Variability and Reproducibility of Directed and Undirected Functional MRI Connectomes in the Human Brain

Allegra Conti [1,*] , Andrea Duggento [2] , Maria Guerrisi [2], Luca Passamonti [3,4,*], Iole Indovina [1,5] and Nicola Toschi [2,6,*]

1   Laboratory of Neuromotor Physiology, IRCCS Santa Lucia Foundation, 00179 Rome, Italy
2   Department of Biomedicine and Prevention, University of Rome Tor Vergata, 00133 Rome, Italy
3   Department of Clinical Neurosciences, University of Cambridge, Cambridge CB2 0QQ, UK
4   Institute of Bioimaging and Molecular Physiology, National Research Council, 20090 Milano, Italy
5   Saint Camillus International University of Health and Medical Sciences, 00131 Rome, Italy
6   Department of Radiology, Athinoula A. Martinos Center for Biomedical Imaging, Boston, MA 02129, USA
*   Correspondence: a.conti@hsantalucia.it (A.C.); lp337@medschl.cam.ac.uk (L.P);
     toschi@med.uniroma2.it (N.T.)

check for updates

Abstract: A growing number of studies are focusing on methods to estimate and analyze the functional connectome of the human brain. Graph theoretical measures are commonly employed to interpret and synthesize complex network-related information. While resting state functional MRI (rsfMRI) is often employed in this context, it is known to exhibit poor reproducibility, a key factor which is commonly neglected in typical cohort studies using connectomics-related measures as biomarkers. We aimed to fill this gap by analyzing and comparing the inter- and intra-subject variability of connectivity matrices, as well as graph-theoretical measures, in a large (n = 1003) database of young healthy subjects which underwent four consecutive rsfMRI sessions. We analyzed both directed (Granger Causality and Transfer Entropy) and undirected (Pearson Correlation and Partial Correlation) time-series association measures and related global and local graph-theoretical measures. While matrix weights exhibit a higher reproducibility in undirected, as opposed to directed, methods, this difference disappears when looking at global graph metrics and, in turn, exhibits strong regional dependence in local graphs metrics. Our results warrant caution in the interpretation of connectivity studies, and serve as a benchmark for future investigations by providing quantitative estimates for the inter- and intra-subject variabilities in both directed and undirected connectomic measures.

Keywords: functional networks; functional magnetic resonance imaging; connectome; connectivity matrices; graphs; reproducibility; granger causality; transfer entropy

1. Introduction

The interest in studying directed and undirected interactions between different regions in the human brain (i.e., the functional 'connectome') is growing exponentially [1–7], and the advent of graph-theoretical applications to neuroscience has provided additional avenues to represent, analyze and interpret information contained in complex, possibly dynamic networks like the human connectome [8–12]. Functional connectome estimates are typically derived from neuromonitoring data (e.g., resting state functional MRI—rsfMRI), and established methods for computing whole-brain functional connectivity matrices include both undirected and directed estimators [13–15]. As is well known, the reproducibility of MRI data varies widely across modalities, and rsfMRI data are known to exhibit significant inter- and intra-subject fluctuations [16–18]. While this can significantly bias and

hamper the interpretation of functional connectivity studies, which do not typically include targeted scan-rescan experiments to assess reproducibility, the impact of this variability on connectome matrices and related measures has not yet been systematically investigated.

In this study, we explore, quantify and compare the intra- and inter-subject variability (and hence reproducibility) of both directed and undirected resting state functional connectivity measures in a large (1003 subjects) database of high-quality rsfMRI data consisting of 4 scan sessions per subject. In detail, we investigate the reproducibility of whole-brain adjacency matrices derived from Pearson Correlation (PearC), Partial Correlation (PartC) as well as multivariate Granger Causality (mGC) and multivariate Transfer Entropy (mTE). In addition, we investigate the reproducibility of both global (whole-brain) and local (node-wise) graphs metrics calculated for all four types of adjacency matrices.

While PearC and PartC estimate zero-lag, non-directional associations between time-series, mGC and mTE inherently measure directed information transfer. Such directed measures have already been successfully employed in brain connectivity studies [5,19–23] and, while it is known that mTE and mGC are equivalent in the limit of Gaussian data [24], these two estimators may differ when applied when the data distribution is skewed and/or the number of data points is extremely limited. mGC is based on choice of model parameters [25], while TE is model-free and possibly a superior tool for estimating nonlinear interactions [15] like the ones commonly found in biological signals.

2. Materials and Methods

2.1. rsfMRI Data

We employed rsfMRI data from 1003 subjects, part of the Human Connectome Project (HCP) (S1200 PTN release [26]). Each subject underwent a total of 4 resting state scans (2 sessions on 2 different days, 2 scans per session, 1200 timepoints/scan, TR 720 ms, TE 33.1 ms, flip angle 52, FOV 208 × 180, thickness 2.0 mm; 72 slices; 2.0 mm isotropic voxel size, multiband factor 8, echo spacing 0.58 ms, BW 2290 Hz/Px, 14 minutes of acquisition time/scan). Subjects were scanned on a customized Siemens Skyra 3 T scanner [26] (Smith et al. 2013) with higher spatial and temporal resolutions as compared to common fMRI acquisition protocols. Data pre-processing was state-of-the-art and included gradient distortion and motion correction, field map pre-processing using spin echo field map (specific for each scanning day), intensity normalization and bias field removal. Residual artifacts were removed through an automatic classifier specifically trained to this particular dataset, which is able to remove measurement noise, additional motion or physiological artifacts with extremely high accuracy. Further details regarding data processing can be found in the HCP S1200 Release reference manual [https://www.humanconnectome.org/storage/app/media/documentation/s1200/HCP_S1200_Release_Reference_Manual.pdf].

In this paper, we processed subject- and scan-wise fMRI timeseries (1200 points each) resulting from group independent component analysis (gICA) at dimensionality 15 (made available by the HCP consortium). Exemplary slices of the resulting node (i.e., component, also called "subnetworks" in this paper) maps, along with their physiological interpretations [27], are shown in Figure 1. To cater for confounding elements introduced by locally varying hemodynamic response functions (HRFs) [28], fMRI timeseries were preprocessed using a the blind deconvolution approach [29] (maximum lag 10 s = 14 time points, threshold: 1 standard deviation) implemented in a publicly available toolbox (https://users.ugent.be/~dmarinaz/HRF_deconvolution.html). Stationarity of all signals was verified using the Augmented Dickey-Fuller test. All signals were successively standardized prior to additional processing.
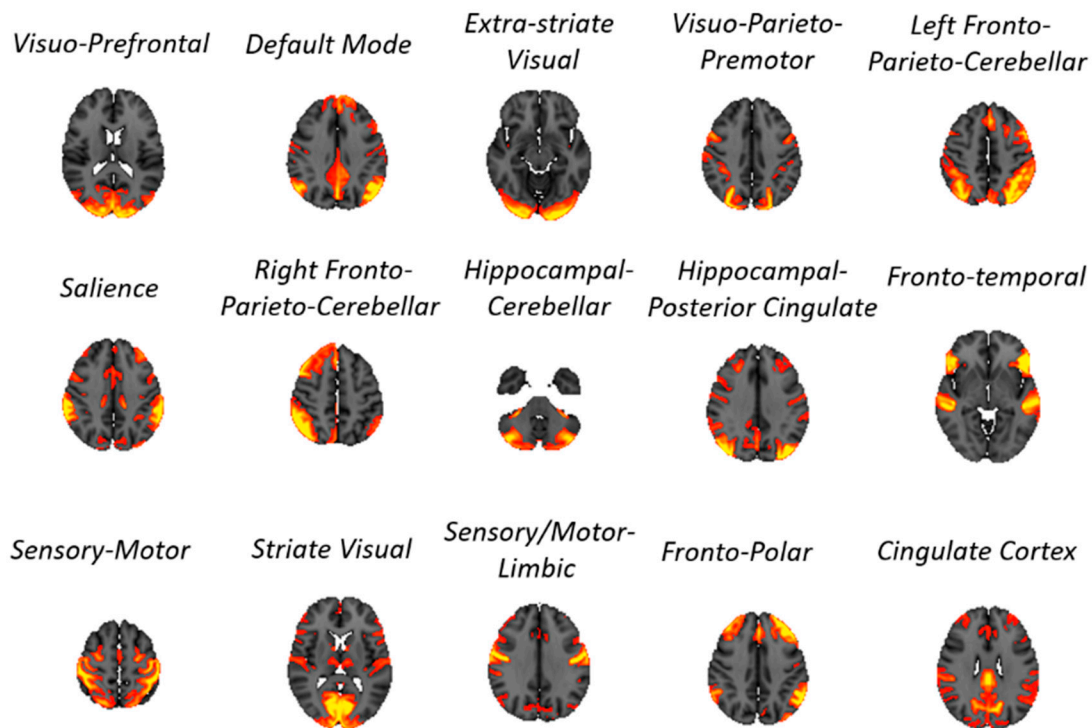
Figure 1. Brain functional subnetworks (i.e., independent components) identified by group independent component analysis (gICA) in 1003 subjects drawn from the Human Connectome Project database. The subject-wise 15-node specific time-series relative to these components were employed for all analyses in this paper.

2.2. Estimation of Adjacency Matrices

Starting from the 15 subject-, scan- and node-specific timeseries (see Figure 2A for an example), subject-wise undirected adjacency matrices were obtained through both PearC and PartC between all pairs of timeseries through in-house code written in in MATLAB (v. 2018a). Negative correlations were set to zero. Directed adjacency matrices where obtained through mGC estimates, in its most recent state-space formulation [30,31], and through mTE.

mGC and mTE were calculated through in-house modified versions of the Matlab Tools for the computation of multiscale Granger Causality (http://www.lucafaes.net/msGC.html) [31] and of multivariate Transfer Entropy (http://www.lucafaes.net/cTE.html) [32]. In the case of GC, the autoregressive order was chosen my minimizing the median Schwartz criterion across all subjects (order = 5). In mTE, conditioning vectors are formed according to a non-uniform building scheme selecting past terms up to a maximum lag of 5 time points while minimizing the conditional entropy [32]. Examples of adjacency matrices (which are symmetrical for undirected estimators, and asymmetrical for directed ones) obtained through all four methods are shown in Figure 2B.
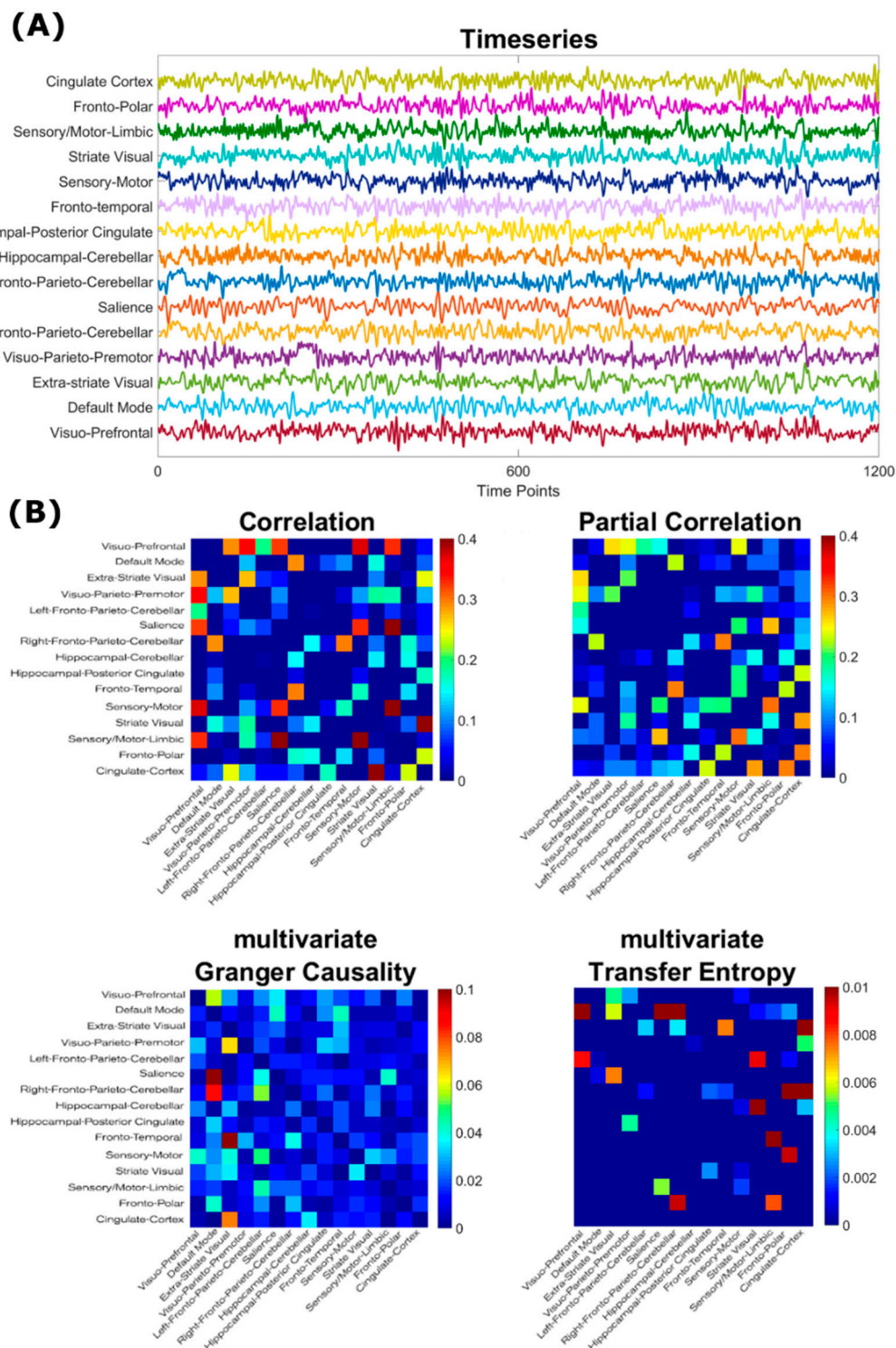
Figure 2. (**A**) Example node-wise timeseries from a single scan, used to compute subject-, session- and scan-specific adjacency matrices; (**B**) Examples of adjacency matrices obtained from both undirected (first row) and directed (second row) connectivity estimators for one subject. For each method, the median across 4 rsfMRI sessions from a single subject is shown. Diagonal elements are set to zero.

2.3. Global and Local Graph Metric Estimation

Starting from the adjacency matrices, we computed both global and local graph-theoretical indices of functional connectivity for all three methods/estimators. All graph measures were computed via the Brain Connectivity Toolbox [33] (https://sites.google.com/site/bctnet/), by using functions available for

weighted adjacency matrices for both directed (in case of mGC and mTE) and undirected (for pearC and partC) graph metrics.

For each subject and for each scan, we calculated graph metrics quantifying the centrality of a node within a network (local strength and betweenness centrality), its ability to transmit information at local level (local efficiency) and its integration properties (clustering coefficient) [33]. Also, since global graph metrics characterize the overall organization of a network, we computed the global strength, the global efficiency and the global clustering coefficient as the average of the respective local metrics of all nodes. In addition, we calculated graph transitivity (a property related to the existence of tightly connected communities of nodes). Figure 3 exemplifies the meaning and calculation of these graph-theoretical metrics.
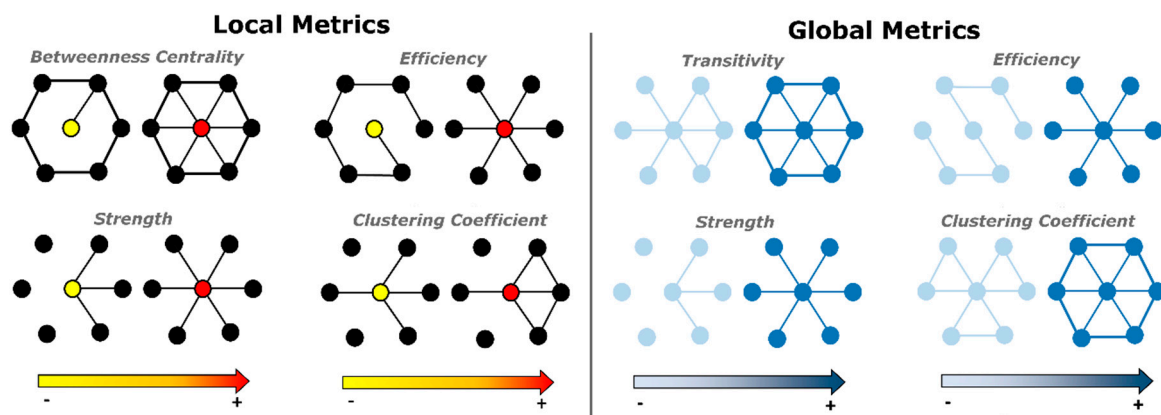

Figure 3. Exemplification of local (**left**) and global (**right**) graph metrics as a function of node neighborhood connectivity. Local metrics (**left**): taking the yellow/red center node as a reference, its betweenness centrality and local efficiency increase as the average shortest path between the center node and all its neighbors decreases. Strength/degree is proportional to the number of connections, the clustering coefficient increases with the ratio between the number of links actually present between the vertices within node neighborhood and the number of links that could possibly exist between them. Global metrics (**right**): transitivity and clustering coefficient both reflect prevalence of clustered connectivity around nodes, while efficiency and strength increase with the increase of the local efficiency and strength of each node in the graph [33].

2.4. Inter- and Intra-Subject Variability Distributions

Adjacency matrix weights (both directed and undirected) and both local and global graph-theoretical measures, have their own native scales and dimensionalities and, in the case of local graph metrics, this scale also depends on the specific brain region. The same issue, which results in dimensional and extremely different effect sizes across estimators, also affects graph-theoretical metrics calculated from different estimators. Therefore, in order to compare inter- and intra-subject fluctuations in adjacency matrices, as well as graph-theoretical metrics, we employed dimensionless, normalized pairwise quantifiers of asymmetry, along with their estimated distributions.

In detail, we defined the pairwise normalized difference/asymmetry (ND) as:

$$\text{ND} = \frac{a - b}{a + b} \tag{1}$$

where a and b refer to two distinct scans (either intra- or inter-subject). ND is both normalized (i.e., a-dimensional) and bounded (i.e., between -1 and 1).

In the case of scalar quantities like global or local graph-theoretical metrics, Equation (1) can be applied directly. When comparing adjacency matrices, Equation (1) was modified by using Frobenius distances between (in this case) matrices *a* and *b* as follows:

$$\text{ND} = \sqrt{\frac{trace((a-b)\cdot(a-b)')}{trace((a+b)\cdot(a+b)')}} \tag{2}$$

To empirically estimate (and successively compare) the intra- and inter-subject distributions in ND, we proceeded as follows. From each set of 4 subject-wise scans, we sampled all 6 possible pairs of scans, computed the ND for each pair, and repeated this step for all 1003 subjects. For each method, this resulted in $1003 \times 6 = 6018$ distinct ND values referring to intra-subject variability. Subsequently, in order to build a comparable distribution related to inter-subject variability, we randomly and repeatedly sampled (without replacement) four scans from four distinct subjects, hence constructing set of 4 "inter-subject" scans from which 6 distinct pairs could be constructed and 6 distinct ND values could be calculated as above. To control for possible differences arising from intra-subject habituation, the random sampling was performed so that each "inter-subject" set of four scans contained exactly 1 sample from the 1003-sized set of first, second, third and fourth scans. This procedure was repeated 1003 times, resulting in $1003 \times 6 = 6018$ distinct ND values referring to inter-subject variability. Figure 4 summarizes the sampling process.
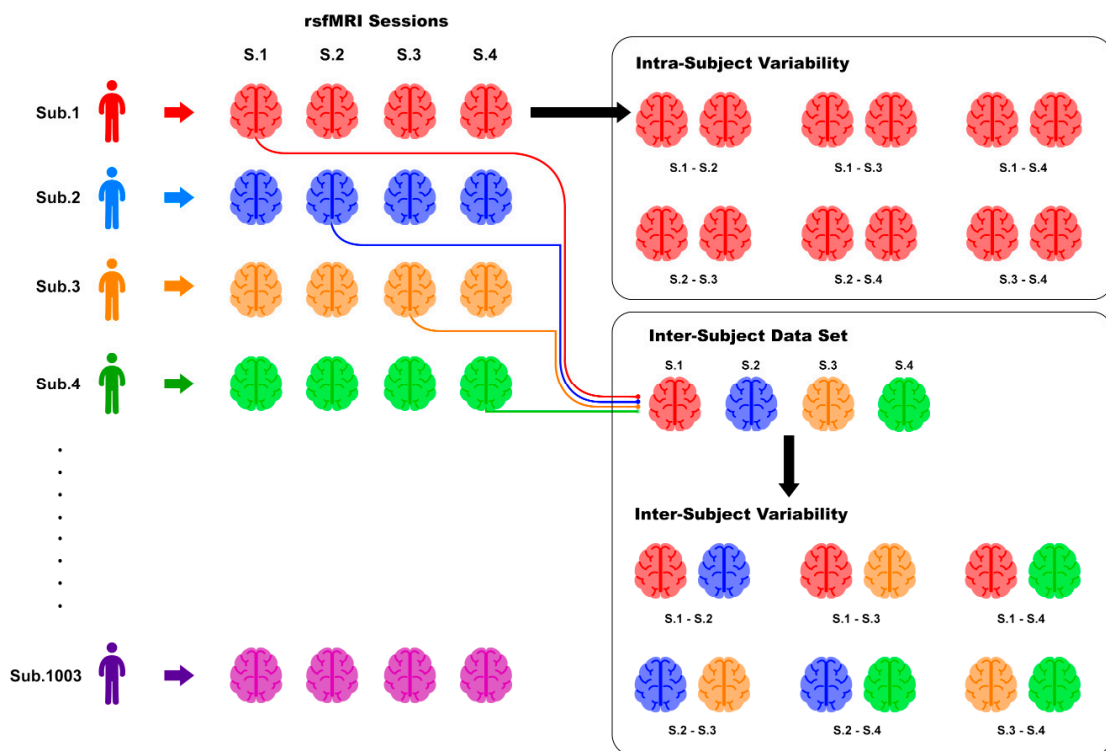

Figure 4. Sampling strategy used to construct intra- and inter-subject pairs and subsequent distributions of ND (Equations (1) and (2)). The constructed samples for the basis for computing the distributions are shown, e.g., in Figures 5 and 6.

2.5. Statistical Analysis

After sampling and computation of all ND values for each graph-theoretical metric and for all four types of adjacency matrix, the resulting "inter" and "intra" distribution medians of ND were compared through nonparametric Mann-Whitney-U tests. Effect Size (ES) were evaluated as the difference between the medians of the "inter" and "intra" distributions in ND (i.e., ES = "inter" − "intra").

The result of each comparison is therefore a *p*-value with an associated ES. Additionally, in order to explore possible differences in variability across nodes (i.e., brain regions), we pooled ND values for all local graph-theoretical metrics across all four methods (4 metrics × four methods = 16 values per node), and tested the effect of anatomical localization using a nonparametric Kruskal-Wallis test. Whenever a significant effect was found, pairwise comparison between nodes was performed in order to further examine which pairwise differences were the main drivers of the overall effect. This procedure was repeated separately for inter- and intra-subject variability.

3. Results

Figure 5 shows the "inter" and "intra" distribution for ND obtained when comparing pairs of adjacency matrices (Equation (2)) for all four methods (PearC, PartC, mGC and mTE), along with effect sizes. For all methods, as hypothesized, median intra-subject variability is significantly lower than median inter subject variability ($p < 10^{-20}$ for all methods). The largest ES was associated with mGC and PearC (ES = 0.03 for both), followed by PartC and mTE (ES = 0.02 for both). Qualitatively, mTE showed the largest median variabilities (both intra- and inter-subject), whereas PearC and PartC showed similar median variabilities.
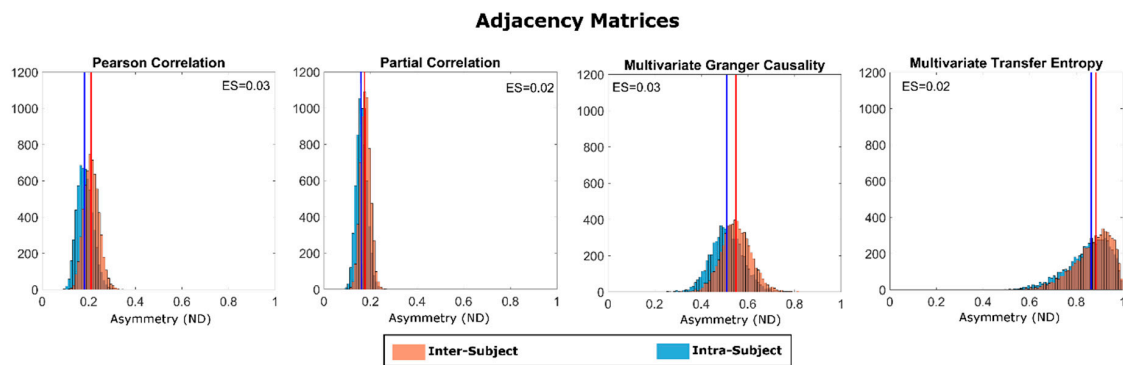

Figure 5. Intra- (blue) and inter- (orange) subject distributions of the Normalized Differences (ND), for all four connectivity estimation methods. Effect size (ES) (i.e., the difference between the median values of the inter- and intra-subject ND distributions) are shown in the insets. Mann-Whitney U-tests for comparing medians returned $p < 10^{-20}$ for all four methods.

Figure 6 shows the "inter" and "intra" distribution for ND (Equation (1)) obtained when comparing pairs of global graph-theoretical metrics, along with the results of statistical testing in terms of *p*-values and ES. Interestingly, intra- and inter-subject fluctuations are distinguishable only in some cases. In particular, we found statistically significant differences ($p < 0.05$) for all metrics estimated through PearC (named "Bivariate Correlation" in figure) and PartC. However, in these cases, intra-subject fluctuations were larger than inter-subject fluctuations (i.e., negative ES). On the other hand, when looking global metrics calculated through mGC, significant differences between intra- vs. inter-subject fluctuations were mainly associated with positive ES. No differences were found between ND distributions for global-graphs metrics evaluated through mTE. Notably, in this analysis, all significant differences were associated with minimal (compared to unity) effect sizes (of the order of between 10^{-2} and 10^{-3}). In addition, the widths of intra- and inter-subject variability distributions were qualitatively similar for all methods. The main analyses presented in Figures 5 and 6 were also repeated using a finer parcellation (50 subnetworks—see supplementary materials (See Figures A1 and A2)). The results were fully compatible (both in effect sizes and directions) with the main results obtained using 15 subnetworks.
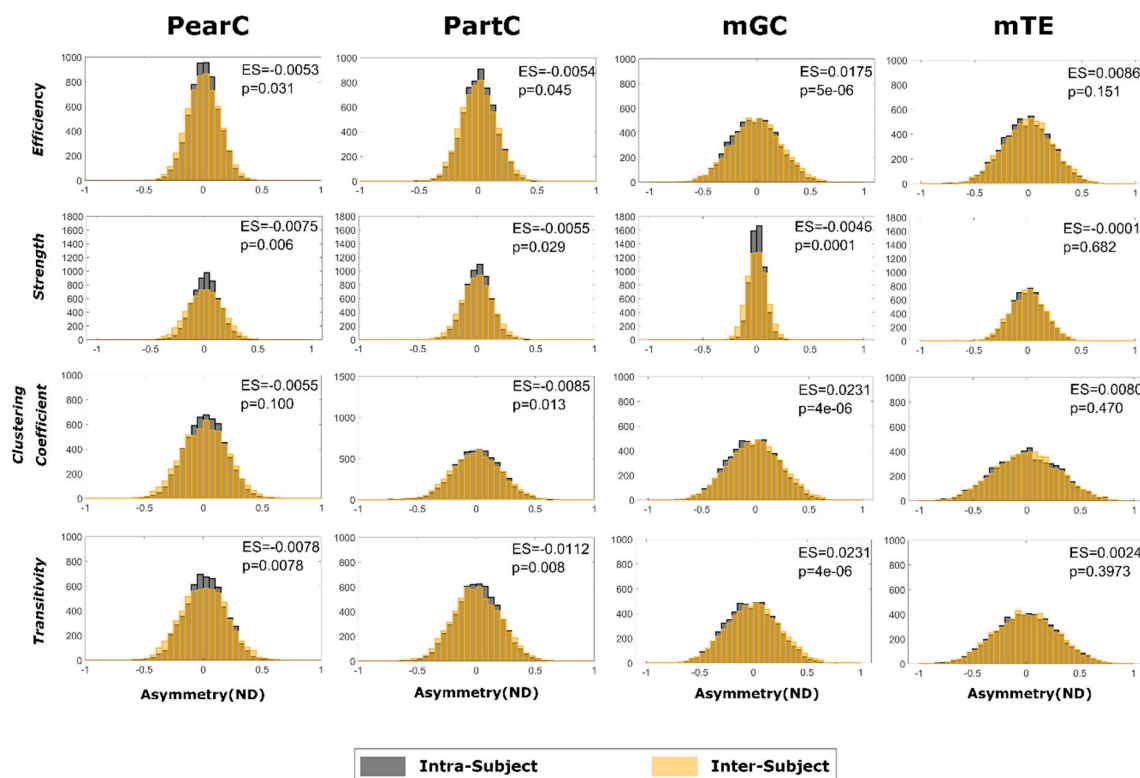
Figure 6. Intra- (gray) and inter- (orange) subject distributions of the Normalized Differences (ND), for all four connectivity estimation methods, along with the p-value resulting from the corresponding Mann-Whitney-U test. Effect size (ES) (i.e., the difference between the median values of the inter- and intra-subject ND distributions) are shown in the insets.

Figure 7 summarizes the results of comparing ND distributions computed from local graph-theoretical metrics, along with the results of statistical testing as well as ES. In this case, an intricate, anatomically dependent pattern emerged, where the statistical difference between median intra- and inter-subject fluctuations depends on the specific node, with some nodes exhibiting higher intra-subject (as compared to inter-subject) fluctuations. Asterisks indicate nodes in which differences between inter- and intra-subject variabilities resulted in be different from 0 ($p < 0.05$, Mann-Whitney U test). These results highlight the strong spatial dependence across the brain in reproducibility of all local graph-theoretical indices. Only the betweenness centrality of the Hippocampal-Posterior Cingulate subnetwork show statistically higher intra-subject (as opposed to inter-subject) variability across all methods. Still, ES in local efficiency and strength estimated in the same subnetwork tend to the same direction in all methods. Higher intra-subject (as opposed to inter-subject) variability of the betweenness centrality and local efficiency and strength was also found for the salience and the sensory motor subnetworks. On the other hand, the default mode subnetwork displayed higher intra-subject (as opposed to inter-subject) stability in both betweenness centrality and local clustering coefficient across methods. The same hold for the betweenness centrality of the fronto-temporal subnetwork. The results shown in Figure 7 allow also to qualitatively compare the variability of the local metrics obtained though undirected (PearC and PartC) and directed (mGC and mTE) methods. In particular, local efficiency, strength and clustering coefficient in the visuo-prefrontal subnetwork are highly reproducible within the same subject when estimated using mGC and mTE.
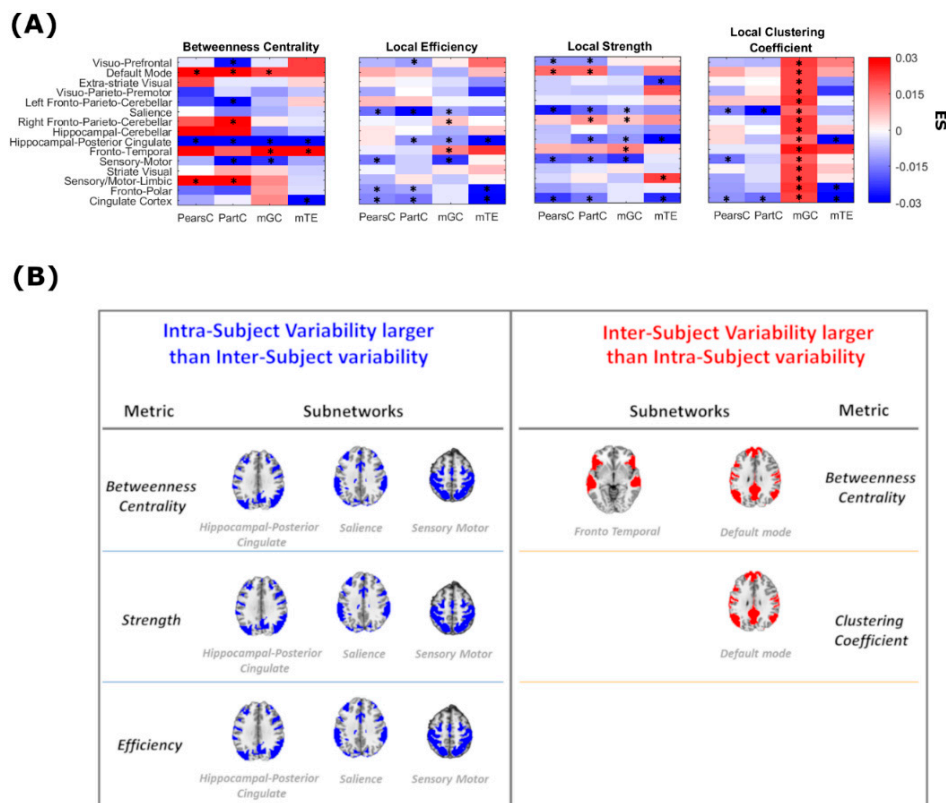
Figure 7. (**A**) Effect size, i.e., differences between median intra- and inter-subject distributions in local graph metrics for each node and for each connectivity estimation method (PearC, PartC, mGC and mTE). For each metric, darker red and blue cells correspond to higher and lower ES values, respectively. The nodes in which differences between inter- and intra-subject variabilities are statistically different from 0 ($p < 0.05$) are indicated with an asterisk; (**B**) Subnetworks (i.e., components) with larger intra-subject than inter-subject variability in various metrics are shown in blue, while subnetworks showing a larger intra-subject reproducibility are shown in red.

With reference to Figure 7, Table 1 summarizes the number of nodes where statistically significant intra- vs. inter-subject differences were found. Overall, across all four methods (PearC, PartC, mGC and mTE) the percentage of nodes (out of 15) where inter- and intra-subject variability were statistically different varied between 28% and 38% (bottom row of Table 1), and in 3–35% of the 15 nodes, the inter-subject variability is significantly higher than the intra-subject variability. mGC is the method with the lowest average number of nodes (across metrics) where the above-mentioned difference is significantly different from zero. Also, node strength is the metric where the above-mentioned difference is significantly different from zero in the highest average number of nodes (across methods). Importantly, in the absence of a ground-truth, it is not possible to determine which source of variability dominates the overall effects. This could only be ascertained with synthetic data which accurately models neuronal spiking and cardiovascular coupling [9].

Figure 8 shows the results the Kruskal-Wallis test performed separately on the "intra" and on the "inter" ND values. In both intra- and inter-subject variability we found a significant effect of node ($p = 1.1 \times 10^{-4}$). In post-hoc comparisons (Figure 8B) we found that overall intra-subject variability in local graph-theoretical metrics of the fronto-temporal subnetwork is lower as compared to the hippocampal-posterior cingulate, the cingulate cortex and the fronto-polar subnetworks. Also, overall inter-subject variability in local graph-theoretical metric is lower in the salience subnetwork as compared to the hippocampal-cerebellar, fronto-temporal and sensory/motor-limbic subnetworks.

Table 1. Number of nodes showing statistically different intra- and inter-subject variabilities. In brackets: (number of nodes showing higher inter- vs. intra-subject variability/number of nodes showing higher intra- vs. inter-subject variability). The right column shows the average percentage of nodes (standard deviation in brackets), across metrics, where the inter-subject variability is significantly higher than the intra-subject one. The bottom row shows the percentage count of the first number in each cell.

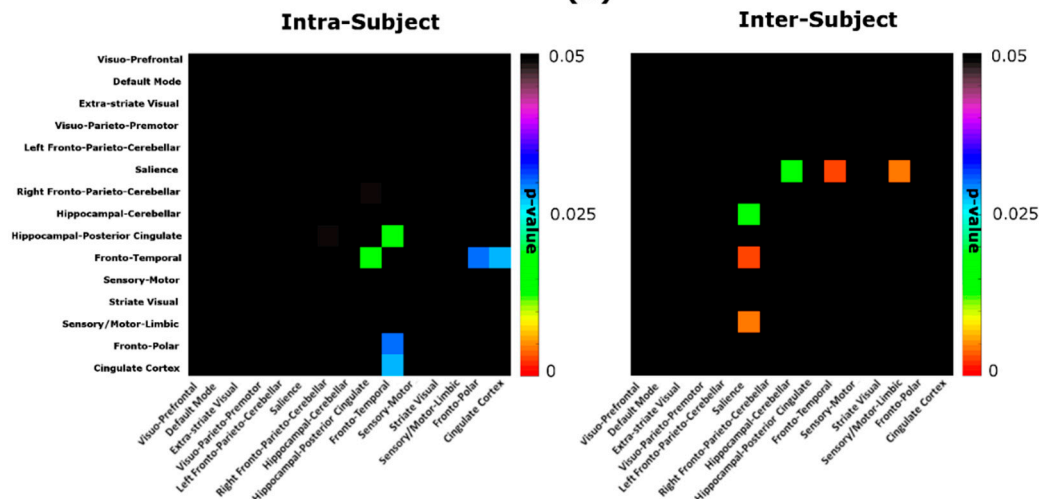| Method	Betweenness Centrality	Local Efficiency	Strength	Clustering Coefficient	*Mean(sd) % out of 15*
PearsC	3(2/1)	4(0/4)	5(1/4)	3(0/3)	*5%(6%)*
PartC	7(3/4)	5(0/5)	7(2/5)	2(0/2)	*8%(9%)*
mGC	4(2/2)	5(2/3)	5(2/3)	15(15/0)	*35%(38%)*
mTE	3(1/2)	3(0/3)	4(1/3)	3(0/3)	*3%(3%)*
Mean(sd) % out of 15	*28%(13%)*	*28%(6%)*	*35%(8%)*	*38%(41%)*	

(A)
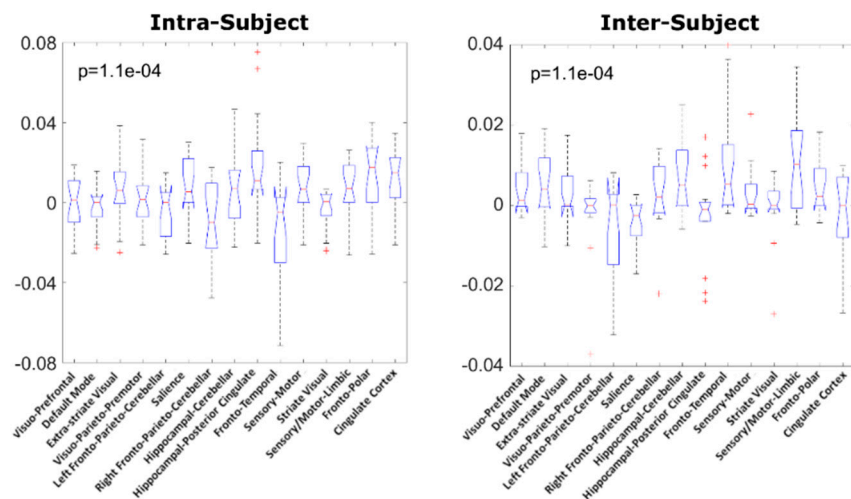
(B)

Figure 8. (**A**) Intra and inter ND values (across 1003 subjects; left: intra-subject variability, right: inter-subject variability) for all local metrics (**B**) *p*-values obtained from pairwise comparison of node-specific, intra- and inter-subject variability.

4. Discussion

In this paper, we have studied the inter- and intra-subject variability of directed and undirected connectivity matrices as well as of global as well as local graph-theoretical metrics derived from resting state functional MRI data in a large sample (n = 1003) of resting state fMRI data, which also has the unique property of including scan-rescan sessions. While previous reproducibility studies exist, they are commonly focused on specific data processing [34] or statistical [35] aspects, and are usually based on limited sample sizes.

We employed a fully data-driven parcellation, derived from group independent component analysis performed by the HCP consortium and which is hence independent from any anatomical priors commonly present in classical parcellations. To further eliminate any unwanted sources of variability and improve the comparability between distributions, we adopted a bootstrap approach followed by nonparametric statistics on a normalized difference metric, which also allowed us to pool across methods and/or nodes and make qualitative inter-method comparisons. In addition, in order to investigate the dependence of our results on ICA dimensionality (and hence coarseness/fineness of the resulting parcellation), we repeated the main analyses using data based on a dimensionality of 50 instead of 15. Interestingly, all effect direction and sizes are highly comparable to the main results ad dimensionality 15.

Our results show that while adjacency matrix weights generated using conventional methods like PearC or PartC exhibit 1) a qualitatively lower variability as compared to directed metrics and 2) a lower intra-subject vs. inter-subject variability; these differences virtually disappear when looking at global and local graphs-metrics values, where re-scanning the same subject results in the same statistical fluctuation as scanning a different subject. The overall higher variability in adjacency matrices obtained from directed metrics may be partially due to the higher complexity of both estimators and the multiple degrees of freedom (e.g., model order, embedding dimensions) in their computation, whose exploration was beyond the scope of this study. Also, the appropriateness of using of mGC in fMRI has been the object of discussion [36–42], mainly because of the confounding elements introduced by locally varying neurovascular coupling and the interaction of these confounds with, e.g., parameter choices, and hence the overall validity of the estimators. To this end, we applied a blind deconvolution approach that estimates a local HRF for each signal and returns an estimate of neural activity. Still, the application of this type of pipeline is not yet widespread in fMRI studies using causality methods, and that it also involves parameter choices whose effect on reproducibility merit investigation in a separate paper. Similarly, while partial and Pearson correlation do not involve parameter optimization, both mGC and mTE estimations are based on parameter choices inherent to each method. In this sense, the main aim of our work was to not to directly compare across methods, but rather to provide an estimation, in a high-quality and unique data sample, of overall the intra- and inter-subject reproducibility of a variety of connectomic metrics which are commonly used in the fMRI community.

Interestingly, we found an articulate pattern of "intra" vs. "inter" variability differences as a function of metrics, as well as node (i.e., anatomical localization), which points towards a certain degree of unpredictability in the reproducibility of results obtained using each possible composite pipeline. Interestingly, in several nodes and metrics, median intra-subject variability was found to be larger than median inter-subject variability. In this context, previous studies have highlighted the higher intra-subject (as compared to inter-subject) variability of functional connectivity in specific regions of the brain [43–46]. Indeed, intra-subject variability depends on several factors such as diurnal rhythms [47,48], cognitive/behavioral contexts [49–51], caffeine intake [43], sleepiness [52–54] and attention [46,55], which may explain the high inter-subject fluctuations we found in this paper.

Limitations and Future Perspectives

We employed public HCP time-series data stemming from a group ICA at dimensionality 15. Several previous investigations have extracted useful information employing "small" networks and graph metrics [10,20,56,57]. Also, additional analyses performed at higher dimensionality confirm

the main results of this paper. However, we cannot exclude that the fluctuations quantified in this study may change as a function of parcellation, and should therefore be investigated ad hoc on the basis of data specifications and parcellation details in any specific study where a reproducibility assessment is warranted, e.g., in longitudinal investigations. Also, in view of the advent of novel connectivity estimators, like, e.g., Generalized Recurrent Neural Network-based Dynamic Causal Modeling [58] or model-based whole-brain effective connectivity (MOU-EC) [59], future investigations are needed to characterize the repeatability and robustness of the connectivity matrices generated by such strategies. In addition, since dynamical connectivity measures and connectivity density are also likely to be impacted by reproducibility issues [60], a study of the robustness of dynamic functional connectomes, based on nonparametric distribution-building approaches like the one we employed in this paper, appears warranted. Finally, future investigation of the reproducibility of the regional dependence of functional connectivity by using large-scale brain network models would also be useful for better understanding the causes of intra- and inter-subject fluctuations in conventional connectivity estimators we observed in this study.

5. Conclusions

In this paper, we have studied inter- and intra-subject variability of directed and undirected connectivity matrices, as well as of global and local graph-theoretical metrics derived from resting state functional MRI data in a large sample (n = 1003) of resting state fMRI data released by the HCP. We analyzed both directed (Granger Causality and Transfer Entropy) and undirected (Pearson Correlation and Partial Correlation) association measures and derived graph-theoretical measures.

Our results show that adjacency matrices exhibit a higher intra-subject (as compared to inter-subject) reproducibility. However, intra- and inter-subject dispersion of both global and local graphs-metrics resulted as being comparable. We therefore demonstrate that, even with large volumes of densely sampled rsfMRI data, reproducibility is an issue that should be constantly kept in mind and that connectomics studies should, whenever possible, include a scan-rescan validation arm, at least in a subset of subjects. Our results also serve as benchmarks for future investigations aiming to estimate the variability in other directed and undirected connectomic databases, and can also be employed for prospective power calculations in planning functional connectomics experiments.

Appendix A Variability and Reproducibility of Directed and Undirected Functional MRI Connectomes in the Human Brain
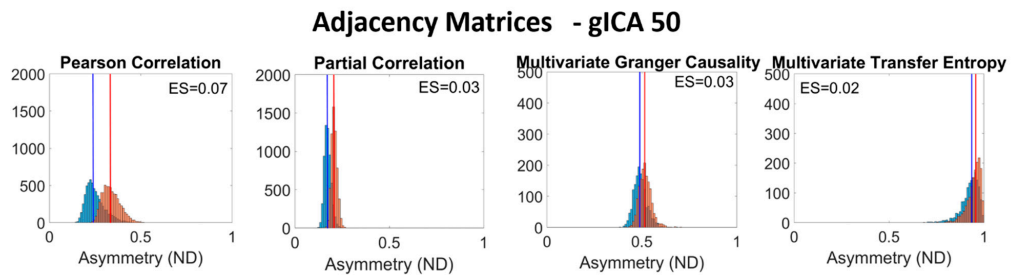

Figure A1. Intra- (blue) and inter- (orange) subject distributions of the Normalized Differences (ND), for adjacency matrices obtained from the ICA decomposition with dimensionality of 50, for all four connectivity estimation methods. Effect size (ES) (i.e., the difference between the median values of the inter- and intra- subject ND distributions) are shown in the insets. Mann-Whitney U-tests for comparing medians returned $p < 10^{-20}$ for all four methods.

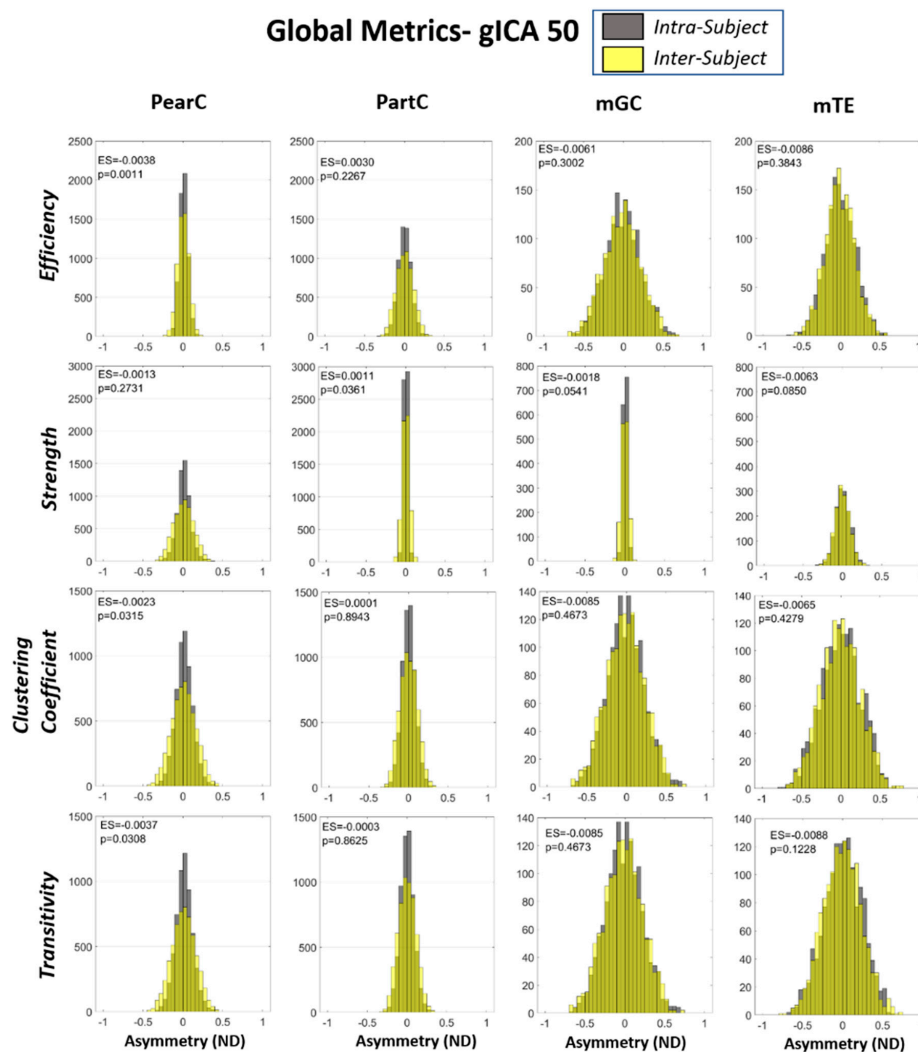**Figure A2.** Intra- (black) and inter- (yellow) subject distributions of the Normalized Differences (ND), for all four connectivity estimation methods, using ICA-derived time-series at dimensionality 50. Effect size (ES) (i.e., the difference between the median values of the inter- and intra- subject ND distributions) together with the *p*-values resulting from the Mann-Whitney-U tests, are shown in the insets.

References

1. Wang, H.E.; Bénar, C.G.; Quilichini, P.P.; Friston, K.J.; Jirsa, V.K.; Bernard, C. A systematic framework for functional connectivity measures. *Front. Mol. Neurosci.* **2014**, *8*, 8. [CrossRef] [PubMed]

2. Blinowska, K.J. Review of the methods of determination of directed connectivity from multichannel data. *Med. Boil. Eng.* **2011**, *49*, 521–529. [CrossRef] [PubMed]

3. Bullmore, E.; Sporns, O.; Bullmore, E. The economy of brain network organization. *Nat. Rev. Neurosci.* **2012**, *13*, 336–349. [CrossRef] [PubMed]

4. Fornito, A.; Zalesky, A.; Bullmore, E.T. (Eds.) Chapter 3—Connectivity Matrices and Brain Graphs. In *Fundamentals of Brain Network Analysis*; Academic Press: San Diego, CA, USA, 2016; pp. 89–113.

5. Schmidt, C.; Pester, B.; Schmid-Hertel, N.; Witte, H.; Wismüller, A.; Leistritz, L. A Multivariate Granger Causality Concept towards Full Brain Functional Connectivity. *PLoS ONE* **2016**, *11*, 0153105. [CrossRef] [PubMed]

6. Schoffelen, J.-M.; Hultén, A.; Lam, N.; Marquand, A.F.; Uddén, J.; Hagoort, P. Frequency-specific directed interactions in the human brain network for language. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 8083–8088. [CrossRef] [PubMed]

7. Salvador, R.; Suckling, J.; Schwarzbauer, C.; Bullmore, E.; Bullmore, E. Undirected graphs of frequency-dependent functional connectivity in whole brain networks. *Philos. Trans. R. Soc. B Boil. Sci.* **2005**, *360*, 937–946. [CrossRef]

8. Sporns, O. Graph theory methods: Applications in brain networks. *Dialogues Clin. Neurosci.* **2018**, *20*, 111–121.

9. Duggento, A.; Passamonti, L.; Valenza, G.; Barbieri, R.; Guerrisi, M.; Toschi, N. Multivariate Granger causality unveils directed parietal to prefrontal cortex connectivity during task-free MRI. *Sci. Rep.* **2018**, *8*, 5571. [CrossRef]

10. Toschi, N.; Riccelli, R.; Indovina, I.; Terracciano, A.; Passamonti, L. Functional connectome of the five-factor model of personality. *Pers. Neurosci.* **2018**, *1*, 1. [CrossRef]

11. Vecchio, F.; Miraglia, F.; Rossini, P.M. Connectome: Graph theory application in functional brain network architecture. *Clin. Neurophysiol. Pract.* **2017**, *2*, 206–213. [CrossRef]

12. Fornito, A.; Zalesky, A.; Bullmore, E.T. (Eds.) Chapter 1—An Introduction to Brain Networks. In *Fundamentals of Brain Network Analysis*; Academic Press: San Diego, CA, USA, 2016; pp. 1–35.

13. Deshpande, G.; LaConte, S.; James, G.A.; Peltier, S.; Hu, X. Multivariate Granger causality analysis of fMRI data. *Hum. Brain Mapp.* **2009**, *30*, 1361–1373. [CrossRef] [PubMed]

14. Baccala, L.A.; Sameshima, K. Partial directed coherence: A new concept in neural structure determination. *Boil. Cybern.* **2001**, *84*, 463–474. [CrossRef] [PubMed]

15. Vicente, R.; Wibral, M.; Lindner, M.; Pipa, G. Transfer entropy—A model-free measure of effective connectivity for the neurosciences. *J. Comput. Neurosci.* **2011**, *30*, 45–67. [CrossRef] [PubMed]

16. Zhang, N.; Wang, F.; Turner, G.H.; Gore, J.C.; Avison, M.J.; Chen, L.M. Intra- and inter-subject variability of high field fMRI digit maps in somatosensory area 3b of new world monkeys. *Neuroscience* **2010**, *165*, 252–264. [CrossRef] [PubMed]

17. Smith, S.M.; Beckmann, C.F.; Ramnani, N.; Woolrich, M.W.; Bannister, P.R.; Jenkinson, M.; Matthews, P.M.; Mcgonigle, D.J. Variability in fMRI: A re-examination of inter-session differences. *Hum. Brain Mapp.* **2005**, *24*, 248–257. [CrossRef] [PubMed]

18. Seghier, M.L.; Price, C.J. Interpreting and Utilising Intersubject Variability in Brain Function. *Trends Cogn. Sci.* **2018**, *22*, 517–530. [CrossRef]

19. Vakorin, V.A.; Ross, B.; Krakovska, O.; Bardouille, T.; Cheyne, D.; McIntosh, A.R. Complexity analysis of source activity underlying the neuromagnetic somatosensory steady-state response. *NeuroImage* **2010**, *51*, 83–90. [CrossRef]

20. Baig, M.Z.; Kavakli, M. Connectivity Analysis Using Functional Brain Networks to Evaluate Cognitive Activity during 3D Modelling. *Brain Sci.* **2019**, *9*, 24. [CrossRef]

21. Shovon, M.H.; Nandagopal, N.; Vijayalakshmi, R.; Du, J.T.; Cocks, B. Directed Connectivity Analysis of Functional Brain Networks During Cognitive Activity Using Transfer Entropy. *Neural Process. Lett.* **2017**, *45*, 807–824. [CrossRef]

22. Duggento, A.; Bianciardi, M.; Passamonti, L.; Wald, L.L.; Guerrisi, M.; Barbieri, R.; Toschi, N. Globally conditioned Granger causality in brain–brain and brain–heart interactions: A combined heart rate variability/ultra-high-field (7 T) functional magnetic resonance imaging study. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2016**, *374*, 20150185. [CrossRef]

23. Marinazzo, D.; Liao, W.; Chen, H.; Stramaglia, S. Nonlinear connectivity by Granger causality. *NeuroImage* **2011**, *58*, 330–338. [CrossRef] [PubMed]

24. Barnett, L.; Barrett, A.B.; Seth, A.K. Granger Causality and Transfer Entropy Are Equivalent for Gaussian Variables. *Phys. Rev. Lett.* **2009**, *103*, 238701. [CrossRef] [PubMed]

25. Seth, A.K.; Barrett, A.B.; Barnett, L. Granger Causality Analysis in Neuroscience and Neuroimaging. *J. Neurosci.* **2015**, *35*, 3293–3297. [CrossRef] [PubMed]

26. Van Essen, D.C.; Smith, S.M.; Barch, D.M.; Behrens, T.E.; Yacoub, E.; Ugurbil, K. WU-Minn HCP Consortium The WU-Minn Human Connectome Project: An overview. *NeuroImage* **2013**, *80*, 62–79. [CrossRef] [PubMed]

27. Toschi, N.; Duggento, A.; Passamonti, L. Functional Connectivity in Amygdalar-Sensory/(Pre)Motor networks at rest: New evidence from the Human Connectome Project. *Eur. J. Neurosci.* **2017**, *45*, 1224–1229. [CrossRef] [PubMed]

28. Deshpande, G.; Sathian, K.; Hu, X. Effect of hemodynamic variability on Granger causality analysis of fMRI. *Neuroimage* **2010**, *52*, 884–896. [CrossRef] [PubMed]

29. Wu, G.-R.; Liao, W.; Stramaglia, S.; Ding, J.-R.; Chen, H.; Marinazzo, D. A blind deconvolution approach to recover effective connectivity brain networks from resting state fMRI data. *Med. Image Anal.* **2013**, *17*, 365–374. [CrossRef] [PubMed]

30. Barnett, L.; Seth, A.K. Granger causality for state-space models. *Phys. Rev. E* **2015**, *91*, 040101. [CrossRef]

31. Faes, L.; Nollo, G.; Stramaglia, S.; Marinazzo, D. Multiscale Granger causality. *Phys. Rev. E* **2017**, *96*, 042150. [CrossRef]

32. Faes, L.; Nollo, G.; Porta, A. Information Domain Approach to the Investigation of Cardio-Vascular, Cardio-Pulmonary, and Vasculo-Pulmonary Causal Couplings. *Front. Physiol.* **2011**, *2*, 2. [CrossRef]

33. Rubinov, M.; Sporns, O. Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage* **2010**, *52*, 1059–1069. [CrossRef] [PubMed]

34. Song, X.; Panych, L.P.; Chou, Y.-H.; Chen, N.-K. A Study of Long-Term fMRI Reproducibility Using Data-Driven Analysis Methods. *Int. J. Imaging Syst. Technol.* **2014**, *24*, 339–349. [CrossRef] [PubMed]

35. Chen, X.; Lu, B.; Yan, C.-G. Reproducibility of R-fMRI metrics on the impact of different strategies for multiple comparison correction and sample sizes. *Hum. Brain Mapp.* **2018**, *39*, 300–318. [CrossRef] [PubMed]

36. Chen, Y.; Rangarajan, G.; Feng, J.; Ding, M. Analyzing multiple nonlinear time series with extended Granger causality. *Phys. Lett. A* **2004**, *324*, 26–35. [CrossRef]

37. Benhmad, F. Modeling nonlinear Granger causality between the oil price and U.S. dollar: A wavelet based approach. *Econ. Model.* **2012**, *29*, 1505–1514. [CrossRef]

38. Montalto, A.; Stramaglia, S.; Faes, L.; Tessitore, G.; Prevete, R.; Marinazzo, D. Neural networks with non-uniform embedding and explicit validation phase to assess Granger causality. *Neural Netw.* **2015**, *71*, 159–171. [CrossRef] [PubMed]

39. Schreiber, T. Measuring Information Transfer. *Phys. Rev. Lett.* **2000**, *85*, 461–464. [CrossRef] [PubMed]

40. Wibral, M.; Vicente, R.; Lizier, J.T. (Eds.) *Directed Information Measures in Neuroscience*; Understanding Complex Systems; Springer: Berlin/Heidelberg, Germany, 2014.

41. Wen, X.; Rangarajan, G.; Ding, M. Is Granger Causality a Viable Technique for Analyzing fMRI Data? *PLoS ONE* **2013**, *8*, e67428. [CrossRef] [PubMed]

42. Ramsey, J.; Hanson, S.; Hanson, C.; Halchenko, Y.; Poldrack, R.; Glymour, C.; Halchenko, Y. Six problems for causal inference from fMRI. *NeuroImage* **2010**, *49*, 1545–1558. [CrossRef]

43. Laumann, T.O.; Gordon, E.M.; Adeyemo, B.; Snyder, A.Z.; Joo, S.J.; Chen, M.-Y.; Gilmore, A.W.; McDermott, K.B.; Nelson, S.M.; Dosenbach, N.U.; et al. Functional system and areal organization of a highly sampled individual human brain. *Neuron* **2015**, *87*, 657–670. [CrossRef]

44. Mueller, S.; Wang, D.; Fox, M.D.; Yeo, B.T.; Sepulcre, J.; Sabuncu, M.R.; Shafee, R.; Lu, J.; Liu, H. Individual Variability in Functional Connectivity Architecture of the Human Brain. *Neuron* **2013**, *77*, 586–595. [CrossRef] [PubMed]

45. Chen, B.; Xu, T.; Zhou, C.; Wang, L.; Yang, N.; Wang, Z.; Dong, H.-M.; Yang, Z.; Zang, Y.-F.; Zuo, X.-N.; et al. Individual Variability and Test-Retest Reliability Revealed by Ten Repeated Resting-State Brain Scans over One Month. *PLoS ONE* **2015**, *10*, e0144963. [CrossRef] [PubMed]

46. Kong, R.; Li, J.; Sun, N.; Sabuncu, M.R.; Schaefer, A.; Zuo, X.-N.; Holmes, A.J.; Eickhoff, S.; Yeo, B.T.T. Controlling for Intra-Subject and Inter-Subject Variability in Individual-Specific Cortical Network Parcellations. *bioRxiv* **2017**, 213041. [CrossRef]

47. Hodkinson, D.J.; O'Daly, O.; A Zunszain, P.; Pariante, C.M.; Lazurenko, V.; O Zelaya, F.; A Howard, M.; Williams, S.C.R. Circadian and homeostatic modulation of functional connectivity and regional cerebral blood flow in humans under normal entrained conditions. *Br. J. Pharmacol.* **2014**, *34*, 1493–1499. [CrossRef] [PubMed]

48. Shannon, B.J.; Dosenbach, R.A.; Su, Y.; Vlassenko, A.G.; Larson-Prior, L.J.; Nolan, T.S.; Snyder, A.Z.; Raichle, M.E. Morning-evening variation in human brain metabolism and memory circuits. *J. Neurophysiol.* **2013**, *109*, 1444–1456. [CrossRef] [PubMed]

49. Gordon, E.M.; Breeden, A.L.; Bean, S.E.; Vaidya, C.J. Working memory-related changes in functional connectivity persist beyond task disengagement. *Hum. Brain Mapp.* **2014**, *35*, 1004–1017. [CrossRef] [PubMed]

50. Lewis, C.M.; Baldassarre, A.; Committeri, G.; Romani, G.L.; Corbetta, M. Learning sculpts the spontaneous activity of the resting human brain. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 17558–17563. [CrossRef]

51. Tambini, A.; Ketz, N.; Davachi, L. Enhanced Brain Correlations during Rest Are Related to Memory for Recent Experiences. *Neuron* **2010**, *65*, 280–290. [CrossRef]

52. Yeo, B.T.; Tandi, J.; Chee, M.W. Functional connectivity during rested wakefulness predicts vulnerability to sleep deprivation. *NeuroImage* **2015**, *111*, 147–158. [CrossRef]

53. Laumann, T.O.; Snyder, A.Z.; Mitra, A.; Gordon, E.M.; Gratton, C.; Adeyemo, B.; Gilmore, A.W.; Nelson, S.M.; Berg, J.J.; Greene, D.J.; et al. On the Stability of BOLD fMRI Correlations. *Cereb. Cortex* **2017**, *27*, 4719–4732. [CrossRef]

54. Wang, C.; Ong, J.L.; Patanaik, A.; Zhou, J.; Chee, M.W.L. Spontaneous eyelid closures link vigilance fluctuation with fMRI dynamic connectivity states. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 9653–9658. [CrossRef] [PubMed]

55. Shine, J.M.; Koyejo, O.; Poldrack, R.A. Temporal metastates are associated with differential patterns of time-resolved connectivity, network topology, and attention. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 9888–9891. [CrossRef] [PubMed]

56. Vecchio, F.; Miraglia, F.; Marra, C.; Quaranta, D.; Vita, M.G.; Bramanti, P.; Rossini, P.M. Human brain networks in cognitive decline: A graph theoretical analysis of cortical connectivity from EEG data. *J. Alzheimer's Dis.* **2014**, *41*, 113–127. [CrossRef] [PubMed]

57. Anastasiadou, M.N.; Christodoulakis, M.; Papathanasiou, E.S.; Papacostas, S.S.; Hadjipapas, A.; Mitsis, G.D. Graph Theoretical Characteristics of EEG-Based Functional Brain Networks in Patients with Epilepsy: The Effect of Reference Choice and Volume Conduction. *Front. Mol. Neurosci.* **2019**, *13*, 221. [CrossRef] [PubMed]

58. Wang, Y.; Wang, Y.; Lui, Y.W. Generalized Recurrent Neural Network accommodating Dynamic Causal Modeling for functional MRI analysis. *NeuroImage* **2018**, *178*, 385–402. [CrossRef] [PubMed]

59. Gilson, M.; Zamora-Lopez, G.; Pallares, V.; Adhikari, M.H.; Senden, M.; Campo, A.T.; Mantini, D.; Corbetta, M.; Deco, G.; Insabato, A. MOU-EC: Model-based whole-brain effective connectivity to extract biomarkers for brain dynamics from fMRI data and study distributed cognition. *bioRxiv* **2019**, 531830. [CrossRef]

60. Preti, M.G.; Bolton, T.A.; Van De Ville, D. The dynamic functional connectome: State-of-the-art and perspectives. *NeuroImage* **2017**, *160*, 41–54. [CrossRef]