UCSF
UC San Francisco Previously Published Works

Title

Test–Retest Reliability of Computational Network Measurements Derived from the Structural Connectome of the Human Brain

Permalink

https://escholarship.org/uc/item/2p2860zn

Journal

Brain Connectivity, 3(2)

ISSN

2158-0014

Authors

Owen, Julia P
Ziv, Etay
Bukshpun, Polina
et al.

Publication Date

2013-04-01

DOI

10.1089/brain.2012.0121

Peer reviewed

Test–Test Reliability of Computational Network Measurements Derived from the Structural Connectome of the Human Brain

Julia P. Owen,[1,2] Etay Ziv,[1] Polina Bukshpun,[3] Nicholas Pojman,[3] Mari Wakahiro,[3] Jeffrey I. Berman,[4] Timothy P.L. Roberts,[4] Eric J. Friedman,[5] Elliott H. Sherr,[3] and Pratik Mukherjee[1,2]

Abstract

Structural magnetic resonance (MR) connectomics holds promise for the diagnosis, outcome prediction, and treatment monitoring of many common neurodevelopmental, psychiatric, and neurodegenerative disorders for which there is currently no clinical utility for MR imaging (MRI). Before computational network metrics from the human connectome can be applied in a clinical setting, their precision and their normative intersubject variation must be understood to guide the study design and the interpretation of longitudinal data. In this work, the reproducibility of commonly used graph theoretic measures is investigated, as applied to the structural connectome of healthy adult volunteers. Two datasets are examined, one consisting of 10 subjects scanned twice at one MRI facility and one consisting of five subjects scanned once each at two different facilities using the same imaging platform. Global graph metrics are calculated for unweighed and weighed connectomes, and two levels of granularity of the connectome are evaluated: one based on the 82-node cortical and subcortical parcellation from FreeSurfer and one based on an atlas-free parcellation of the gray–white matter boundary consisting of 1000 cortical nodes. The consistency of the unweighed and weighed edges and the module assignments are also computed for the 82-node connectomes. Overall, the results demonstrate good-to-excellent test–retest reliability for the entire connectome-processing pipeline, including the graph analytics, in both the intrasite and intersite datasets. These findings indicate that measurements of computational network metrics derived from the structural connectome have sufficient precision to be tested as potential biomarkers for diagnosis, prognosis, and monitoring of interventions in neurological and psychiatric diseases.

Key words: connectome; diffusion tensor imaging (DTI); graph theory; magnetic resonance imaging (MRI); network; structural connectivity; tractography; white matter

Introduction

O VER THE PAST 30 years, magnetic resonance (MR) imaging has revolutionized the diagnosis and clinical management of neurological disease, including major illnesses such as stroke, brain tumors, multiple sclerosis, and epilepsy. However, despite this record of progress, most of the worldwide morbidity and death from neurological and psychiatric diseases remain unaddressed by current diagnostic imaging techniques. Current clinical MR imaging (MRI) modalities are not typically helpful for diagnosis, outcome prediction, or treatment monitoring of many common neurodevelop-

mental, psychiatric, and neurodegenerative disorders. Prominent examples include autism, schizophrenia, bipolar disorder, major depression, Parkinson's disease, and Alzheimer's disease. Collectively, they have much greater prevalence than those neurological diseases for which MRI is routinely useful, and the incidence of disorders such as autism and Alzheimer's disease continues to rise (Rosen and Napadow, 2011).

The pathophysiology of many neurodevelopmental, psychiatric, and neurodegenerative diseases is thought to be structurally diffuse, unlike that of more focal disorders such as stroke and brain tumors for which MRI has proven clinical

Departments of [1]Radiology & Biomedical Imaging, [2]Bioengineering & Therapeutic Sciences, and [3]Department of Neurology, University of California, San Francisco, California.
[4]Department of Radiology, Children's Hospital of Philadelphia, University of Pennsylvania, Philadelphia.
[5]International Computer Science Institute, University of California, Berkeley, California.

utility. Therefore, advances in diagnosis and prognosis depend on a better understanding of the brain at a systems level. Recently, there has been much interest in measuring the structural and functional connectivity of brain regions using diffusion MR tractography and resting-state functional MRI, respectively. This has in turn led to the new science of MR connectomics, a burgeoning research methodology that applies graph theory to whole-brain structural and functional networks, which are referred to as connectomes (Bullmore and Sporns, 2009; Gong et al., 2009; Hagmann et al., 2007, 2010; Iturria-Medina et al., 2007; Kuceyeski et al., 2011; Li et al., 2012a, 2012b; Seung, 2012; Sporns et al., 2005; Sporns, 2011; van den Heuvel and Sporns, 2011; Zalesky et al., 2010). Graph theoretic principles can be used to ascertain properties such as the local and global efficiency of a network, the similarity of different networks, and the modular organization of distinct subnetworks within a larger network. The connectome framework has already yielded novel insights into the structural and functional organization of the normal human brain and has begun to be applied to the study of human brain development and of neurological and psychiatric diseases (Fan et al., 2011; Fornito and Bullmore, 2012; Hagmann et al., 2012; Irimia, et al., 2012; Owen et al., 2012; Shu et al., 2009; Shu et al. 2012; Tymofiyeva et al., 2012; Verstraete et al., 2011; Yan et al., 2011; Yap et al., 2011).

For clinical application of MR connectomics, there is a need to determine the measurement precision of whole-brain network metrics and their normative intersubject variation. This information, in conjunction with effect size, can be used to determine statistical power and sample sizes for cross-sectional studies of group differences between patients and matched controls, as well as for longitudinal studies of development, disease evolution, and treatment efficacy.

There have been several prior studies that examine the aspects of the test–retest reliability of structural connectome reconstruction methods and the application of graph analytics to the resulting networks (Bassett et al., 2011; Cammoun et al., 2012; Cheng et al., 2012; Dennis et al., 2012; Hagmann et al., 2008; Vaessen et al., 2010). However, many important details of the connectome reconstruction pipeline differ across these prior studies, with no consensus on the best approach. Further, only fair-to-good levels of reproducibility have been achieved for measurements of connectome graph metrics in these previous reports.

In this article, we investigate the test–retest reliability and intersubject variation of network metric measurements from the normal adult structural connectome using acquisition and analysis methods suitable for clinical research. This includes a standard 3T diffusion MRI protocol and widely used postprocessing software. Connectomic analysis was carried out at both low- (82 regions) and high- (1000 regions) parcellation scales. We explore weighed connectomes, in which connection strength information is preserved as edge weights, and unweighed connectomes, in which the edges are binarized as being present above a connection strength threshold or absent otherwise. In addition to summary metrics of the entire network, such as mean degree, characteristic path length, and mean clustering coefficient, we also examine measures at higher levels of granularity, including the modular organization of subnetworks as well as the consistency of individual network connections in edge space.

We demonstrate that the proposed connectome-processing pipeline produces good-to-excellent reliability of network metric measurements, both for a single imaging site as well as across two geographically distant sites, employing the same type of MR scanner. These results represent an initial step toward validating structural brain network metrics as quantitative imaging biomarkers of neurologic and psychiatric diseases, including for multicenter clinical trials.

Materials and Methods

Subjects

Ten healthy control subjects (mean age 26.7 ± 5.9 years, five men, nine right-handed) were scanned at Site 1 twice with an average of 30.4 ± 2.7 days between scans. Five healthy controls subjects (mean age 34.6 ± 10.7 years, four men, four right-handed) were scanned once at Site 1 and once at Site 2, with an average of 60.8 ± 33.6 days between scans. All study procedures were approved by the institutional review boards at our medical centers and are in accordance with the ethics standards of the Helsinki Declaration of 1975, as revised in 2008.

Image acquisition

All MRI was performed on a 3T TIM Trio MR scanner (Siemens, Erlangen, Germany) at each site, using 32-channel head phased-array radiofrequency head coils. High-resolution structural MRI of the brain was performed with an axial 3D magnetization prepared rapid-acquisition gradient-echo (MPRAGE) T1-weighted sequence (echo time [TE] = 1.64 ms, repetition time [TR] = 2530 ms, TI = 1200 ms, flip angle of 7°) with a 256-mm field of view (FOV), and 160 1.0-mm contiguous partitions at a 256×256 matrix.
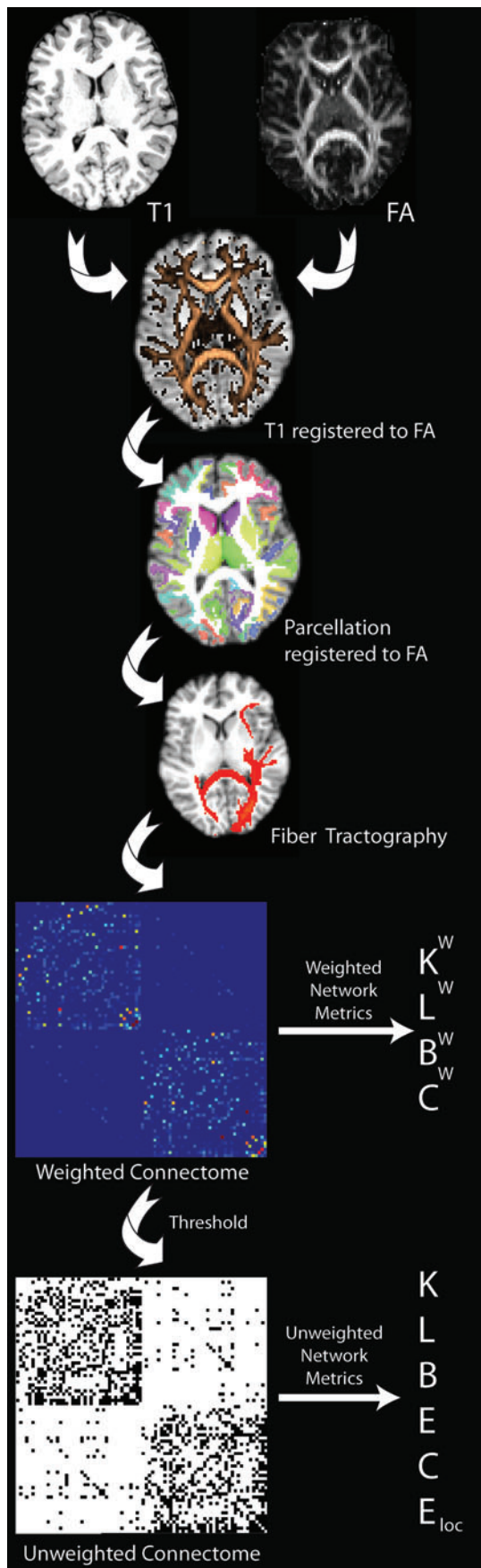
Whole-brain diffusion was performed with a multislice 2D single-shot twice-refocused spin-echo echo-planar sequence with 30 diffusion-encoding directions, the iPAT technique for parallel imaging with a reduction factor of 2, a diffusion-weighing strength of $b = 1000 \, s/mm^2$; TE/TR = 88/10300 ms; number of excitations [NEX] = 1; interleaved 2-mm axial sections with no gap; inplane resolution of 2×2 mm with a 128×128 matrix; and an FOV of 256 mm. An additional image set was acquired with no diffusion weighing ($b = 0 \, s/mm^2$). The total acquisition time for diffusion imaging was 8.5 min.

Data preprocessing

After the nonbrain tissue was removed using the Brain Extraction Tool (Smith, 2002), the diffusion-weighed images were corrected for motion and eddy currents using the functional MRI of the brain (FMRIB) linear-image registration tool (FLIRT) with a 12-parameter linear image registration (Jenkinson et al., 2002) using the $b = 0 \, s/mm^2$ image as the reference. The fractional anisotropy (FA) image was calculated using FSL's DTIFIT.

Cortical parcellation

The T1-weighted MR images were automatically segmented using FreeSurfer 5.1.0 (Fischl et al., 2004) with the default settings of recon-all, resulting in 68 cortical regions and 14 subcortical regions. The 68 cortical regions

FIG. 1. Cartoon depicting the 82-node connectome-processing pipeline. First, the diffusion and structural magnetic resonance images are aligned; then, the FreeSurfer labels are transformed to the diffusion space to seed the tractography. Tractography is performed with each of the 82 seed regions to construct the structural connectome.

were transformed to the gray–white matter boundary (GWB) using FreeSurfer. These 82 regions represent the nodes of the low-resolution connectome. In addition to this atlas-based coarse parcellation, an atlas-free finer parcellation of the cortical GWB was performed. One thousand clusters of voxels were grown along the 3D boundary as described in the first phase of the boundary partition in Hagmann et al. (2007). These 1000 regions represent the nodes in the high-resolution cortical connectome. These nodes, however, are purely cortical, as the GWB is not easily segmented for the subcortical structures.

Fiber tractography and connectome reconstruction

Using FLIRT, the affine transform from diffusion to structural space was calculated by registering the FA volume to the T1 volume. Each of the cortical GWB volumes and the subcortical volumes was registered to the diffusion space to be used as seeds for the tractography. Probabilistic tractography was performed with probtrackx2 (Behrens et al., 2007), with 2000 streamlines initiated from each seed voxel using the default options. The number of streamlines from each seed to each of the other seeds, called targets, was summed across the voxels to obtain a connection strength between each seed and target pair. This connection strength was then divided by the sum of voxels in the seed and target region to account for differences in volume between the various cortical or subcortical regions. Since tractography cannot determine directionality due to the antipodal symmetry of diffusion imaging, the normalized connection strength between each seed and target pair in both directions was summed, and the connection strength of a seed with itself was set to zero. This processing pipeline (Fig. 1) closely follows the M2 method, described by Li et al. (2012b), and is also described in detail in Owen et al. (2012).

Weighed and unweighed connectomes were used in this investigation. The weighed connectomes do not require thresholding, but a threshold value for the connection strength has to be used to binarize the weighed network for the unweighed metrics. A threshold of 200 was applied to all the 82-node connectomes; this value was chosen such that the mean degree of the 10 intrasite subjects was ~ 12, following van den Heuvel and Sporns (2011). The same connection strength threshold of 200 was also used to binarize all of the unweighed 1000-node connectomes.

Network analysis: graph metrics

The reproducibility of summary graph metrics was investigated for the unweighed and weighed whole-brain structural networks.

The unweighed metrics used to assess the stability of the global network structure include mean degree (K), characteristic path length (L), mean global efficiency (E), and mean normalized betweenness centrality (B). The degree of a

node is the total number of suprathreshold connections it makes with other nodes in the network. Characteristic path length is calculated by taking the average of all the shortest paths between all pairs of nodes in the network; it is related to the speed at which information can be disseminated through a network. The normalized betweenness of a node is the number of shortest paths that pass through it and is normalized by the total number of possible shortest paths. Normalized betweenness provides a measure of dispersion of the shortest paths. Global efficiency is computed by taking the mean inverse path length between all pairs of nodes. This measure is similar to the characteristic path length, but the inversion reduces the disproportionate effect of long or infinite path lengths that occur in very sparse or disconnected networks, respectively.

Local connectivity metrics include the mean clustering coefficient (C) and the mean local efficiency (E_{loc}). The latter provides a measure of the efficiency of the local environment of a node. The clustering coefficient is the ratio of closed triangles between triplets of nodes and the total number of connected triplets. The triangular pattern of connectivity is assumed to be a feature of networks with strong local integration of information.

We also test the following weighed metrics: mean strength (K^w), weighed characteristic path length (L^w), mean normalized weighed betweenness (B^w), and mean weighed clustering coefficient (C^w), which have similar interpretations as their unweighed analogs. The unweighed and weighed metrics were applied to the 82-node and 1000-node connectomes. A comprehensive discussion of these network metrics and their significance can be found in Rubinov and Sporns (2010). All network metrics were implemented in MatLab, and the code is a part of an open-source software package, the Brain Connectivity Toolbox (Rubinov and Sporns, 2010).

Measurement precision of graph metrics

To quantify the stability of the network metrics, we use the pooled within-group percentage coefficient of variation (CV%) and the intraclass correlation coefficient (ICC). CV% is defined as the ratio of the mean intrasubject standard deviation (SD) to the overall measurement mean (Lachin, 2004; Vaessen et al., 2010). CV% measures the precision of a metric for all subjects. ICC is the ratio of intersubject variance to the sum of intersubject and intrasubject variance. According to the well-established guidelines for clinical research (Fleiss, 1986; Tooth et al., 2005), ICC below 0.4 are considered poor reproducibility; ICC values between 0.4 and 0.75 are considered fair-to-good reproducibility; and ICC values above 0.75 are considered excellent reproducibility.

Since ICC quantifies the intersubject variation that is not accounted for by the intrasubject variation, a large intrasubject variance can be tempered by an even greater intersubject variance. CV% and ICC are not independent measures (the intrasubject variance is part of both), but they do not always agree, because they reflect different aspects of the test–retest variation. Low CV% indicates that the measure is precise, whereas a low CV% and a high ICC indicate that the measure is precise and captures individual variability (i.e., the intersubject variation is much larger than the intrasubject variation). However, a low CV% and a low ICC indicate that, while the measure is precise, the variation over the subjects is on par with the variation within a subject and is not highly sensitive to individual differences. Conversely, a high CV% and high ICC reveal that, while a measure is not precise, it does reflect the individual variation. As such, these two measures are best interpreted in conjunction with each other.

To determine statistical significance, we calculated the p values for ICC with a nonparametric, resampling method. The subject labels are randomly reassigned, and ICC is recomputed for the permuted data; a total of 5000 permutations were performed. The number of ICC values for the permuted data that are greater than the value obtained for the original data is divided by the total number of permutations. If the value obtained for the original data was greater than all values from the permutated data, the p value is reported as $p < 0.0002$. The null hypothesis is that the data labels are randomly assigned. This procedure tests whether the ICC can be improved upon by random reassignment, with a significance threshold of $p < 0.05$.

Effect of connectome processing parameters

To be sure that the results are not biased by the threshold level selected for probabilistic tractography, we explore the effect of this threshold on the reliability of the global and local network measures. CV% and ICC were recalculated for the unweighed metrics for the 82- and 1000-node connectomes at thresholds varying over a range of values centered on 50 and 200 for the 82- and 1000-node connectomes, respectively. In a separate examination, we vary the number of streamlines initiated per voxel to assess the sensitivity of the results to this parameter. The tractography was rerun using 250, 500, 1000, and 4000 streamlines per voxel to compare with the default value of 2000 streamlines per voxel used in all other analyses presented in this article. ICC and CV were calculated for each number of streamlines for the 82- and 1000-node connectomes and for the unweighed and weighed network metrics. For the unweighed metrics, the threshold was set for each number of streamlines such that the mean degree over all networks was equivalent to that obtained with 2000 streamlines per voxel. Due to the computational expense involved, the tractography was only rerun with 250 and 4000 streamlines per voxel for the 1000-node connectomes. These two experiments, investigating the dependence on threshold and on seeding density, were only done with the intrasite data.

Network analysis: edge consistency

Edge consistency of the unweighed and weighed 82-node networks was investigated. To assess the consistency of unweighed connectomes, the number of edges either present or absent in any pair of networks was normalized by the total number of possible edges; we call this measure the edge agreement (Hagmann et al., 2008; Cammoun et al., 2012). For the weighed connectomes, the correlation coefficient was computed between the weighed edges of two networks. For both edge agreement and edge correlation, the metrics are computed on an intrasubject and intersubject level as described in the section entitled Intrasubject and Intersubject Calculations. A second measure of connection strength consistency was performed by calculating the CV% for every suprathreshold edge (>50) present in both scans for each intrasite subject. Then, we calculated the average number of
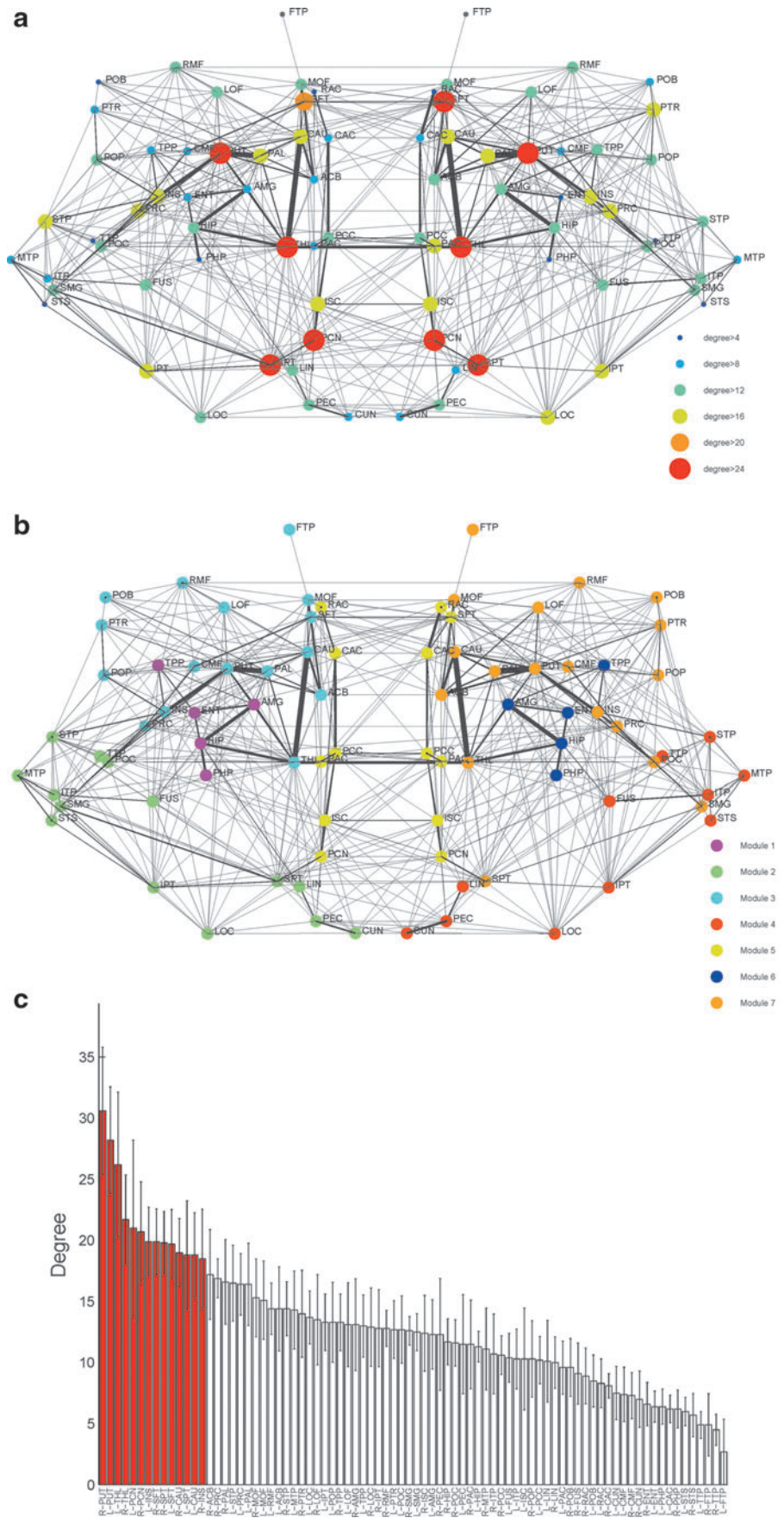
FIG. 2. **(a)** Representative connectome, **(b)** representative module partition, and **(c)** mean degree distribution across 10 intrasite subjects.

High Resolution Cortical Connectome: Degree

LH   RH

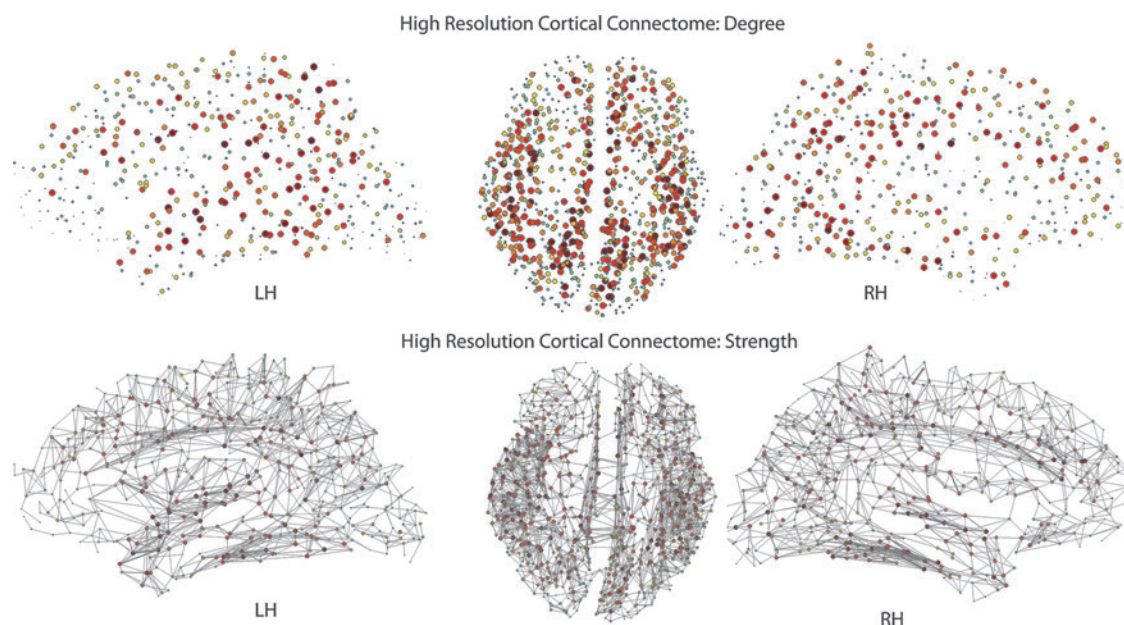High Resolution Cortical Connectome: Strength

LH   RH

FIG. 3. Representative high-resolution cortical connectome: degree and strength of nodes.

steps between the two nodes corresponding to each of these suprathreshold edges using the distance-weighing option in probtrackx2. The correlation coefficient and p value were calculated for the CV% and the distance between the nodes to test the hypothesis that long-range connections have less-reliable connection strengths.

Network analysis: module assignment

To evaluate the reproducibility of the community structure within connectomes, we measured the test–retest reliability of the module assignments provided by a community detection algorithm proposed in Blondel et al. (2008) as implemented in the Brain Connectivity Toolbox. Both unweighed and weighed 82-node connectomes, all thresholded at the same level, but with suprathreshold edge-weight information retained in the latter, were decomposed into modules, and the consistency of the modular assignment was quantified using the Hubert Rand index (HRI) as defined in Hubert and Baker (1977). First, we investigated the stability of the algorithm, since its community assignments are sensitive to the ordering of the nodes; hence, multiple repetitions of the algorithm can yield different assignments. We ran the algorithm 100 times, while randomly reordering the nodes, and computed the mean and SD of the HRI across runs for each connectome as a measure of algorithmic stability. We also computed the mean and SD of the HRI for random community assignments. A total of six communities were randomly assigned to 82 nodes; we chose six, as this was the most frequent number of modules detected. Then, we assessed the test–retest reliability of the module assignments calculating both the intrasubject and intersubject consistency as described in the next section.

Intrasubject and intersubject calculations

For the edge consistency metrics and the HRI, we compute a mean intrasubject value for each subject and a mean inter-

subject value, averaged over all subjects. To compare the intrasubject consistency for the edge-consistency metrics, we compute the edge agreement and correlation coefficient for each subjects' two scans and then compare these values to the intersubject consistency by computing the mean edge agreement and mean correlation coefficient for all pairs of intersubject scans. This procedure is applied to both the intra- and intersite data sets. With the intersite data, the inter-subject edge agreement and correlation coefficient were only computed for the pairs of scans obtained at different sites. As with the edge consistency analysis, we computed the HRI for the module assignments for each subject's test–retest data and the intrasubject module consistency, and compared that to the intersubject module consistency by computing the HRI for all pairs of intersubject scans. This procedure is applied to both the intra- and intersite data sets. With the inter-site data, the intersubject module consistency was only computed for pairs of scans obtained at different sites. To account for the variation introduced by the nondeterministic community detection algorithm, we computed the module assignments 100 times for each connectome, and thus the HRI for each pair of community assignments was averaged over 100 repetitions.

Results

Connectomes, degree distributions, and modular assignments

In Fig. 2, we show an example of an 82-node connectome in a single normal adult volunteer as well as the modular assignments for the same subject. The locations of the high-degree nodes, also known as hubs, are consistent with those reported in the literature (Hagmann et al., 2008; van den Heuvel and Sporns, 2011). The modules detected include the structural core (Module 5) across both the cerebral hemispheres and three other modules in each hemisphere. All seven modules exhibit bilateral symmetry, and the six that are contained within

a single cerebral hemisphere segregate into the anterior, posterior, and subcortical/inferior subnetworks. In Fig. 2c, the degree distribution across the 82 nodes is plotted. Only the first scans of the intrasite data set were used in this calculation (a total of 10 subjects). The hub regions, defined as those nodes with a degree greater than one SD above the mean, are colored in red. These hubs correspond closely to those reported in Li et al. (2012b). In Fig. 3, the degrees and strengths are shown for the 1000-node cortical connectome for the same individual illustrated in Fig. 2. For clarity, only the top 1% of edges by strength are displayed in the bottom row of Fig. 3.

Test–retest reliability of summary graph metrics for the 82-node connectomes

In Fig. 4, the summary network metrics of the 82-node connectomes are plotted for the intrasite and intersite datasets. In Table 1, the mean, SD for Session 1 and 2, CV%, and ICC, as well as the p values for ICC, are provided for these metrics of the unweighed 82-node connectomes across all 10 intrasite and five intersite subjects. The CV% values for the unweighed metrics computed for the intrasite data are all $<3.5\%$, indicating a low intrasubject variation with respect to the mean. The
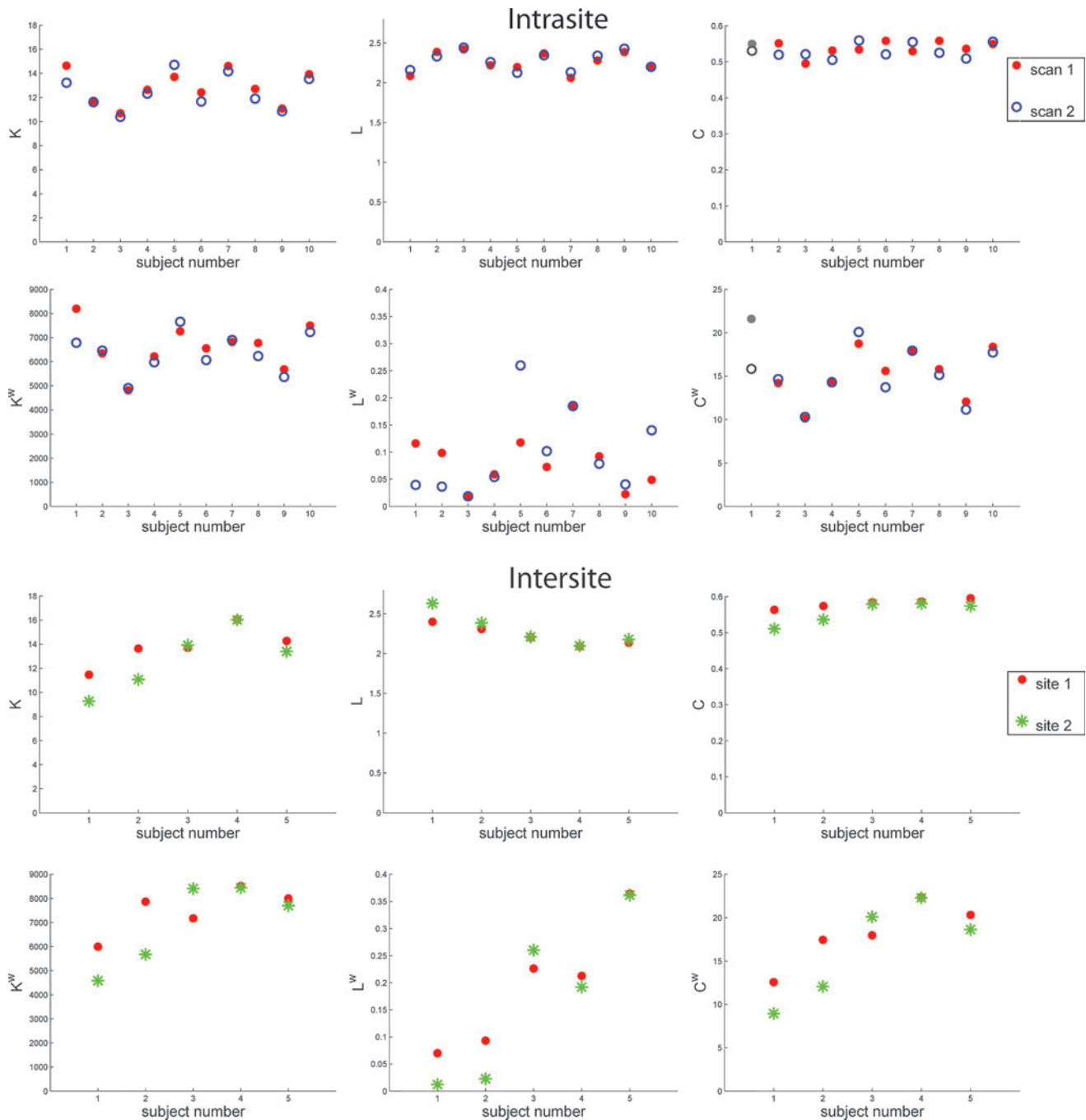

FIG. 4. Selected unweighed and weighed metrics: K, L, C, K^w, L^w, and C^w are plotted for the 82-node connectomes across the intrasite and intersite scans.

TABLE 1. UNWEIGHTED METRICS, 82-NODE CONNECTOME

Intrasite	Session 1 Avg. ± SD	Session 2 Avg. ± SD	CV%	ICC	p Value ICC
K	12.8 ± 1.4	12.4 ± 1.4	3.21	0.89	**0.0002**
L	2.26 ± 0.13	2.28 ± 0.12	1.44	0.92	**0.0002**
C	0.54 ± 0.02	0.53 ± 0.02	3.40	0.51	0.46
B	0.032 ± 0.003	0.033 ± 0.003	2.73	0.90	**< 0.0002**
E	0.51 ± 0.02	0.51 ± 0.02	1.18	0.92	**< 0.0002**
E_{loc}	0.75 ± 0.01	0.75 ± 0.02	1.79	0.54	0.32

Intersite	Site 1 Avg. ± SD	Site 2 Avg. ± SD	CV%	ICC	p Value ICC
K	13.8 ± 1.6	12.7 ± 2.6	6.26	0.79	**0.02**
L	2.22 ± 0.13	2.30 ± 0.21	2.26	0.82	**< 0.0002**
C	0.59 ± 0.01	0.55 ± 0.03	3.05	0.59	**0.03**
B	0.031 ± 0.003	0.033 ± 0.005	4.00	0.82	**< 0.0002**
E	0.52 ± 0.03	0.50 ± 0.04	1.94	0.84	**< 0.0002**
E_{loc}	0.78 ± 0.01	0.76 ± 0.02	1.61	0.63	**0.02**

The p values for ICC less than the significance threshold of 0.05 are given in boldface.
CV%, percentage coefficient of variation; ICC, intraclass correlation coefficient; SD, standard deviation; L, characteristic path length; C, mean clustering coefficient; B, mean normalized betweenness centrality; E, mean global efficiency; E_{loc}, mean local efficiency.

ICC values for the summary network metrics, with the exception of C and E_{loc}, are ≥ 0.89, and their corresponding p values are ≤ 0.0002. For the unweighed intersite data, the CV% values, while higher than for the unweighed intersite data, are still < 6.3%. The unweighed summary network metrics have excellent reproducibility across sites, with ICC values > 0.75, once again with the exception of C and E_{loc}.

The reproducibility statistics for the weighed summary network metrics are presented in Table 2. The weighed metrics, in general, are not quite as highly reproducible as the unweighed metrics for both the intrasite and intersite data. K^w and C^w have excellent reproducibility for the intrasite data with CV% of 4.3% and 5.5%, respectively, and ICC values of 0.84 and 0.83, respectively, with p values for ICC reflecting a strong statistical significance. For the intersite data, however, L^w and C^w emerge as the most reproducible with CV% values of 14.5% and 10.6% and ICC values 0.94 and 0.81, respectively, and p values reflecting significance.

The other weighed network metrics have a higher CV%, a lower ICC, or are not statistically significant. The histograms for the node metrics k_i, l_i, c_i, b_i, k^w_i, l^w_i, c^w_i, and b^w_i are presented in Supplementary Fig. S1 (Supplementary Data are available online at www.liebertpub.com/scd). In general, the variation for the node metrics is higher than the variation for the summary network metrics, and the unweighed metrics are less variable than the weighed metrics.

Test–retest reliability of summary graph metrics for the 1000-node connectomes

In Fig. 5, the summary network metrics of the high-resolution cortical connectomes are plotted for the intrasite and intersite data sets. The intrasite data show a high test–retest reliability for all metrics. For the intersite data, K^w and L^w are highly consistent, while C^w is less reproducible in the intersite cohort as compared to the intrasite data. In Table 3, the mean and SD

TABLE 2. WEIGHTED METRICS, 82-NODE CONNECTOME

Intrasite	Session 1 Avg. ± SD	Session 2 Avg. ± SD	CV%	ICC	p Value ICC
K^w	6600 ± 950	6400 ± 840	4.31	0.84	**0.001**
L^w	0.083 ± 0.05	0.096 ± 0.08	34.75	0.67	0.05
C^w	15.9 ± 3.4	15.1 ± 3.0	5.50	0.83	**0.001**
B^w	0.067 ± 0.005	0.069 ± 0.004	4.58	0.52	0.34

Intersite	Site 1 Avg. ± SD	Site 2 Avg. ± SD	CV%	ICC	p Value ICC
K^w	7500 ± 970	7000 ± 1700	10.17	0.69	0.14
L^w	0.190 ± 0.12	0.170 ± 0.15	14.46	0.94	**< 0.0002**
C^w	18.1 ± 3.7	16.4 ± 5.7	10.57	0.81	**0.01**
B^w	0.066 ± 0.003	0.067 ± 0.006	2.63	0.70	0.10

The p values for ICC less than the significance threshold of 0.05 are given in boldface.
B^w, mean normalized weighed betweenness; K^w, mean strength; C^w, mean weighed clustering coefficient; L^w, weighed characteristic path length.
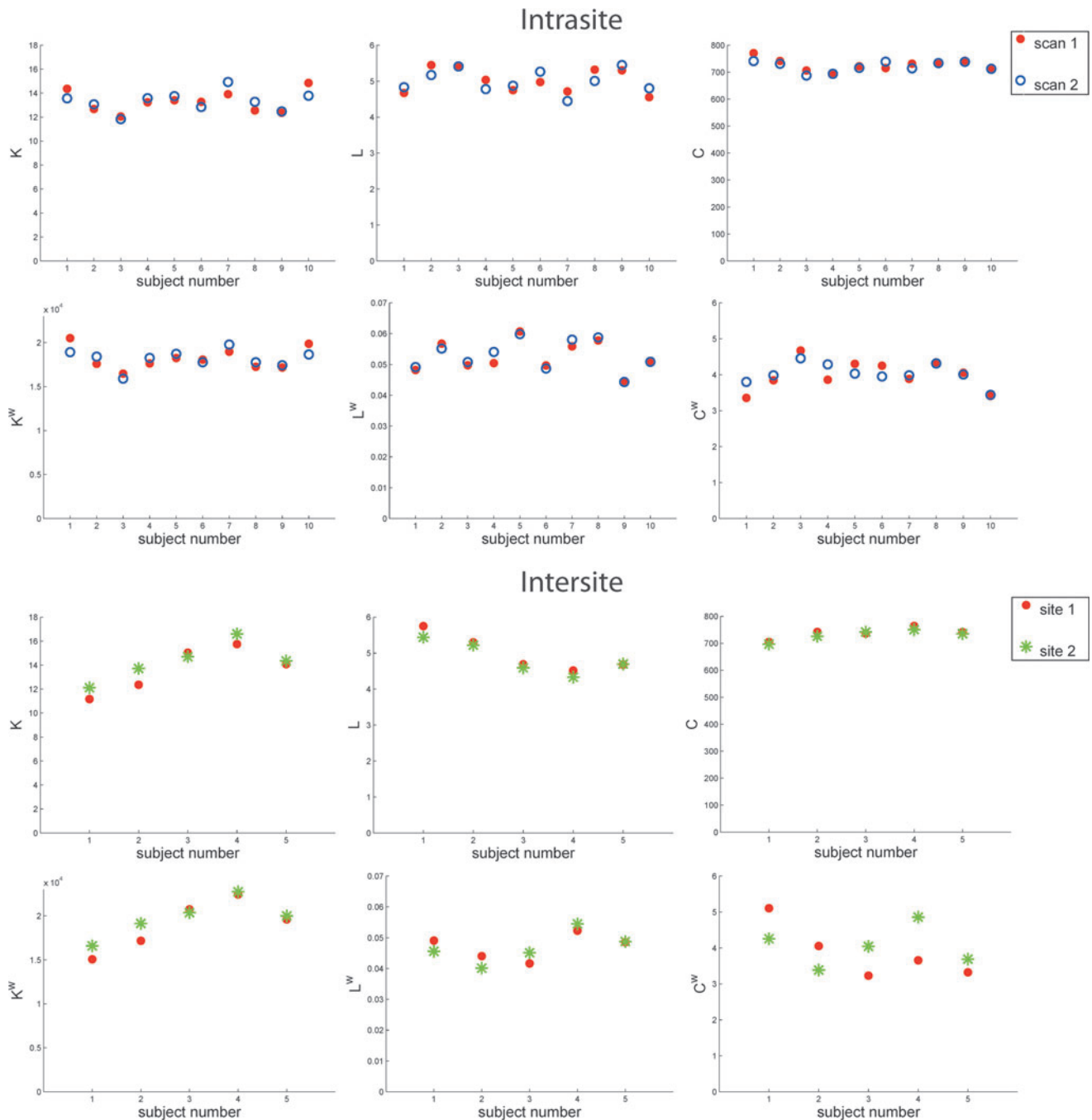
FIG. 5. Selected unweighed and weighed metrics: K, L, C, K^w, L^w, and C^w are plotted for the high-resolution cortical connectomes across intrasite and intersite scans.

for Session 1 and 2 and the CV% and ICC values for the unweighed metrics are reported for the high-resolution connectomes. In general, the CV% values are lower than 5% for all unweighed summary metrics, both intrasite and intersite, and the ICC values are all higher than 0.75, with many values exceeding 0.8. The ICC *p* values for all unweighed metrics show highly significant reproducibility. In Table 4, the reproducibility statistics are presented for the weighed summary metrics of the 1000-node cortical connectomes. The CV% values for all weighed metrics applied to the intrasite data are < 3.6%, and the ICC values are ≥ 0.79. The *p* values

for ICC are again strongly statistically significant. The ICC values for the intersite data reflect excellent reproducibility of K^w and L^w, with CV% < 4% and ICC values ≥ 0.82. However, C^w and B^w are not significantly reproducible across sites.

The effect of threshold and number of streamlines on ICC and CV%

To address the possible confounds introduced by fixing the connectivity threshold across subjects for the calculation of

TABLE 3. UNWEIGHED METRICS, 1000-NODE CONNECTOME

Intrasite	Session 1 Avg. ± SD	Session 2 Avg. ± SD	CV%	ICC	p Value ICC
K	13.3 ± 0.9	13.3 ± 0.8	2.89	0.78	**0.0034**
L	5.0 ± 0.3	5.0 ± 0.3	2.96	0.79	**0.004**
C	730 ± 20	720 ± 20	1.12	0.78	**0.001**
B	0.0061 ± 0.0007	0.0061 ± 0.0006	4.89	0.79	**0.005**
E	0.23 ± 0.01	0.23 ± 0.01	2.58	0.78	**0.005**
E_{loc}	0.72 ± 0.01	0.72 ± 0.01	0.42	0.75	**0.0072**
Intersite	Site 1 Avg. ± SD	Site 2 Avg. ± SD	CV%	ICC	p Value ICC
K	13.7 ± 1.9	14.3 ± 1.6	3.81	0.89	**0.0004**
L	5.0 ± 0.5	4.9 ± 0.5	1.98	0.94	**0.001**
C	740 ± 20	730 ± 20	0.94	0.88	**0.02**
B	0.006 ± 0.001	0.006 ± 0.001	3.31	0.94	**0.002**
E	0.23 ± 0.02	0.24 ± 0.2	1.74	0.94	**< 0.0002**
E_{loc}	0.72 ± 0.01	0.72 ± 0.01	0.42	0.87	**0.01**

The *p* values for ICC less than the significance threshold of 0.05 are given in boldface.

the unweighed network metrics, we investigate CV% and ICC as a function of threshold (Fig. 6). The threshold was varied over an order of magnitude from 10 to 100 for the 82-node connectomes and from 50 to 500 for the 1000-node connectomes. We found that for the 82-node connectomes, the ICC and CV% values for the graph metrics that assess the global structure (K, L, B, and E) were very stable. On the other hand, the metrics that quantify the local structure (C and E_{loc}) have more variable reliability, with ICC values ranging from ∼0.5 to 0.8. In fact, the threshold of 50 that was chosen as the default value is actually a local minimum of the ICC versus threshold curve. The ICC and CV% values for the 1000-node connectomes exhibit a similar resilience to threshold this is true for all metrics, both global and local. Figure 7 shows the results for varying the number of streamlines initiated from each voxel for both unweighed and weighed metrics, at both the 82- and 1000-node resolution. To conserve space, we only plot ICC; the CV% results are similar and can be found in Supplementary Fig. S2. ICC and CV% are very stable for all numbers of streamlines and metrics, with weighed betweenness and weighed characteristic path length

showing the most variability in ICC and CV, respectively, at the 82-node parcellation.

Test–retest reliability of network edges

Figure 8 illustrates the test–retest reliability of the network edges. The top row displays the results for the unweighed 82-node connectomes. The points are the edge agreement for the intrasubject data, and the red line denotes the mean intersubject edge agreement with the SD plotted as the dashed line above and below the mean. The intrasubject edge agreement for all subjects in both the intrasite and intersite cohorts is greater than the mean intersubject agreement, and all, but one, of the intrasubject points are greater than the mean plus one SD for the intersubject edge agreement. The bottom row of Fig. 8 displays the correlation coefficients for the edge weights for the intrasite and intersite weighed 82-node connectomes. Here, the data points represent the intrasubject correlation coefficient, and the red line marks the mean intersubject correlation coefficient with the SD interval indicated as dashed lines. The general trend is similar to that

TABLE 4. WEIGHED METRICS, 1000-NODE CONNECTOME

Intrasite	Session 1 Avg. ± SD	Session 2 Avg. ± SD	CV%	ICC	p Value ICC
K^w	18200 ± 1300	18200 ± 1000	2.78	0.79	**0.002**
L^w	0.052 ± 0.005	0.053 ± 0.0051	1.69	0.95	**< 0.0002**
C^w	4.0 ± 0.4	4.0 ± 0.3	3.53	0.80	**0.003**
B^w	1200 ± 66	1200 ± 69	1.26	0.93	**< 0.0002**
Intersite	Site 1 Avg. ± SD	Site 2 Avg. ± SD	CV%	ICC	p Value ICC
K^w	19000 ± 2900	20000 ± 2200	3.33	0.90	**< 0.0002**
L^w	0.047 ± 0.004	0.047 ± 0.005	3.99	0.82	**0.01**
C^w	3.9 ± 0.8	4.0 ± 0.6	13.91	0.54	0.31
B^w	1200 ± 120	1100 ± 50	4.54	0.69	0.11

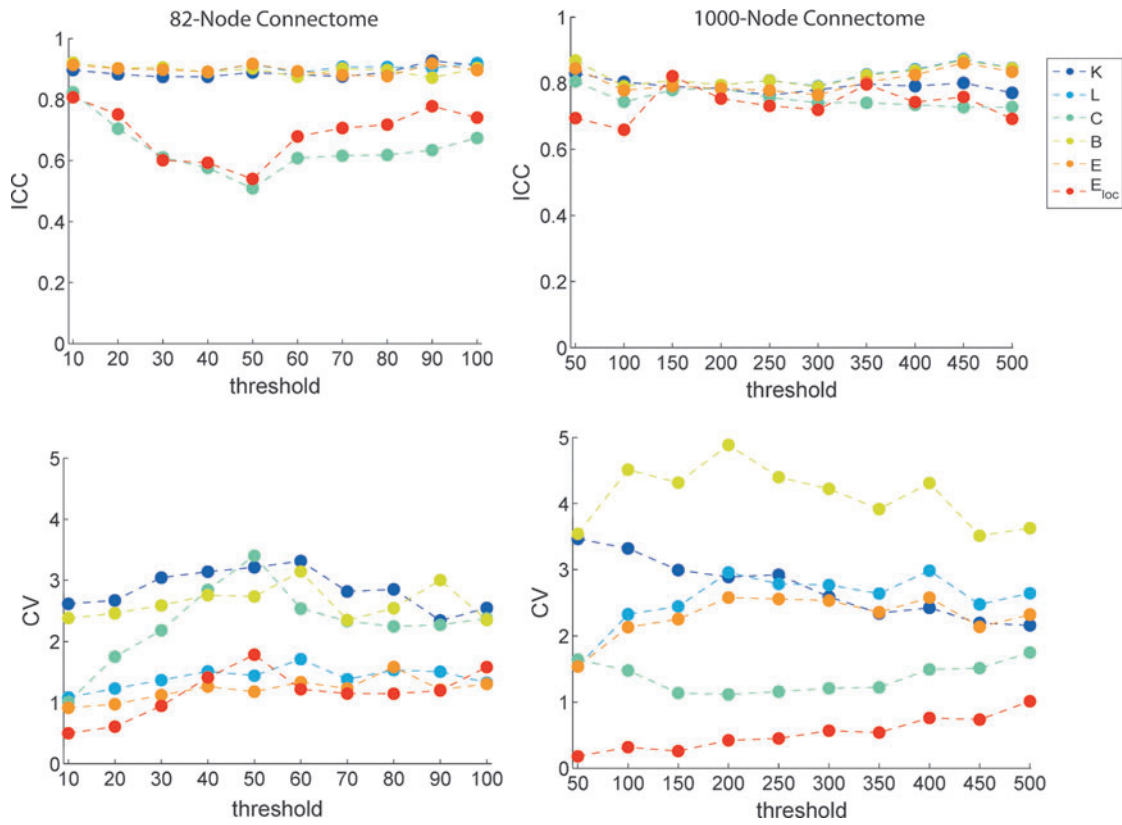The *p* values for ICC less than the significance threshold of 0.05 are given in boldface.

FIG. 6. The effect of threshold on intraclass correlation coefficient (ICC) and percentage coefficient of variation (CV%) for the graph metrics computed for the 82- and 1000-node connectomes.
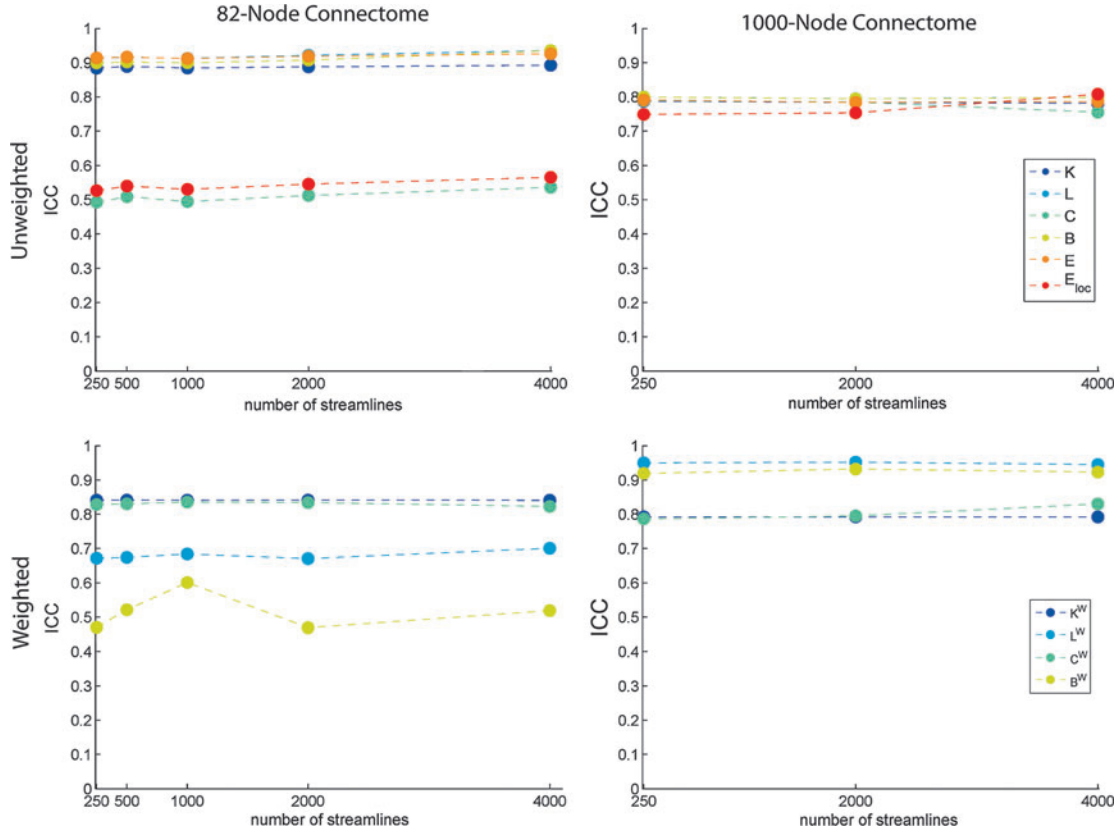

FIG. 7. The effect of number of streamlines initiated from every seed voxel on ICC for the unweighed and weighed graph metrics computed for the 82- and 1000-node connectomes.
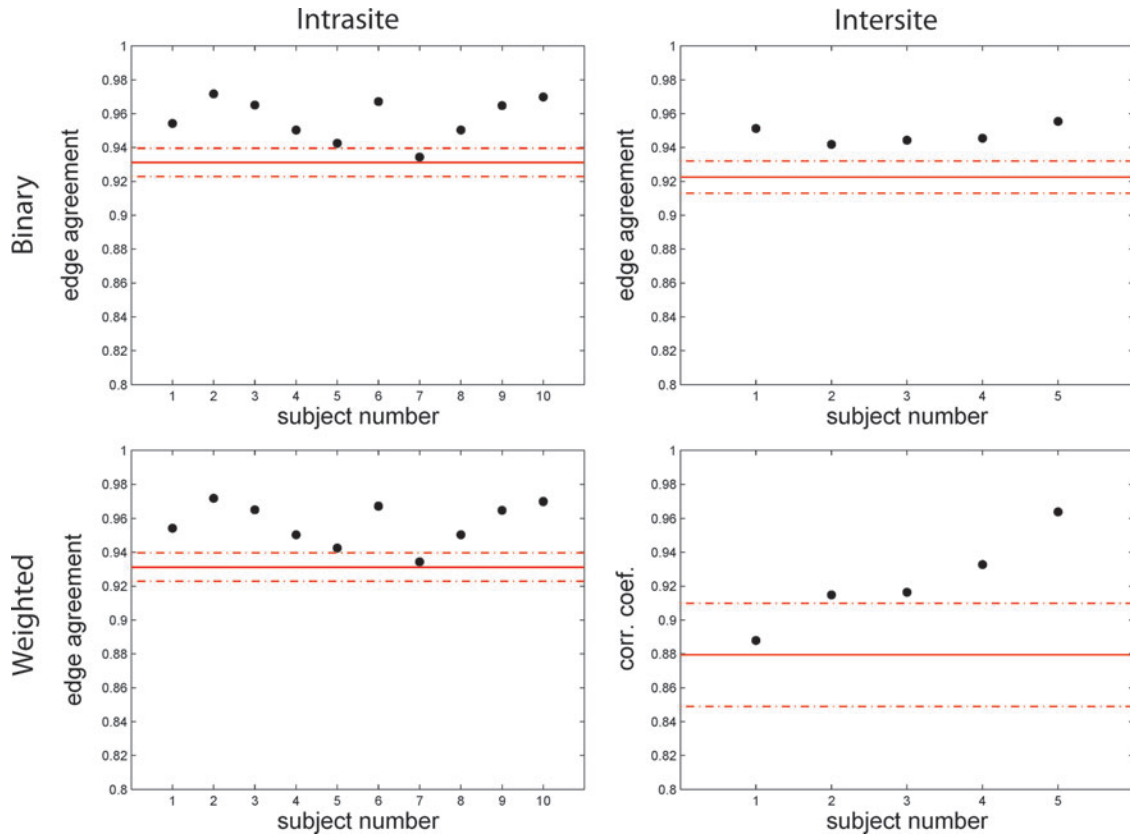
FIG. 8. Edge consistency for binary and weighed connectomes: the edge agreement and correlation coefficient are plotted for pairs of intrasite and intersite scans. In each plot, the black points denote the intrasubject variability, and the red line denotes the mean intersubject variability with a dotted line showing the standard deviation (SD).

observed for the binarized edge agreement, with less intra-subject variability than intersubject variability. In Fig. 9a, a histogram of the CV% values for the individual connection strengths is shown; most connection strengths have a CV% ranging from 10% to 40%, demonstrating that the connection strengths between pairs of nodes are less reliable than the summary graph metrics. In Fig. 9b, a scatter plot of CV% of the connection strengths and the step distance between

nodes are provided for an individual connectome; the correlation coefficient is 0.32 ($p < 0.000001$). This illustrates that there is a significant correlation between the reproducibility of connection strength and the distance along the white-matter tract between the two connected nodes, as inferred over all edges within a single connectome. On average, longer white-matter tracts tend to have less-reproducible connectivity than shorter white-matter tracts. However, the tract length
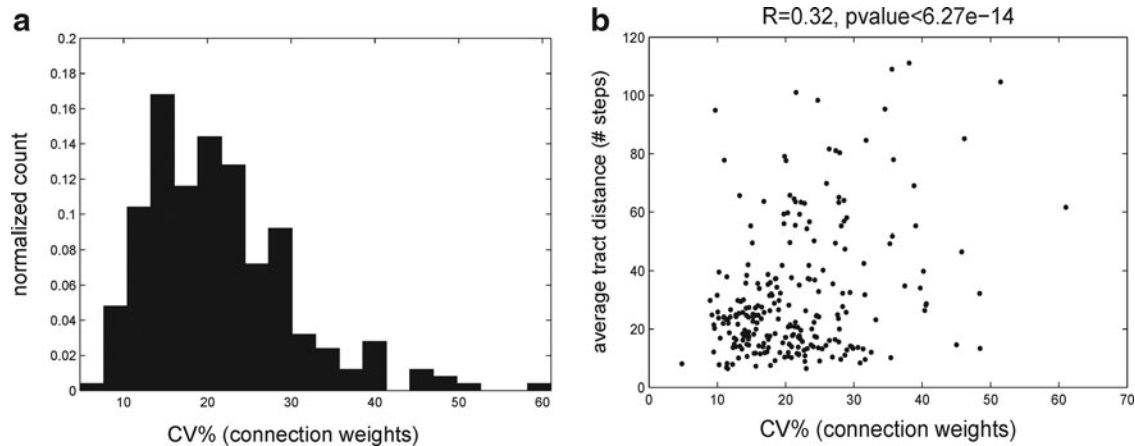

FIG. 9. Consistency of individual connection weights: **(a)** Histogram of the CV% for all suprathreshold connection weights and **(b)** correlation of CV% of node-to-node connection weights and the number of tractography steps between the two nodes
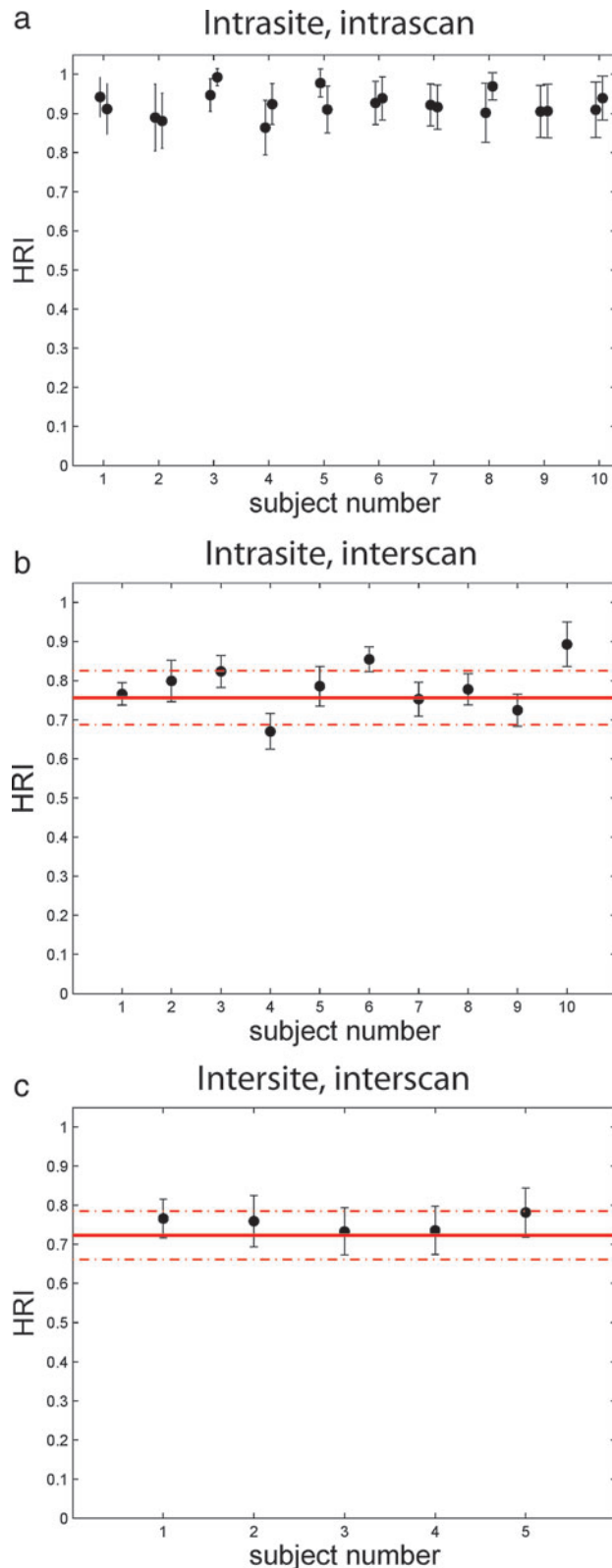
FIG. 10. Module assignment consistency: the mean Hubert Rand index (HRI) is plotted for **(a)** intrasite, intrascan data to measure the reliability of the community detection algorithm, **(b)** intrasite, interscan, and **(c)** intrasite, interscan. In **(b)** and **(c)**, the black points denote the intrasubject variability with SD error bars, and the red line denotes the mean intersubject variability with a dotted line for the SD.

accounts for <10% of the variance in connection strength reliability.

Test–retest reliability of modular organization

In Fig. 10, we present the results from the test–retest reliability analysis of the module assignments. In Fig. 10a, the mean and SD of the intrascan HRI are plotted for each 82-node connectome in the intrasite cohort. There is some variability in the reproducibility from scan to scan, but all HRI values are above 0.8, giving bounds for the amount of variability introduced by the community detection algorithm. We computed the mean and SD of HRI for random community assignments to be 0.44 ± 0.01. In Fig. 10b, the mean and SD of the interscan HRI are plotted for each intrasite subject, and the red line denotes the mean HRI between all intersubject pairs. All, but one, of the intrasubject HRI values are within one SD of the mean intersubject HRI. In Fig. 10c, we show the intersite data, plotted in the same fashion as the intrasite data. Here, we also find that the intrasubject HRI values are within one SD of the mean intersubject HRI.

Discussion

Methodological considerations and comparison with the prior literature

We have focused on the reproducibility and interindividual variability of the commonly used computational network metrics applied to the normal adult structural connectome, as reconstructed from a diffusion MR acquisition employing parameters in wide current use for clinical research (Mukherjee et al., 2008) and freely available postprocessing software such as FSL and FreeSurfer that are popular worldwide. Overall, we obtained the ICC values for the unweighed global graph metrics, such as K, L E, and B (Tables 1 and 3), which are in the range (ICC > 0.75) considered to be excellent reproducibility (Fleiss, 1986; Tooth et al., 2005). This compares favorably to all prior studies of the human structural connectome, which find only fair-to-good reproducibility (ICC = 0.40–0.75) of these global metrics (Bassett et al., 2011; Cheng et al., 2012; Dennis et al., 2012; Vaessen et al., 2010). Our results for reproducibility of the local network metrics, C and E_{loc} (Tables 1 and 3), are in the same fair-to-good range found by these prior reports. However, the dependence of these measures on connectivity threshold indicates that a different choice of threshold would produce an excellent reliability of these local metrics (Fig. 6). In addition, we report the CV% and edge-weight consistency values in agreement with those found previously (Bassett et al., 2011; Cammoun et al., 2012; Cheng et al., 2012; Dennis et al., 2012; Hagmann et al., 2008; Vaessen et al., 2010).

Since the graph analytics are computed from structural networks reconstructed from the diffusion data, our investigation evaluates the entire diffusion acquisition and connectome reconstruction pipeline. Our robust results might possibly be attributed to several advantages of the methodology employed herein: (1) a relatively short diffusion MR acquisition time, which limits motion artifacts; (2) HARDI reconstruction of crossing fibers, as opposed to a simpler DTI model; (3) probabilistic tractography, instead of a deterministic streamline tractography; (4) cortical and subcortical segmentation in the subjects' native space, and not a common

atlas space; and (5) a fixed connectivity threshold across subjects, rather than a common level of network sparsity across subjects. Although it would be ideal to systematically optimize the diffusion MR acquisition and connectome processing pipeline while accounting for all of these various factors, this is currently impractical because of the large parameter space involved and the considerable computational expense involved in reconstructing each structural connectome. Therefore, we can only present a subjective appraisal of the variables that might have affected the performance of our proposed methodology, compared to the existing literature.

Our diffusion imaging acquisition time is 8.5 min as compared to ~26 min for diffusion spectrum imaging as investigated in Bassett et al. (2011) and used exclusively in Cammoun et al. (2012). Arguably, a shorter scan time would produce more-reliable results, especially for clinical patients. From these diffusion MR data, we use bedpostx for HARDI reconstruction to estimate multiple tensors (up to 2) in every voxel allowing for sensitivity to crossing fibers, in contrast to the single-fiber approach used in Vaessen et al. (2010), Bassett et al. (2011), and Cheng et al. (2012). The use of probabilistic tractography could also potentially increase the reliability of the connectome reconstruction, as it was found to be more robust and sensitive than deterministic streamline tractography in an *ex vivo* mouse model (Moldrich et al., 2010). Among prior studies, only Vaessen et al. (2010) and Dennis et al. (2012) employ probabilistic tractography algorithms. We found that the reproducibility of the network metrics is consistent across an order of magnitude of the number of streamlines initiated, indicating that the choice to use 2000 streamlines per voxel is well within a stable regime (Fig. 7).

We segment the cortical and subcortical regions in the subjects' native space with FreeSurfer as opposed to registering each subject to a common atlas space. Due to the individual variability in cortical sulcation, using regions defined for each subject would likely lead to more reproducible results. Of the previous studies, Dennis et al. (2012) and Cammoun et al. (2012) also use FreeSurfer to generate the seed and target regions for the connectome, whereas the other studies use segmentation in a group atlas space. Further, our study is the only one to include the 7 subcortical areas segmented by FreeSurfer. To examine the reliability of the connectome reconstruction at a more fine-grained scale, we segment the GWB of each subject into 1000 cortical ROIs and process the connectomes entirely in the subjects' native space. Other articles have also addressed the reliability of high-resolution connectomes (Bassett et al., 2011; Cammoun et al., 2012), but their subsampled parcels are derived after registration to a template, rather than in the subject's native space.

Lastly, the technique used to binarize the connectomes, either fixing the connectivity threshold or fixing the sparsity across all subjects, could affect the intersubject variation. If a common sparsity is enforced across subjects, as done in Vaessen et al. (2010), Bassett et al. (2011), and Dennis et al. (2012), then it follows that the ICC values could be decreased due to reduced intersubject variation. We instead take the approach of a common threshold to allow variability in the network sparsity. We also investigate the effect of threshold on ICC and CV%, finding that the majority of graph metrics are reliable regardless of threshold at both the low- and high-parcellation scales. Cheng et al. (2012) also used a fixed threshold, although they found worse ICC values

than we do, likely due to the other aforementioned differences in the processing pipelines. The fact that we find the variability of these summary network metrics to be much smaller within subjects than between subjects suggests that the interindividual differences in connectome sparsity may be biologically meaningful. The sparsity values of the intrasite unweighed 82-node connectomes in our study have a mean of 0.84 ± 0.02 and range from 0.82 to 0.87. This range of network sparsity values is not wide enough to appreciably bias the comparisons of graph metrics (Anderson et al., 1999).

Reproducibility and intersubject variation in weighed low-resolution connectomes

In addition to analyzing binarized connectomes, we extend the evaluation of test–retest reliability to include weighed connectomes. For the 82-node connectomes, the intersubject variation of weighed network metrics is similar to that of their unweighed counterparts (Fig. 4). The ICC values of the weighed network metrics show them to have good-to-excellent reproducibility (Table 2). Regarding local network measures, the test–retest reliability of K^w is similar to that of K, and the reliability of C^w is better than that of its unweighed equivalent. The latter is because there is a greater intersubject variation in C^w than in C. However, the reproducibility of global network metrics such as L and B is considerably less for weighed than unweighed measures (Tables 1 and 2). This is because the intrasubject variability of the weighed global measures is higher than the unweighed measures (Fig. 4). Hence, for the 82-node connectomes, global network measures appear to be more reliable for binarized networks, whereas local network measures, excluding K, appear to be more reliable in weighed networks. However, choosing a different threshold for binarization can result in excellent reliability of the unweighed local graph metrics (Fig. 6), similar to that of their weighed counterparts.

Reproducibility and intersubject variation in high-resolution connectomes

Another direction encompassed by our study is to examine the test–retest reliability of connectomes at high-parcellation scales. In addition to providing more detailed maps of white-matter connectivity than low-resolution connectomes, the atlas-free segmentation is preferable for patients whose brain morphology may not fit the templates based on normal anatomy, such as those with mass lesions or structural anomalies. All network metrics were found to be highly reproducible for the intrasite 1000-node connectomes, including the unweighed and weighed measures, with strong statistical significance for ICC (Tables 3 and 4).

Reproducibility and intersubject variation of connectomes across sites

We also conducted a preliminary investigation in five subjects of test–retest connectome reliability across two different sites, one on each coast of the United States, to establish feasibility for multicenter studies. The results are qualitatively similar to those of the intrasite data, but, in general, slightly worse quantitatively (Tables 1–4). This resembles the findings of multicenter DTI studies that show a slightly lower reliability of intersite data than intrasite data for white-matter

microstructural measures such as FA (Heiervang et al., 2006; Pfefferbaum et al., 2003; Vollmar et al., 2010).

Consistency of network edges and modules in unweighed and weighed connectomes

The edge consistency, the most commonly analyzed measure of reliability for the connectome, is high for both the unweighed and weighed 82-node networks (Fig. 8). However, the reproducibility of individual node-to-node connection strengths (i.e., edge weights) is much lower than that of summary network metrics (Fig. 9a), in agreement with the results of Vaessen et al. (2010). Bassett et al. (2011), Cammoun et al. (2012), and Cheng et al. (2012) have all reported similar intrasubject correlation coefficients of ~0.9 and intersubject correlation coefficients from 0.6 to 0.8. Hagmann et al. (2008) also examined the test–retest repeatability of connection strengths in one subject and reported an intrasubject correlation coefficient of 0.78 and a mean intersubject correlation coefficient of 0.65. Bloy et al. (2012) recently introduced a new framework for computing structural connectivity from diffusion tractography data that produce more reliable connection strengths as well as more reproducible nodal connection distributions, an analog of node strength for weighed connectomes.

We found that individual edge weights are harder to reproduce across sites. This fact could explain the overall higher CV% and lower ICCs for the weighed metrics applied to the intersite scans.

Another interesting finding of our work is that longer tracts tend to have less-reproducible connection strengths than shorter tracts, although the effect of the tract length was small (Fig. 9b). The relationship between tract length and connection strength reliability may vary with the approach used to reconstruct the connectome, with possibly different effects of methods that seed only at the GWB versus those that seed throughout the white matter. This remains a direction for future research.

Network modules are generally consistent across healthy individuals both within and between sites, with slightly increased reproducibility for the intrasubject comparisons (Fig. 10). It is not expected that modular assignments would differ greatly across healthy adults at the level of granularity of the 82-node connectomes; therefore, it is reassuring that there is not a large effect of subject or site on the network partitions.

Limitations and future directions

The within-site test–retest interval used in this study assumes that the structural connectome is stable over 1 month in healthy adults. However, the interscan interval was, on average, twice as long for the intersite data as for the intrasite data. This could contribute to the increased variation observed in the intersite data compared to that within a site, although the intersite data still demonstrate precision within an acceptable range for clinical research.

We have kept the scanner model and acquisition parameters constant for all scans; therefore, this work does not investigate of the effect of an MRI scanner model or scanning parameters on the connectome. Vaessen et al. (2010) investigated the effect of diffusion directions on summary network metrics. Future work is warranted to study the effect of such variables as scanner model, spatial resolution, and tractography algorithm on the reproducibility of graph metrics.

One drawback to the 1000-node connectomes is that many of the analyses performed on the 82-node connectomes, such as module assignments, overall edge consistency, and the consistency of individual connection weights, cannot be adapted to the 1000-node connectomes, since these smaller nodes cannot be mapped one to one between scans. Registering the gray–white matter boundary to a template would allow for identical parcellations across scans. Using a template, though, would not be atlas free, which we feel is a strong advantage for the clinical populations whose brains may not fit established norms. The 1000-node connectome contains only cortical nodes; a future direction will be to include subcortical regions in the high-resolution connectomes, which require parcellating the gray–white matter boundary for the subcortical structures.

Conclusion

Taken as a whole, the results presented here further validate the connectome framework as having sufficient precision for clinical research using common diffusion scan parameters on a modern 3T system. Most network metrics are highly reproducible and could be employed to track the progression of disease or to assess the effectiveness of a course of treatment, if it can be shown that they have sufficient sensitivity to the underlying pathology and to the effect of therapy. Given the compelling results of reliability studies and initial investigations in clinical populations, the evidence points toward MR connectomics having a major impact on the scientific understanding and diagnostic evaluation of neurological and psychiatric disorders for which MRI does not currently have clinical utility.

Acknowledgments

Author Disclosure Statement

References

Anderson BS, Butts C, Carley K. 1999. The interaction of size and density with graph level indices. Social Networks 21:239–267.
Bassett D, Brown J, Deshpande V, Carlson Jean, Grafton S. 2011. Conserved and variable architecture of human white matter connectivity. Neuroimage 54:1262–1279.

Behrens TEJ, Berg HJ, Jbabdi S, Rushworth MF, Woolrich MW. 2007. Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? Neuroimage 34:144–155.

Blondel V, Guillame JL, Lambiotte R, Lefebvre E. 2008. Fast unfolding of communities in large networks. J Stat Mech 2008:P10008.

Bloy L, Ingalhalikar M, Batmanghelich N, Schultz R, Roberts TP, and Verma R. 2012. An integrated framework for high angular resolution diffusion imaging-based investigation of structural connectivity. Brain Connect 2:69–79.

Bullmore E, Sporns O. 2009. Complex brain networks: graph theoretical analysis of structural and functional systems. Nat Rev Neurosci 10:186–198.

Cammoun L, Gigandet X, Meskaldji D, Thiran JP, Sporns O, Do K, Maeder P, Meuli R, Hagmann P. 2012. Mapping the human connectome at multiple scales with diffusion spectrum MRI. J Neurosci Methods 203:386–397.

Cheng H, Wang Y, Sheng J, Kronenberger W, Mathews V, Hummer T, Saykin A. 2012. Characteristics and variability of structural networks derived from diffusion tensor imaging. Neuroimage 61:1153–1164.

Dennis E, Jahanshad N, Toga A, McMahon K, Zubicaray G, Martin N, Wright M, Thompson P. 2012. Test-retest reliability of graph theory measures of structural brain connectivity. MICCAI 2012, Part III, LNCS 7512:305–312.

Fan Y, Shi F, Smith JK, Lin W, Gilmore JH, Shen D. 2011. Brain anatomical networks in early human brain development. Neuroimage 54:1862–1871.

Fischl B, van der Kouwe A, Destrieux C, Halgren E, Segonne F, Salat DH, et al. 2004. Automatically parcellating the human cerebral cortex. Cereb Cortex 14:11–22.

Fleiss JL. 1986. *The Design and Analysis of Clinical Experiments*. New York: John Wiley & Sons.

Fornito A, Bullmore ET. 2012. Connectomic intermediate phenotypes for psychiatric disorders. Front Psychiatry 3:32.

Gong G, He Y, Concha L, Lebel C, Gross D, Evans A, et al. 2009. Mapping anatomical connectivity patterns of human cerebral cortex using *in vivo* diffusion tensor imaging tractography. Cereb Cortex 19:524–536.

Hagmann P, Cammoun L, Gigandet X, Gerhard S, Grant PE, Wedeen V, et al. 2010. MR connectomics: principles and challenges. J Neurosci Methods 194:34–45.

Hagmann P, Cammoun L, Gigandet X, Meuli R, Honey C, Wedeen V, et al. 2008. Mapping the structural core of human cerebral cortex. PLoS Biol 6:e159.

Hagmann P, Grant PE, Fair DA. 2012. MR connectomics: a conceptual framework for studying the developing brain. Front Syst Neurosci 6:43.

Hagmann P, Kurant M, Gigandet X, Thiran P, Wedeen VJ, Meuli R, et al. 2007. Mapping human whole-brain structural networks with diffusion MRI. PLoS One 2:e597.

Heiervang E, Behrens TEJ, Mackay CE, Robson MD, Johansen-Berg H. 2006. Between session reproducibility and between subject variability of diffusion MR and tractography measures. Neuroimage 33:867–877.

Hubert LJ, Baker FB. 1977. The comparison and fitting of given classification schemes. J Math Psychol 16:233–253.

Irimia A, Chambers M, Torgerson C, Filippou M, Hovda D, Alger J, et al. 2012. Patient-tailored connectomics visualization for the assessment of white matter atrophy in traumatic brain injury. Front Neurol 3:10.

Iturria-Medina Y, Canales-Rodríguez EJ, Melie-García L, Valdés-Hernández PA, Martínez-Montes E, Alemán-Gómez Y, et al. 2007. Characterizing brain anatomical connections using diffusion weighted MRI and graph theory. Neuroimage 36:645–660.

Jenkinson M, Bannister P, Brady M, Smith S. 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17:825–841.

Kuceyeski A, Maruta J, Niogi SN, Ghajar J, Raj A. 2011. The generation and validation of white matter connectivity importance maps. Neuroimage 58:109–121.

Lachin J. 2004. The role of measurement reliability in clinical trials. Clin Trials 1:553–566.

Li L, Rilling JK, Preuss TM, Glasser MF, Damen FW, Hu X. 2012a. Quantitative assessment of a framework for creating anatomical brain networks via global tractography. Neuroimage 61:1017–1030.

Li L, Rilling JK, Preuss TM, Glasser MF, Hu X. 2012b. The effects of connection reconstruction method on the interregional connectivity of brain networks via diffusion tractography. Hum Brain Mapp 33:1894–1913.

Moldrich RX, Pannek K, Hoch R, Rubenstein JL, Kurniawan ND, Richards LJ. 2010. Comparative mouse brain tractography of diffusion magnetic resonance imaging. Neuroimage 51:1027–1036.

Mukherjee P, Chung SW, Berman JI, Hess CP, Henry RG. 2008. Diffusion tensor MR imaging and fiber tractography: technical considerations. AJNR Am J Neuroradiol 29:843–852.

Owen JP, Li YO, Ziv E, et al. 2013. The structural connectome of the human brain in agenesis of the corpus callosum. Neuroimage 70:340–355.

Pfefferbaum A, Adalsteinsson E, Sullivan E. 2003. Replicability of diffusion tensor imaging measurements of fractional anisotropy and trace in brain. J Magn Reson Imaging 18:427–433.

Rosen BR, Napadow V. 2011. Quantitative markers for neuropsychiatric disease: Give it a rest. Radiology 259:17–19.

Rubinov M, Sporns O. 2010. Complex network measures of brain connectivity: uses and interpretations. Neuroimage 52:1059–1069.

Seung S. 2012. *Connectome: How the Brain's Wiring Makes Us Who We Are*. New York: Houghton Mifflin Harcourt.

Shu N, Liang Y, Li H, Zhang J, Li X, Wang L, et al. 2012. Disrupted topological organization in white matter structural networks in amnestic mild cognitive impairment: relationship to subtype. Radiology 265:518–527.

Shu N, Liu Y, Li J, Yonghui L, Chunshui Y, Jiang T. 2009. Altered anatomical network in early blindness revealed by diffusion tensor tractography. PLoS One 4:e7228.

Smith S. 2002. Fast robust automated brain extraction. Hum Brain Mapp 17:143–155.

Sporns O. 2011. The human connectome: a complex network. Ann NY Acad Sci 1224:109–125.

Sporns O, Tononi G, Kotter R. 2005. The human connectome: a structural description of the human brain. PLoS Comput Biol 1:e42.

Tooth L, Bennett S, McCluskey A, Hoffmann T, McKenna K, Lovarini M. 2005. Appraising the quality of randomized controlled trials: inter-rater reliability for the OTseeker evidence database. J Eval Clin Pract 11:547–555.

Tymofiyeva O, Hess C, Ziv E, Tian N, Bonifacio S, McQuillen P, et al. 2012. Towards the baby connectome: mapping the structural connectivity of the newborn brain. PLoS One 7:e31029.

Vaessen MJ, Hofman PAM, Tijssen HN, Aldenkamp AP, Jansen JFA, Backes WH. 2010. The effect and reproducibility of different clinical DTI gradient sets on small world brain connectivity measures. Neuroimage 51:1106–1116.

van den Heuvel MP, Sporns O. 2011. Rich club organization of the human connectome. J Neurosci 31:15775–15786.

Verstraete E, Veldink JH, Mandl RCW, van den Berg LH, van den Heuvel MP. 2011. Impaired structural motor connectome in amyotrophic lateral sclerosis. PLoS One 6:e24239.

Vollmar C, O'Muircheartaigh J, Barker G, Symms M, Thompson P, Kumari V, et al. 2010. Identical, but not the same: Intra-site and inter-site reproducibility of fractional anisotropy measures on two 3.0 T scanners. Neuroimage 51:1384–1394.

Yan C, Gong G, Wang J, Wang D, Liu D, Zhu C, et al. 2011. Sex- and brain size-related small-world structural cortical networks in young adults: a DTI tractography study. Cereb Cortex 21:449–458.

Yap PT, Fan Y, Chen Y, Gilmore JH, Lin W, Shen D. 2011. Development trends of white matter connectivity in the first years of life. PLoS One 6:e24678.

Zalesky A, Fornito A, Harding IH, Cocchi L, Yücel M, Pantelis C, et al. 2010. Whole-brain anatomical networks: does the choice of nodes matter? Neuroimage 50:970–983.

Address correspondence to:
Pratik Mukherjee
Department of Radiology and Biomedical Imaging
Center for Molecular and Functional Imaging
University of California, San Francisco
UCSF Box 0946
185 Berry Street, Suite 350
San Francisco, CA 94107

E-mail: pratik.mukherjee@ucsf.edu