Frontiers in Neurosciences February 6, 2014

Reproducibility of graph metrics of human brain structural networks

Jeffrey T. Duda [1,*], **Philip A. Cook** [1] **and James C. Gee** [1]

[1] *Penn Image Computing and Science Laboratory, University of Pennsylvania, Department of Radiology, Philadelphia, PA, USA*

Correspondence*:
Jeffrey T. Duda
Penn Image Computing and Science Laboratory, University of Pennsylvania, Department of Radiology, 3600 Market Street, Suite 370, Philadelphia, PA, USA, jtduda@seas.upenn.edu

Neuroinformatics with the Insight ToolKit

ABSTRACT

Recent interest in human brain connectivity has led to the application of graph theoretical analysis to human brain structural networks, in particular white matter connectivity inferred from diffusion imaging and fiber tractography. While these methods have been used to study a variety of patient populations, there has been less examination of the reproducibility of these methods. A number of tractography algorithms exist and many of these are known to be sensitive to user-selected parameters. The methods used to derive a connectivity matrix from fiber tractography output may also influence the resulting graph metrics. Here we examine how these algorithm and parameter choices influence the reproducibility of proposed graph metrics on a publicly available test-retest dataset consisting of 21 healthy adults. The dice coefficient is used to examine topological similarity of constant density subgraphs both within and between subjects. Seven graph metrics are examined here: mean clustering coefficient, characteristic path length, largest connected component size, assortativity, global efficiency, local efficiency, and rich club coefficient. These reproducibility of these network summary measures is examined using the intraclass correlation coefficient (ICC). Graph curves are created by treating the graph metrics as functions of a parameter such as graph density. Functional data analysis techniques are used to examine differences in graph measures that result from the choice of fiber tracking algorithm. The graph metrics consistently showed good levels of reproducibility as measured with ICC, with the exception of some instability at low graph density levels. The global and local efficiency measures were the most robust to the choice of fiber tracking algorithm.

Keywords: Structure, Tractography, Connectivity, Brain, Network, Reproducibility, Graph

1 INTRODUCTION

Combining magnetic resonance imaging (MRI) of the human brain with graph theory analysis has emerged as a powerful approach to studying large-scale networks of both structural and functional connectivity. In the case of structural connectivity, the use of diffusion weighted MRI and associated white matter fiber tractography methods provide the ability to identify the long-range pathways that connect cortical regions and form a network architecture (Basser et al., 2000; Lazar et al., 2003; Hagmann et al., 2003; Xue et al., 1999). The use of graph theoretical analysis to study the topology and structure of these large scale networks is an increasingly active topic of research (Hagmann et al., 2008; Zalesky et al., 2010; Cheng et al., 2012a; Bastiani et al., 2012; Irimia and Van Horn, 2012; Sporns, 2011; Fornito et al., 2012). These methods have been used to examine the structural consequences of neurological disorders (Xie and

He, 2012; Guye et al., 2010; Martin, 2012) as well as the relationship between structure and function (Hagmann et al., 2008; Honey et al., 2009, 2007).

Previous studies examining the reproducibly of graph-based metrics in functional networks have shown good levels of reproducibly in MEG (Deuker et al., 2009), fMRI using BOLD contrast (Telesford et al., 2010; Braun et al., 2012; Schwarz and McGonigle, 2011; Liang et al., 2012; Weber et al., 2013) and arterial spin labeling (Weber et al., 2013). A number of studies have also examined reproducibly in structural networks, each focusing on various aspects of the complex processing pipeline that is a prerequisite for these measures. These have included studies of diffusion spectrum imaging (Cammoun et al., 2012; Bassett et al., 2011) and high angular resolution diffusion imaging (Dennis et al., 2012). Some studies have examined probabilistic tractography (Owen et al., 2013; Vaessen et al., 2010). Diffusion tensor imaging (DTI) based studies using deterministic tractography have included the examination of tractography seed density (Cheng et al., 2012b), anatomic label density (Bassett et al., 2011), and studies examining a variety of network measures (Cheng et al., 2012b; Irimia and Van Horn, 2012). Further exploration of the reproducibility of these graph metrics as it relates to their utility in longitudinal studies (Telesford et al., 2013).

In the paper we constrain our analysis to DTI-based deterministic fiber tractography. Within this constraint, we examine multiple algorithms for computing streamlines to examine their influence on the final graph metrics. A set of manually defined cortical parcellations (Klein and Tourville, 2012) is used along with a more common template-based parcellation scheme (Tzourio-Mazoyer et al., 2002). The intraclass correlation coefficient (ICC) is used to examine the reproducibility of network summary measures that results from combinations of fiber tracking algorithm and anatomical label set. The dice coefficient provides a measure of topographical similarity to examine the reproducibility of subgraphs extracted as a function of graph density. Graph curves are constructed for a variety of metrics and functional data analysis is used to examine how these metrics differ as a function of graph density or other parameters that are specific to a given metric. We use freely available data and software to create a framework that facilitates future extensions that may examine additional aspects of the processing as well as the comparison to, or addition of, multiple imaging modalities.

2 MATERIALS & METHODS

2.1 NEUROIMAGING DATA

The Multi-Modal MRI Reproducibility Resource (Landman et al., 2011), informally known as the Kirby dataset (`http://www.nitrc.org/projects/multimodal`), provides a publicly available test-retest data set consisting of 21 healthy control subjects (11 males). The mean age is 31.76 ± 9.35 with a range of [22,61]. This data set provides a multitude of MR image types, but here only the T1-weighted anatomical images and diffusion tensor images are examined. The T1 images have a resolution of $1.2 \times 1.0 \times 1.0mm$. The diffusion images have a resolution of $0.828125 \times 0.828125 \times 2.2mm$. The diffusion data includes a single $b = 0$ volume and 34 directional diffusion weighted images acquired with $b = 700s/mm^2$.

2.2 ANATOMICAL LABELING

A graph consists of nodes and the edges that connect those nodes. To construct a graph from a brain, a set of anatomical labels are used to define the nodes of the graph. To determine if manually defined cortical labels would provide an inherent advantage in reproducibility we used of the Mindboggle dataset which provides a set of manually drawn cortical regions (DKT31) along with a skull-stripped image for a single time point for each subject in the Kirby data set (Klein and Tourville, 2012). To utilize these labels in network creation we performed an intra subject registration between each subject's two T1 images. A brain mask was created from the provided skull-stripped T1 image by thresholding and a morphological closing. This mask was warped into the unlabeled T1 image space and used to create a skull-stripped image. For each time, a transformation was found between the skull-stripped T1 image and the b=0 image, acquired as part of the DTI acquisition. In all subjects, the manually defined labels were propagated into the DTI space for both time points using the appropriate composed transform.

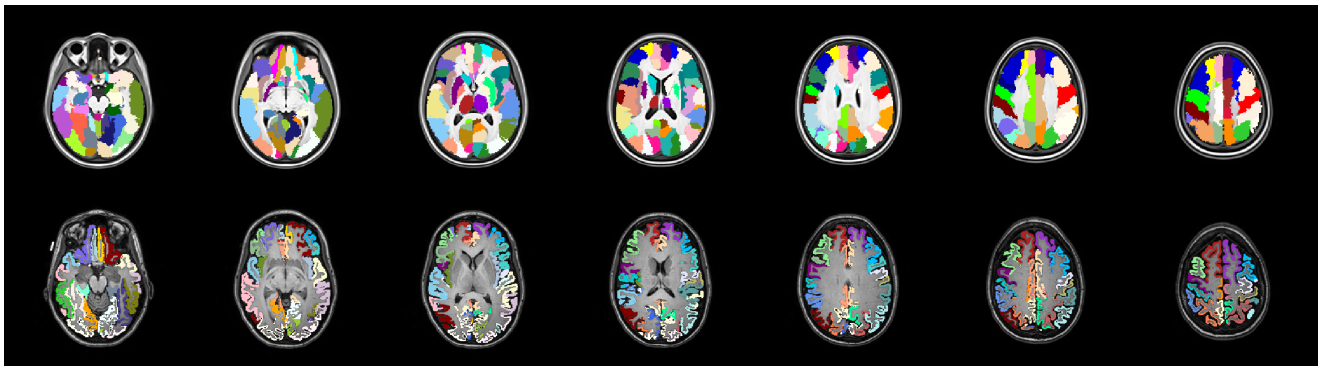

Figure 1. Two sets of anatomical labels are used to define the networks. The template based AAL labels (top) and The DKT31 manually defied labels that are provided via Mindboggle (bottom).

One of the most common label sets used in studies of both functional and structural connectivity is the AAL label set (Tzourio-Mazoyer et al., 2002) which is a template based label set. An existing multivariate template had been created from the Kirby dataset using the `antsMultivariateTemplateConstruction.sh` tool, part of the Advanced Normalization Tools (ANTs) software package (Avants et al., 2009). The `antsRegistration` tool was used to find a deformable mapping between the T1 template image distributed with the AAL label and the population specific template created from the Kirby data. In order to transform these labels into each subject's DTI space, it was necessary to find a transform from the template to each subject's T1 and from T1 to DTI within each subject. For the template-to-T1 transform, the `antsCorticalThickness.sh` tool was used. This software first applied a bias correction using the N4 algorithm (Tustison et al., 2010). Next a registration based skull stripping was performed to provide a cerebrum mask of the T1 image. This was followed by a final cerebrum-only registration to the template. These transforms were composed with the T1-to-DTI transforms, providing a single transform that was used to warp the the AAL labels into DTI space using nearest neighbor interpolation. Labels of structures outside of the cerebrum were removed. Many AAL labels include both gray and white matter, here the labels were masked to only include voxels that were identified as cortical gray matter by the DKT31 labels described in the previous section. The AAL labels for deep gray structures (e.g. thalamus) were not masked but used in their entirety. Both label set are illustrated in figure 1, while the entire processing scheme is illustrated in figure 2.

2.3 DIFFUSION DATA PREPROCESSING

The Camino toolkit (Cook et al., 2006) was used to calculate diffusion tensor images via a weighted linear fitting (Basser et al., 1994; Salvador et al., 2005), and was used for subsequent deterministic tractography. The brain masks defined in T1 space were warped into DTI space and used to prevent tracking outside the brain. Fractional anisotropy (FA) images were calculated and a tractography seed-map was created to include all voxels in the cerebrum with an FA of at least 0.2.

One of the primary differences among the various approaches to deterministic tractography is the algorithm used to determine the direction that a streamline should proceed from a given point. Here we examine four different approaches:

1. Fiber Assignment by Continuous Tracking (FACT) - The primary direction of diffusion (PDD) is followed until the streamline enters a new voxel (Xue et al., 1999).
2. Euler - The PDD is followed for a constant step size (Basser et al., 2000).
3. Rourth-order Runge-Kutta (RK4) - The direction of the step is determined by taking and averaging a weighted series of partial steps (Basser et al., 2000).
4. Tensor Deflection (TEND) - The local fiber trajectory is a function of the previous direction and the local diffusion tensor (Lazar et al., 2003)

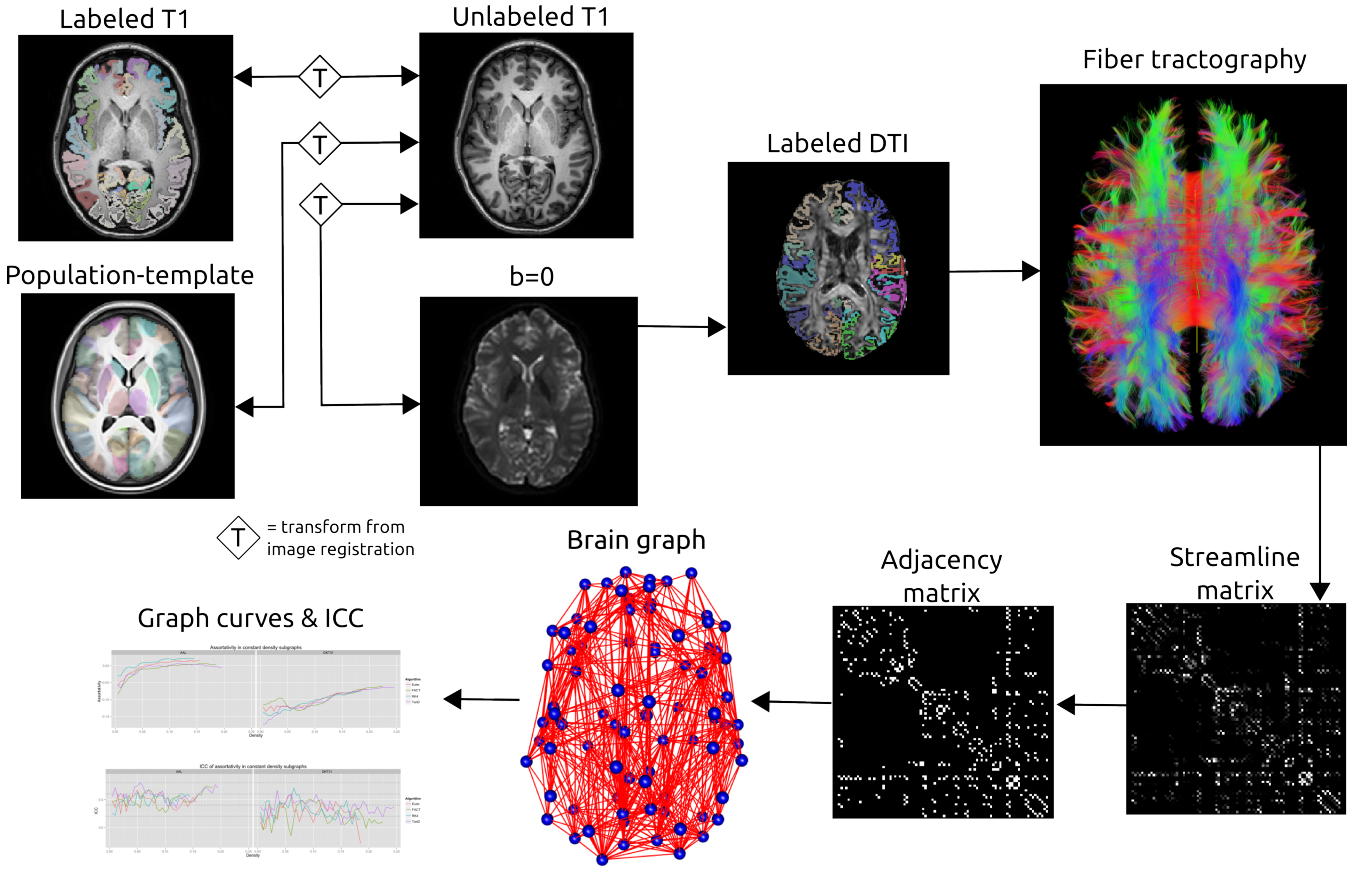

Figure 2. Schematic of the network processing scheme. Image registration is used to find transformations between the T1 image and: the T1 image for that subject's other time point; the population template; the b=0 image acquired as part of the DTI acquisition. Labels are transformed into the DTI space where fiber tractography is performed. A matrix is created that records the number of streamline connecting each pair of labeled regions. This matrix is thresholded as constant density to create an adjacency matrix which defines connections in a brain graph. Graph curves are generate by calculating network summary measures over a range of density values and ICC plots are used to examine the reproducibility of the metrics.

Shared parameters used in the fiber tracking were held constant as follows

1. Streamlines were terminated if curvature of more than 90 degrees over 5 steps was detected.
2. Streamlines were terminated if an FA value of less of 0.2 was encountered.
3. A step size of 0.5mm was used.
4. Linear interpolation of the primary direction of diffusion was used for Euler and RK4.

2.4 GRAPH GENERATION

While the nodes of a graph were defined using anatomical labels, the edges of the graph were defined by using fiber tractography to identify white matter pathways that connect brain regions. For a given set of streamlines, the `connmat` tool provided by the Camino toolkit was used to generate a connectivity matrix that records how many streamlines connect each pair of regions in a given set of target regions. This program starts at the seed point for a streamline and proceeds outward in each direction to determines the two target regions encountered. Only streamlines that connect two unique regions are retained and a given streamline may be only be counted as connecting a single pair of target regions. Fiber tractography

does not provide a measure of directionality (i.e. neither node can be considered a starting point or ending point) so the resulting matrices resulting in undirected graphs.

Graph are often compared by first ensuring that they have the same density (Achard et al., 2006; Bassett et al., 2006), where density for an undirected graph is defined as:

$$D(G) = \frac{\|E(G)\|}{(\|N(G)\|(\|N(G)\| - 1))}$$

where $N(G)$ is the set of all nodes in graph G and $E(G)$ is the set of all edges in G. The number of nodes in the graph and the desired density determine the number of edges that the graph should contain. Edges of higher weights are given priority and lower weighted edges are removed to obtain the desired density level. The weights of the remaining edges are then set to 1 for a final binarized graph. This cumulative thresholding provides a normalized method for comparing network measures as it results in the comparison of graphs with an equal percentage of significant connections. Graphs are typically compared over a range of density levels. Here, we only directly compare measures obtained from graphs with an equal number of nodes and thus an equal number of edges after density thresholding.

2.5 NETWORK METRICS

A large number of graph metrics are available for quantifying properties of binary, undirected networks (Rubinov and Sporns, 2010). Here we examine a number that are common in current literature: largest connected component size (Bassett et al., 2011), assortativity(Newman, 2006; Bassett et al., 2008), clustering coefficient (Watts and Strogatz, 1998), characteristic path length(Watts and Strogatz, 1998), global and local efficiency (Latora and Marchiori, 2001), and rich club coefficient(Collin et al., 2013). An ITK module named Petiole (`https://github.com/jeffduda/Petiole`) was created to calculate these network measures from 2D connectivity matrices. This module incorporates and extends an existing implementation of a graph class (Tustison et al., 2008) and provides ITK functions for a variety of graph metrics while using the matlab-based Brain Connectivity Toolkit (Rubinov and Sporns, 2010) for algorithmic guidance. While many of these metrics include implementations for weighted graphs and/or directed graphs, here we focus on their application to unweighted, undirected graphs. Summaries and equations for these metrics are provided here:

Size of Largest Connected Component. A connected component of a graph is a subset of the graph, G_i, where there exists a path between all pairs of nodes and for which no path exist to additional nodes in G. The largest connected component is the G_i with the greatest number of nodes, $\|N(G_i\|$. This measure relates to the global level of connectivity within a subject's brain network Bassett et al. (2011).

Assortativity. The degree of a node is the number of neighboring nodes that it connects to (i.e. shares an edge with). Assortativity measures how preferentially nodes of similar degree connect to one another (Newman, 2006) and is defined as:

$$A = \frac{\frac{1}{E}\sum_i j_i k_i - [\frac{1}{E}\sum_i \frac{1}{2}(j_i + k_i)]^2}{\frac{1}{E}\sum_i \frac{1}{2}(j_i^2 + k_i^2) - [\frac{1}{E}\sum_i \frac{1}{2}(j_i + k_i)]^2}$$

where j_i, k_i are the degrees of the nodes connected by edge i and $E = \|E(G)\|$. High assortativity suggests higher network resilience, making a network less vulnerable to attack (Newman, 2002).

Clustering Coefficient. This measure quantifies how likely is that two nodes with a common neighbor are connected to one another (Watts and Strogatz, 1998). Here we calculate the clustering coefficient at each node and calculate the mean over all nodes in the network for our final network summary measure. The clustering coefficient at node i is given by:

$$C_i = \frac{e_i}{\|K_i\|(\|K_i\| - 1)}$$

where K_i is the set of all nodes that share an edge with i and e_i is the set of all edges that connect nodes in K_i.

Characteristic Path Length. The pathlength, L_{ij}, that connects two nodes, i and j, is defined as the minimum number of edges that must be traversed to travel from i to j (Dijkstra, 1959). The characteristic path length is the average pathlength over all possible pairs of connections in a graph. In an undirected graph this is:

$$L = \frac{1}{\|N(G)\|(\|N(G)\| - 1)} \sum_{ij \in G, i \neq j} L_{ij}$$

This measure is only defined for fully connected graphs. Here, we apply the density thresholding first and then extract the largest connected component in order to calculate the characteristic path length.

Global Efficiency. This measure is related to the characteristic path length, in that it attempts to quantify the mean efficiency between any two nodes in the graph. Unlike the characteristic path length, this metric is defined for both connected and unconnected graphs (Latora and Marchiori, 2001).

$$F_{glob} = \frac{1}{\|N(G)\|(\|N(G)\| - 1)} \sum_{i \neq j \in G} 1/L_{ij}$$

Local Efficiency. This metric relates to fault tolerance and examines efficiency between neighbors on a node i, if that node were removed from the graph (Latora and Marchiori, 2001).

$$F_{loc} = \frac{1}{\|N(G)\|} \sum_{i \in n} F(G_i)$$

where G_i is the subgraph of G that results from removing node i.

Rich Club Coefficient This measures quantifies how preferentially the high-degree nodes (i.e. rich nodes) in a graph connect to other high-degree nodes (Colizza et al., 2006).

$$R(G, k) = \frac{\|E(G, k)\|}{\|N(G, k)\|(\|N(G, k)\| - 1)}$$

where $N(G, k)$ is the set of nodes of degree k or higher and $E(G, k)$ is the set of edges connecting two nodes in $N(G, k)$.

2.6 GRAPH CURVES

The metrics listed above are all applied to thresholded binary graphs. As discussed earlier, this binary graph result from thresholding at a constant density. These metrics may then be treated as functional curves of metric vs. graph density. By doing this, we are able to compare binary graphs in a way that incorporates the continuous structure of the original connectivity matrices. The rich club coefficient however is dependent upon two parameters, the graph density and, k, the degree threshold used to determined what constitutes a rich-node. For this metric we threshold at the highest density common to all graphs and explore how the value changes with k. For all other metrics, we examine their curves as a function of graph density.

2.7 STATISTICAL ANALYSIS

Before examining how the graph metrics change with density it is necessary to examine the maximum density of the graphs to determine the range over which graphs may be compared. Additionally, it is interesting to examine the topological similarity in the thresholded graphs. This is done using the dice coefficient which measures similarity between two graphs as:

$$Dice(x, y) = \frac{2\|E(x) \cap E(y)\|}{\|E(x)\| + \|E(y)|}$$

where edges are considered equal if they connect the same two nodes. This is equivalent to treating each connectivity matrix as a binarized 2D image and using the Dice metric to measure overlap. The mean

intra- and inter-subject topological similarity was computed over a range of densities for each combination of tracking algorithm and anatomical label sets. This allows us to examine the reproducibility of within-subject topography compared to between subject topography. This metric is limited to lie in the range $[0, 1]$ and can be interpreted as a measure of degree of overlap between graphs. This provides a stricter metric than measuring overlap between sets of nodes as complete node-overlap is a necessary but incomplete condition for complete edge-overlap.

Graph curves are used to examine the reproducibility of the graph metrics as a function of an independent parameter, typically graph density. At each point along the curve, reproducibility of the metric is quantified using the ICC:

$$ICC = \frac{\sigma_{bs}^2}{\sigma_{bs}^2 + \sigma_{ws}^2}$$

where σ_{bs}^2 is the between-subject variance and σ_{ws}^2 is the within subject variance. The 'ICC' package for R is used for this calculation. The ICC is plotted along with the mean graph metrics for each combination of algorithm and label set. At points where little to no variance exists in a graph metric, the ICC is not calculated as it becomes unstable under those conditions. The following guidelines may be used to interpret ICC values: ICC $<$ 0.2 'poor agreement'; 0.21 - 0.40 'fair agreement'; 0.41-0.60 moderate agreement; 0.61-0.80 'strong agreement'; ICC $>$ 0.8 'near perfect agrement (Telesford et al., 2010; Montgomery et al., 2002). Dashed lines indicating the boundaries of these categories have been included on all ICC plots to aid interpretation.

To identify group diferences that result from fiber tracking algorithm we incorporated methods from functional data analysis (FDA) which treats each curve as a function. For each group, the set of all curves were averaged to create a single mean curve. While there are a variety of methods for computing the difference between two curves, here we chose the simplest method, the non-parametric permutation test. Each mean curve was treated a function and the area between the group mean curves was found. Individual group assignments were then permuted using random sampling without replacement and then used to calculate mean curves . The area between the random-group mean curves was calculated. This was performed iteratively (i=10000). We recorded, x, the number of times area between the mean curves from the randomly assigned groups is larger than the area between the true group mean curves. The p-value for the true group difference is then defined as x/i. We report these differences for between-algorithm curves as they derive from graph of equal size, but do compare curves that derive from different anatomical label sets.

3 RESULTS

3.1 NETWORK DENSITY

Maximal densities for connectivity matrices across all tracking algorithm ranged from 0.17 to 0.30 for the AAL labels and from 0.20 to 0.41 for DKT31. Maximal densities in the DTK31 data was generally higher than in the AAL as illustrated in figure 3. Both label sets had the same lowest-to-highest ordering of mean maximal density within algorithms: RK4 $<$ Euler $<$ FACT $<$ TEND.

3.2 NETWORK TOPOLOGY

Dice coefficients for intra-subject similarity ranged from 0.70 to 0.81 for the AAL labels and from 0.59 to 0.82 for the DTK31 labels. Inter-subject similarity ranged from 0.51 to 0.71 for AAL labels and from 0.32 to 0.71 for the DTK31 labels. For all algorithm-label pairings, intra-subject overlap was greater than inter-subject overlap across the range of densities as illustrated in figure 4. Permutation testing of intra-subject dice vs. density curves did not reveal any significant differences between algorithms for either label set. However, a number of differences were found in the inter-subject comparions. The resulting p-values are listed in table 1.

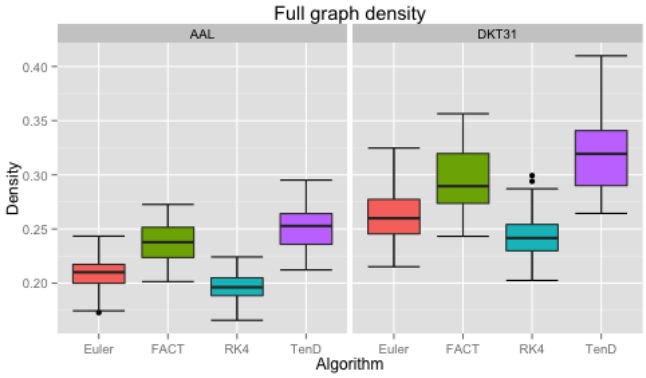

Figure 3. Boxplots illustrating the density values for unthresholded connectivity matrices for all subjects and all time points, grouped by fiber tracking algorithm (Euler,FACT,RK4,TEND) and anatomical label set (AAL,DKT31).

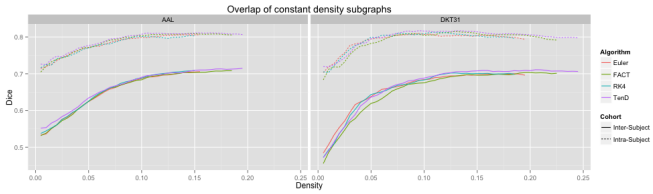

Figure 4. Connectivity matrices were thresholded over a range of density values. At each density level, consistency of network topography was estimated by calculating the mean dice overlap for both intra subject and intersubject pairs.

Table 1. Functional data analysis is used along with permutation testing to look for differences in dice overlap measures between graph generated from different fiber tracking algorithms. Upper triangular p-values are for the AAL labels, while lower triangular are for the DTK31 label set (* indicated significance).

	Intra Subject				Inter Subject			
Euler		0.9555	0.9864	0.6549		0.7770	0.2675	0.0351*
FACT	0.4186		0.7970	0.9524	0.0002*		0.2586	0.0355*
RK4	0.8780	0.6179		0.5358	0.0632	0.0014*		0.1335
TEND	0.3952	0.6655	0.7564		0.0001*	0.0003*	0.0838	
	Euler	FACT	RK4	TEND	Euler	FACT	RK4	TEND

3.3 NETWORK SUMMARY MEASURES OVER GRAPH DENSITY

For each combination of tracking and label set, the mean curves that were calculated to examine how the metrics change as a function of graph density are illustrated in figure 5 along with the ICC curves that quantify reproducibility. Only the characteristic path length curves exhibit a different shape between label sets, and only at low density values. This is likely a results of the smaller number of regions in DKT31 label set. Clustering coefficient, and global and local efficiency exhibit the most similarity across label sets. Comparing within metric and within label set, the fiber tracking algorithms appear consistent as far as shape. Functional data analysis, along with permutation testing does reveal a number of significant

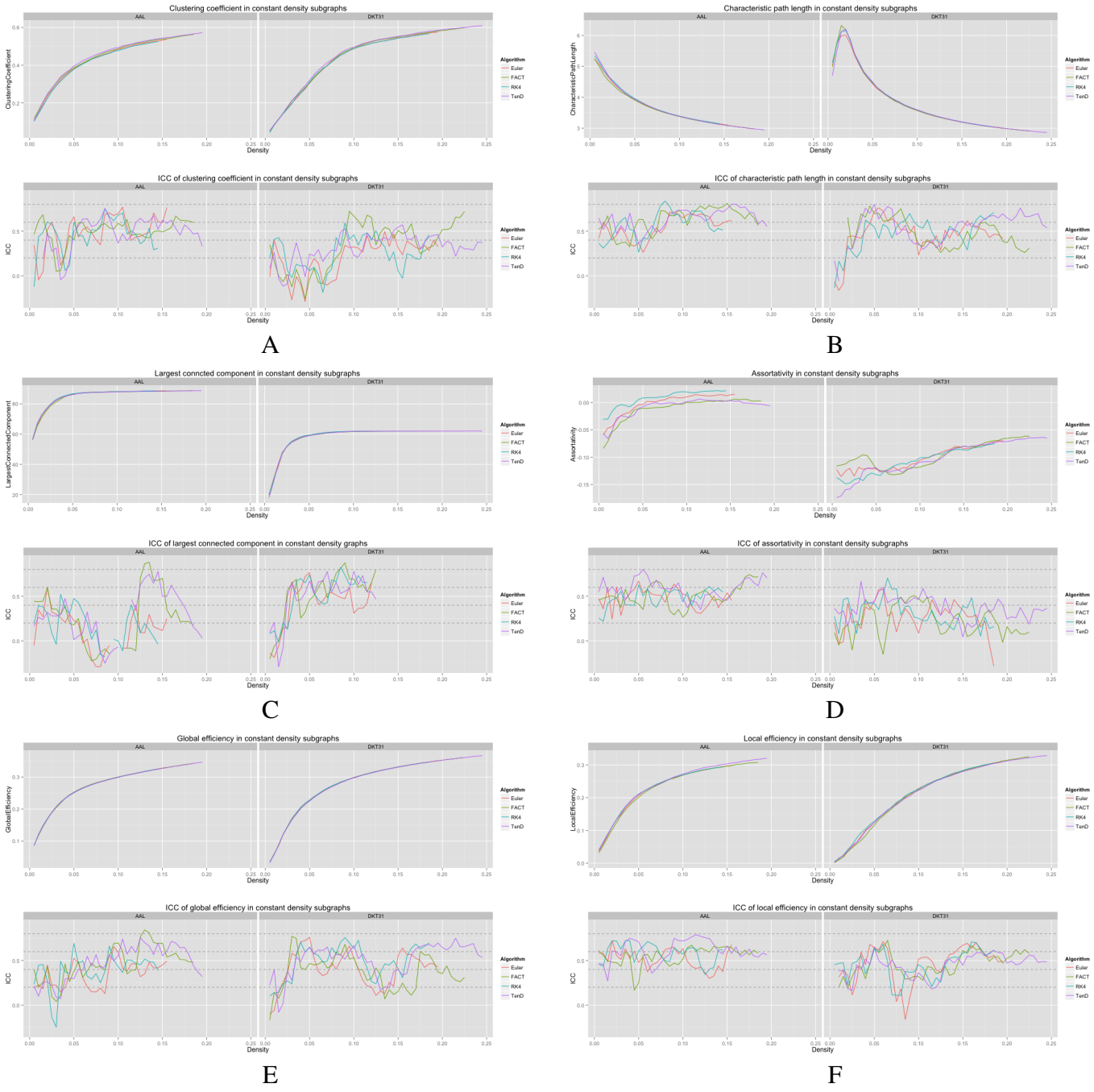

Figure 5. Graph metric vs. graph density plots along with corresponding ICC plots for A) Mean clustering coefficient B) Characteristic path length C) Largest connected component size D) Assortativity E) Global efficiency and F) Local efficiency

differences between graph curves however, as listed in table 2. No significant differences were found between tracking algorithms using the DKT31 labels. Within the AAL labels, significant differences were found between RK4 and TEND for four of the six metrics examined.

Table 2. Functional data analysis is used along with permutation testing to look for pair-wise differences in graph-metric vs. graph-density curves that result from different fiber tracking algorithms and label sets. Only the first time-point for each subject is used. For each metric, the upper-triangular values are for p-values for the AAL labels while the lower-triangular values were generated with the DKT31 label set (* indicated significance).

	Clustering Coefficient				Characteristic Path Length			
Euler		0.2164	0.5296	0.0346*		0.2786	0.2389	0.4728
FACT	0.6962		0.0246*	0.2822	0.9982		0.0145*	0.1235
RK4	0.9927	0.7958		0.0049*	0.8465	0.9199		0.3031
TEND	0.1327	0.8858	0.2061		0.4854	0.6459	0.8234	
	Connected Component Size				Assortativity			
Euler		0.3471	0.9324	0.7556		0.3680	0.3294	0.4651
FACT	0.9447		0.0468*	0.3025	0.6361		0.0270*	0.8877
RK4	0.9998	0.7610		0.4748	0.9326	0.3250		0.0666
TEND	0.7912	0.8336	0.7269		0.8272	0.4021	0.9895	
	Global Efficiency				Local efficiency			
Euler		0.8617	0.9861	0.7272		0.4579	0.9227	0.6065
FACT	0.8677		0.6882	0.7667	0.8794		0.2486	0.1230
RK4	0.6295	0.9415		0.8677	0.9745	0.4413		0.7557
TEND	0.9977	0.9633	0.7438		0.8752	0.7987	0.4775	
	Euler	FACT	RK4	TEND	Euler	FACT	RK4	TEND

3.4 RICH CLUB COEFFICIENT OVER NODE-DEGREE

Because the rich club coefficient requires the selection of multiple parameters, we chose to examine how this metric changes as a function of k, the node degree that determines what is considered a 'rich' node. The plots for the mean graph curves and ICC coefficients are illustrated in figure 6. The results are similar to the examinations over graph density in that the same shape appears for both label sets, but with an scaling difference and the tracking algorithms have similar shapes but within the AAL networks, differences were found in the RK4-FACT (p=0.0338) and RK4-TEND (p=0.0252) comparisons. The p-values for all comparisons are listed in table 3

4 DISCUSSION

4.1 NETWORK TOPOLOGY

Although a number of studies have examined the reproducibility of graph metrics on structural brain networks derived from DTI-based fiber tractography, there are no known papers that focus on the selection of deterministic tracking algorithm. To facilitate later examination of graph metrics as a function of graph density, we first examined the reliability of identifying subgraphs by thresholding. Using the dice coefficient as a measure of overlap we demonstrated that the intra subject agreement was much higher than the inter subject agreement across all tracking algorithms and label sets. No significant differences were found for intra-subject comparisons. The inter-subject comparions indicate that the TEND method most consistently identifies similar subgraphs at a given density. However, further analysis of additional tracking parameters is necessary to determine the full set of conditions under which this result holds.

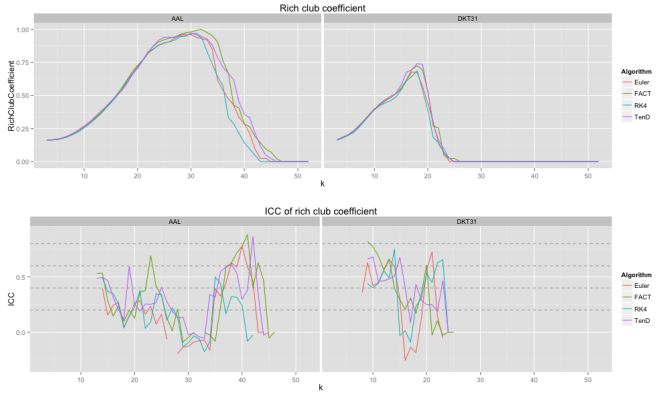

Figure 6. Rich club coefficient was examined over a range of levels, k, and a constant graph density of 0.15

Table 3. Functional data analysis is used along with permutation testing to look for differences in rich club coefficients generated from different fiber tracking algorithms. Upper triangular p-values are for the AAL labels, while lower triangular are for the DTK31 label set (* indicated significance).

Euler		0.3359	0.6830	0.2064
FACT	0.8944		0.0338*	0.8537
RK4	0.9282	0.7219		0.0252*
TEND	0.7548	0.8800	0.3360	
	Euler	FACT	RK4	TEND

4.2 NETWORK SUMMARY MEASURES OVER GRAPH DENSITY

Global and local efficiency are the most robust to choice of fiber tracking algorithm, and have high levels of reproducibility across density levels. Assortativity and characteristic path length are highly reproducible across density levels, but are sensitive to choice of fiber tracking algorithm. In general, the portions of the graph curves at low density value and less reproducible than the segments at high density. For many metrics, the graph curves strongly converge at high density values suggesting that examining the metrics at those densities may be of little use. Examining both the mean graph curves and ICC plots may provide guidance for the range of densities that should be looked at in a group comparison study.

4.3 RICH CLUB COEFFICIENT OVER NODE-DEGREE

The examination of rich club coefficient as a function of degree-level demonstrates the use of graph curves over a parameter other than graph density. Consistency appears to have a somewhat inverse relationship to the coefficient as a function of node degree level. This is a result of the fact that the rich club coefficient values converge at high and low densities. Here, all graphs were thresholded at the maximum density achievable by all graphs. Because network size is held constant here, the average node degree would drop with lowered density, and additional work is required to more understand the relationship between graph-density and degree-level that would provide the most reproducible results.

4.4 LIMITATIONS AND FUTURE DIRECTIONS

The are a number of methodological limitations to the work presented here. We limited the fiber tracking to deterministic methods and used constant shared parameters for these methods. The influence of these parameters on individual tracking algorithms and the resulting graph metrics demands further exploration. In the choice of anatomical label sets, we limited the analysis to a set of manually defined labels, and a often used set of template-based labels. In each case we used the labels 'as-is' without upsampling to a higher number of regions. This may reduce the reproducibility of the networks, but provides more interpretable results if one wishes to examine individual connections or a subset of connections (e.g. default mode network) since the labels have well defined anatomical associations.

An additional limitation of this work is the use of streamline count matrices as the basis for thresholding to create constant density graphs. Multiple options exist for normalizing the streamline count matrices using the volumes of the target cortical regions and/or the average length of the streamlines the connect two regions. The volume based normalization may accomodate the differences that are seen between graph curves that were generated using the different anatomical labels. However, the focus here was on the influence of the fiber tracking and no direct comparisons were made between graph curves generated from the different label sets. A number of additional options exist for creating a weighted connectivity matrix including the average FA of fibers that connect two regions. Since the data set examined also includes magnetization transfer data, the average magnetization transfer ratio along streamlines could potentially be useful as it directly related to myelin content in white matter. These issues were beyond the scope of the current study but would made for an intriguing extension of the current work.

The selection of graph metrics for analysis is another limitation of the study. An exhaustive examination of all possible graph metrics was not feasible so metrics that have been studied previously were chosen to give additional context to existing work. Many of the metrics examined have alternate formulations for weighted graphs. Here, only unweighted graph metrics were examined as they are prevalent in current literature. The creation of a testing framework that relies upon a public data set and open-source code was intended to facilitate the further exploration of the issues listed here.

4.5 CONCLUSION

This study evaluate the reproducibility of graph summary metrics in structural brain networks derived from DTI based deterministic fiber tractography. Four different fiber tracking algorithms were examined along with two different anatomical label set. A number of graph metrics were examined by creating graph curves that capture how a metric changes over a parameter such as graph density. ICC plots were used to evaluate the reproducibility of the metrics and FDA was used to identify significant differences between graph curves generated using different fiber tracking algorithms. While differences between the tracking algorithms were not drastic, they were significant in many cases, suggesting that future studies should give careful consideration to the choice of fiber tracking algorithm based upon the graph metric that will be analyzed.

4.6 DATA SHARING

Free, publicly-available data and software was used throughout. The scripts used to generate the data and figures are available at: `https://github.com/jeffduda/StructConnRepro`. This repository contains the configuration file that, when added to ITK, will download and compile Petiole which builds the executables that were used to generate the graph metrics examined in this study. The template with labels is available at `http://figshare.com/articles/Kirby_multivariate_template/852989`, the final segmentations used as the target regions for fiber tracking are available at `http://figshare.com/articles/MMRR21_DTI_Targets/850369` to provide a convenient starting point for reproducing or extending the methods presented here.

DISCLOSURE/CONFLICT-OF-INTEREST STATEMENT

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

REFERENCES

289 Basser, P. J., Pajevic, S., Pierpaoli, C., Duda, J., and Aldroubi, A. (2000) In vivo fiber tractography using
290 dt-mri data. *Magn Reson Med* 44 625–632.

291 Lazar, M., Weinstein, D. M., Tsuruda, J. S., Hasan, K. M., Arfanakis, K., Meyerand, M. E., et al.
292 (2003) White matter tractography using diffusion tensor deflection. *Hum Brain Mapp* 18 306–321.
293 doi:10.1002/hbm.10102.

294 Hagmann, P., Thiran, J.-P., Jonasson, L., Vandergheynst, P., Clarke, S., Maeder, P., et al. (2003) Dti
295 mapping of human brain connectivity: statistical fibre tracking and virtual dissection. *Neuroimage* 19
296 545–554.

297 Xue, R., van Zijl, P. C., Crain, B. J., Solaiyappan, M., and Mori, S. (1999) In vivo three-dimensional
298 reconstruction of rat brain axonal projections by diffusion tensor imaging. *Magn Reson Med* 42 1123–
299 1127.

300 Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., Wedeen, V. J., et al. (2008) Mapping
301 the structural core of human cerebral cortex. *PLoS Biol* 6 e159. doi:10.1371/journal.pbio.0060159.

302 Zalesky, A., Fornito, A., Harding, I. H., Cocchi, L., Yücel, M., Pantelis, C., et al. (2010) Whole-brain
303 anatomical networks: does the choice of nodes matter? *Neuroimage* 50 970.

304 Cheng, H., Wang, Y., Sheng, J., Sporns, O., Kronenberger, W. G., Mathews, V. P., et al. (2012a) Opti-
305 mization of seed density in dti tractography for structural networks. *J Neurosci Methods* 203 264–272.
306 doi:10.1016/j.jneumeth.2011.09.021.

307 Bastiani, M., Shah, N. J., Goebel, R., and Roebroeck, A. (2012) Human cortical connectome reconstruc-
308 tion from diffusion weighted mri: the effect of tractography algorithm. *Neuroimage* 62 1732–1749.
309 doi:10.1016/j.neuroimage.2012.06.002.

310 Irimia, A. and Van Horn, J. D. (2012) The structural, connectomic and network covariance of the human
311 brain. *Neuroimage* 66C 489–499. doi:10.1016/j.neuroimage.2012.10.066.

312 Sporns, O. (2011) The non-random brain: efficiency, economy, and complex dynamics. *Front Comput*
313 *Neurosci* 5 5. doi:10.3389/fncom.2011.00005.

314 Fornito, A., Zalesky, A., Pantelis, C., and Bullmore, E. T. (2012) Schizophrenia, neuroimaging and
315 connectomics. *Neuroimage* 62 2296–2314. doi:10.1016/j.neuroimage.2011.12.090.

316 Xie, T. and He, Y. (2012) Mapping the alzheimer's brain with connectomics. *Frontiers in Psychiatry* 2.
317 doi:10.3389/fpsyt.2011.00077.

318 Guye, M., Bettus, G., Bartolomei, F., and Cozzone, P. J. (2010) Graph theoretical analysis of structural
319 and functional connectivity mri in normal and pathological brain networks. *MAGMA* 23 409–421.
320 doi:10.1007/s10334-010-0205-z.

321 Martin, G. (2012) Network analysis and the connectopathies: current research and future approaches.
322 *Nonlinear Dynamics Psychol Life Sci* 16 79–90.

323 Honey, C. J., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J. P., Meuli, R., et al. (2009) Predicting
324 human resting-state functional connectivity from structural connectivity. *Proc Natl Acad Sci U S A* 106
325 2035–2040. doi:10.1073/pnas.0811168106.

326 Honey, C. J., Ktter, R., Breakspear, M., and Sporns, O. (2007) Network structure of cerebral cortex
327 shapes functional connectivity on multiple time scales. *Proc Natl Acad Sci U S A* 104 10240–10245.
328 doi:10.1073/pnas.0701519104.

329 Deuker, L., Bullmore, E. T., Smith, M., Christensen, S., Nathan, P. J., Rockstroh, B., et al. (2009)
330 Reproducibility of graph metrics of human brain functional networks. *Neuroimage* 47 1460–1468.

331 Telesford, Q. K., Morgan, A. R., Hayasaka, S., Simpson, S. L., Barret, W., Kraft, R. A., et al. (2010)
332 Reproducibility of graph metrics in fmri networks. *Frontiers in neuroinformatics* 4.

333 Braun, U., Plichta, M. M., Esslinger, C., Sauer, C., Haddad, L., Grimm, O., et al. (2012) Test–retest reli-
334 ability of resting-state connectivity network characteristics using fmri and graph theoretical measures.
335 *Neuroimage* 59 1404–1412.

336 Schwarz, A. J. and McGonigle, J. (2011) Negative edges and soft thresholding in complex network analy-
337 sis of resting state functional connectivity data. *Neuroimage* 55 1132–1146. doi:10.1016/j.neuroimage.
338 2010.12.047.

Liang, X., Wang, J., Yan, C., Shu, N., Xu, K., Gong, G., et al. (2012) Effects of different correlation metrics and preprocessing factors on small-world brain functional networks: a resting-state functional mri study. *PLoS one* 7 e32766.

Weber, M. J., Detre, J. A., Thompson-Schill, S. L., and Avants, B. B. (2013) Reproducibility of functional network metrics and network structure: a comparison of task-related bold, resting asl with bold contrast, and resting cerebral blood flow. *Cogn Affect Behav Neurosci* 13 627–640. doi:10.3758/s13415-013-0181-7.

Cammoun, L., Gigandet, X., Meskaldji, D., Thiran, J. P., Sporns, O., Do, K. Q., et al. (2012) Mapping the human connectome at multiple scales with diffusion spectrum mri. *J Neurosci Methods* 203 386–397. doi:10.1016/j.jneumeth.2011.09.031.

Bassett, D. S., Brown, J. A., Deshpande, V., Carlson, J. M., and Grafton, S. T. (2011) Conserved and variable architecture of human white matter connectivity. *Neuroimage* 54 1262–1279. doi:10.1016/j.neuroimage.2010.09.006.

Dennis, E. L., Jahanshad, N., Toga, A. W., McMahon, K. L., de Zubicaray, G. I., Martin, N. G., et al. (2012) Test-retest reliability of graph theory measures of structural brain connectivity. *Med Image Comput Comput Assist Interv* 15 305–312.

Owen, J. P., Ziv, E., Bukshpun, P., Pojman, N., Wakahiro, M., Berman, J. I., et al. (2013) Test-retest reliability of computational network measurements derived from the structural connectome of the human brain. *Brain Connect* 3 160–176. doi:10.1089/brain.2012.0121.

Vaessen, M. J., Hofman, P. A. M., Tijssen, H. N., Aldenkamp, A. P., Jansen, J. F. A., and Backes, W. H. (2010) The effect and reproducibility of different clinical dti gradient sets on small world brain connectivity measures. *Neuroimage* 51 1106–1116. doi:10.1016/j.neuroimage.2010.03.011.

Cheng, H., Wang, Y., Sheng, J., Kronenberger, W. G., Mathews, V. P., Hummer, T. A., et al. (2012b) Characteristics and variability of structural networks derived from diffusion tensor imaging. *Neuroimage* 61 1153–1164. doi:10.1016/j.neuroimage.2012.03.036.

Telesford, Q. K., Burdette, J. H., and Laurienti, P. J. (2013) An exploration of graph metric reproducibility in complex brain networks. *Front Neurosci* 7 67. doi:10.3389/fnins.2013.00067.

Klein, A. and Tourville, J. (2012) 101 labeled brain images and a consistent human cortical labeling protocol. *Front Neurosci* 6 171. doi:10.3389/fnins.2012.00171.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002) Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage* 15 273–289. doi:10.1006/nimg.2001.0978.

Landman, B. A., Huang, A. J., Gifford, A., Vikram, D. S., Lim, I. A. L., Farrell, J. A. D., et al. (2011) Multi-parametric neuroimaging reproducibility: a 3-t resource study. *Neuroimage* 54 2854–2866. doi:10.1016/j.neuroimage.2010.11.047.

Avants, B. B., Tustison, N., and Song, G. (2009) Advanced normalization tools (ants). *Insight J* .

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., et al. (2010) N4itk: improved n3 bias correction. *IEEE Trans Med Imaging* 29 1310–1320. doi:10.1109/TMI.2010.2046908.

Cook, P., Bai, Y., Nedjati-Gilani, S., Seunarine, K., Hall, M., Parker, G., et al., Camino: Open-source diffusion-mri reconstruction and processing. *14th scientific meeting of the international society for magnetic resonance in medicine* (2006), volume 2759.

Basser, P. J., Mattiello, J., and LeBihan, D. (1994) Estimation of the effective self-diffusion tensor from the nmr spin echo. *J Magn Reson B* 103 247–254.

Salvador, R., Pea, A., Menon, D. K., Carpenter, T. A., Pickard, J. D., and Bullmore, E. T. (2005) Formal characterization and extension of the linearized diffusion tensor model. *Hum Brain Mapp* 24 144–155. doi:10.1002/hbm.20076.

Achard, S., Salvador, R., Whitcher, B., Suckling, J., and Bullmore, E. (2006) A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *The Journal of Neuroscience* 26 63–72.

Bassett, D. S., Meyer-Lindenberg, A., Achard, S., Duke, T., and Bullmore, E. (2006) Adaptive reconfiguration of fractal small-world human brain functional networks. *Proceedings of the National Academy of Sciences* 103 19518–19523.

Rubinov, M. and Sporns, O. (2010) Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* 52 1059–1069. doi:10.1016/j.neuroimage.2009.10.003.
Newman, M. E. J. (2006) Modularity and community structure in networks. *Proc Natl Acad Sci U S A* 103 8577–8582. doi:10.1073/pnas.0601602103.
Bassett, D. S., Bullmore, E., Verchinski, B. A., Mattay, V. S., Weinberger, D. R., and Meyer-Lindenberg, A. (2008) Hierarchical organization of human cortical networks in health and schizophrenia. *J Neurosci* 28 9239–9248. doi:10.1523/JNEUROSCI.1929-08.2008.
Watts, D. J. and Strogatz, S. H. (1998) Collective dynamics of 'small-world' networks. *Nature* 393 440–442. doi:10.1038/30918.
Latora, V. and Marchiori, M. (2001) Efficient behavior of small-world networks. *Physical review letters* 87 198701.
Collin, G., Sporns, O., Mandl, R. C., and van den Heuvel, M. P. (2013) Structural and functional aspects relating to cost and benefit of rich club organization in the human cerebral cortex. *Cerebral Cortex* .
Tustison, N., Yushkevich, P., Song, Z., and Gee, J. (2008) Graph cuts, caveat utilitor, and euler's bridges of konigsberg. *Insight Journal* July http://hdl.handle.net/1926/1503.
Newman, M. E. (2002) Assortative mixing in networks. *Physical review letters* 89 208701.
Dijkstra, E. W. (1959) A note on two problems in connexion with graphs. *Numerische mathematik* 1 269–271.
Colizza, V., Flammini, A., Serrano, M. A., and Vespignani, A. (2006) Detecting rich-club ordering in complex networks. *Nature physics* 2 110–115.
Montgomery, A. A., Graham, A., Evans, P. H., and Fahey, T. (2002) Inter-rater agreement in the scoring of abstracts submitted to a primary care research conference. *BMC Health Services Research* 2 8.