NeuroImage 145 (2017) 118–129

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Consistency-based thresholding of the human connectome

James A. Roberts[a,b,*], Alistair Perry[a,c,d], Gloria Roberts[d,e], Philip B. Mitchell[d,e], Michael Breakspear[a,b,f]

[a] Systems Neuroscience Group, QIMR Berghofer Medical Research Institute, Herston, QLD 4006, Australia
[b] Centre for Integrative Brain Function, QIMR Berghofer Medical Research Institute, Herston, QLD 4006, Australia
[c] Centre for Healthy Brain Ageing (CHeBA), School of Psychiatry, University of New South Wales, Sydney, NSW, Australia
[d] School of Psychiatry, University of New South Wales, Sydney, NSW 2052, Australia
[e] Black Dog Institute, Prince of Wales Hospital, Hospital Road, Randwick, NSW 2031, Australia
[f] Metro North Mental Health Service, Royal Brisbane and Women's Hospital, Brisbane, QLD, Australia

ABSTRACT

Densely seeded probabilistic tractography yields weighted networks that are nearly fully connected, hence containing many spurious fibers. It is thus necessary to prune spurious connections from probabilistically-derived networks to obtain a more reliable overall estimate of the connectivity. A standard method is to threshold by weight, keeping only the strongest edges. Here, by measuring the consistency of edge weights across subjects, we propose a new thresholding method that aims to reduce the rate of false-positives in group-averaged connectivity matrices. Close inspection of the relationship between consistency, weight, and distance suggests that the most consistent edges are in fact those that are strong for their length, rather than simply strong overall. Hence retaining the most consistent edges preserves more long-distance connections than traditional weight-based thresholding, which penalizes long connections for being weak regardless of anatomy. By comparing our thresholded networks to mouse and macaque tracer data, we also show that consistency-based thresholding exhibits the species-invariant exponential decay of connection weights with distance, while weight-based thresholding does not. We also show that consistency-based thresholding can be used to identify highly consistent and highly inconsistent subnetworks across subjects, enabling more nuanced analyses of group-level connectivity than just the mean connectivity.

Introduction

Tractography is a widely-used method for inferring white matter connectivity from diffusion imaging data, and is central to the field of connectomics (Fornito et al., 2015). Various algorithms use voxel-based estimates of water diffusion to infer the likely paths of white matter bundles, either deterministically or probabilistically (Behrens et al., 2003; Descoteaux et al., 2009). Regardless of the method, there is uncertainty over which connections are "true" connections and which are spurious (de Reus and van den Heuvel, 2013; Girard et al., 2014; Smith et al., 2012). In particular, while deterministic tractography has high specificity at the cost of sensitivity to crossing fibers and hence has a high rate of false negatives, probabilistic algorithms yield inherently noisy connection matrices, at least at the single subject level, and hence likely contain numerous false positives (Thomas et al., 2014).

Pooling data over subjects is a common way to reduce the signal to noise ratio, such as by averaging connectivity matrices across subjects (Hagmann et al., 2008; Perry et al., 2015), or determining a consensus connectivity by selecting edges that appear in at least some fraction of the subjects (de Reus and van den Heuvel, 2013; van den Heuvel and Sporns, 2011). Seeking a consensus in this way is problematic for networks derived from densely-seeded probabilistic tractography, where all individual subject-wise networks are densely connected. The most common method in this setting is to "threshold" networks to some desired density by keeping only the strongest links (Rubinov and Sporns, 2010). This method is applicable to dense networks, but it is not at all clear that the strongest links are always the most accurate for inferring white matter connectivity (Gigandet et al., 2008).

Besides reducing spurious connections, thresholding connectivity matrices also plays an important role in graph-based characterization of connection topology (Bullmore and Bassett, 2011; Van Wijk et al., 2010). Thresholding is used to determine binary adjacency matrices associated with weighted networks, enabling use of the full armory of graph-theoretic tools for unweighted networks (Rubinov and Sporns, 2010). Thresholding can also be used to identify subnetworks composed of the strongest (or weakest) edges, whether for analysis of these

* Corresponding author at: Systems Neuroscience Group, QIMR Berghofer Medical Research Institute, Herston, QLD 4006, Australia.
E-mail address: james.roberts@qimrberghofer.edu.au (J.A. Roberts).

http://dx.doi.org/10.1016/j.neuroimage.2016.09.053
Received 16 March 2016; Accepted 21 September 2016
Available online 22 September 2016
1053-8119/

subnetworks or simply for ease of visualization.

In this technical note, we propose a hybrid thresholding method that seeks a group consensus connectivity by thresholding the averaged network to retain only those connections whose weights are the most consistent across the group. Unlike the popular weight-based thresholding, our consistency-based thresholding pays heed to the within-group intersubject variability when deriving a group-averaged matrix. We show that our consistency-based approach avoids the "hard threshold edge" imposed by traditional thresholding and preserves the role of long-range connections. We also study the influence of thresholding strategy on the network topology of the ensuing structural connectome.

Methods

We derived estimates of whole brain structural connectivity from diffusion images of 75 healthy subjects (aged 17–30 years, 47 females). The structural connectivity matrices were derived in a recent study; we briefly present the methods here but see Roberts et al. (2016) for full details on the acquisition and tractography details.

Diffusion MRI data were acquired from all participants on a Philips 3 T Achieva Quasar Dual MRI scanner (Philips Medical System, Best, The Netherlands) using a single-shot echo-planar imaging (EPI) sequence (TR=7767 ms, TE=68 ms). For each diffusion scan, 32 gradient directions (b=1000 s/mm²) and a non-diffusion-weighted acquisition (b=0 s/mm²) were acquired over a 96×96 image matrix (field of view 240 mm×240 mm×137.5 mm), with a slice thickness of 2.5 mm and no gap, reconstructed to yield 1 mm×1 mm×2.5 mm voxels (where the longer dimension is along the dorsoventral axis). Two sets of diffusion scans were acquired for each subject.

We employed a probabilistic streamline algorithm (Tournier et al., 2012) to generate high-resolution whole-brain fiber tracks. The fiber orientation distribution (FOD) within each voxel was estimated using MRtrix software (Tournier et al., 2012) by performing constrained spherical deconvolution (Tournier et al., 2007) with a maximum spherical harmonic order (*lmax*) of 6. As an intermediate step to constrain the spherical deconvolution, a single-fiber response kernel was estimated from all white matter voxels with fractional anisotropy FA > 0.7. Streamlines were seeded using the skull-stripped brain mask together with a restriction to voxels with FOD amplitude > 0.1. Streamlines will not start outside this region and terminate if they reach the boundary. Tractograms were generated using a probabilistic streamlines algorithm (Tournier et al., 2012), which produces a set of connection trajectories by randomly sampling from the orientation uncertainty inherent in each FOD along the streamline paths. Although non-isotropic voxels were used within the analysis, we subsequently checked the resultant fiber orientations and tractograms, finding no issue with potential biases on quality. To confirm this we repeated our analysis in an independent dataset acquired with isotropic voxels (Supplementary information 1).

Our connectivity matrices were reconstructed from densely seeded tractography (10^8 seeds) and parcellated into a relatively fine representation of 513 uniformly sized cortical and sub-cortical regions (Zalesky et al., 2010). The resulting weighted, undirected matrices were nearly fully connected in each subject. The weights are the number of streamlines linking each pair of regions. The spatial connection distance between all nodes was obtained using a streamline-based quantification of distance, in addition to the more traditionally-used Euclidean distance. In six subjects, for each pathway in the connectome, the shortest streamline between the node pair was found, and its length determined; these minimum lengths were then averaged across the six subjects. The minimum streamline length per subject was used here as streamlines are more likely to erroneously continue beyond the length of the connecting pathway (and hence overestimate the actual connection length) rather than provide an erroneously short pathway. In previous work (Roberts et al., 2016) it was observed that the length distribution converged after averaging over a small number of subject streamline lengths. Indeed, the precise details of the length distribution do not influence our results, as demonstrated in an independent dataset where we used every individual's own set of streamlines (Supplementary information 1).

This combination of streamline generation and anatomical parcellation yields a weighted structural connectivity graph within each subject, which we denote W, and a corresponding matrix of streamline lengths, F. Within each W, a weighted connection w_{ij} represents the number of streamlines from region i terminating within a 2 mm radius of region j, with corresponding streamline length f_{ij}. The 2 mm radius ensures that fiber terminations near the gray-matter boundary, where the diffusion signal becomes noisier and weaker, are adequately captured. While this could theoretically lead to streamlines being counted twice, our 513 node parcellation is sufficiently coarse that this occurs very infrequently. To improve signal to noise ratio, the corresponding w_{ij} were summed across each subject's two diffusion scans. The non-directional nature of tractography implies that W is symmetric – that is $w_{ij}=w_{ji}$. The larger number of likely random seeds located along longer fiber bundles is well known to result in over-defined fiber densities (Smith et al., 2013). To reduce this confounding effect, w_{ij} were adjusted by dividing the raw count by the streamline length f_{ij} between nodes i and j, $w_{ij} \rightarrow w_{ij}/f_{ij}$ (cf. Hagmann et al., 2008). That is, because we seed the white matter uniformly, a tract that is twice as long will have received twice as many seeds. Its weight (streamline count) will thus have been biased relative to a shorter tract with the same true fiber density.

We note that the very dense seeding of our probabilistic tractography yields connectivity matrices that are fully connected (or very nearly fully connected) in all subjects. Whilst the biological connectome at this level of resolution is likely not fully connected, estimates of connection density in the field range very broadly, from < 5% (Hagmann et al., 2008) to 13–36% for the entire brain and 32–52% for cortico-cortical connections in the mouse (Oh et al., 2014). Seeding probabilistic tractography densely, then setting a post-hoc threshold, allows investigation of topology over a range of connection densities, and not that dictated by the acquisition and reconstruction technique.

We estimated the consistency of every edge weight by measuring the coefficient of variation across subjects. We then compared networks thresholded by weight to networks thresholded by consistency. To quantify differences in network topology, we calculated graph metrics (clustering, rich club, modularity) using the Brain Connectivity Toolbox (Rubinov and Sporns, 2010).

Results

We begin by characterizing the consistency of edge weights across subjects. We quantify consistency by calculating the coefficient of variation (CV) of the weights (SD/mean) across subjects. A low CV corresponds to a high consistency. Edge CV broadly decreases with increasing edge weight (Fig. 1), thus showing that consistency increases with edge weight. This is the widely-assumed justification for weight-based thresholding; here we verify that the very strongest edges are indeed the most consistent. However, this inverse relationship between CV and weight does not follow a simple linear trend. Rather, edges with weights two orders of magnitude below the strongest are almost as consistent.

This slow fall-off of consistency with weight is particularly clear when taking the spatial dimension into account. Weights decrease roughly exponentially with streamline length as shown previously (Roberts et al., 2016). However, CV increases more slowly with distance, such that some of the longest connections are in fact the most consistent (Fig. 2A). In a sense this is unsurprising: true long connections would be expected to have consistent weights between subjects. Grouping the edges by quartiles in consistency (Fig. 2A–D) shows that each quartile's cloud of points spans almost the full range of

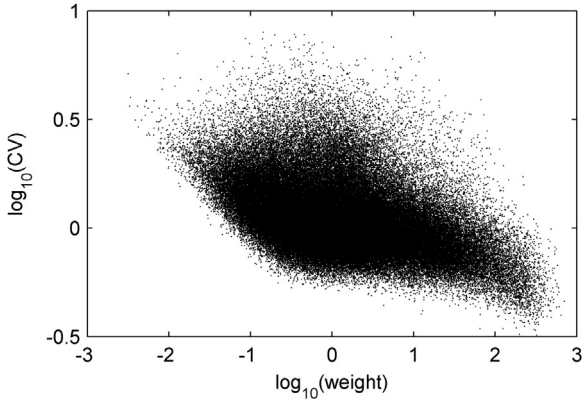

Fig. 1. Consistency of edges across subjects versus weight. Note double logarithmic coordinates. Points with low coefficient of variation (CV) are the most consistent.

streamline lengths – there are both long and short consistent connections and long and short inconsistent connections. The main difference between the quartiles is that the lower-consistency edges tend to be of lower weight relative to the background distance effect (Fig. 2E).

These differences between weight and consistency are important for deciding which edges one should keep when thresholding. Typical thresholding methods select edges on the basis of weight alone (Fig. 3A), preserving the edges with the highest weights up to a desired network density. For very sparse connectivity (< 1% density) the chosen edges are indeed highly likely to correspond to true connections, but these will also be very short connections. Similarly, the very weakest connections are also likely to be the most unreliable. But in between these two extremes there is much more variability than is widely appreciated. For moderate densities it is unlikely that the strongest connections will always be the most anatomically accurate. If we instead threshold by consistency – preserving the edges with the lowest CV up to a desired network density – the set of retained edges

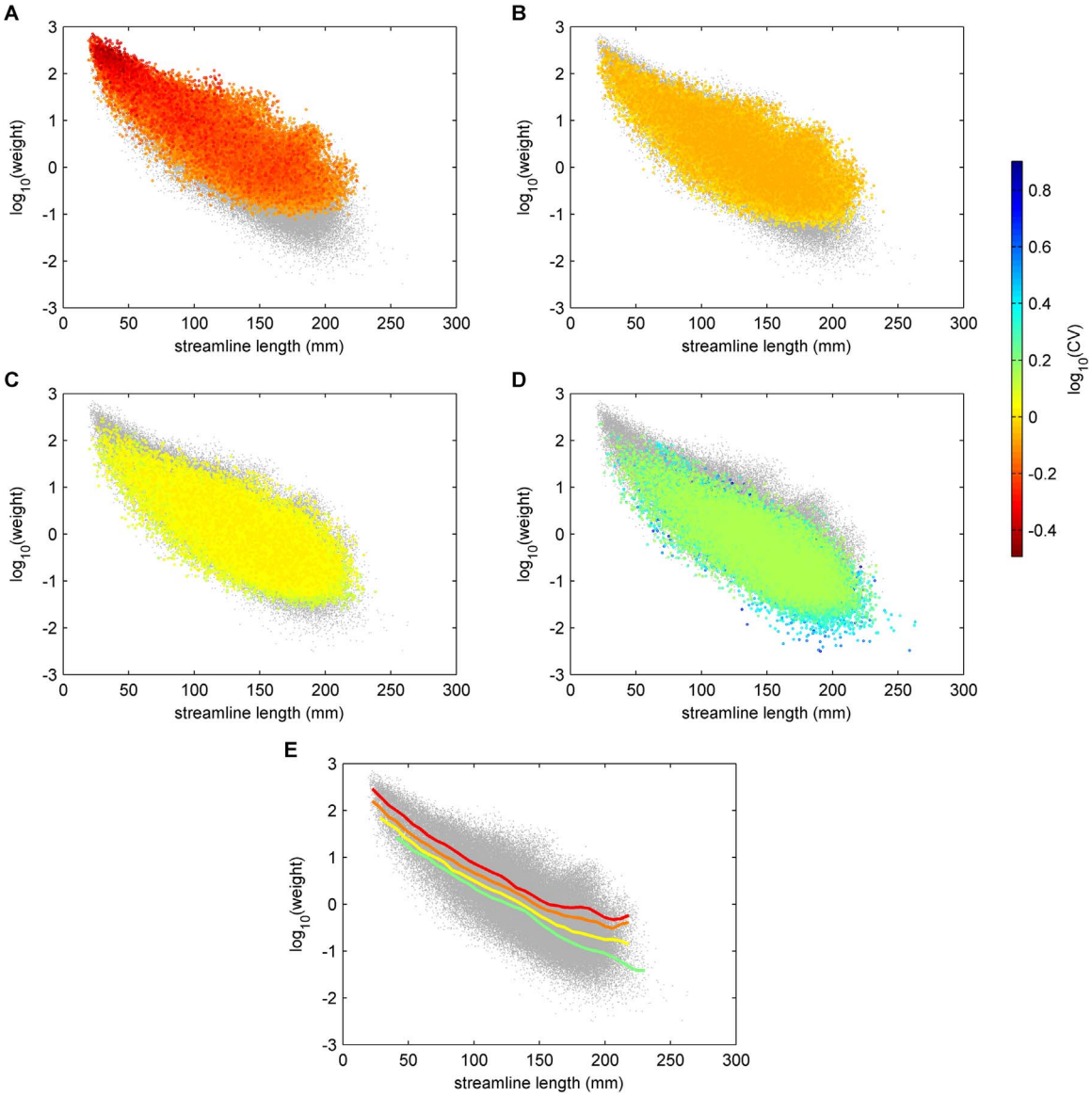

Fig. 2. Consistency of edge weights as a function of weight and streamline length. (A–D) Edges grouped by consistency in decreasing quartiles 1–4, respectively, with colored points showing the coefficient of variation (CV) of edges in the corresponding quartile, and gray points showing the overall set of edges. Low CV=more consistent, and warmer colors overplot cooler colors. (E) Mean weight versus streamline length for the four quartiles (red, orange, yellow, and green in decreasing order of consistency), calculated using 40 uniform bins (plotted where these bins contain at least 50 edges). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

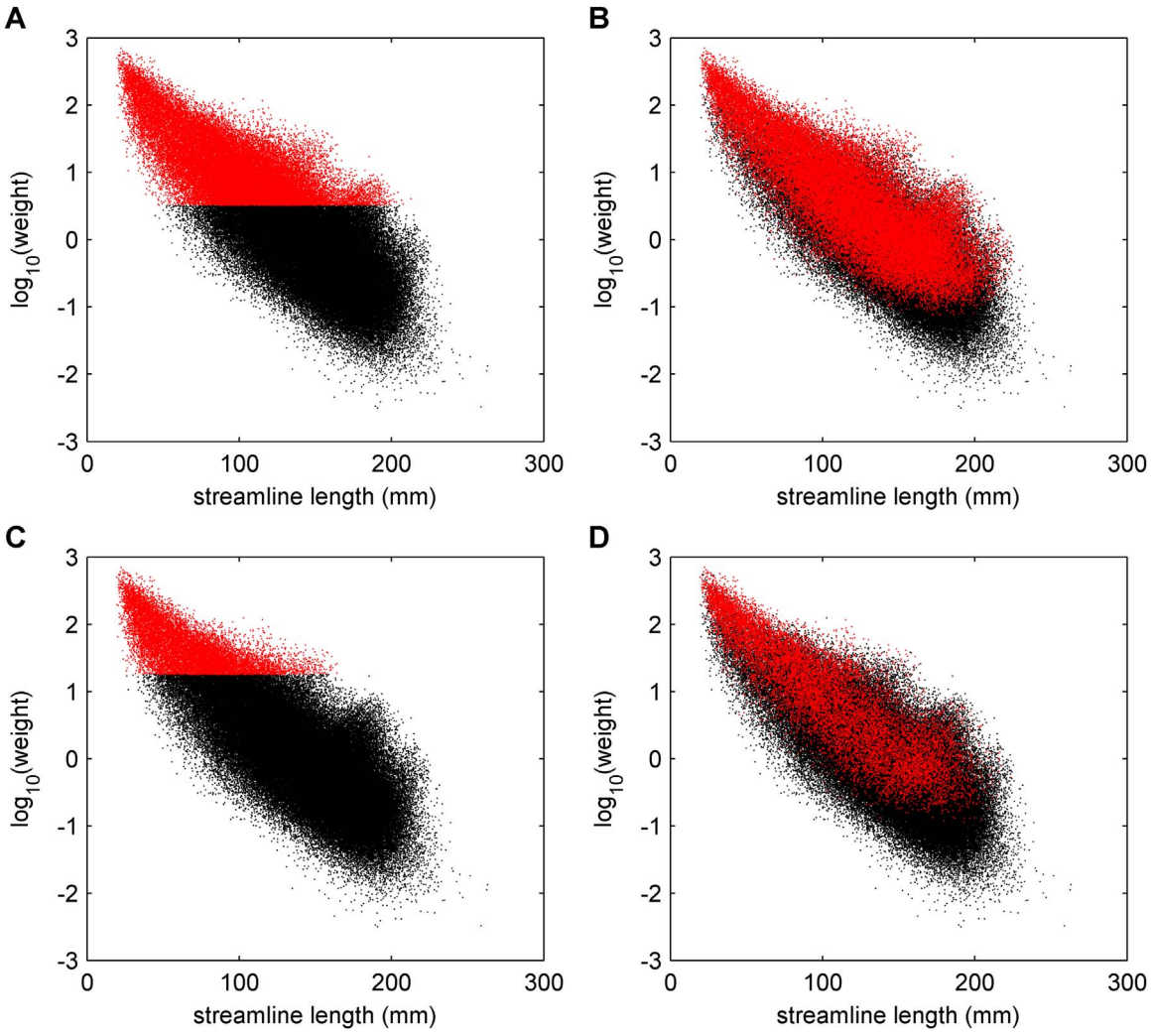

Fig. 3. Comparison of thresholding methods. (A) Effect of weight-based thresholding on the connectome's weight-distance relationship at 30% density, showing all edges retained (red) and discarded (black). (B) Consistency-based thresholding at 30% density. (C, D) Threshold comparison for 10% density. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

includes more numerous long fibers (Fig. 3B, red). This is true also for sparser connectivity, e.g. top 10% (Fig. 3C and D). Implementation of this consistency-based thresholding is a straightforward modification of the standard method for weight-based thresholding (Rubinov and Sporns, 2010), ranking edges by consistency rather than weight. We provide MATLAB code at http://www.sng.org.au/Downloads. The code can be trivially extended to threshold by any desired measure beyond our chosen coefficient of variation; another example could be the quartile coefficient of dispersion.

Inspection of the clouds of weights versus distance (Figs. 2 and 3) shows that the consistent edges tend to be those edges that are strong for their length; conversely, the inconsistent edges are weak for their length. Differences between the two thresholding methods are stark in the distributions of edge weights (Fig. 4): the standard edge threshold yields a very sharp (and arbitrary) cut-off in the relationship between weights and streamline lengths (Fig. 4A, red), which does not arise when using the consistency threshold (Fig. 4A, green), which retains a greater proportion of long edges (Fig. 4B).

The preservation of more numerous long range edges by the consistency-based thresholding leads to more numerous inter-hemispheric connections: these are likely under-represented after weight thresholding due to their lengths. This difference in inter-hemispheric density is apparent in the connectivity matrices (Fig. 5). For weight-based thresholding (to 30% overall density) the inter-hemispheric connections (off-diagonal block in Fig. 5A) make up 13% of the edges after thresholding, versus 32% for consistency-based thresholding (Fig. 5B). Thus the notion of simply retaining the strongest edges should arguably be modified to take into account the fact that the brain is a spatial network. Longer edges are inherently weaker, partly due to anatomy (Henderson and Robinson, 2014; Roberts et al., 2016) and partly due to biases in tractography algorithms such as the accumulation of errors that cause streamlines to veer off course (Li et al., 2012; Zalesky, 2008; Zalesky and Fornito, 2009). Since randomly-accumulated errors are unlikely to accumulate in the same away across subjects, this suggests that consistency-based thresholding likely mitigates the influence of distance-dependent biases.

The relationship between consistency and the connection weight-length relationship can be further quantified by estimating the probability that a connection of a given weight and length is consistent. A simple way to estimate consistency probability as a function of connection weight and length is to first estimate the density of all edges by calculating a 2-D histogram (Fig. 6A), then apply a threshold (here 30% density) by consistency and count the number of surviving edges (Fig. 6B). The density of the points in the full network (Fig. 6A) shows that the apparently-uniform dense cloud of points in Figs. 2 and 3 obscures a peak ~2/3 of the way along the cloud. The density of consistent points (Fig. 6B) exhibits a peak at short lengths (as expected from Fig. 2) and is relatively uniform otherwise, and shifted slightly

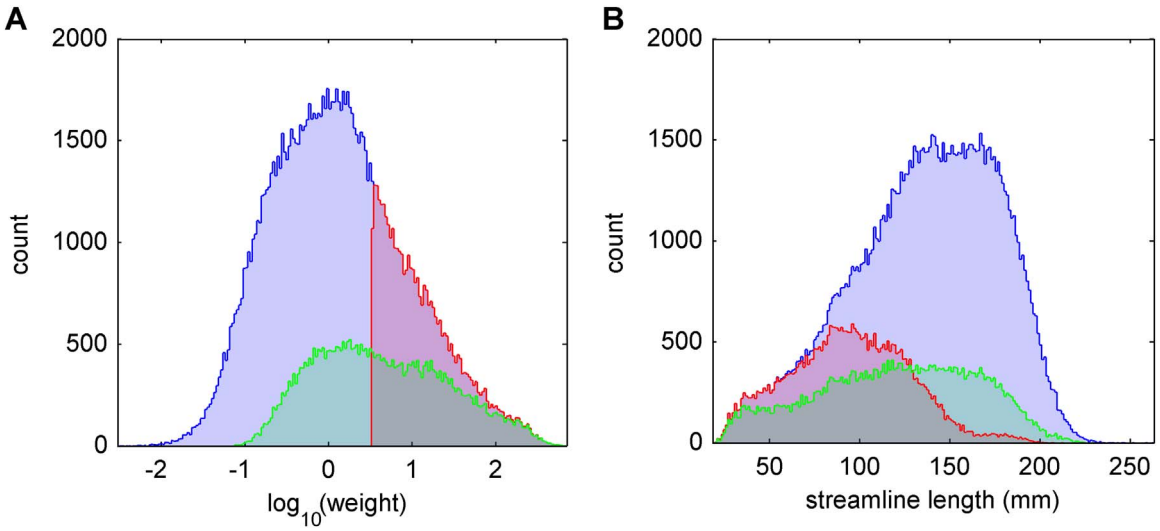

Fig. 4. (A) Distributions of weights for unthresholded (blue), weight-thresholded (red), and consistency-thresholded (green) networks. (B) Corresponding streamline length distributions. Networks thresholded to 30% density. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

toward higher weights. The ratio of these two sets of counts is thus an estimate of the probability of a connection being consistent (Fig. 6C). We find that the probability of consistency increases with weight but does not fall off with distance as quickly as weight does. This result thus quantifies the fact that edges that are strong for their length are most likely to be consistent between subjects.

It is important to note that all thresholding methods that depend on the weights will depend on the definition of these weights. Another method for estimating connection weights is the mean FA along the connection. Analysis of this quantity using the elderly dataset (Supplementary information 1, Figure S1.2) reveals that mean FA increases with length. For long high-FA connections, consistency does not depend strongly on FA or distance. For medium- and short-range connections, higher-FA connections have higher consistency, in agreement with our finding for streamline counts.

Network structure

We next studied the influence of threshold strategy on the network structure of the ensuing group structural connectome. Weight-based and consistency-based thresholding yield networks that share many of the same hub regions (hubs are regions that are densely connected to other regions), but show substantial differences in regional connectivity, largely reflecting differences in the long fibers (Fig. 7A and B). In particular, the consistency-thresholded network includes denser inter-hemispheric and caudal-rostral connections. Both thresholded networks contain high-degree hub nodes in posterior, temporal, frontal, and midline areas, though the consistency-thresholded networks contain denser clusters in these last two regions. The picture is also similar when studying rich nodes by strength (Fig. 7C and D). Again there is greater inter-hemispheric connectivity in the consistency-thresholded networks than the weight-thresholded ones. In this case, however, the positions of the hub nodes are similar for the two threshold methods. The differences are less dramatic because the node strengths are dominated by the strongest edges, which tend to be retained by both thresholding methods, as compared to the binary case in which all the edges are weighted equally.

To further elucidate the similarities and differences between these networks we partitioned edges into three subsets: strong-consistent, strong-inconsistent, and weak-consistent. This analysis also demonstrates how consistency-based thresholding can be combined with other thresholding methods to uncover subtle features of connectivity. We focus on those edges that are strong (top 5% by weight), consistent (top 5% by consistency), weak (bottom 87% by weight), and inconsistent (bottom 87% by consistency) – these values are chosen such that the intersections strong-consistent, weak-consistent, and strong-inconsistent all have density 2%. The strong-consistent network (Fig. 8A) consists of nodes and edges that are primarily at superficial

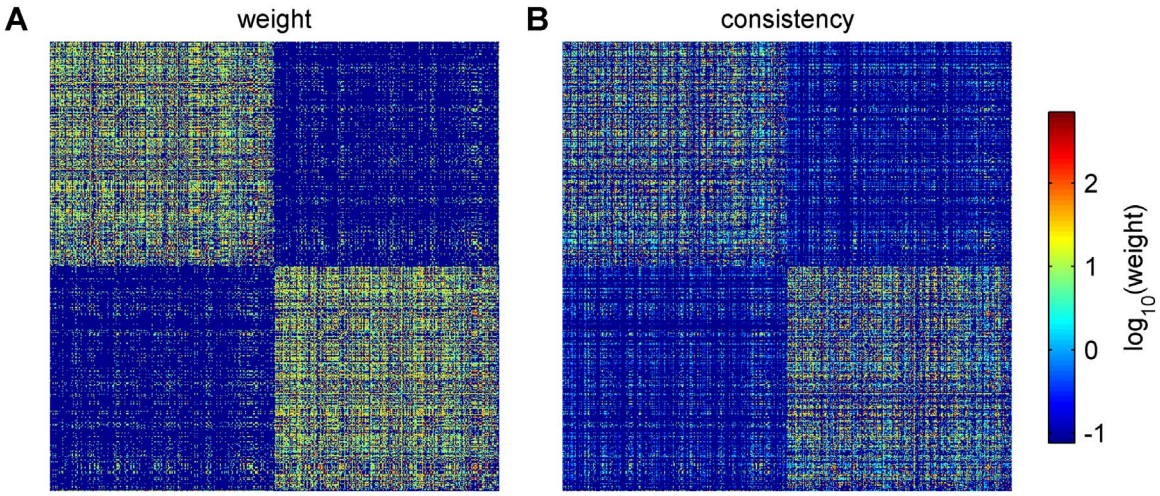

Fig. 5. Connectivity matrices at 30% density. (A) Weight-based thresholding. (B) Consistency-based thresholding.

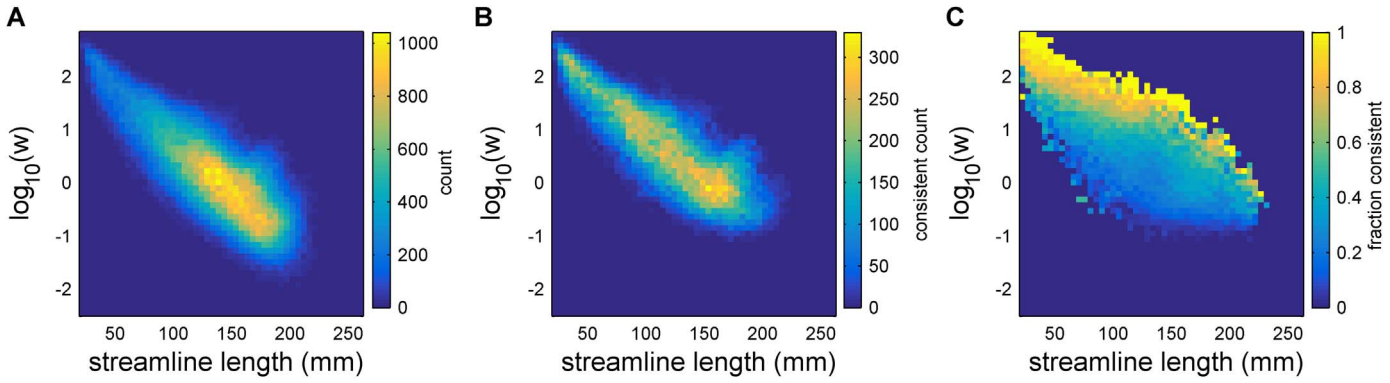

Fig. 6. Probability of consistency for a given weight and length. A: Histogram of log-weights versus streamline lengths, using 50 bins in each dimension. B: Histogram of consistent edges after discarding all but the top 30% most consistent. C: Fraction of edges that are consistent as a function of log-weight and streamline length.

sites away from the center of the brain. Most of the inter-hemispheric edges of this subnetwork interconnect frontal areas. The weak-consistent network (Fig. 8B) differs strikingly from the strong ones: it is primarily composed of inter-hemispheric connections, with connections primarily involving midline nodes. This network has the same density as the strong-consistent network but appears denser because its mean wiring length is much longer: 131 mm vs 48 mm. The strong-inconsistent network (Fig. 8C) has similar wiring length (53 mm) to the strong-consistent network, and predominantly makes short-range intra-hemispheric connections among parietal and inferior temporal areas.

Topological properties

We next compared the topology of the consistency-thresholded networks to the weight-thresholded networks. Following from the visualization of hub connectivity (Fig. 7), it is natural to ask whether hubs belong to a rich club – that is, are hub-to-hub connections denser than expected by chance (van den Heuvel and Sporns, 2011)? We calculated the binary rich club coefficient (Rubinov and Sporns, 2010) for both thresholded networks, and compared each to degree-preserving random surrogate networks. Both weight-based (Fig. 9A) and consistency-based (Fig. 9B) thresholded networks have rich clubs, as evidenced by rich club coefficients higher than the surrogate values for a wide range of node degrees. However, the networks thresholded by consistency are substantially richer than those thresholded by weight, and are topologically enriched over a wider range of degrees (Fig. 9C).

Consistency-thresholded networks also have lower mean node strength and lower clustering than weight-thresholded networks (Fig. 9D and E). Lower mean node strength is due to the lower mean weight (as is clear from the weight distributions in Fig. 4A), and occurs despite the fact that equal threshold densities impose the same mean degree on both methods, and despite consistency-based thresholding giving higher maximum degree for the most well-connected hubs (Fig. 9B). The lower clustering is primarily due to consistency-based threshold retaining a greater proportion of long-range fibers, making these networks less lattice-like than the weight-based thresholded networks. The thresholding strategy does not exert an influence on network modularity index (Fig. 9F).

Repeated analysis without rescaling the weights by distance

The decrease in weights with distance and the interaction between the weights-vs-distance trends with consistency could conceivably be affected by the $1/d$ rescaling of the weights. To verify that our results do not derive from this rescaling, we repeated our entire analysis without the $1/d$ rescaling (Supplementary information 2). All of our findings hold. The reason for this is that the main distance effect on the weights is a roughly exponential decay; an additional polynomial factor is only a (relatively) small change to the weight-vs-distance relationship.

Repeated analysis in an independent dataset

To verify that our consistency-based thresholding method is applicable to other data sets, we analyzed intersubject consistency on an independent dataset: the elderly connectome (Perry et al., 2015; Sachdev et al., 2010; Tsang et al., 2013). Reproducing our results regarding the dependence of intersubject-consistency on edge weight and distance would show that they generalize to other datasets, and are not due to idiosyncrasies in our dataset. For details on the acquisition and tractography analysis of the elderly connectome, we direct the reader to Perry et al. (2015). The only differences between that paper and the elderly dataset used in this paper is that here we have seeded more densely (100 million seeds, as in our main younger adult dataset), used a slightly newer MRtrix version (v3-12 versus v3-9; cf. v2 for our younger adult dataset), and have used 94 subjects (aged 76–93 years, 55 females) rather than the original 115 subjects (including only those who meet strict criteria for classification as “cognitively normal”). As in Fig. 2, the relationship between weights and distances is again roughly exponential (Supplementary information 1, Supplementary Fig. S1.1). This reproduces the finding of our previous study (Roberts et al., 2016). Grouping the connections by consistency again shows that both consistent and inconsistent edges span almost the full range of streamline lengths. Crucially, our main finding, that the most consistent connections are those that are strong for their length, holds in the elderly connectome (Supplementary information 1).

Comparison with mouse and macaque tract tracing data

There is currently no ground truth whole-brain human connectome. Tracer-based connectomes from other species remain the widely accepted ground truth for dMRI-based validation. We thus compared the weight-vs-distance relationship in our data to tracer-based connectomes compiled from mouse and macaque (Horvát et al., 2016). In that recent work, Horvát et al. found a species-invariant exponential distance rule governing the organization of cortical wiring, such that connection weights decay exponentially with distance, falling by several orders of magnitude over the range of connection lengths. To test whether our consistency-based thresholded data exhibit an exponential distance rule, and whether the decay rates show quantitative agreement with tracer data from other mammals, we fitted the exponential decay of connection weights for the mouse, macaque, and human connectomes. For the human data we compared weight-based and consistency-based thresholding. We find that consistency-based thresholding is in excellent agreement with the exponential decay rates in the tracer-derived mouse and macaque data, while weight-based thresholding decays far too slowly (Fig. 10A–D). Thresholding by weight removes

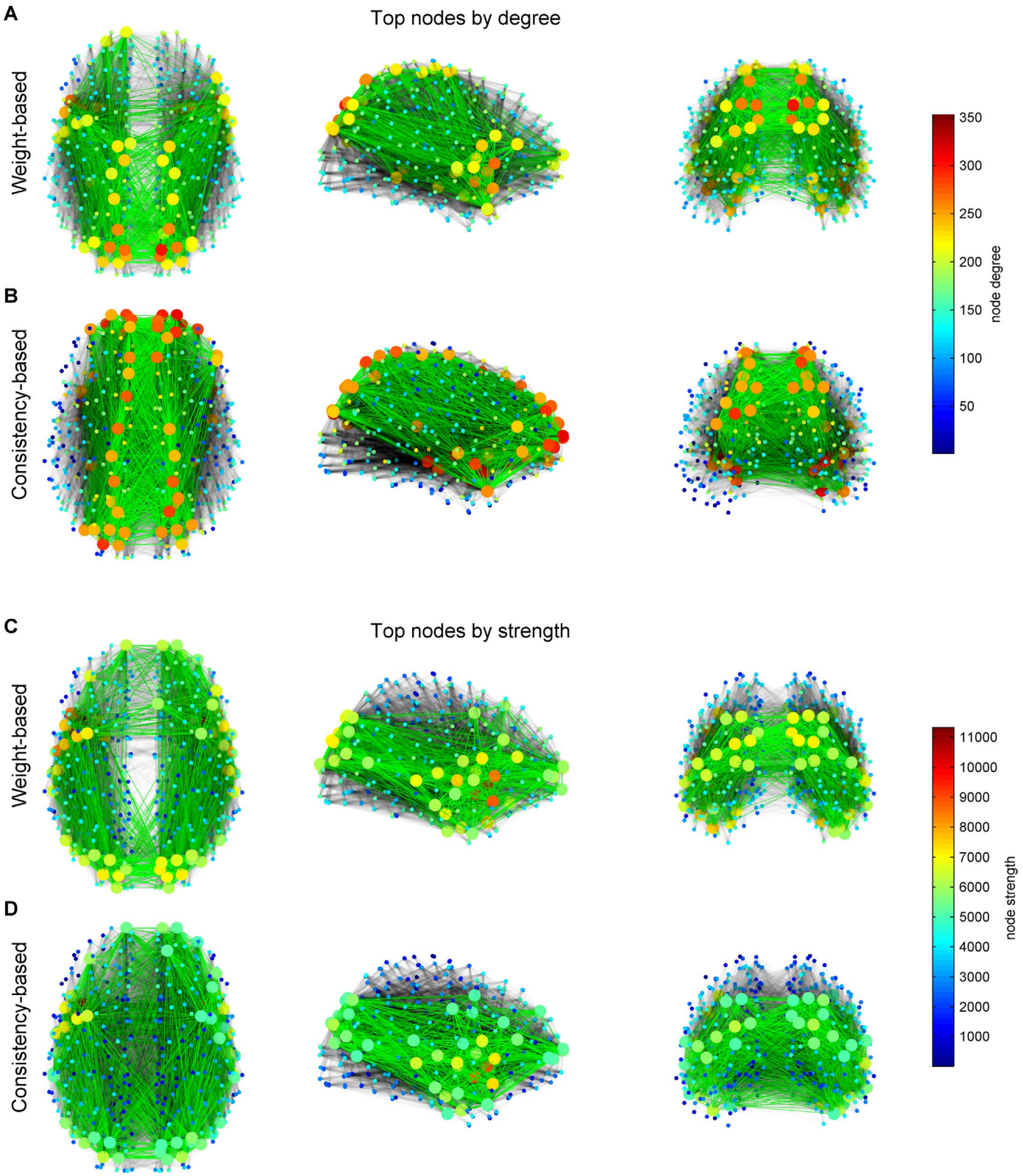

Fig. 7. Comparison of threshold methods for topology of hubs. (A) Brain connectivity after weight-based thresholding, showing the top 75 nodes by degree (large circles) and their connectivity to each other (green lines) and to all other nodes (gray lines, small circles). Nodes are colored by degree (red=high, blue=low). Networks shown after thresholding to 30% density and viewed from the top, side, and back (columns 1–3). (B) High-degree hubs after consistency-based thresholding. (C) Brain connectivity after weight-based thresholding, showing the top 75 nodes by strength. (D) High-strength hubs after consistency-based thresholding. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

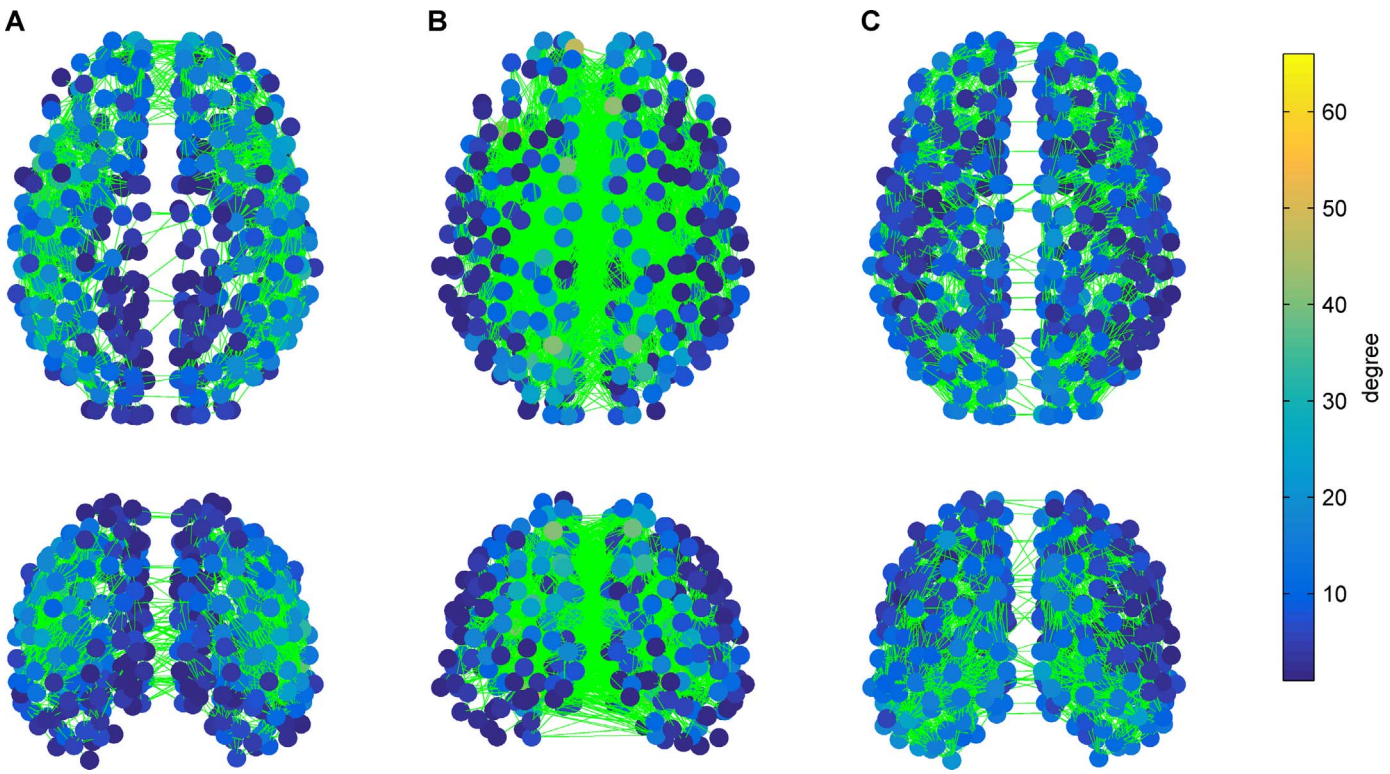

Fig. 8. Subnetworks identified by using both weight-based and consistency-based thresholding together, viewed from the top (top row) and back (bottom row). (A) Strong-consistent network. (B) Weak-consistent network. (C) Strong-inconsistent network. Isolated nodes with no edges in the subnetwork (i.e., with degree=0) are not plotted; the three networks in A–C involve 91%, 83%, and 99% of the nodes, respectively.

too many weak connections to exhibit a substantial exponential decay with distance (falling only ~2 orders of magnitude, Fig. 10D). To study the convergence to this rule as a function of the numbers of subjects, we performed a bootstrap resampling on ensembles of subjects drawn from our sample (Fig. 10E). Exemplar plots of weight-vs-distance for smaller samples are shown in Inline Supplementary Figure 1. For moderate (and larger) ensemble sizes, the consistency-based thresholded networks exhibit an exponential distance rule that is strikingly close to the macaque and mouse data. Weight-based thresholding is relatively insensitive to sample size (except for very small samples), but yields consistently-low decay rates. This quantitative match with mouse and macaque tracer data supports the use of consistency-based thresholding and also provides an intriguing further insight into species-invariant principles governing the organization of cortical wiring.

Dependence on sample size

As shown in the previous section, it is possible to explore the dependence of sample size on the thresholded networks. This analysis shows that weight-based thresholding converges relatively quickly, but to a very shallow decay rate that shows a poor qualitative fit to the data and is in disagreement with exponents from tracer-based data. While consistency-based thresholding converges quickly to an exponential form of decay (Inline Supp. Fig. 1E), the numerical exponent only converges to these independent estimates for samples of $N > \sim 25$.

Discussion

Thresholding is a widely-used method for pruning weighted networks with the aim of eliminating spurious edges and facilitating analysis and visualization. Connections are most commonly discarded according to their weights, on the assumption that weaker connections are more likely to be spurious. We systematically investigated thresholding probabilistic tractography-derived connectivity by consistency of weights across a group of networks, on the basis that the edges that are most consistent are the least likely to be spurious false positives arising in noisy, subject specific data. We found that thresholding by weight strength alone biases networks toward shorter connections, discarding weaker long connections that are more consistent across subjects. Our results suggest that thresholding by consistency preserves connections that are strong for their length, rather than simply strong overall. Moreover, consistent edges exhibit the exponential decay of connection weight with distance seen in tracer-derived connectomes, which is not obtained when retaining only the strongest edges.

The central idea of estimating the edges that are most consistent across an ensemble of networks is by no means restricted to structural brain networks derived from probabilistic tractography. For example, this approach would be useful for identifying the consistent core across a group of functional connectivity matrices. Indeed, quantifying and improving the reliability and reproducibility of functional connectomes from resting-state fMRI is a major challenge facing neuroimaging (Zuo et al., 2014; Zuo and Xing, 2014). Functional connectivity data also shows a strong distance-dependence on weight (Lord et al., 2012; Salvador et al., 2005). Future work will aim to validate our method against functional data. In fact our method can be used to address problems involving groups of networks in network science more broadly, such as the identification of gene regulatory networks (Marbach et al., 2012).

Delineation of a network ensemble into those edges that are consistent and those edges that are inconsistent across subjects serves two main purposes. First, this delineation identifies the consistent core that is conserved across the group. This is useful for reducing noise (whether from the analysis or processing pipelines) and for identifying nodes that are presumably more “fundamental” to the group (Hagmann et al., 2008). Identification of such core nodes is a central aim of network science, with many methods targeting this within a single network, such as backbones using the “disparity filter” (Foti et al., 2011; Serrano et al., 2009), k- and s-

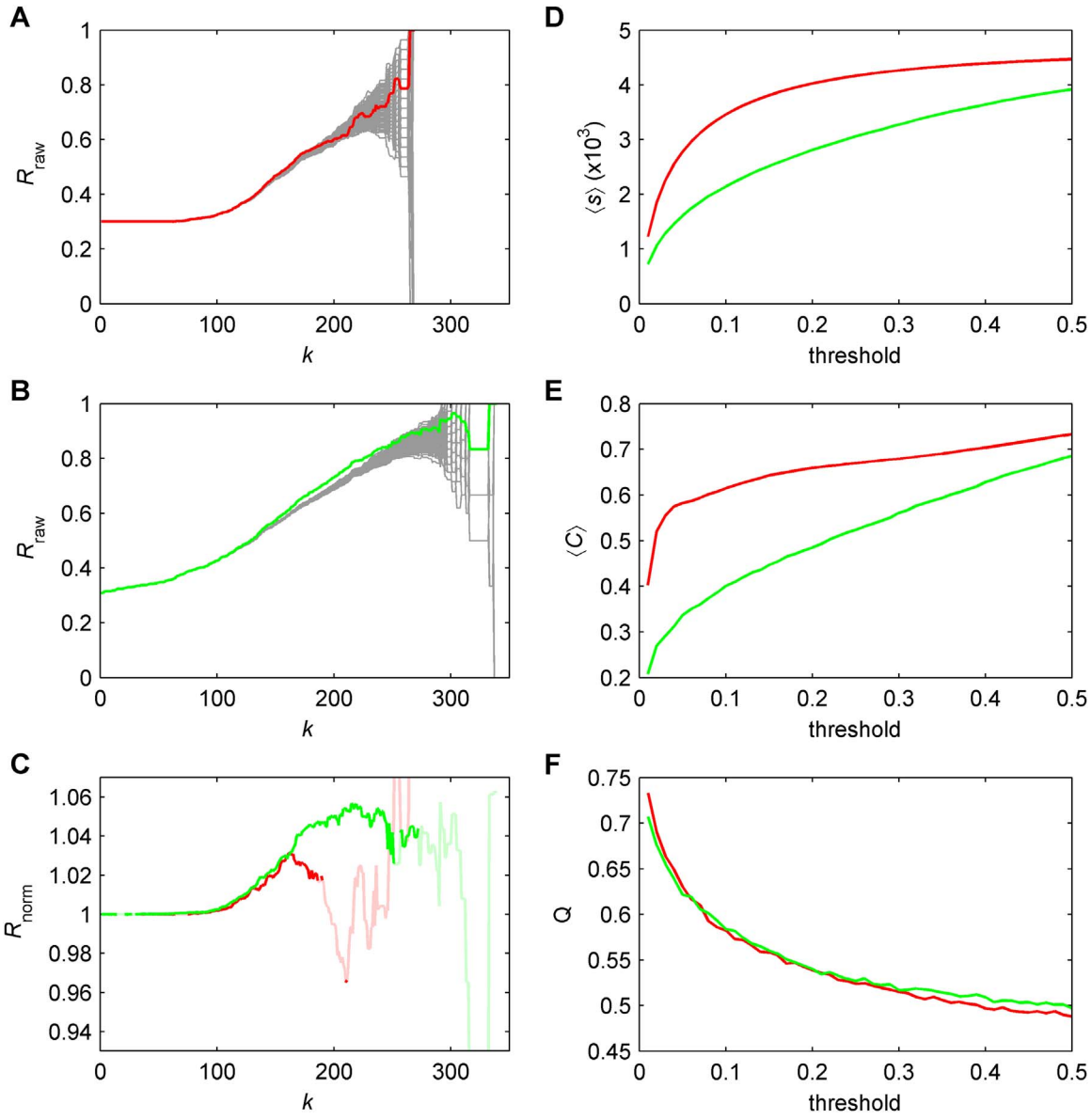

Fig. 9. Rich club effect and node statistics. Left column: Binary unnormalized rich club coefficients R_{raw} versus degree k for (A) weight-based thresholding (red) and (B) consistency-based thresholding (green), with corresponding R_{raw} curves for 1000 degree-preserving random surrogate networks. (C) Normalized rich club coefficients R_{norm} for weight-based thresholding (red) and consistency-based thresholding (green), with dark colors denoting values of k where R_{raw} differed significantly from the surrogate ensemble (by being ranked in the top or bottom 2.5% of the ensemble; i.e., a two-sided test at the p=0.05 level). Right column: Comparison of network measures for weight-based thresholding (red) and consistency-based thresholding (green). (D) Mean node strength $< s >$. (E) Mean binary clustering coefficient $< C >$. (F) Modularity index Q. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

cores (Dorogovtsev et al., 2006), rich clubs (Colizza et al., 2006), core-periphery structure (Bassett et al., 2013), multi-resolution window-based thresholding (Lohse et al., 2014), and minimum spanning tree approaches (Alexander-Bloch et al., 2010). These methods are sensitive to various features of network topology, and offer various possibilities for identifying prominent subnetworks. The consistent subnetwork across a group of networks is thus a simple cross-subject method within this broad class. This idea has also been used to identify consistent motif structures (Iakovidou et al., 2013). Second, finding inconsistent edges reveals which parts of the network are most susceptible to noise, which could aid the development of tractography methods to better deal with these regions. Inconsistent edges are also potentially informative for identifying network correlates of inter-subject differences measured in any modality. For a set of networks whose edges are known with high confidence, it may be that those edges that differ most between individuals carry the greatest potential to co-vary with specific physiological differences (Bassett et al., 2011).

Our method is complementary to the consensus-based thresholding that has been used on networks derived from deterministic tractography (de Reus and van den Heuvel, 2013). For example, van den Heuvel and Sporns (2011) selected edges on the basis of their appearance in at least 75% (for 82 nodes) or 50% (for 1170 nodes) of subjects. Seeking a consensus in the adjacency matrices requires that the individual subject matrices be reasonably sparse – this is not feasible for densely connected weighted matrices such as the ones used here, which were derived from probabilistic tractography. The consensus approach has also recently been used in a distance-preserving manner by first binning by distance then retaining the most commonly-occurring edges within each bin (Mišić et al., 2015). Although binning heuristics are somewhat arbitrary, our findings provide a principled justification for distance-preserving thresholds due to the fact that highly consistent edges occur across the entire length distribution. Indeed our finding that edges that are strong for their length are more consistent could be

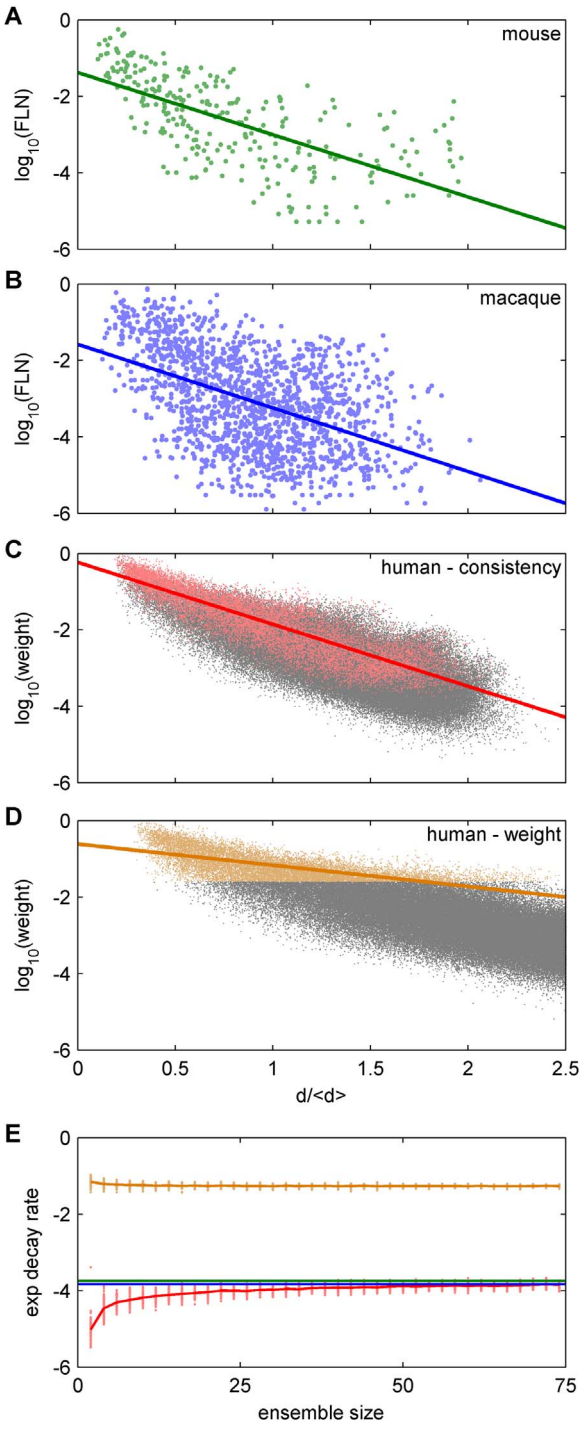

Fig. 10. Exponential decay in connection weights across species. (A) Fraction of labeled neurons (FLN) versus normalized connection distance for tracer-derived mouse connectome of Oh et al. (2014) and Zingg et al. (2014), as collated by Horvát et al. (2016). Dots correspond to individual connections; line is a linear least-squares fit. (B) Tracer-derived macaque connectome of Markov et al. (2012), as collated by Horvát et al. (2016). (C) Consistency-based thresholding of probabilistic tractography-derived human connectome data. Gray dots correspond to the fully-connected network; red dots to the 10% most consistent edges. (D) Weight-based thresholding, yellow dots correspond to the 10% strongest edges. In panels C and D, weights are rescaled by the maximum weight and distances are rescaled by the means of the post-threshold distributions. (E) Exponential decay rates for consistency-based (red) and weight-based (yellow) thresholding as a function of number of subjects, with mouse (green) and macaque (blue) decay rates for comparison. Dots correspond to fitted exponents from 100 bootstrap ensembles for each ensemble size, red and yellow lines correspond to ensemble means. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

used to (at least partly) justify other methods for applying distance-preserving thresholds. If one's goal was to reduce biases against long connections, one example could be to threshold by weight×length (or some other function of length), which could be interpreted as retaining edges with high wiring cost.

One limitation of our results is that consistent edges could be biased by the specifics of data acquisition, pre-processing, and tractography algorithms (Bastiani et al., 2012; Girard et al., 2014). This is a difficult matter to resolve because there is as yet no ground truth human connectome. We have compared our thresholded networks to rodent and non-human primate tractography, for which in situ tracing data is available, showing that a species-invariant exponential decay of connection weights is reproduced by consistency-based thresholding. Future work could extend this analysis to other graph metrics. The lack of ground truth also makes it difficult to verify whether weight-based thresholding yields the most accurate connections. Comparison of the consistent subnetworks obtained from different analysis methods may reveal biases inherent to specific methods, thereby helping to resolve their respective strengths and weaknesses. Also, here we have restricted attention to intersubject consistency for networks defined on a fixed set of nodes based on subject-specific parcellations. A more general approach would be to analyze the interaction between consistency of edge weights with the consistency of node locations, which depends on the individual-subject parcellations. Indeed individual-level anatomy has recently been used to inform estimates of group-level connectivity (Lefranc et al., 2016; Roca et al., 2010).

Another limitation is that consistency-based thresholding is primarily aimed at analyzing group-averaged connectivity, so its ability to threshold individual-subject networks is limited to identifying the weighted individual subnetworks with the same (consistent) adjacency matrix. It is possible to relax this restriction in some cases. As one possibility, the probabilistic threshold derived from the group-level probability of consistency (Fig. 6C) could be applied to other connectivity matrices. This would enable, for example, the thresholding of random matrices to yield surrogate networks drawn from a population with the same consistency profile. Such random consistent networks would also preserve, on average, the post-thresholded relationship between weight and distance, so could be considered a form of geometric surrogate network (Roberts et al., 2016; Samu et al., 2014). This idea of modeling (in a statistical sense) the population ensemble from which the subject networks are assumed to be drawn has also been used to argue for approaches that go beyond taking only the group mean as the representative group network (Simpson et al., 2012). Instead of seeking to compare networks at the individual level, another potential use for our method is to apply the consistency threshold to each group and compare the group-level networks. Such a procedure would need to be supplemented with a method to allow statistical testing of between group effects, such as permutation testing (e.g., by permuting the group labels).

A potential application for consistency-based thresholding is to understand the dependence of graph metrics on group size. Such an analysis could compare different thresholding methods and different methods for identifying the group-level network (e.g. mean vs median). We presented such an analysis in this direction in Fig. 10E. Using consistency-based thresholding, we find rapid convergence (even for small samples of N~5) to a connectome that shows a clear exponential distance effect (Inline Supp. Fig. 1E). Convergence of the exponent to the value estimated in tracer studies is slower, requiring approximately N=25 data sets. The rate of convergence likely depends on the acquisition details, and so may occur more quickly using more advanced diffusion sequences and longer acquisition times. Although weight-based thresholding on small samples (or even single subjects) may be precise we caution that it is likely inaccurate, biasing the network toward strong short-range connectivity.

As tractography methods progress, the need for post-hoc thresholding to overcome technical limitations will diminish. For example,

filtering algorithms such as Spherical-deconvolution Informed Filtering of Tractograms (SIFT) (Smith et al., 2015) can be used to reduce the probability of spurious edges for data imaged using appropriate acquisition parameters (this was not the case for the networks studied here). Residual noisy edges that survive the streamline filtering could again be pruned by consistency. The further development of metrics for analyzing weighted networks will also reduce (but not eliminate) the need for thresholding techniques. Methods for reducing the influence of false positives will still be useful for bringing the estimated networks closer to the assumed ground truth. Sparse networks are currently widely used in the literature, so more principled identification of these reduced-density networks is crucial. We contend that a systematic approach to intersubject consistency offers an important step toward solving this problem.

Acknowledgments

This work was supported by the National Health and Medical Research Council (Program Grant 1037196) and the Australian Research Council (Center of Excellence for Integrative Brain Function CE140100007).

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.neuroimage.2016.09.053.

References

Alexander-Bloch, A.F., Gogtay, N., Meunier, D., Birn, R., Clasen, L., Lalonde, F., Lenroot, R., Giedd, J., Bullmore, E.T., 2010. Disrupted modularity and local connectivity of brain functional networks in childhood-onset schizophrenia. Front. Syst. Neurosci. 4, 147.

Bassett, D.S., Brown, J.A., Deshpande, V., Carlson, J.M., Grafton, S.T., 2011. Conserved and variable architecture of human white matter connectivity. Neuroimage 54, 1262–1279.

Bassett, D.S., Wymbs, N.F., Rombach, M.P., Porter, M.A., Mucha, P.J., Grafton, S.T., 2013. Task-based core-periphery organization of human brain dynamics. PLoS Comput. Biol. 9, e1003171.

Bastiani, M., Shah, N.J., Goebel, R., Roebroeck, A., 2012. Human cortical connectome reconstruction from diffusion weighted MRI: the effect of tractography algorithm. Neuroimage 62, 1732–1749.

Behrens, T., Woolrich, M., Jenkinson, M., Johansen-Berg, H., Nunes, R., Clare, S., Matthews, P., Brady, J., Smith, S., 2003. Characterization and propagation of uncertainty in diffusion-weighted MR imaging. Magn. Reson. Med. 50, 1077–1088.

Bullmore, E.T., Bassett, D.S., 2011. Brain graphs: graphical models of the human brain connectome. Annu. Rev. Clin. Psychol. 7, 113–140.

Colizza, V., Flammini, A., Serrano, M.A., Vespignani, A., 2006. Detecting rich-club ordering in complex networks. Nat. Phys. 2, 110–115.

de Reus, M.A., van den Heuvel, M.P., 2013. Estimating false positives and negatives in brain networks. Neuroimage 70, 402–409.

Descoteaux, M., Deriche, R., Knösche, T.R., Anwander, A., 2009. Deterministic and probabilistic tractography based on complex fibre orientation distributions. Med. Imaging IEEE Trans. 28, 269–286.

Dorogovtsev, S.N., Goltsev, A.V., Mendes, J.F.F., 2006. K-core organization of complex networks. Phys. Rev. Lett. 96, 040601.

Fornito, A., Zalesky, A., Breakspear, M., 2015. The connectomics of brain disorders. Nat. Rev. Neurosci. 16, 159–172.

Foti, N.J., Hughes, J.M., Rockmore, D.N., 2011. Nonparametric sparsification of complex multiscale networks. PLoS One 6, e16431.

Gigandet, X., Hagmann, P., Kurant, M., Cammoun, L., Meuli, R., Thiran, J.-P., 2008. Estimating the confidence level of white matter connections obtained with MRI tractography. PLoS One 3, e4006.

Girard, G., Whittingstall, K., Deriche, R., Descoteaux, M., 2014. Towards quantitative connectivity analysis: reducing tractography biases. Neuroimage 98, 266–278.

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

Henderson, J.A., Robinson, P.A., 2014. Relations between the geometry of cortical gyrification and white-matter network architecture. Brain Connect. 4, 112–130.

Horvát, S., Gămănuţ, R., Ercsey-Ravasz, M., Magrou, L., Gămănuţ, B., Van Essen, D.C., Burkhalter, A., Knoblauch, K., Toroczkai, Z., Kennedy, H., 2016. Spatial embedding and wiring cost constrain the functional layout of the cortical network of rodents and primates. PLoS Biol. 14, e1002512.

Iakovidou, N.D., Dimitriadis, S.I., Laskaris, N.A., Tsichlas, K., Manolopoulos, Y., 2013. On the discovery of group-consistent graph substructure patterns from brain networks. J. Neurosci. Methods 213, 204–213.

Lefranc, S., Roca, P., Perrot, M., Poupon, C., Le Bihan, D., Mangin, J.-F., Rivière, D., 2016. Groupwise connectivity-based parcellation of the whole human cortical surface using watershed-driven dimension reduction. Med. Image Anal. 30, 11–29.

Li, L., Rilling, J.K., Preuss, T.M., Glasser, M.F., Hu, X., 2012. The effects of connection reconstruction method on the interregional connectivity of brain networks via diffusion tractography. Hum. Brain Mapp. 33, 1894–1913.

Lohse, C., Bassett, D.S., Lim, K.O., Carlson, J.M., 2014. Resolving anatomical and functional structure in human brain organization: identifying mesoscale organization in weighted network representations. PLoS Comput. Biol. 10, e1003712.

Lord, A., Horn, D., Breakspear, M., Walter, M., 2012. Changes in community structure of resting state functional connectivity in unipolar depression. PLoS One 7, e41282.

Marbach, D., Costello, J.C., Küffner, R., Vega, N.M., Prill, R.J., Camacho, D.M., Allison, K.R., Kellis, M., Collins, J.J., Stolovitzky, G., 2012. Wisdom of crowds for robust gene network inference. Nat. Methods 9, 796–804.

Markov, N.T., Ercsey-Ravasz, M.M., Gomes, A.R.R., Lamy, C., Magrou, L., Vezoli, J., Misery, P., Falchier, A., Quilodran, R., Gariel, M.A., 2012. A weighted and directed interareal connectivity matrix for macaque cerebral cortex. Cereb. Cortex, bhs270.

Mišić, B., Betzel, Richard, F., Nematzadeh, A., Goñi, J., Griffa, A., Hagmann, P., Flammini, A., Ahn, Y.-Y., Sporns, O., 2015. Cooperative and competitive spreading dynamics on the human connectome. Neuron 86, 1518–1529.

Oh, S.W., Harris, J.A., Ng, L., Winslow, B., Cain, N., Mihalas, S., Wang, Q., Lau, C., Kuan, L., Henry, A.M., Mortrud, M.T., Ouellette, B., Nguyen, T.N., Sorensen, S.A., Slaughterbeck, C.R., Wakeman, W., Li, Y., Feng, D., Ho, A., Nicholas, E., Hirokawa, K.E., Bohn, P., Joines, K.M., Peng, H., Hawrylycz, M.J., Phillips, J.W., Hohmann, J.G., Wohnoutka, P., Gerfen, C.R., Koch, C., Bernard, A., Dang, C., Jones, A.R., Zeng, H., 2014. A mesoscale connectome of the mouse brain. Nature 508, 207–214.

Perry, A., Wen, W., Lord, A., Thalamuthu, A., Roberts, G., Mitchell, P., Sachdev, P., Breakspear, M., 2015. The organisation of the elderly connectome. Neuroimage 114, 414–426.

Roberts, J.A., Perry, A., Lord, A.R., Roberts, G., Mitchell, P.B., Smith, R.E., Calamante, F., Breakspear, M., 2016. The contribution of geometry to the human connectome. Neuroimage 124, 379–393.

Roca, P., Tucholka, A., Riviere, D., Guevara, P., Poupon, C., Mangin, J.-F., 2010. Inter-subject connectivity-based parcellation of a patch of cerebral cortex. Med. Image Comput. Comput.-Assist. Interv., 347–354.

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

Sachdev, P.S., Brodaty, H., Reppermund, S., Kochan, N.A., Trollor, J.N., Draper, B., Slavin, M.J., Crawford, J., Kang, K., Broe, G.A., 2010. The Sydney Memory and Ageing Study (MAS): methodology and baseline medical and neuropsychiatric characteristics of an elderly epidemiological non-demented cohort of Australians aged 70–90 years. Int. Psychogeriatr. 22, 1248–1264.

Salvador, R., Suckling, J., Coleman, M.R., Pickard, J.D., Menon, D., Bullmore, E., 2005. Neurophysiological architecture of functional magnetic resonance images of human brain. Cereb. Cortex 15, 1332–1342.

Samu, D., Seth, A.K., Nowotny, T., 2014. Influence of wiring cost on the large-scale architecture of human cortical connectivity. PLoS Comput. Biol. 10, e1003557.

Serrano, M.Á., Boguná, M., Vespignani, A., 2009. Extracting the multiscale backbone of complex weighted networks. Proc. Natl. Acad. Sci. USA 106, 6483–6488.

Simpson, S.L., Moussa, M.N., Laurienti, P.J., 2012. An exponential random graph modeling approach to creating group-based representative whole-brain connectivity networks. Neuroimage 60, 1117–1126.

Smith, R.E., Tournier, J.-D., Calamante, F., Connelly, A., 2013. SIFT: spherical-deconvolution informed filtering of tractograms. Neuroimage 67, 298–312.

Smith, R.E., Tournier, J.-D., Calamante, F., Connelly, A., 2015. SIFT2: enabling dense quantitative assessment of brain white matter connectivity using streamlines tractography. Neuroimage 119, 338–351.

Smith, R.E., Tournier, J.D., Calamante, F., Connelly, A., 2012. Anatomically-constrained tractography: improved diffusion MRI streamlines tractography through effective use of anatomical information. Neuroimage 62, 1924–1938.

Thomas, C., Frank, Q.Y., Irfanoglu, M.O., Modi, P., Saleem, K.S., Leopold, D.A., Pierpaoli, C., 2014. Anatomical accuracy of brain connections derived from diffusion MRI tractography is inherently limited. Proc. Natl. Acad. Sci. USA 111, 16574–16579.

Tournier, J.-D., Calamante, F., Connelly, A., 2007. Robust determination of the fibre orientation distribution in diffusion MRI: non-negativity constrained super-resolved spherical deconvolution. Neuroimage 35, 1459–1472.

Tournier, J., Calamante, F., Connelly, A., 2012. MRtrix: diffusion tractography in crossing fiber regions. Int. J. Imaging Syst. Technol. 22, 53–66.

Tsang, R.S.M., Sachdev, P.S., Reppermund, S., Kochan, N.A., Kang, K., Crawford, J., Wen, W., Draper, B., Trollor, J.N., Slavin, M.J., 2013. Sydney memory and ageing study: an epidemiological cohort study of brain ageing and dementia. Int. Rev. Psychiatry 25, 711–725.

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

Van Wijk, B.C., Stam, C.J., Daffertshofer, A., 2010. Comparing brain networks of different size and connectivity density using graph theory. PloS One 5, e13701.

Zalesky, A., 2008. DT-MRI fiber tracking: a shortest paths approach. Med. Imaging IEEE Trans. 27, 1458–1471.

Zalesky, A., Fornito, A., 2009. A DTI-derived measure of cortico-cortical connectivity. Med. Imaging IEEE Trans. 28, 1023–1036.

Zalesky, A., Fornito, A., Harding, I.H., Cocchi, L., Yücel, M., Pantelis, C., Bullmore, E.T., 2010. Whole-brain anatomical networks: does the choice of nodes matter? Neuroimage 50, 970–983.

Zingg, B., Hintiryan, H., Gou, L., Song, M.Y., Bay, M., Bienkowski, M.S., Foster, N.N., Yamashita, S., Bowman, I., Toga, A.W., 2014. Neural networks of the mouse neocortex. Cell 156, 1096–1111.

Zuo, X.-N., Anderson, J.S., Bellec, P., Birn, R.M., Biswal, B.B., Blautzik, J., Breitner, J.C.S., Buckner, R.L., Calhoun, V.D., Castellanos, F.X., Chen, A., Chen, B., Chen, J., Chen, X., Colcombe, S.J., Courtney, W., Craddock, R.C., Di Martino, A., Dong, H.-M., Fu, X., Gong, Q., Gorgolewski, K.J., Han, Y., He, Y., He, Y., Ho, E., Holmes, A., Hou, X.-H., Huckins, J., Jiang, T., Jiang, Y., Kelley, W., Kelly, C., King, M., LaConte, S.M., Lainhart, J.E., Lei, X., Li, H.-J., Li, K., Li, K., Lin, Q., Liu, D., Liu, J., Liu, X., Liu, Y.,

Lu, G., Lu, J., Luna, B., Luo, J., Lurie, D., Mao, Y., Margulies, D.S., Mayer, A.R., Meindl, T., Meyerand, M.E., Nan, W., Nielsen, J.A., O'Connor, D., Paulsen, D., Prabhakaran, V., Qi, Z., Qiu, J., Shao, C., Shehzad, Z., Tang, W., Villringer, A., Wang, H., Wang, K., Wei, D., Wei, G.-X., Weng, X.-C., Wu, X., Xu, T., Yang, N., Yang, Z., Zang, Y.-F., Zhang, L., Zhang, Q., Zhang, Z., Zhang, Z., Zhao, K., Zhen, Z., Zhou, Y., Zhu, X.-T., Milham, M.P., 2014. An open science resource for establishing reliability and reproducibility in functional connectomics. Sci. Data 1, 140049.

Zuo, X.-N., Xing, X.-X., 2014. Test-retest reliabilities of resting-state FMRI measurements in human brain functional connectomics: a systems neuroscience perspective. Neurosci. Biobehav. Rev. 45, 100–118.