: The principal-component analysis of large (PET) data sets. *J. Cereb. Blood Flow Metab.* **1993**, *13*, 5–14. [[CrossRef](#)] [[PubMed](#)]
4. Van den Heuvel, M.P.; Hulshoff Pol, H.E. Exploring the brain network: A review on resting-state fMRI functional connectivity. *Eur. Neuropsychopharmacol.* **2010**, *20*, 519–534. [[CrossRef](#)] [[PubMed](#)]
5. Bullmore, E.; Sporns, O. The economy of brain network organization. *Nat. Rev. Neurosci.* **2012**, *13*, 336–349. [[CrossRef](#)]
6. Watts, D.J.; Strogatz, S.H. Collective dynamics of ‘small-world’ networks. *Nature* **1998**, *393*, 440–442. [[CrossRef](#)]
7. Latora, V.; Marchiori, M. Efficient behavior of small-world networks. *Phys. Rev. Lett.* **2001**, *87*, 198701. [[CrossRef](#)]

8. Bassett, D.S.; Bullmore, E. Small-world brain networks. *Neuroscientist* **2006**, *12*, 512–523. [[CrossRef](#)]
9. Liao, X.; Vasilakos, A.V.; He, Y. Small-world human brain networks: Perspectives and challenges. *Neurosci. Biobehav. Rev.* **2017**, *77*, 286–300. [[CrossRef](#)]
10. Ajilore, O.; Lamar, M.; Leow, A.; Zhang, A.; Yang, S.; Kumar, A. Graph theory analysis of cortical-subcortical networks in late-life depression. *Am. J. Geriatr. Psychiatry* **2014**, *22*, 195–206. [[CrossRef](#)]
11. Bruno, J.; Hosseini, S.M.; Kesler, S. Altered resting state functional brain network topology in chemotherapy-treated breast cancer survivors. *Neurobiol. Dis.* **2012**, *48*, 329–338. [[CrossRef](#)] [[PubMed](#)]
12. Hayasaka, S.; Laurienti, P.J. Comparison of characteristics between region-and voxel-based network analyses in resting-state fMRI data. *Neuroimage* **2010**, *50*, 499–508. [[CrossRef](#)]
13. Luo, C.Y.; Guo, X.Y.; Song, W.; Chen, Q.; Cao, B.; Yang, J.; Gong, Q.Y.; Shang, H.F. Functional connectome assessed using graph theory in drug-naïve Parkinson’s disease. *J. Neurol.* **2015**, *262*, 1557–1567. [[CrossRef](#)] [[PubMed](#)]
14. Sanz-Arigita, E.J.; Schoonheim, M.M.; Damoiseaux, J.S.; Rombouts, S.A.R.B.; Maris, E.; Barkhof, F.; Scheltens, P.; Stam, C.J. Loss of ‘Small-World’ Networks in Alzheimer’s Disease: Graph Analysis of fMRI Resting-State Functional Connectivity. *PLoS ONE* **2010**, *5*, e13788. [[CrossRef](#)] [[PubMed](#)]
15. Tarchi, L.; Damiani, S.; Fantoni, T.; Pisano, T.; Castellini, G.; Politi, P.; Ricca, V. Centrality and interhemispheric coordination are related to different clinical/behavioral factors in attention deficit/hyperactivity disorder: A resting-state fMRI study. *Brain Imaging Behav.* **2022**, *16*, 2526–2542. [[CrossRef](#)] [[PubMed](#)]
16. Bennett, C.M.; Miller, M.B. How reliable are the results from functional magnetic resonance imaging? *Ann. N. Y. Acad. Sci.* **2010**, *1191*, 133–155. [[CrossRef](#)] [[PubMed](#)]
17. Noble, S.; Scheinost, D.; Constable, R.T. A decade of test-retest reliability of functional connectivity: A systematic review and meta-analysis. *Neuroimage* **2019**, *203*, 116157. [[CrossRef](#)]
18. Andellini, M.; Cannata, V.; Gazzellini, S.; Bernardi, B.; Napolitano, A. Test-retest reliability of graph metrics of resting state MRI functional brain networks: A review. *J. Neurosci. Methods* **2015**, *253*, 183–192. [[CrossRef](#)]
19. Braun, U.; Plichta, M.M.; Esslinger, C.; Sauer, C.; Haddad, L.; Grimm, O.; Mier, D.; Mohnke, S.; Heinz, A.; Erk, S.; et al. Test-retest reliability of resting-state connectivity network characteristics using fMRI and graph theoretical measures. *Neuroimage* **2012**, *59*, 1404–1412. [[CrossRef](#)]
20. Wang, J.H.; Zuo, X.N.; Gohel, S.; Milham, M.P.; Biswal, B.B.; He, Y. Graph theoretical analysis of functional brain networks: Test-retest evaluation on short- and long-term resting-state functional MRI data. *PLoS ONE* **2011**, *6*, e21976. [[CrossRef](#)]
21. Xiang, J.; Xue, J.; Guo, H.; Li, D.; Cui, X.; Niu, Y.; Yan, T.; Cao, R.; Ma, Y.; Yang, Y.; et al. Graph-based network analysis of resting-state fMRI: Test-retest reliability of binarized and weighted networks. *Brain Imaging Behav.* **2019**, *14*, 1361–1372. [[CrossRef](#)] [[PubMed](#)]
22. Caviness, V.S., Jr.; Meyer, J.; Makris, N.; Kennedy, D.N. MRI-Based Topographic Parcellation of Human Neocortex: An Anatomically Specified Method with Estimate of Reliability. *J. Cogn. Neurosci.* **1996**, *8*, 566–587. [[CrossRef](#)] [[PubMed](#)]
23. Dosenbach, N.U.; Visscher, K.M.; Palmer, E.D.; Miezin, F.M.; Wenger, K.K.; Kang, H.C.; Burgund, E.D.; Grimes, A.L.; Schlaggar, B.L.; Petersen, S.E. A core system for the implementation of task sets. *Neuron* **2006**, *50*, 799–812. [[CrossRef](#)] [[PubMed](#)]
24. Fan, L.; Li, H.; Zhuo, J.; Zhang, Y.; Wang, J.; Chen, L.; Yang, Z.; Chu, C.; Xie, S.; Laird, A.R.; et al. The Human Brainnetome Atlas: A New Brain Atlas Based on Connectional Architecture. *Cereb. Cortex* **2016**, *26*, 3508–3526. [[CrossRef](#)] [[PubMed](#)]
25. Power, J.D.; Cohen, A.L.; Nelson, S.M.; Wig, G.S.; Barnes, K.A.; Church, J.A.; Vogel, A.C.; Laumann, T.O.; Miezin, F.M.; Schlaggar, B.L.; et al. Functional network organization of the human brain. *Neuron* **2011**, *72*, 665–678. [[CrossRef](#)]
26. Tzourio-Mazoyer, N.; Landeau, B.; Papathanassiou, D.; Crivello, F.; Etard, O.; Delcroix, N.; Mazoyer, B.; Joliot, M. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* **2002**, *15*, 273–289. [[CrossRef](#)]
27. Termenon, M.; Jaillard, A.; Delon-Martin, C.; Achard, S. Reliability of graph analysis of resting state fMRI using test-retest dataset from the Human Connectome Project. *Neuroimage* **2016**, *142*, 172–187. [[CrossRef](#)]
28. Evans, A.C.; Janke, A.L.; Collins, D.L.; Baillet, S. Brain templates and atlases. *Neuroimage* **2012**, *62*, 911–922. [[CrossRef](#)]
29. Macey, P.M.; Macey, K.E.; Kumar, R.; Harper, R.M. A method for removal of global effects from fMRI time series. *Neuroimage* **2004**, *22*, 360–366. [[CrossRef](#)]
30. Murphy, K.; Birn, R.M.; Handwerker, D.A.; Jones, T.B.; Bandettini, P.A. The impact of global signal regression on resting state correlations: Are anti-correlated networks introduced? *Neuroimage* **2009**, *44*, 893–905. [[CrossRef](#)]
31. Aguirre, G.K.; Zarahn, E.; D’Esposito, M. Empirical analyses of BOLD fMRI statistics. II. Spatially smoothed data collected under null-hypothesis and experimental conditions. *Neuroimage* **1997**, *5*, 199–212. [[CrossRef](#)]
32. Aguirre, G.K.; Zarahn, E.; D’Esposito, M. The inferential impact of global signal covariates in functional neuroimaging analyses. *Neuroimage* **1998**, *8*, 302–306. [[CrossRef](#)] [[PubMed](#)]
33. Rubinov, M.; Sporns, O. Complex network measures of brain connectivity: Uses and interpretations. *Neuroimage* **2010**, *52*, 1059–1069. [[CrossRef](#)]
34. Janes, A.C.; Nickerson, L.D.; Frederick Bde, B.; Kaufman, M.J. Prefrontal and limbic resting state brain network functional connectivity differs between nicotine-dependent smokers and non-smoking controls. *Drug Alcohol Depend.* **2012**, *125*, 252–259. [[CrossRef](#)]

35. Khazaie, H.; Veronese, M.; Noori, K.; Emamian, F.; Zarei, M.; Ashkan, K.; Leschziner, G.D.; Eickhoff, C.R.; Eickhoff, S.B.; Morrell, M.J.; et al. Functional reorganization in obstructive sleep apnoea and insomnia: A systematic review of the resting-state fMRI. *Neurosci. Biobehav. Rev.* **2017**, *77*, 219–231. [[CrossRef](#)] [[PubMed](#)]
36. Wang, T.; Yan, J.; Li, S.; Zhan, W.; Ma, X.; Xia, L.; Li, M.; Lin, C.; Tian, J.; Li, C.; et al. Increased insular connectivity with emotional regions in primary insomnia patients: A resting-state fMRI study. *Eur. Radiol.* **2017**, *27*, 3703–3709. [[CrossRef](#)] [[PubMed](#)]
37. Wang, Z.; Suh, J.; Li, Z.; Li, Y.; Franklin, T.; O'Brien, C.; Childress, A.R. A hyper-connected but less efficient small-world network in the substance-dependent brain. *Drug Alcohol Depend.* **2015**, *152*, 102–108. [[CrossRef](#)] [[PubMed](#)]
38. Patriat, R.; Molloy, E.K.; Meier, T.B.; Kirk, G.R.; Nair, V.A.; Meyerand, M.E.; Prabhakaran, V.; Birn, R.M. The effect of resting condition on resting-state fMRI reliability and consistency: A comparison between resting with eyes open, closed, and fixated. *Neuroimage* **2013**, *78*, 463–473. [[CrossRef](#)]
39. Zou, Q.H.; Long, X.Y.; Zuo, X.N.; Yan, C.G.; Zhu, C.Z.; Yang, Y.H.; Liu, D.Q.; He, Y.; Zang, Y.F. Functional Connectivity Between the Thalamus and Visual Cortex Under Eyes Closed and Eyes Open Conditions: A Resting-State fMRI Study. *Hum. Brain Mapp.* **2009**, *30*, 3066–3078. [[CrossRef](#)]
40. Biswal, B.B.; Mennes, M.; Zuo, X.N.; Gohel, S.; Kelly, C.; Smith, S.M.; Beckmann, C.F.; Adelstein, J.S.; Buckner, R.L.; Colcombe, S.; et al. Toward discovery science of human brain function. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 4734–4739. [[CrossRef](#)]
41. Jovicich, J.; Minati, L.; Marizzoni, M.; Marchitelli, R.; Sala-Llonch, R.; Bartres-Faz, D.; Arnold, J.; Benninghoff, J.; Fiedler, U.; Roccatagliata, L.; et al. Longitudinal reproducibility of default-mode network connectivity in healthy elderly participants: A multicentric resting-state fMRI study. *Neuroimage* **2016**, *124*, 442–454. [[CrossRef](#)] [[PubMed](#)]
42. Buysse, D.J.; Nofzinger, E.A.; Germain, A.; Meltzer, C.C.; Wood, A.; Ombao, H.; Kupfer, D.J.; Moore, R.Y. Regional brain glucose metabolism during morning and evening wakefulness in humans: Preliminary findings. *Sleep* **2004**, *27*, 1245–1254. [[CrossRef](#)]
43. Hodkinson, D.J.; O'Daly, O.; Zunszain, P.A.; Pariante, C.M.; Lazurenko, V.; Zelaya, F.O.; Howard, M.A.; Williams, S.C. Circadian and homeostatic modulation of functional connectivity and regional cerebral blood flow in humans under normal entrained conditions. *J. Cereb. Blood Flow Metab.* **2014**, *34*, 1493–1499. [[CrossRef](#)] [[PubMed](#)]
44. Trefler, A.; Sadeghi, N.; Thomas, A.G.; Pierpaoli, C.; Baker, C.I.; Thomas, C. Impact of time-of-day on brain morphometric measures derived from T1-weighted magnetic resonance imaging. *Neuroimage* **2016**, *133*, 41–52. [[CrossRef](#)] [[PubMed](#)]
45. Du, H.X.; Liao, X.H.; Lin, Q.X.; Li, G.S.; Chi, Y.Z.; Liu, X.; Yang, H.Z.; Wang, Y.; Xia, M.R. Test-retest reliability of graph metrics in high-resolution functional connectomics: A resting-state functional MRI study. *CNS Neurosci. Ther.* **2015**, *21*, 802–816. [[CrossRef](#)]
46. Jin, D.; Xu, K.; Liu, B.; Jiang, T.; Liu, Y. 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, USA, 18–21 July 2018; Volume 2018, pp. 1028–1031. [[CrossRef](#)]
47. Fang, Z.; Spaeth, A.M.; Ma, N.; Zhu, S.; Hu, S.; Goel, N.; Detre, J.A.; Dinges, D.F.; Rao, H. Altered salience network connectivity predicts macronutrient intake after sleep deprivation. *Sci. Rep.* **2015**, *5*, 8215. [[CrossRef](#)]
48. Yang, F.N.; Xu, S.; Chai, Y.; Basner, M.; Dinges, D.F.; Rao, H. Sleep deprivation enhances inter-stimulus interval effect on vigilant attention performance. *Sleep* **2018**, *41*, zsy189. [[CrossRef](#)]
49. Wang, J.; Qiu, S.; Xu, Y.; Liu, Z.; Wen, X.; Hu, X.; Zhang, R.; Li, M.; Wang, W.; Huang, R. Graph theoretical analysis reveals disrupted topological properties of whole brain functional networks in temporal lobe epilepsy. *Clin. Neurophysiol.* **2014**, *125*, 1744–1756. [[CrossRef](#)] [[PubMed](#)]
50. Wang, J.; Wang, X.; Xia, M.; Liao, X.; Evans, A.; He, Y. GREYNA: A graph theoretical network analysis toolbox for imaging connectomics. *Front. Hum. Neurosci.* **2015**, *9*, 386. [[CrossRef](#)]
51. Medaglia, J.D. Graph Theoretic Analysis of Resting State Functional MR Imaging. *Neuroimaging Clin. N. Am.* **2017**, *27*, 593–607. [[CrossRef](#)]
52. Cao, H.; Plichta, M.M.; Schafer, A.; Haddad, L.; Grimm, O.; Schneider, M.; Esslinger, C.; Kirsch, P.; Meyer-Lindenberg, A.; Tost, H. Test-retest reliability of fMRI-based graph theoretical properties during working memory, emotion processing, and resting state. *Neuroimage* **2014**, *84*, 888–900. [[CrossRef](#)] [[PubMed](#)]
53. McGraw, K.O.; Wong, S.P. Forming inferences about some intraclass correlation coefficients. *Psychol. Methods* **1996**, *1*, 30–46. [[CrossRef](#)]
54. Shrout, P.E.; Fleiss, J.L. Intraclass correlations: Uses in assessing rater reliability. *Psychol. Bull.* **1979**, *86*, 420–428. [[CrossRef](#)]
55. Winer, B.J. *Statistical Principles in Experimental Design*, 2nd ed.; McGraw-Hill: New York, NY, USA, 1971.
56. Yan, C.G.; Liu, D.Q.; He, Y.; Zou, Q.H.; Zhu, C.Z.; Zuo, X.N.; Long, X.Y.; Zang, Y.F. Spontaneous Brain Activity in the Default Mode Network Is Sensitive to Different Resting-State Conditions with Limited Cognitive Load. *PLoS ONE* **2009**, *4*, e5743. [[CrossRef](#)]
57. Birn, R.M. The role of physiological noise in resting-state functional connectivity. *Neuroimage* **2012**, *62*, 864–870. [[CrossRef](#)] [[PubMed](#)]
58. Zuo, X.N.; Xing, X.X. Test-retest reliabilities of resting-state FMRI measurements in human brain functional connectomics: A systems neuroscience perspective. *Neurosci. Biobehav. Rev.* **2014**, *45*, 100–118. [[CrossRef](#)] [[PubMed](#)]
59. Shehzad, Z.; Kelly, A.M.; Reiss, P.T.; Gee, D.G.; Gotimer, K.; Uddin, L.Q.; Lee, S.H.; Margulies, D.S.; Roy, A.K.; Biswal, B.B.; et al. The resting brain: Unconstrained yet reliable. *Cereb. Cortex* **2009**, *19*, 2209–2229. [[CrossRef](#)] [[PubMed](#)]
60. Deuker, L.; Bullmore, E.T.; Smith, M.; Christensen, S.; Nathan, P.J.; Rockstroh, B.; Bassett, D.S. Reproducibility of graph metrics of human brain functional networks. *Neuroimage* **2009**, *47*, 1460–1468. [[CrossRef](#)]

61. Matheson, G.J. We need to talk about reliability: Making better use of test-retest studies for study design and interpretation. *PeerJ* **2019**, *7*, e6918. [[CrossRef](#)]
62. Guo, C.C.; Kurth, F.; Zhou, J.; Mayer, E.A.; Eickhoff, S.B.; Kramer, J.H.; Seeley, W.W. One-year test-retest reliability of intrinsic connectivity network fMRI in older adults. *Neuroimage* **2012**, *61*, 1471–1483. [[CrossRef](#)]
63. Liang, X.; Wang, J.; Yan, C.; Shu, N.; Xu, K.; Gong, G.; He, Y. Effects of different correlation metrics and preprocessing factors on small-world brain functional networks: A resting-state functional MRI study. *PLoS ONE* **2012**, *7*, e32766. [[CrossRef](#)] [[PubMed](#)]
64. Liao, X.H.; Xia, M.R.; Xu, T.; Dai, Z.J.; Cao, X.Y.; Niu, H.J.; Zuo, X.N.; Zang, Y.F.; He, Y. Functional brain hubs and their test-retest reliability: A multiband resting-state functional MRI study. *Neuroimage* **2013**, *83*, 969–982. [[CrossRef](#)] [[PubMed](#)]
65. Schwarz, A.J.; McGonigle, J. Negative edges and soft thresholding in complex network analysis of resting state functional connectivity data. *Neuroimage* **2011**, *55*, 1132–1146. [[CrossRef](#)] [[PubMed](#)]
66. Song, J.; Desphande, A.S.; Meier, T.B.; Tudorascu, D.L.; Vergun, S.; Nair, V.A.; Biswal, B.B.; Meyerand, M.E.; Birn, R.M.; Bellec, P.; et al. Age-related differences in test-retest reliability in resting-state brain functional connectivity. *PLoS ONE* **2012**, *7*, e49847. [[CrossRef](#)]
67. Noble, S.; Scheinost, D.; Constable, R.T. A guide to the measurement and interpretation of fMRI test-retest reliability. *Curr. Opin. Behav. Sci.* **2021**, *40*, 27–32. [[CrossRef](#)]
68. Herting, M.M.; Gautam, P.; Chen, Z.; Mezher, A.; Vetter, N.C. Test-retest reliability of longitudinal task-based fMRI: Implications for developmental studies. *Dev. Cogn. Neurosci.* **2018**, *33*, 17–26. [[CrossRef](#)]
69. Jiang, L.; Xu, T.; He, Y.; Hou, X.H.; Wang, J.; Cao, X.Y.; Wei, G.X.; Yang, Z.; He, Y.; Zuo, X.N. Toward neurobiological characterization of functional homogeneity in the human cortex: Regional variation, morphological association and functional covariance network organization. *Brain Struct. Funct.* **2015**, *220*, 2485–2507. [[CrossRef](#)]
70. Raichle, M.E. Behind the scenes of functional brain imaging: A historical and physiological perspective. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 765–772. [[CrossRef](#)]
71. Meindl, T.; Teipel, S.; Elmouden, R.; Mueller, S.; Koch, W.; Dietrich, O.; Coates, U.; Reiser, M.; Glaser, C. Test-retest reproducibility of the default-mode network in healthy individuals. *Hum. Brain Mapp.* **2010**, *31*, 237–246. [[CrossRef](#)]
72. Fazlollahi, A.; Bourgeat, P.; Liang, X.; Meriaudeau, F.; Connelly, A.; Salvado, O.; Calamante, F. Reproducibility of multiphase pseudo-continuous arterial spin labeling and the effect of post-processing analysis methods. *Neuroimage* **2015**, *117*, 191–201. [[CrossRef](#)]
73. Hodkinson, D.J.; Krause, K.; Khawaja, N.; Renton, T.F.; Huggins, J.P.; Vennart, W.; Thacker, M.A.; Mehta, M.A.; Zelaya, F.O.; Williams, S.C.; et al. Quantifying the test-retest reliability of cerebral blood flow measurements in a clinical model of on-going post-surgical pain: A study using pseudo-continuous arterial spin labelling. *Neuroimage Clin.* **2013**, *3*, 301–310. [[CrossRef](#)] [[PubMed](#)]
74. Jahng, G.H.; Song, E.; Zhu, X.P.; Matson, G.B.; Weiner, M.W.; Schuff, N. Human brain: Reliability and reproducibility of pulsed arterial spin-labeling perfusion MR imaging. *Radiology* **2005**, *234*, 909–916. [[CrossRef](#)] [[PubMed](#)]
75. Yang, F.N.; Xu, S.; Spaeth, A.; Galli, O.; Zhao, K.; Fang, Z.; Basner, M.; Dinges, D.F.; Detre, J.A.; Rao, H. Test-retest reliability of cerebral blood flow for assessing brain function at rest and during a vigilance task. *Neuroimage* **2019**, *193*, 157–166. [[CrossRef](#)] [[PubMed](#)]
76. Telesford, Q.K.; Morgan, A.R.; Hayasaka, S.; Simpson, S.L.; Barret, W.; Kraft, R.A.; Mozolic, J.L.; Laurienti, P.J. Reproducibility of graph metrics in FMRI networks. *Front. Neuroinform.* **2010**, *4*, 117. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.