Frontiers in Neurosciences November 13, 2013

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. These graph metrics typically derive from fiber tractography, however 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 young adults. Network summary measures are examined using the intraclass correlation coefficient (ICC), and the dice coefficient is used to examine overlap of constant density subgraphs. Functional data analysis techniques are used to examine differences in network measures that result from the choice of fiber tracking algorithm. 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 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; **?**). The use of graph theoretical analysis to study the topology of these structural networks has increasingly been used to examine the structural consequences of neurological disorders (Xie and He, 2012; **?**) as well as the relationship between structure and function ().

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; **?**) and arterial spin labeling (**?**). 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). DTI-based studies using deterministic tractography have included the examination of tractography seed density (Cheng et al., 2012), anatomic label density (Bassett et al., 2011), and studies examining a variety of network measures (Cheng et al., 2012; Irimia and Van Horn, 2012).

In the paper we constrain our analysis DTI-based deterministic fiber tractography. Within this constraint, we examine multiple algorithms for computing streamlines and their required parameters 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 (**?**). The intraclass correlation coefficient is used to examine the reproducibility of network summary measures that results from combinations of fiber tracking algorithm and anatomical label set. 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) 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.

2.2 ANATOMICAL LABELING

The Mindboggle dataset provides a set of manually drawn cortical regions (DKT31) and a skull-stripped image for a single time point for each subject (Klein and Tourville, 2012). To utilize these labels in network creation we must perform an intrasubject registration between T1 images and an inter modality registration from the T1 to DTI space for each time point. An affine registration is performed between T1 images. The brainmask is then warped into the non-labeled time point. The Mindboggle brainmask include cerebellum and are applied to each T1 data set. The b=0 volume, acquired as part the diffusion tensor acquisition sequence is registered to the T1 image from the same acquisition. By composing these transforms, the DKT31 labels are transferred into the DTI space for each time point using nearest neighbor interpolation.

In addition to the DKT31 labels, the AAL label set was examined, due to it's prevalence in the examination of both structural and functional connectivity (Tzourio-Mazoyer et al., 2002). A similar registration-based approach is used to transform the AAL label in the DTI space. An existing multivariate template had been created from the MMRR dataset using the antsMultivariateTemplateConsturction.sh. Here we only use the T1 component of this population-specific template. The AAL labels were transferred into the space of this template via template-to-template registration using the antsRegistration tool. AAL labels that lie outside of the cerebrum were excluded from later analysis. In order to transform these labels into each subject's T1 space, the antsCorticalThickness.sh tool was used. This software first applies bias correction using the N4 algorithm. Next a registration based skull stripping is performed to provide a cerebrum mask of the T1 image. This is followed by a final cerebrum-only registration to the template. Composing the appropriate transforms again provides a method to transfer these labels in the DTI space for each time point. As the AAL labels include both gray and white matter, they are masked to only include voxels that have been identified as cortical gray matter by the DKT31 labels.

2.3 DIFFUSION DATA PREPROCESSING

The Camino toolkit () is used to calculate diffusion tensor images via a weighted linear fitting (**??**), and is also used for subsequent deterministic tractography. The brainmasks defined in anatomical space are

warped into DTI space and are used to restrict the tractography to eliminate false positives that could results from streamlines that leave and then reenter the brain. Fractional anisotropy (FA) images are calculated and a tractography seed-map is created by including 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 (**?**).
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 entire diffusion tensor is used to deect the estimated ber trajectory (Lazar et al., 2003)

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 in 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

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.

To compare graphs it is necessary to first ensure that the graphs are of equal density, 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. To obtain constant density graphs, the M highest weighted edges are retained and all other edges are removed where M is fully determined by the desired density and number of nodes in the 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. Here, we only directly compare measures obtained from graphs with an equal number of nodes.

2.5 NETWORK METRICS

A large number of metric 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 (**?**), assortativity(**??**), 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(**?**). 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 (**?**) and provides ITK functions for a variety of graph metrics while using the matlab-based Brain Connectivity Toolkit (**?**) 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 **?**.

Assortativity. Assortativity measures how preferentially nodes of similar degree connect to one another (**?**) 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 $E = \|E(G)\|$.

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 (**?**). 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. However, this metric is defined for both connected 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 (**?**).

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

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 graph curves 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. 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 reproducibly 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:

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

where σ^2_{bs} is the between-subject variance and σ^2_{ws} 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; **?**). 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 which treats each curve as a function. While there are a variety of methods for computing the difference between two curves, here we choose the simplest method, the non-parametric permutation test. We first treat each mean curve as a function and find the area between a pair of curves. We then permute group assignment of individuals used to calculate the mean curves using random sampling without replacement. We then calculate new mean curves from the randomly assigned groups and measure the area between these curves. This it performed iteratively (i=100000). We record, 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.

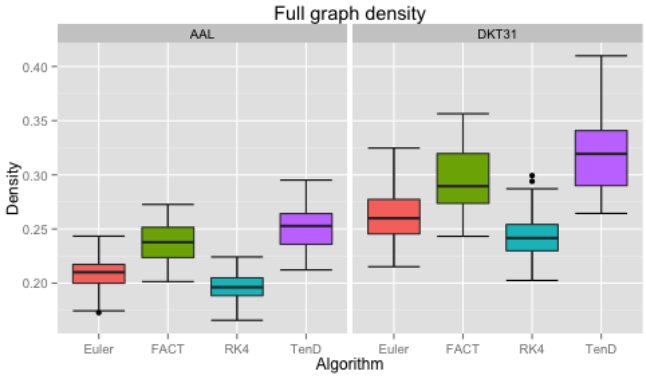

Figure 1. Boxplots illustrating the density values for unthresholded connectivity matrices for all subjects and all time points, grouped by fiber tracking algorithm (Euler,FACT,RK4,TEND) and anatomical label set (AAL,DKT31).

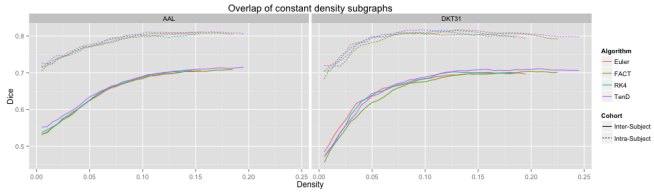

Figure 2. Connectivity matrices were thresholded over a range of density values. At each density level, consistency of network topography was estimated by calculating the mean dice overlap for both intra subject and intersubject pairs.

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 1. 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 2. 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.

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 3 along with the ICC curves that quantify reproducibility. While no direct comparisons were done between graph curves generated from different label sets, the scales of the plots are equal to allow for convenient visual comparison. 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

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 values are for the AAL labels, while lower triangular are for the DTK31 label set.

	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

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.

	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

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 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.

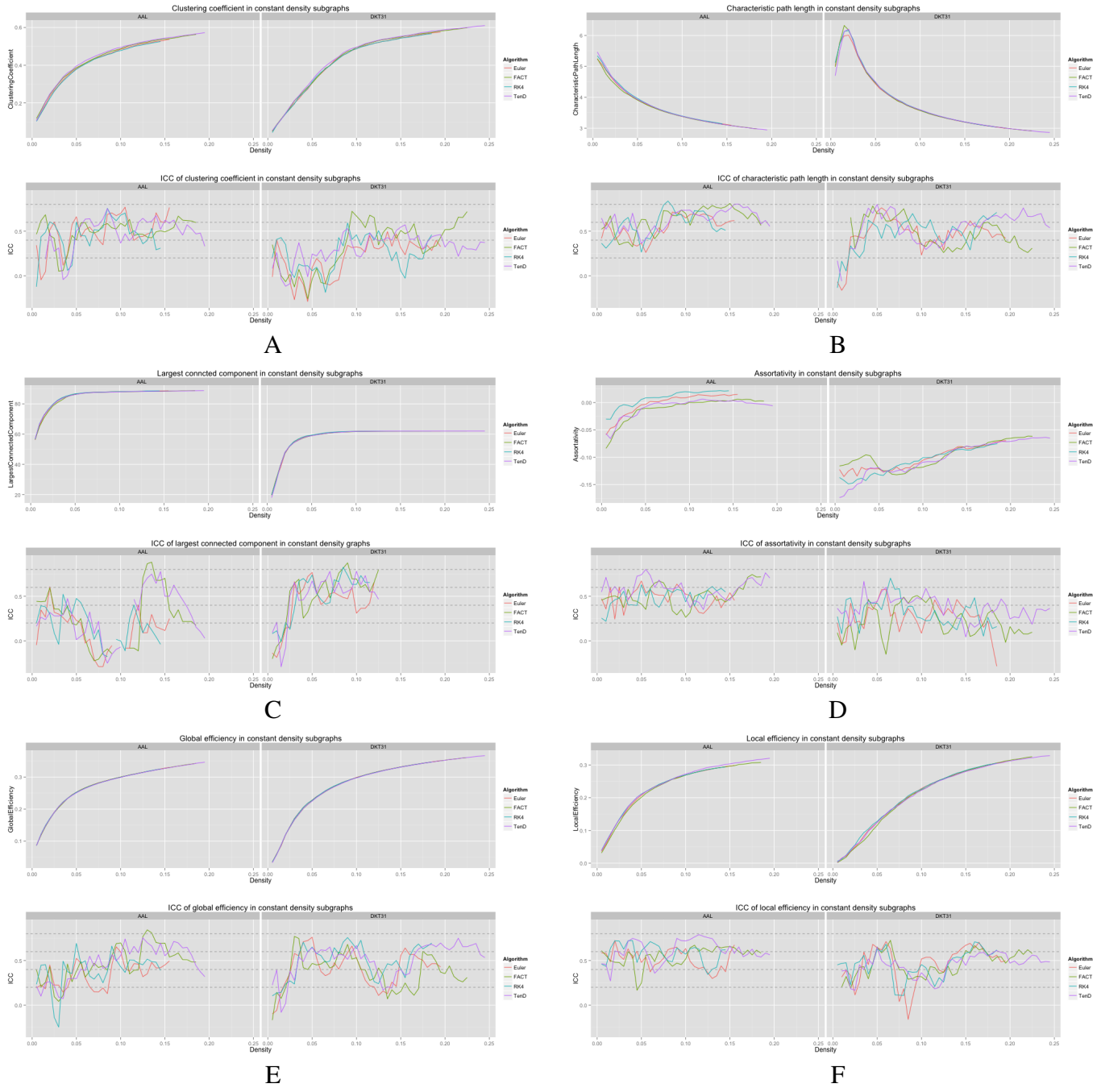

Figure 3. Graph metric vs. graph density plots along with corresponding ICC plots for A) Mean clustering coefficient B) Characteristic path length C) Largest connected component size D) Assortativity E) Global efficiency and F) Local efficiency

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 4. The results are similar to the examinations over graph density in that the same shape appears for both label sets, but

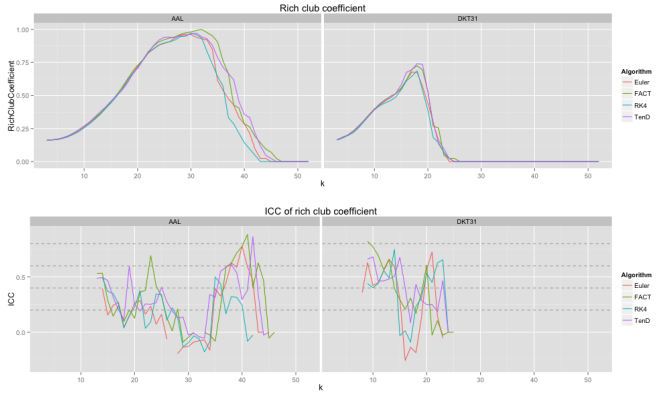

Figure 4. Rich club coefficient was examined over a range of levels, k, and a constant graph density of 0.15

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 values are for the AAL labels, while lower triangular are for the DTK31 label set.

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

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. Using the AAL label set we demonstrated differences in topology between FACT and TEND which suggest that the TEND algorithm may provide a more reproducible measure of network topology across subjects. This is significant as the AAL labels are prevalent throughout all forms of connectivity studies and the FACT algorithm is the most common choice among deterministic tractography methods. However, further analysis of additional tracking parameters is necessary to determine the full set of conditions under which this difference holds.

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 reprocibily across density levels. Assortativity and characteristic path length are highly repoducible 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.

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 x 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 densitites. Here, all graphs were thresholded at the maximum density achievable by all graphs. As the average node degree would drop with lowered density, additional work is required to more thoroughly 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 label 'as-is' without upsampling to a higher number of regions.

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 inluding the averge FA of fibers that connect two regions. Since the data set examined also includes magenetization 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 prevelant 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 differenent 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 reprociblity of the metrics and FDA was used to indentify significant differences between graph curves generated using different fiber tracking algorithms. While differenences 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. Also, the final segmentations used as the target regions for fiber tracking are included in this repository 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

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

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

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

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

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

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

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

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

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

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.

289 Owen, J. P., Ziv, E., Bukshpun, P., Pojman, N., Wakahiro, M., Berman, J. I., et al. (2013) Test-retest reli-
290 ability of computational network measurements derived from the structural connectome of the human
291 brain. *Brain Connect* 3 160–176. doi:10.1089/brain.2012.0121.
292 Vaessen, M. J., Hofman, P. A. M., Tijssen, H. N., Aldenkamp, A. P., Jansen, J. F. A., and Backes,
293 W. H. (2010) The effect and reproducibility of different clinical dti gradient sets on small world brain
294 connectivity measures. *Neuroimage* 51 1106–1116. doi:10.1016/j.neuroimage.2010.03.011.
295 Cheng, H., Wang, Y., Sheng, J., Kronenberger, W. G., Mathews, V. P., Hummer, T. A., et al. (2012) Char-
296 acteristics and variability of structural networks derived from diffusion tensor imaging. *Neuroimage* 61
297 1153–1164. doi:10.1016/j.neuroimage.2012.03.036.
298 Irimia, A. and Van Horn, J. D. (2012) The structural, connectomic and network covariance of the human
299 brain. *Neuroimage* 66C 489–499. doi:10.1016/j.neuroimage.2012.10.066.
300 Klein, A. and Tourville, J. (2012) 101 labeled brain images and a consistent human cortical labeling
301 protocol. *Front Neurosci* 6 171. doi:10.3389/fnins.2012.00171.
302 Landman, B. A., Huang, A. J., Gifford, A., Vikram, D. S., Lim, I. A. L., Farrell, J. A. D., et al. (2011)
303 Multi-parametric neuroimaging reproducibility: a 3-t resource study. *Neuroimage* 54 2854–2866. doi:
304 10.1016/j.neuroimage.2010.11.047.
305 Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002)
306 Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of
307 the mni mri single-subject brain. *Neuroimage* 15 273–289. doi:10.1006/nimg.2001.0978.
308 Watts, D. J. and Strogatz, S. H. (1998) Collective dynamics of 'small-world' networks. *Nature* 393
309 440–442. doi:10.1038/30918.
310 Latora, V. and Marchiori, M. (2001) Efficient behavior of small-world networks. *Physical review letters*
311 87 198701.
312 Motter, A. E., Zhou, C., and Kurths, J. (2005) Enhancing complex-network synchronization. *EPL*
313 *(Europhysics Letters)* 69 334.
314 Newman, M. E. (2002) Assortative mixing in networks. *Physical review letters* 89 208701.
315 Ravasz, E. and Barabási, A.-L. (2003) Hierarchical organization in complex networks. *Physical Review*
316 *E* 67 026112.
317 Achard, S. and Bullmore, E. (2007) Efficiency and cost of economical brain functional networks. *PLoS*
318 *Comput Biol* 3 e17. doi:10.1371/journal.pcbi.0030017.
319 Colizza, V., Flammini, A., Serrano, M. A., and Vespignani, A. (2006) Detecting rich-club ordering in
320 complex networks. *Nature physics* 2 110–115.
321 van Wijk, B. C. M., Stam, C. J., and Daffertshofer, A. (2010) Comparing brain networks of different size
322 and connectivity density using graph theory. *PLoS One* 5 e13701. doi:10.1371/journal.pone.0013701.
323 Rubinov, M. and Sporns, O. (2010) Complex network measures of brain connectivity: uses and
324 interpretations. *Neuroimage* 52 1059–1069. doi:10.1016/j.neuroimage.2009.10.003.