Published in final edited form as:

J Neurosci Methods. 2012 January 15; 203(1): 264–272. doi:10.1016/j.jneumeth.2011.09.021.

Optimization of seed density in DTI tractography for structural networks

Hu Cheng^{a,*}, Yang Wang^b, Jinhua Sheng^b, Olaf Sporns^a, William G. Kronenberger^c, Vincent P. Mathews^b, Tom A. Hummer^c, and Andrew J. Saykin^{a,b,c}

^aDepartment of Psychological and Brain Sciences, Indiana University, Bloomington, IN 47405, USA

^bCenter for Neuroimaging, Department of Radiology and Imaging Sciences, Indiana University School of Medicine, Indianapolis, IN 46202, USA

^cDepartment of Psychiatry, Indiana University School of Medicine, Indianapolis, IN 46202, USA

Abstract

Diffusion tensor imaging (DTI) has been used for mapping the structural network of the human brain. The network can be constructed by choosing various brain regions as nodes and fiber tracts connecting those regions as links. The structural network generated from DTI data can be affected by noise in the scans and the choice of tractography algorithm. This study aimed to examine the effect of the number of seeds in tractography on the variance of structural networks. The variance of the network was characterized using an approach similar to the National Electrical Manufacturers Association (NEMA) standards for measurement of image noise. It was shown that the variance of the network is inversely related to the square root of seed density. Consequently, the number of seeds has a large impact on local characteristics and metrics of the network architecture. As the number of seeds increased, increased stability of structural network metrics was observed. However, more seeds can also lead to more spurious fibers and thus affect nodal degrees and edge weights, and proper thresholding may be necessary to create an appropriate weighted network. Because the variance of the network is also influenced by other imaging factors, further increase in the number of seeds has little effect in reducing the network variance. The selection of the seed number should be a balance between the network variance and computational effort.

Keywords

Diffusion tensor imaging; Structural brain network; Tractography; Seeds; Variance

1. Introduction

Fiber tractography analysis of diffusion tensor imaging (DTI) scan data has been widely used to map white matter fiber bundles. A recent application of this technique was in the construction of the structural network of the human brain, by which the topological architecture of the human brain can be revealed non-invasively (Bassett et al., 2010; Gong et al., 2009). Mathematically a network can be characterized as a collection of nodes and links (also called edges) between pairs of nodes (Sporns, 2010). For the human brain network, the nodes are often identified as anatomically parcellated regions, while the edges of the

network are the fiber pathways between those regions. Depending on the property of the edge, a network can be classified as binary or weighted. A network is called a weighted network if different weights are assigned to the edges to characterize the connectivity strengths. Alternatively, a network is called a binary network if the weights are either 1 or 0, indicating connected or not connected, respectively. Weighted networks have previously been used to describe the inter-cortical connections of the human brain, in which weights were calculated from the tractography data (e.g. Hagmann et al., 2008). Weights of fiber pathways are determined by tractography which is typically classified as either deterministic, meaning any two voxels are either connected by streamlines or not, or probabilistic, in which the connection between any two voxels is characterized by a probability (Mori and van Zijl, 2002). The deterministic approach is frequently employed by the researchers for quantifying structural networks on account of its simplicity and computation speed. With deterministic tractography, the number of fibers can be directly counted. The fiber density (number of fibers normalized by seed density) is one important measure to characterize the strength of inter-regional connections, and thus the topology of a weighted structural network. Despite distinct weighting schemes adopted by different groups (Bassett et al., 2010; Hagmann et al., 2007, 2010a; Lo et al., 2010; Shu et al., 2011), the number of fibers is generally recognized as a key component in the definition of network weights. Hence, the reliability of the constructed network can be significantly influenced by the performance of the fiber tracking algorithm.

In a deterministic tractography approach, the tracking starts from a seed selected in a voxel. The seed can be selected from anywhere within a voxel. A fiber is generated and propagated from that seed following some rules. The tracking process stops if certain deterministic criteria are met. Usually, one or more seeds are selected in each voxel for whole brain tractography. The outcome of the tracking is influenced by many factors such as the signal to noise ratio (SNR) of the DTI data and seeds locations. Because of limited SNR of the DTI data, the uncertainty of fiber tracking can be accumulated in the fiber tracking process, resulting in dispersions of the projectile of the fiber path (Anderson, 2001; Lazar and Alexander, 2003). A small perturbation to the tracking condition such as different noise or initial seed positions will result in different fibers. To reduce the variability of fibers caused by seed positions, more seeds should be randomly distributed within the voxels. Although it has been observed that a larger number of fibers leads to a larger number of edges (Hagmann et al., 2007), the effect of the number of seeds on the reliability of measures characterizing structural networks has not yet been addressed. In addition, network metrics can quantify different aspects of network topology using local metrics (e.g. degree, strength, clustering) and global metrics (e.g. paths and distances), and differential effects on these metrics warrant investigation.

A major potential application of network analysis is the comparison of different groups that have potentially altered network structures. For instance, topological differences reflected in the clustering coefficient and path length for schizophrenia patients in comparison to a healthy control group have been identified (van den Heuvel et al., 2010; Zalesky et al., 2011). In addition, alteration of structural connectivity was recently observed in the developing human brain (Hagmann et al., 2010b). Two recent articles have used just one seed in the deterministic fiber tracking (Bassett et al., 2010; Gong et al., 2009), which could introduce significant variances in the final networks. In contrast, Hagmann et al. used 30 seeds in their fiber tracking of diffusion spectrum imaging data (Hagmann et al., 2007). To improve reliability and reduce intra-subject variance a larger number of seeds originating from each voxel in tractography may be desired in order to reduce the variance from fiber tracking. Since more seeds generate more fibers and consequently lengthen the computation time for network construction significantly, knowing how many seeds are sufficient for a specific application can be very useful in practice. We hypothesize that, theoretically, the

variance of the number of tracks connecting two brain regions is proportional to the number of seeds per voxel (NSPV). To validate the relation, structural networks were constructed from real DTI data of human subjects. The NSPV was varied from 1 to 40 in the tractography algorithm. The nodes were obtained from gyral-based parcellation of high resolution anatomical images. We also propose a simple approach to evaluate the relation between the number of random seeds and the variance of the structural network. A practical method is proposed to estimate optimal NSPV for fiber tracking of DTI data. The effect of NSPV on network metrics is also discussed.

2. Theory

Streamline based tractography selects a seed in a voxel and then propagates a fiber from that seed. The seed can be placed anywhere within a voxel. For given nodes, a seed may end up with a fiber connecting two nodes, or a fiber not connecting any pairs of nodes, or no fiber at all. It has been shown that the dispersion of a fiber appears to be linearly related to $1/\text{SNR}$ (Anderson, 2001; Lazar and Alexander, 2003). Because of this uncertainty of fibers, by placing a large number of seeds randomly within one voxel, we may obtain a set of fibers crossing that voxel that show some statistical distribution. Each constructed fiber is associated with a single seed of origin, but each seed may not necessarily end up with a fiber. With multiple seeds for each voxel, the mapping between fibers and voxels is typically many-to-one.

Now we consider the number of fibers between any two regions of interest (ROIs). We can denote the probability for a fiber connecting two ROIs (ROI_i and ROI_j) through voxel k as P_{ij}^k . Then the number of fibers connecting ROI_i and ROI_j (F_{ij}) using N random initial seeds in each voxel is a product of N and the sum of the probability of all voxels, as shown in Eq. (1).

$$\langle F_{ij} \rangle = \sum_k N P_{ij}^k. \quad (1)$$

The variance can be calculated from the statistics of a binomial distribution:

$$\sigma_{ij}^2 = \sum_k N P_{ij}^k (1 - P_{ij}^k). \quad (2)$$

Therefore, the standard deviation of fiber density between any two ROIs is a function of N (Eq. (3)).

$$\frac{\sigma_{ij}}{N} = \frac{\beta_{ij}}{\sqrt{N}}, \quad (3)$$

where $\beta_{ij} = \sqrt{\sum_k P_{ij}^k (1 - P_{ij}^k)}$ is a constant determined by the probabilities. If using the weighting scheme proposed by Bassett et al. (2010), the weight is the number of fibers. Then the standard deviation of the weights follows Eq. (3). If using the definition of weights proposed by Hagmann et al. (2007), the weight between any two ROIs (nodes) is defined as the fiber density scaled by the volume of the two ROIs, which is mathematically described in Eq. (4).

$$w_{ij} = \frac{2}{N(n_i + n_j)} \sum_m \frac{1}{L_{ij}^m}, \quad (4)$$

where n_i denotes the number of voxels in ROI_i, and L_{ij}^m denotes the length of the m th track between ROI_i and ROI_j. Because longer fibers have more seeds in it and therefore produce more tracts from the fiber tracking, the $1/L_{ij}^m$ term can effectively offset this bias.

Assuming the mean length between the two ROIs is L_0 , then the weights w_{ij} is approximately proportional to the ratio between F_{ij} and L_0 (Eq. (5)).

$$w_{ij} = \frac{2}{N(n_i + n_j)} \sum_m \frac{1}{L_0 + \Delta L_{ij}^m} \approx \frac{2}{N(n_i + n_j)} \sum_m \left(\frac{1}{L_0} + \frac{\Delta L_{ij}^m}{L_0^2} \right) \quad (5)$$

$$\approx \frac{2F_{ij}}{N(n_i + n_j)L_0}$$

Therefore, its standard deviation is also inversely related to $\sqrt{N}$, as the standard deviation of number of tracks described in Eq. (3).

A network is composed of nodes and edges. Assuming that the nodes are fixed, the variance of a network comes from the variance of the weights. Although the variance of weights may be different for each edge, the variance of a network can be characterized using a single estimate based on the principle employed in the National Electrical Manufacturers Association (NEMA) standards for measuring SNR of MRI images (NEMA, 2008). In the primary measurement procedure for SNR, two sets of images are acquired under the same conditions. A difference image is obtained by performing a subtraction of the two images. The noise can be calculated from the standard deviation of a ROI in the difference image. Similar to noise quantification in NEMA standard, a difference network can be derived by subtracting the weights between corresponding edges of two networks obtained separately but under the same conditions. To avoid confusion with the term ‘variance’ in probability and statistics, a new term ‘Variance Of Network’ (VON) is introduced here. The variance of network is defined as the standard deviation of all the edge weights in the difference network, normalized by the mean of all edge weights. As a simple example, given two networks

$$A = \begin{pmatrix} 0 & 2.0 & 3.0 \\ 2.0 & 0 & 4.0 \\ 3.0 & 4.0 & 0 \end{pmatrix}, B = \begin{pmatrix} 0 & 2.1 & 3.2 \\ 2.1 & 0 & 3.9 \\ 3.2 & 3.9 & 0 \end{pmatrix}, \quad \text{the difference is} \quad \text{dif} = \begin{pmatrix} 0 & 0.1 & 0.2 \\ 0.1 & 0 & -0.1 \\ 0.2 & -0.1 & 0 \end{pmatrix}$$

The standard deviation of the edge weights is 0.137 and the mean edge weight is 3.0. Then we have $VON = 0.137/3.0 = 0.046$.

The NEMA approach can also be used to investigate the variance from a specific cause. For instance, with two DTI scans acquired with the same protocol, the inter-scan variability of a network from DTI images can be explored. VON is calculated from the two networks constructed from the two DTI image sets using the same fiber tracking methods and same parcellation.

To obtain seed-related VON denoted as VON_s , two networks *Net1* and *Net2* can be constructed in the same way by performing fiber tracking twice on the same DTI data. A network can be represented by a matrix where the elements are the weights. Then we have

$$VON_s = \frac{\text{std}(Net1 - Net2)}{\langle Net1 + Net2 \rangle / 2}. \quad (6)$$

The operation in Eq. (6) only applies to the non-zero elements in matrix *Net1* and *Net2*. Because binomial distribution for large number can be approximated by a Gaussian distribution, each non-zero element in the difference matrix follows a Gaussian distribution of zero mean. If the variance is the same for all the elements, then VON_s is proportional to $1/\sqrt{N}$ based on Eqs. (3) and (5). Practically the variance is edge-dependent, i.e., the variance of the elements may differ from each other. Because the variance has a limited range, the weights can be re-ordered based on the value of variance, i.e.,

$$N_w = \int_{\sigma_{\min}}^{\sigma_{\max}} n(\sigma) d\sigma, \quad (7)$$

where N_w is the total number of nonzero weights, $n(\sigma)$ is the density of edges at standard deviation of σ , which varies from $\sigma_{\min}$ to $\sigma_{\max}$. As an approximation, Eq. (7) can be rewritten in a discrete format by dividing the weights into m groups with the variances in each group being close to each other. We have

$$\sum_{j=1}^m n_j = N_w, \quad (8)$$

with n_j denoting the number of weights in j th group. Assuming the standard deviation in the j th group can be approximated by a constant σ_j , then the variance (σ^2) of nonzero elements of the difference matrix is approximately equal to the sum of the variances of all the groups:

$$\sigma^2 = \frac{\sum_{j=1}^m n_j \sigma_j^2}{\sum_{j=1}^m n_j}. \quad (9)$$

Because σ_j should be a function of N similar to Eq. (3),

$$\sigma_j = \beta_j \sqrt{N}, \quad (10)$$

where β_j is a constant for group j , then we have

$$\sigma^2 = \frac{N \sum_{j=1}^m n_j \beta_j^2}{\sum_{j=1}^m n_j}. \quad (11)$$

Therefore, the VON should also exhibit a linear relationship with $1/\sqrt{N}$

$$VON_s \propto \frac{1}{\sqrt{N}}. \quad (12)$$

The integrity of this linear relationship depends on the number of edges (the larger the better) and the range of variance (the smaller the better).

Besides seed-related variance, the constructed network involves several procedures that may introduce errors and variances. The DTI images have noises and artifacts (Rohde et al., 2004). The fiber tracking can introduce errors because of algorithm hypotheses, limitation of

spatial resolution, partial volume effects, and poor SNR (Mori and van Zijl, 2002). There are errors in node ROIs as well. For instance, the parcellation can have some errors due to contrast or SNR limitations of the T1 weighted image and the algorithm itself. In addition, the registration between T1-weighted image and DTI can also have errors due to image distortion and partial volume effects. Each of these factors will affect the reliability of a network. The seed-related variance can be reduced by increasing the number of seeds according to Eq. (12), but will have little effect up to a point when variance of other sources becomes dominant. We propose that the total VON can be decomposed as seed-related VON and VON from other factors independent of seed selection. The relation can be written as

$$VON_t^2 = VON_s^2 + VON_o^2, \quad (13)$$

where VON_t denotes the total VON, and VON_o the VON from all other factors. If $VON_t \gg VON_s$, there is no need to increase the number of seeds, so this can be used as a guideline for choosing optimal NSPV. While VON_t cannot be calculated from the DTI data in a straightforward way, it can be estimated using the NEMA approach by splitting the data into two subsets. DTI data for tractography typically have more than 20 gradient directions. The data set can be divided into two groups with balanced gradient directions, and two networks can be constructed from the two DTI subsets. Then VON_t can be calculated, and the value is not much different than that for the full dataset (Cheng et al., 2011). We use the following criterion to determine the optimal number of seeds:

$$VON_s^2 < 0.1 \times VON_t^2. \quad (14)$$

Because of Eqs. (12) and (13), one only needs to calculate VON_s and VON_t for one specific NSPV. If we choose NSPV = 10 to balance between variability of VON and computational time, the optimal NSPV is determined by

$$NSPV \geq 100VON_{s=0}^2 / VON_t^2. \quad (15)$$

3. Methods

3.1. MRI Images

Ten young male adults with an average age of 24.0 ± 3.2 years were included in this study. They were all healthy volunteers with no history of neurological and psychiatric disorders. The MR data were acquired on a 3.0T TIM Trio scanner using a 12 channel head coil. A SE-EPI DTI sequence was performed using the following parameters: matrix = 128×128 ; FOV = $256 \text{ mm} \times 256 \text{ mm}$; TR/TE = 8300/77 ms; 68 transversal slices with 2 mm thickness; 48 diffusion gradient directions with $b = 1000 \text{ s/mm}^2$, and 8 samplings at $b = 0$. In addition, each session included a high resolution T1-weighted MP-RAGE imaging as anatomical reference for subsequent parcellation and coregistration.

The DTI data was first preprocessed with FDT toolbox of FSL (<http://www.fmrib.ox.ac.uk/fsl/>) to correct for artifacts induced by head motion and eddy current. This was done by registering all the image volumes to the first b0 image via an affine transformation. The processed DTI images were then output to Diffusion Toolkit (<http://trackvis.org/>) for tensor reconstruction and fiber tracking, using the FACT (fiber assignment by continuous tracking) streamline tractography algorithm (Mori et al., 1999). The FACT algorithm initializes fibers from many seed points randomly distributed within each voxel all over the brain and propagates these fibers along the vector of the largest principle axis of diffusion tensor within each voxel until certain termination criteria are met. In our case, stop angle threshold

was set to 35°, which means if the angle change between two voxels is greater than 35°, the tracking process stops. A spline filtering was then applied to smooth the tracks. For a given DTI data set, we varied the number of random seeds from 1 to 40 (i.e., 1, 5, 10, 15, 20, 25, 30, 35, 40) and two tractography data sets were computed for each number.

Anatomical parcellation was performed using FreeSurfer 4.5 (<http://surfer.nmr.mgh.harvard.edu/>) on the high-resolution T1-weighted anatomical image acquired with MP-RAGE sequence. The parcellation was an automated operation on each subject to obtain 68 gyral-based ROIs, with 34 cortical ROIs in each hemisphere. The T1-weighted anatomical image was registered to the low resolution b0 image of DTI data using the FLIRT toolbox in FSL, and the warping parameters were applied to the ROIs so that a new set of ROIs in the DTI image space was created. These new ROIs were used for constructing the structural network.

3.2. Network construction

The topological representation of a network is a collection of nodes and edges between pairs of nodes. In constructing the weighted, undirected network, the nodes were chosen to be the 68 registered ROIs originally obtained from FreeSurfer parcellation. The weight of the edge was defined as the density of the fibers connecting a pair of nodes, which is the number of tracks between two ROIs divided by the mean volume of the two ROIs (Hagmann et al., 2007). Eq. (1) shows that the number of fibers is linearly related to the number of seeds per voxel N . A scaling factor of N was applied in the network construction to offset that effect. The weights were computed according to Eq. (16).

$$w_{ij} = \frac{2}{N(n_i + n_j)} \sum_m \frac{1}{L_{ij}^m}, \quad (16)$$

where n_i denotes the number of voxels in ROI _{i} and L_{ij}^m denotes the length of the m th fiber between ROI _{i} and ROI _{j} . A fiber is considered to connect two ROIs if and only if its end points fall in the two ROIs respectively. The weighted network can be represented by a matrix M_w . The rows and columns correspond to the nodes, and the elements of the matrix correspond to the weights. Hence, $M_w(i,j)$ represents the weight between nodes i and j . Similarly a matrix was also obtained for the number of fibers (M_{nf}) and the mean fiber length (M_{fl}), with $M_{nf}(i,j)$ representing number of fibers and $M_{fl}(i,j)$ representing mean fiber length between nodes i and j .

3.3. Variance of network

For each number of random seeds, two sets of tractography data were obtained. The variance of network was computed for both M_{nf} and M_w . VON of M_{nf} characterizes the variation of fiber quantity and VON of M_w characterizes the variation of network weights. The VONs are expected to be linearly related to the square root of NSPV, as shown in Eq. (17).

$$y_j^N = \frac{C_j}{\sqrt{N}}, \quad (17)$$

where y_j^N denotes the VON of M_{nf} or M_w for N seeds per voxel and subject j , C_j was a constant for subject j and $N = 1, 5, 10, 15, 20, 25, 30, 35, 40$. The coefficient C_j was different for different subject and can be solved as

$$C_j = \frac{\sum_j^N \sqrt{N}}{9}. \quad (18)$$

Eq. (17) can be rewritten as

$$Y_j^N \equiv \frac{VON_j^N}{\left(\sum_j^N \sqrt{N}\right)/9} \approx \frac{1}{\sqrt{N}}. \quad (19)$$

The scaled VON_j^N was plot as a function of the number of random seeds. The same analysis was conducted on the data for all ten human subjects.

The fluctuation of M_{nf} or M_w can also be characterized by the correlation of results from the two tractography data. For each N , cross-correlation coefficients between two M_{nf} or M_w matrices were calculated for all the non-zero elements. The correlation coefficients were then plot against the NSPV.

To estimate how much seed-related variance contributed to the total variance and examine the decomposition hypothesis described by Eq. (13), the 48 direction DTI data were divided into two groups, each with 24 diffusion weighted images and four b0 images. The gradient directions are counter-balanced. Fiber tracking was performed the same way as described before by applying four different numbers of seeds (1, 2, 5, 10) twice. Networks were constructed for each track data, resulting in 16 networks per subject (two subsets, four NSPVs per subset, two fiber trackings per NSPV). The seed-related VONs were computed for each NSPV and subset, and a mean value was taken from two subsets. The total VONs were computed from the corresponding networks from two subsets. A mean value was taken from four VONs for every NSPV.

4. Network thresholding

Three preprocessing procedures were performed before conducting network analysis: normalization, thresholding and renormalization. The normalization procedure divides each network matrix by its total weight so that the total weight after normalization is 1. This procedure can reduce inter-subject variance due to difference in weights and group all the subjects to find a common threshold. The thresholding procedure is aimed to remove the edges with very small weights that are physically implausible. These extremely small weights typically represent two different types of error from fiber tracking: spurious edges, and underestimation of the weights for edges with long physical distance. The spurious edges can be removed by thresholding. Although the procedure will inevitably remove some edges that are true but the weights are under-estimated, it remains attractive because very small weights can cause some network metrics very unstable. Given a network Net , the information of fiber length can be used to find the appropriate threshold to remove spurious edges. By further assuming that one NSPV is sufficient to find the white matter pathways shorter than 10 mm, the thresholding procedure is outlined as follows. First, a network is constructed with one seed per voxel, denoted as $Net0$. Then we extract all the edges that exist in Net but not in $Net0$. That procedure is repeated for each subject to obtain a larger sample of data points. Then we take the logarithm on the edge weights, making the distribution of weights near Gaussian. We calculate the mean value and standard deviation for those with physical distance less than 10 mm. The threshold for the networks is set to be exponential of the mean plus standard deviation. The threshold obtained in this way may be

dependent on NSPV but should be stable for large NSPV. The renormalization procedure divides all the networks of all subjects by the maximal weight so that all network weights are between 0 and 1.

4.1. Network analysis—In order to explore the effect of number of random seeds on network properties, a set of network metrics were computed for the network (Rubinov and Sporns, 2010). The node-wise quantities include node degree, node strength, and betweenness centrality. Node degree is the number of links connected to a node. Node strength is the sum of neighboring link weights of a node. Betweenness centrality is the fraction of all shortest paths in the network that pass through a given node. The global metrics include total strength, mean clustering coefficient, average path length, small-worldness index, maximized modularity, and optimal number of modules. Total strength is the sum of all weights. Mean clustering coefficient γ is the global mean of clustering coefficient of each node, which is equivalent to the fraction of the node's neighbors that connect to each other. The average path length λ is the average shortest path length between all pairs of nodes. Small-worldness index swi makes reference to values of the two key metrics (clustering and path length) in a population of random (randomized) graphs, and is defined as

$$swi = \frac{\gamma/\gamma_{rand}}{\lambda/\lambda_{rand}}, \quad (20)$$

where γ_{rand} and λ_{rand} are the mean clustering coefficient and average path length for a random network with the same weights and degree distribution (Rubinov and Sporns, 2010). Maximized modularity evaluates the density of communities relative to a random model. Optimal module partitions divide the network into modules such that the modularity metric is maximized. Those metrics have been well defined previously and have been used to characterize the structural network of human brains in a number of studies (Rubinov et al., 2009; Rubinov and Sporns, 2010; van den Heuvel et al., 2010). All computation was performed in Matlab (The MathWorks, Inc., Natick, MA, USA) (ver. 2008a) using the Brain Connectivity Tool Box. Forty randomized networks were generated to compute the small-worldness index.

The discrepancy of the node-based quantities from two fiber tracking procedures was analyzed using a similar NEMA approach as in the analysis of network variance but with minor modification. The node-based metric is a vector of 68 elements. By concatenating the vector for all subjects, a 2D matrix was formed, with rows as the node index and columns as the subject index. The discrepancy was calculated as the standard deviation of the difference of those matrices from the two track dataset. For global metrics, the discrepancy was computed as the standard deviation of the difference for all subjects. In the end, all variances were scaled with corresponding mean values.

5. Results

A snap shot of the fibers and parcellated cerebral gray matter are displayed in Fig. 1A and B. Networks can be displayed in matrix form. Examples of the weighted network matrix for a single subject, using one seed or 40 seeds for fiber tracking, are shown in Fig. 1C and D, respectively. The nodes were ordered such that all left-hemisphere nodes occupy the first 34 rows/columns. The two strongly connected sub-blocks along the diagonal exclusively correspond to intra-hemispheric connectivity. Although the two networks appear to be quite similar, the weight differences between the two track data using one seed or 40 seeds are quite different (Fig. 1E). The amplitude of the difference for one seed is much higher than that for 40 seeds, indicating that the fluctuation of the network is larger for fewer seeds.

Variation analysis for M_{nf} and M_w as a function of number of random seeds in fiber tracking demonstrates that variation decreases as NSPV increases (Fig. 2). The variance of network weight is as large as 60% of its mean value for one seed (Fig. 2A). From one seed to 40 seeds, the VON values drop significantly, to less than 10% of the mean value. Despite the large variances observed, even for one seed, the networks from two repeated fiber trackings are highly correlated. The cross-correlation coefficients are higher than 0.95 for both network weights and number of fibers (Fig. 2B). The correlation increases as NSPV increases. The results indicate that the more NSPV in fiber tracking, the less variable the constructed network. The plots also show that the effect of NSPV is lessened as NSPV increases, due to ceiling/floor effects.

The scaled VONs of track numbers and network weights of all subjects are plotted in Fig. 3. For both M_{nf} and M_w , most of the data points are distributed in the vicinity of the solid line, coinciding with the theoretical prediction from Eq. (19). The variances drop rapidly when NSPV increases from 1 to 10. However, there is not much gain for further increase of NSPV.

If no thresholding takes place, the increase of NSPV has a direct effect on the nodal degree of the constructed network (Fig. 4). More seeds give rise to more edges in the network (Fig. 4B). The total degree of the network increases as the NSPV increases, although the increase gets smaller for higher NSPV values (Fig. 4D). Fig. 4A shows that the total degree drops with thresholding. After a very steep drop at the beginning, a nearly linear relationship between the logarithm of the total degree and the thresholds exists for a wide range of thresholding values. The threshold obtained from our thresholding scheme is NSPV-dependent but is very stable for NSPV = 10 as shown in Fig. 4C. By thresholding the networks using the threshold obtained from NSPV = 10, the total weight is almost constant for all NSPV values (Fig. 4D).

Fig. 5 shows the discrepancy of degree, strength, and betweenness centrality from two fiber trackings. These quantities characterize local properties of the network. The discrepancy is scaled by the corresponding mean value. Similar to M_{nf} and M_w , the discrepancies of degree, strength, and betweenness centrality are inversely related to NSPV. However, although the profiles of change are similar to each other and resembles those for M_{nf} and M_w , the relations with NSPV may be different. For instance, from NSPV = 1 to NSPV = 25, discrepancy of degree drops by 53%; discrepancy of strength drops by 78%; discrepancy of betweenness centrality drops by 68%. The theoretical prediction of VON from Eq. (15) is 80%.

Unlike the node-based quantities such as degree and strength, the global measures of the network such as total strength, small-worldness index, mean clustering coefficient, are insensitive to the NSPV in fiber tracking. No clear trend can be perceived from plots of these values (Fig. 6). For some metrics, the insensitivity is also reflected in the small values. When NSPV is greater than 1, the variances are smaller than 0.04 for total strength, mean clustering coefficient, average path length, and number of optimal modular structure.

Total VON increases as seed-related VON increases, and a linear relation exists between VON_s^2 and VON_t^2 (Fig. 7). For one seed, VON_s^2 is almost half of VON_t^2 . By increasing the NSPV to 10, VON_s^2 accounts for less than 10% of VON_t^2 . However, the trend of the dashed line indicates that VON_s has little effect on VON_t beyond 10 NSPV. This curve validates the decomposition hypothesis that can guide the choice of an optimal NSPV. The insert in Fig. 7 shows that the magnitude of VON_s and VON_t can be very different in individual subject, although the trend of a linear relationship is consistent.

6. Discussion

The effect of seed density in fiber tracking on the variance of constructed structural networks and network metrics was investigated using DTI data obtained from 10 human subjects. A novel method was proposed to evaluate the variance, similar to the noise quantification in NEMA standards for SNR measurement of MRI images. The variance of network weights is comparable to the mean value of the weights for small NSPV (Fig. 2), but can be reduced by increasing NSPV in fiber tracking. The experimental data confirmed theoretical analysis that the variance of the network weights is inversely proportional to square root of NSPV in fiber tracking. As the network from $\text{NSPV} = n$ is equivalent to the sum of n networks obtained from $\text{NSPV} = 1$, this relationship shows that the seed-related variance has a similar averaging effect as random noise.

Using more seeds in DTI tractography has mixed effects on the networks. By looking at the relation between fiber length and the weights for new edges (Fig. 4B), we found that when NSPV increases from 1 to 10, a large portion of the new edges has very short fiber length (<10 mm). As NSPV goes higher, most of the new edges have long fiber length (>50 mm). Because the noise in DTI images makes it hard to track long fibers, we hypothesize that the new edges with small fiber lengths are more likely from spurious fibers, while those with long fiber lengths might reveal true white matter connections but with underestimated fiber density. With the thresholding procedure aimed to remove false edges and implausible weights, the total degree drops dramatically (Fig. 4) and stays almost constant for all NSPV values. The new total degrees correspond to the small-slope portion of the curves in Fig. 4A. The amount of spurious edges and false weights depends on the tracking algorithm and parameters (Dauguet et al., 2007), but also could depend on the number of nodes. More nodes result in smaller ROI size, and a greater chance of random edges. Using 1000 nodes, a steeper curve was found by Hagmann et al. for the relation between number of edges (degree) and number of fibers (proportional to NSPV) (Hagmann et al., 2007).

Further analysis of the network properties suggests that the variance affects the local properties of the network, as shown in Fig. 5. At $\text{NSPV} = 1$, the variance of degree and strength is higher than 10% of the corresponding mean values. The betweenness centrality is even higher. Even without variance from other sources, the seed-related variance can impose severe problem when comparing local metrics for different groups or when the difference between two groups only exists locally. Our results in Fig. 6 showed little effect on global metrics for $\text{NSPV} = 5$. This could be due to more average effects of the global metrics.

With that relation shown in Eqs. (12) and (13), the number of seeds in fiber tracking can be chosen appropriately. Eq. (12) indicates that higher NSPV minimizes seed-related variance, while Eq. (13) suggests there is a bottle neck of total variance that is independent of seed related variance. In this particular setting of DTI acquisition and brain parcellation, the VON_s at $\text{NSPV} = 10$ is around 0.25, and the VON_t is around 3.0. According to Eq. (15), the appropriate NSPV should be greater than 8. Therefore, a value for NSPV of 10–15 will be an acceptable choice as a balance between variance and computation time. On the other hand, although the size of ROIs from the parcellation is quite different from each other and the number of fibers between any two ROIs varies significantly, the good fit of the variance to theoretical prediction also indicates our approach using NEMA standard an effective way of estimating the ‘noise’ of network. This approach can be used to estimate the network variance without performing many trials.

Some of our findings can be generalized to networks constructed in different ways. The definition of the network weight is still rather arbitrary. Besides the weighting scheme adopted in our study (Hagmann et al., 2007), other schemes have been proposed to make the

weights meaningful for deterministic tracking data. For instance, Bassett et al. simply defined the weights as the number of fibers between any two regions (Bassett et al., 2010). The constructed networks using that weighting scheme would correspond to M_{nf} in our study. Bassett et al. also studied an alternative weighting scheme which is the number of fibers normalized by the mean volume of the two regions. The constructed network using that weighting scheme is a variant between M_{nf} and M_w . On the other hand, many different parcellation schemes have been used to construct the structural network as well. Besides anatomically based parcellation (Desikan et al., 2006; Hagmann et al., 2010b) utilized here, there are template-based parcellations such as automated anatomical labeling (Gong et al., 2009; Tzourio-Mazoyer et al., 2002) and random parcellation (Hagmann et al., 2007; Zalesky et al., 2010). Different weighting schemes and parcellation schemes will result in different networks. Even with the same weighting scheme, it has been found that the metrics of a structural network are highly dependent on the choice of nodes (Zalesky et al., 2010). It should be emphasized that the relation between network variance and NSPV from our theory is valid for the weighting schemes mentioned earlier because the deduction is independent of the parcellation scheme. The weighting scheme or parcellation scheme will only affect the magnitude of variance.

Parcellating the cerebral cortex into sixty-eight ROIs is very coarse in general. Intuitively, the use of larger ROIs (fewer nodes) is less susceptible to the variance of tractography. This has two implications for the generalization of our findings to finer parcellation. First, as significant variance was found for one seed per voxel in fiber tracking, more seeds are needed for the group comparison of local metrics. Second, the number of seeds may have some effect on the variance of global metrics. The NEMA based approach can be used to estimate the variance and determine if the number of seeds is sufficient for a specific application using the proposed criterion.

In summary, the relation between variance of structural networks and number of seeds in tractography was deducted theoretically and examined using *in vivo* data. Noticeable variances of local metrics were observed for small NSPV. It was shown that the variance of the network is inversely proportional to the square root of number of seeds. While the influence of NSPV on the network measures depends on the way a network is constructed, that relation, together with the methods of comparing seed-related variance and total variance from DTI subsets, can be used for choosing optimal number of seeds in fiber tracking.

Acknowledgments

This work is partly supported by Indiana CTSI grant and the center for successful parenting, Indiana. We also thank Siemens Healthcare for providing work-in-progress diffusion sequence for our DTI data acquisition.

References

- Anderson AW. Theoretical analysis of the effects of noise on diffusion tensor imaging. *Mag Reson Med*. 2001; 46:1174–88.
- Bassett DS, Brown JA, Deshpande V, Carlson JM, Grafton ST. Conserved and variable architecture of human white matter connectivity. *NeuroImage*. 2010 doi:10.1016/j.neuroimage.2010.09.006.
- Cheng, H.; Kim, D-J.; Sporns, O.; Wang, Y.; Sheng, J.; Saykin, A. Effect of SNR of DTI on the structural network; ISMRM 19th Annual Meeting; 2011;
- Dauguet J, Peled S, Berezovskii V, Delzescaux T, Warfield SK, Born R, et al. Comparison of fiber tracts derived from in-vivo DTI tractography with 3D histological neural tract tracer reconstruction on a macaque brain. *NeuroImage*. 2007; 37:530–8. [PubMed: 17604650]

- Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*. 2006; 31:968–80. [PubMed: 16530430]
- Gong G, He Y, Concha L, Lebel C, Gross DW, Evans AC, et al. Mapping anatomical connectivity patterns of human cerebral cortex using in vivo diffusion tensor imaging tractography. *Cerebral Cortex*. 2009; 19:524–36. [PubMed: 18567609]
- Hagmann P, Cammoun L, Gigandet X, Meuli R, Honey CJ, Wedeen VJ, Sporns O. Mapping the structural core of human cerebral cortex. *PLoS Biol*. 2008; 6:e164. [PubMed: 20076716]
- Hagmann P, Kurrant M, Gigandet X, Thiran P, Wedeen VJ, Meuli R, et al. Mapping human whole-brain structural networks with diffusion MRI. *PLoS Biol*. 2007:e597.
- Hagmann P, Sporns O, Madan N, Cammoun L, Pienaar R, Wedeen VJ, et al. White matter maturation reshapes structural connectivity in the late developing human brain. *PLoS Biol*. 2010a; 107:19067–72.
- Hagmann P, Sporns O, Madan N, Cammoun L, Pienaar R, Wedeen VJ, et al. White matter maturation reshapes structural connectivity in the late developing human brain. *Proc Natl Acad Sci USA*. 2010b; 107:19067–72. [PubMed: 20956328]
- Lazar M, Alexander AL. An error analysis of white matter tractography methods: synthetic diffusion tensor field simulations. *NeuroImage*. 2003; 20:1140–53. [PubMed: 14568483]
- Lo CY, Wang PN, Chou KH, Wang J, He Y, Lin CP. Diffusion tensor tractography reveals abnormal topological organization in structural cortical networks in Alzheimer's disease. *J Neurosci*. 2010; 30:16876–85. [PubMed: 21159959]
- Mori S, Crain B, Chacko V, van Zijl P. Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Ann Neurol*. 1999; 45:265–9. [PubMed: 9989633]
- Mori S, van Zijl PCM. Fiber tracking: principles and strategies – a technical review. *NMR Biomed*. 2002; 15:465–80.
- NEMA. MS 6-2008. Determination of Signal-to-Noise Ratio and Image Uniformity for Single-Channel Non-Volume Coils in Diagnostic MR Imaging. 2008
- Rohde GK, Barnett AS, Basser PJ, Marengo S, Pierpaoli C. Comprehensive approach for correction of motion and distortion in diffusion-weighted MRI. *Mag Reson Med*. 2004; 51:103–14.
- Rubinov M, Knock SA, Stam CJ, Micheloyannis S, Harri SAW, Williams LM, et al. Small-world properties of nonlinear brain activity in schizophrenia. *Human Brain Map*. 2009; 30:403–16.
- Rubinov M, Sporns O. Complex network measures of brain connectivity: uses and interpretations. *NeuroImage*. 2010; 52:1059–69. [PubMed: 19819337]
- Shu N, Liu Y, Li K, Duan Y, Wang J, Yu C, et al. Diffusion tensor tractography reveals disrupted topological efficiency in white matter structural networks in multiple sclerosis. *Cerebral Cortex*. 2011 doi:10.1093/cercor/bhr039.
- Sporns, O. *Network of the Brain*. The MIT Press; Cambridge, MA: 2010.
- Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*. 2002; 15:273–89. [PubMed: 11771995]
- van den Heuvel MP, Mandl RC, Stam CJ, Kahn RS, Hulshoff Pol HE. Aberrant frontal and temporal complex network structure in schizophrenia: a graph theoretical analysis. *J Neurosci*. 2010; 30:15915–26. [PubMed: 21106830]
- Zalesky A, Fornito A, Harding IH, Cocchi L, Yücel M, Pantelis C, et al. Whole-brain anatomical networks: does the choice of nodes matter? *NeuroImage*. 2010; 50:970–83. [PubMed: 20035887]
- Zalesky A, Fornito A, Seal ML, Cocchi L, Westin C-F, Bullmore ET, Egan GF, Pantelis C. Disrupted axonal fiber connectivity in schizophrenia. *Biol Psychiatry*. 2011; 69:80–9. [PubMed: 21035793]

Fig. 1.

Demonstration of the network construction and variance using the data of one subject. (A) A snap shot of the tractography result from diffusion toolkit. (B) The gray matter ROIs obtained from FreeSurfer parcellation, with different ROIs labeled in different colors. These ROIs are used as the nodes for structural network. (C) Logarithm of a network matrix constructed from the track data using 1 seed in fiber tracking. (D) Logarithm of a network matrix constructed from the track data using 40 seeds in fiber tracking. (E) Weight difference for the common edges of the network for 1 seed (blue) and 40 seeds (red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Fig. 2.

(A) VON analysis of the number of fibers (circle) between ROIs and network weight (cross) as a function of the NSPV in fiber tracking for all 10 subjects. For each subject, the variance is normalized by the corresponding mean value of number of fibers or weight. (B) Correlation coefficients of the numbers of fibers (circle) between ROIs and network weights (cross) obtained from two track data as a function of the NSPV in fiber tracking for all 10 subjects.

Fig. 3. Scaled VON (see Eq. (19)) for the number of fibers between ROIs (A) and network weight (B) as a function of the NSPV in fiber tracking for all 10 subjects. All the dots are data points of the 10 subjects. The solid curve is the plot of $y=1/\sqrt{N}$.

Fig. 4.

Thresholding effects of the network. (A) Plots of logarithm of the weights against thresholding, each curve is the mean for all subjects with the same NSPV value (the threshold is in unit of minimum mean weights, defined as the total weights divided by number of all possible pairs of nodes). (B) Plots of the fiber length vs. logarithm of the weights for those edges arisen due to NSPV increases (green circle: NSPV from 1 to 10; blue square: NSPV from 10 to 20; brown triangle: NSPV from 20 to 30; red diamond: NSPV from 30 to 40) for all subjects. (C) Logarithm of the threshold obtained from the data obtained using different NSPV. (D) Plots of total degree against NSPV without thresholding (circle) and with thresholding using the threshold for NSPV = 10 in (C). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Fig. 5.

Discrepancy of the node based quantities for different NSPV. The discrepancy is calculated as the standard deviation of the differences of the 68 node based values between two networks obtained from two fiber trackings, normalized by the corresponding mean value. The displayed data are the averaged results for 10 subjects: (A) degree; (B) strength; and (C) betweenness centrality.

Fig. 6.

Discrepancy of the global metrics for different NSPV. The difference of the global metrics between two networks obtained from two fiber trackings is first obtained, the discrepancy is calculated as the standard deviation of the differences of 10 subjects normalized by the corresponding mean value: (A) total strength; (B) small-worldness index; (C) average path length; (D) mean clustering coefficient; (E) maximized modularity; and (F) optimal number of modules.

Fig. 7.

Relation between seed related VON and total VON for networks constructed from two complimentary subsets of the DTI data. The data points from left to right correspond to NSPV of 10, 5, 2, and 1. The large dots are the average of 10 subjects, which are shown as the inserts. All the VONs are normalized by the mean VON_i^2 for display purpose.