A Symmetry-Based Method to Infer Structural Brain Networks from Probabilistic Tractography Data

Kamal Shadi¹, Saideh Bakhshi†*, David A. Gutman², Helen S. Mayberg³ and Constantine Dovrolis¹*

¹ School of Computer Science, Georgia Institute of Technology, Atlanta, GA, USA, ² Department of Neurology, Psychiatry and Biomedical Informatics, Emory University, Atlanta, GA, USA, ³ Psychiatric Neuroimaging and Therapeutics, Department of Psychiatry, Neurology, and Radiology, Emory University, Atlanta, GA, USA

Recent progress in diffusion MRI and tractography algorithms as well as the launch of the Human Connectome Project (HCP)¹ have provided brain research with an abundance of structural connectivity data. In this work, we describe and evaluate a method that can infer the structural brain network that interconnects a given set of Regions of Interest (ROIs) from probabilistic tractography data. The proposed method, referred to as Minimum Asymmetry Network Inference Algorithm (MANIA), does not determine the connectivity between two ROIs based on an arbitrary connectivity threshold. Instead, we exploit a basic limitation of the tractography process: the observed streamlines from a source to a target do not provide any information about the polarity of the underlying white matter, and so if there are some fibers connecting two voxels (or two ROIs) X and Y, tractography should be able in principle to follow this connection in both directions, from X to Y and from Y to X. We leverage this limitation to formulate the network inference process as an optimization problem that minimizes the (appropriately normalized) asymmetry of the observed network. We evaluate the proposed method using both the FiberCup dataset and based on a noise model that randomly corrupts the observed connectivity of synthetic networks. As a case-study, we apply MANIA on diffusion MRI data from 28 healthy subjects to infer the structural network between 18 corticolimbic ROIs that are associated with various neuropsychiatric conditions including depression, anxiety and addiction.

Keywords: connectome, diffusion MRI, tractography, structural network, network analysis

INTRODUCTION

Diffusion MRI has opened a new window at the meso-scale structure of the living brain (Sporns et al., 2005). Clinicians and researchers can now observe and measure the properties of white matter in a non-invasive manner, analyzing the location and density of neuronal fibers at a spatial granularity of 1–2 mm isotropic voxels (Glasser et al., 2016b). Such structural information is important in deciphering how the brain works (Sporns, 2012, 2013), and it also creates new ways
to understand and potentially diagnose (Ciccarelli et al., 2008; Fornito and Bullmore, 2015) or even treat (Mayberg et al., 2005) various brain diseases (Buckner et al., 2005; Bassett et al., 2008).

Processing diffusion MRI data using tractography algorithms is the next step forward: instead of analyzing the properties of white matter at the level of individual voxels, tractography aims to detect individual bundles of neuronal fibers originating or passing through a given "seed" voxel (Jbabdi et al., 2013). Additionally, given a seed voxel and a target ROI, it is now possible to examine the likelihood that some white matter fibers connect the two (referred to as "probabilistic tractography"), and to track the shape of these connections (Behrens et al., 2007). In this paper, we propose a method to further process the noisy connectivity information provided by probabilistic tractography in order to estimate an interconnection network between a given set of gray matter ROIs.

Diffusion MRI data, jointly with deterministic (Mori et al., 1999) or probabilistic tractography methods (Behrens et al., 2007) have been used successfully during the last decade to infer the structure of the human brain between hundreds of ROIs (Hagmann et al., 2007). Various structural properties of these networks have been discovered for the healthy brain (Bassett et al., 2011) and for various psychiatric diseases (McIntosh et al., 2008; Daianu et al., 2013). When combined with fMRI and behavioral or genomic analysis, these non-trivial topological properties provide new insights about the role of individual ROIs in specific networks and the way in which these distinct ROIs exchange information to produce integrated function (Damoiseaux and Greicius, 2009; Greicius et al., 2009).

A major challenge in this research effort is that the inferred brain networks, as well as their topological properties, are often sensitive to the parameters of the tractography process (Rubinov and Sporns, 2010; Bastiani et al., 2012; Duda et al., 2014; Thomas et al., 2014). In probabilistic tractography, the most critical of those parameters is the connectivity threshold $\tau$ that determines whether the tractography-generated streamlines from a given seed voxel to a target ROI occur with sufficiently large probability to indicate the presence of an actual connection (Li et al., 2012b). If $\tau$ is too low the resulting network includes connections that do not exist in reality and the converse happens if $\tau$ is too high. Even a small number of spurious or miss-detected edges can adversely effect the properties of the inferred networks (Van Wijk et al., 2010). Further, the optimal value of $\tau$, i.e., the threshold that would result in the most accurate reconstruction of the underlying "ground truth" network, may vary between different subjects (Gong et al., 2009) and image acquisitions parameters (Jones et al., 2013).

The problem of selecting an appropriate connectivity threshold in either deterministic or probabilistic tractography is not new. One approach has been to select the largest possible threshold (i.e., fewest possible edges) so that the final inferred network remains connected across the majority of subjects (Li et al., 2009). Depending on the selected ROIs, this approach can lead to many miss-detections (if those ROIs are densely interconnected) or false alarms (if the ROIs are not directly connected).

Another approach has been to analyze the tractography data with a wide range of threshold values, hoping that certain qualitative properties are robust and independent of the exact threshold. Li et al. investigated how the connectivity thresholds affects network density and therefore network efficiency metrics (Li et al., 2012b). Duda et al. have shown the importance of the connectivity threshold for various network metrics such as clustering coefficient and characteristic path length (Duda et al., 2014).

This work focuses on the following problem: how to infer the structural network between a given set of gray matter ROIs in a reliable way that does not require an arbitrary choice of the connectivity threshold? The proposed method, referred to as Minimum Asymmetry Network Inference Algorithm (MANIA), exploits a fundamental limitation of diffusion MRI imaging and of the tractography process: diffusion MRI can estimate the orientation of fibers in each voxel but it cannot infer the polarity (afferent vs. efferent) of those fibers (Robinson et al., 2010; Jbabdi and Johansen-Berg, 2011). Similarly, a tractography algorithm can combine those per-voxel orientations to "stitch together" expected connections but it does not provide any information about the direction of those connections (Hagmann et al., 2008; Donahue et al., 2016). Given this limitation, MANIA expects that the presence of an actual connection from voxel X to voxel Y (in that direction) will be detected by the tractography process as a symmetric connection between X and Y. Similarly, if there is no connection between X and Y, the tractography process should not detect a connection in either direction. Based on this principle, MANIA formulates the network inference problem as an optimization over the range of connectivity threshold values: it selects the value of $\tau$ that minimizes the asymmetry of the resulting network. The network asymmetry is normalized relative to the asymmetry that would be expected due to chance alone in a random network of the same density.

MANIA can work in tandem with all probabilistic tractography methods, such as FSL's probtrackx (Behrens et al., 2007), PiCo (Parker et al., 2003), and iDF-PROBA (Descoteaux et al., 2009). It can be also combined with deterministic tractography methods, such as FACT (Mori et al., 1999), but only if a large number of streamlines (in the thousands) are generated from randomly placed seeds within each voxel (Cheng et al., 2012).

We expect that the given set of ROIs primarily reside in gray matter. Dilating a gray matter ROI so that it includes some white matter voxels may result in connectivity errors, especially with cortical ROIs, because of the dense white matter systems just

**Abbreviations:** AD, Addictive Disorder; DTI, Diffusion Tensor Imaging; FACT, Fiber Assignment by Continuous Tracking; FDT, FMRIB's Diffusion Toolbox; fMRI, functional Magnetic Resonance Imaging; FMRIB, Oxford centre for Functional Magnetic Resonance Imaging of the Brain; fODF-PROBA, Fiber Orientation Distribution Function-Probabilistic; FSL, FMRIB Software Library; GAD, Generalized Anxiety Disorder; HCP, Human Connectome Project; MANIA, Minimum Asymmetry Network Inference Algorithm; MDD, Major Depressive Disorder; MNI, Montreal Neurological Institute; MRI, Magnetic Resonance Imaging; OCD, Obsessive Compulsive Disorder; PiCo, Probabilistic Index of Connectivity; PTSD, Post-Traumatic Stress Disorder; ROI, Region Of Interest; WGB, White matter/Gray matter Boundary.
beneath the cortical sheet (Reveley et al., 2015). The selection of ROIs and the estimation of their boundaries is an important issue that is further discussed in the Discussion.

We evaluate the accuracy of MANIA using both the well-known FiberCup dataset and based on synthetically generated data in which the ground-truth network is known. We also compare MANIA with an ideal threshold-based method in which the optimal connectivity threshold is assumed to be known. Further, we show how to associate a confidence level with each edge, and how to apply MANIA in a group of subjects. Finally, as a case-study, we apply MANIA on diffusion MRI data from 28 healthy subjects (Chen et al., 2013) to infer the structural network between 18 corticolimbic ROIs that are implicated with neuropsychiatric disorders such as major depressive disorder (MDD), post-traumatic stress disorder (PTSD), obsessive compulsive disorder (OCD), generalized anxiety disorder (GAD) and addictive disorder (AD) (Seminowicz et al., 2004; Elman et al., 2013; Beucke et al., 2014; Peterson et al., 2014). We note however that, even though these ROIs are generally associated with various psychiatric disorders and the aspects of emotional regulation putatively impacted by these disorders, the objective of this work is not to infer the network that is associated with any particular disorder.

MATERIALS AND METHODS

MANIA Inputs

The proposed network inference method requires the following inputs:

1. A set of $N$ ROIs that represent the nodes of the structural brain network. The $i$th ROI is a spatially connected cluster of $v_i$ voxels ($i = 1 \cdots N$). The selection of ROIs is important (Zalesky et al., 2010; de Reus and van den Heuvel, 2013b) but outside the scope of MANIA. MANIA attempts to find the anatomic network between the given ROIs independent of whether the latter are defined by an expert neuroanatomist or by a data-driven method. For instance, ROIs may correspond to different Brodmann areas or other anatomical atlases (Tzourio et al., 1997; Petrides, 2005). Or, it could be that the spatial extent of ROIs results from the analysis of fMRI data (McKeown et al., 1997; Craddock et al., 2012; Blumensath et al., 2013; Thirion et al., 2014). ROIs can also be defined based on both fMRI analysis and anatomical knowledge (Glasser et al., 2016a).

2. The results of the tractography process between the previous $N$ ROIs. We assume that the tractography results are structured as voxel-to-ROI matrices (i.e., streamlines are generated from each voxel toward each target ROI), instead of voxel-to-voxel or ROI-to-ROI matrices. Specifically, we represent the output of tractography with $N$ matrices $T_i$ ($i = 1 \cdots N$), defined as follows. The $i$th ROI corresponds to a matrix $T_i$ with $v_i$ rows (i.e., the number of voxels in that ROI) and $N$ columns. The element $(j, k)$ of matrix $T_i$ represents the fraction of tractography-generated streamlines that originate from the seed voxel $j$ of the $i$th ROI and reach any voxel of the $k$th ROI. The same number of streamlines is generated for every seed to target pair $(j, k)$. The $i$th column of matrix $T_i$ is set to zero, meaning that we do not consider edges from an ROI back to itself (even if such fibers exist). Figure 1 illustrates this notation. Since we have $N$ ROIs, there will be $N$ input matrices, one for each source ROI.

Connectivity Threshold $\tau$

How can we decide whether voxel $j$ of the $i$th ROI connects to the $k$th ROI given the fraction $T_i(j, k)$? The simplest approach is to examine if $T_i(j, k)$ is larger than a given “connectivity threshold” $\tau$ ($0 < \tau < 1$). In MANIA, $\tau$ is not a given threshold but an optimization variable, as described in the next section.

As in prior work, we assume that the $i$th ROI is connected to the $k$th ROI as long as at least one voxel of the former is connected to the latter (Hagmann et al., 2008; Bassett et al., 2011). This assumption is not central to MANIA however, and it can be easily replaced with a stronger connectivity constraint, depending on the size of the given ROIs and the spatial resolution at which those ROIs are specified.

Network Inference As an Optimization Problem

For a given value of $\tau$, we can identify the voxels of the $i$th ROI that connect to every other ROI. If this process is repeated for every $i$, we can construct a directed network in which the $i$th ROI is connected to the $j$th ROI if there is at least one voxel in the former that is connected to the latter for that value of $\tau$. This network can be represented with an adjacency matrix $G_\tau$, as follows:

$$G_\tau(i, k) = \begin{cases} 1 & \text{if } T_i(j, k) > \tau \text{ for at least one voxel } j \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

So, the element $(i, k)$ of this matrix is equal to one if there is a (directed) edge from the $i$th ROI to the $k$th ROI. The diagonal entries of $G_\tau$ are set to zero because we do not consider streamlines from a source ROI back to itself.

In graph theory, an $N \times N$ adjacency matrix represents a directed and unweighted graph with $N$ nodes as follows: if there is an edge from node $i$ to node $j$ the $(i, j)$ element of the adjacency matrix is 1; otherwise it is 0. The graph (and the corresponding adjacency matrix) are referred to as “weighted” if each edge is associated with a weight, which typically represents the strength of the edge.
We define the asymmetry $\phi(G)$ of a directed network $G$ as the fraction of edges that are present in only one direction,

$$\phi(G) = \frac{\sum_{i=1}^{N} \sum_{k=1}^{N} G(i,k)(1 - G(k,i))}{\sum_{i=1}^{N} \sum_{k=1}^{N} G(i,k)}$$

(2)

The asymmetry of a network $G$ depends on its density $\rho(G)$, defined as the fraction of connected node-pairs,

$$\rho(G) = \frac{\sum_{i=1}^{N} \sum_{k=1}^{N} G(i,k)}{N(N-1)}$$

(3)

The more edges a directed network has, the more likely it becomes that a pair of nodes will be connected in both directions, i.e., the higher the density, the lower the asymmetry.

More formally, consider a directed network with $K$ directed edges and $N$ nodes. The density is $\rho = K/[N(N-1)]$ ($0 < \rho < 1$). To construct such a network randomly, denoted by $G_\rho$, we simply connect $N(N-1)\rho$ randomly selected but distinct pairs of nodes with directed edges. The expected number of edges that exist in only one direction is $N(N-1)(1-\rho)$ and so the expected value of the asymmetry of $G_\rho$ is:

$$\phi(G_\rho) = \frac{N(N-1)\rho(1-\rho)}{N(N-1)\rho} = 1-\rho$$

(4)

To quantify the actual asymmetry of an observed network $G_t$, we normalize $\phi(G_t)$ by the asymmetry that is expected simply due to chance given the density of this network. So, we define the normalized asymmetry of $G_t$ as

$$\Phi(G_t) = \frac{\phi(G_t)}{\phi(G_\rho)}$$

(5)

which is well defined as long as $\rho(G_t) < 1$.

MANIA is based on the following premise: the inferred directed network should be as symmetric as possible. The reason is that the tractography process is unable to infer the actual direction (polarity) of the underlying neural fibers. So, if there are some fibers connecting two voxels X and Y, tractography should be able, in principle, to follow this connection in both directions, from X to Y and from Y to X. We do not claim that two connected ROIs are always attached with both afferent and efferent fibers; instead, we argue that tractography is not able to discover the polarity of those fibers and so the corresponding connection should be trace-able in both directions.

The presence of some streamlines from some voxels in ROI X to ROI Y does not necessarily mean however that the network inference method will detect an edge both from X to Y and from Y to X; this also depends on the parameter $\tau$. Given that we aim to minimize the asymmetry of the inferred network, MANIA aims to select the value of $\tau$ that leads to the lowest possible asymmetry.

The corresponding optimization problem can be stated as follows: determine the adjacency matrix $\hat{G} = G_\tau^*$, where $\tau^*$ is a value of the connectivity threshold that minimizes the normalized asymmetry of $G_\tau$ across all possible values of $\tau$,

$$\tau^* = \arg \min_{0 < \tau < 1} \Phi(G_\tau)$$

(6)

So, MANIA is based on the premise that there is an ideal value (or range of values) of the connectivity threshold that can correctly classify every directed pair of ROIs as either "connection exists" or "connection does not exist." When such a threshold exists, it will result in a completely symmetric network (because a perfectly accurate tractography-based network cannot be asymmetric). On the other hand, if such an ideal threshold does not exist (for instance, it may be that two connected ROIs are too far from each other and tractography cannot "see" their connection, or that it is impossible for streamlines to cross the White matter/Gray matter Boundary (WGB) of a certain ROI in one direction but not in the opposite), then MANIA aims to at least minimize the normalized asymmetry metric, even if the resulting network will not be completely symmetric.

If there is more than one value of $\tau$ that results in the same minimum of the normalized asymmetry (potentially zero), MANIA reports the network with the largest density. The rationale behind this tie-breaker is to avoid trivial solutions that include only a subset of the actual network edges. The previous optimization problem can be solved numerically by scanning the range of $\tau$ values with a given resolution$^3$. The density of the resulting network is denoted by

$$\rho^* = \rho(G_\tau^*)$$

(7)

As an illustration of the previous method, Figure 2 shows how the network asymmetry (both $\phi$ and $\Phi$) varies with the density $\rho$ as well as the relation between $\rho$ and $\tau$ for the dataset that corresponds to one of the subjects in our case-study.

Threshold-Based Network Inference with Post-Symmetrization

A common network inference method is to rely on a given connectivity threshold $\tau$, as shown in Equation (1). This threshold is sometimes chosen to achieve a certain network density or to ensure that the network is connected (Li et al., 2009).

Given that tractography cannot detect the direction of inferred edges, the resulting network is often "post-symmetrized" (Azadbakht et al., 2015; Donahue et al., 2016). Consider two network nodes $i$ and $j$ and suppose that the fraction of streamlines from $i$ to $j$ is denoted by $T_{i,j}$. Suppose that $T_{i,j} > \tau$ but $T_{j,i} < \tau$, where $\tau$ is a given threshold (not necessarily the threshold chosen by MANIA). We can resolve this conflicting evidence between the two directions of the edge by comparing the ratio $(T_{i,j} - \tau)/(1 - \tau)$, which reflects our confidence that the edge from $i$ to $j$ exists, with the ratio $(\tau - T_{j,i})/\tau$ that reflects our confidence that the edge from $j$ to $i$ does not exist. The two nodes are connected with an (undirected) edge only if the former ratio is larger.

If MANIA cannot find a threshold value that gives a completely symmetric network, the aforementioned post-symmetrization method is applied on the network that results from MANIA.

---

$^3$Since we set the number of streamlines to 5000, the minimum resolution is $\tau = 5000 = 0.0002$. 

Frontiers in Neuroinformatics | www.frontiersin.org 4 November 2016 | Volume 10 | Article 46
Shadi et al. Symmetry-Based Structural Brain Network Inference

**FIGURE 2 | (Top)** Network density $\rho$ as a function of the connectivity threshold $\tau$ (plotted for one subject in our DTI dataset). **(Bottom)** Network asymmetry (red) and normalized network asymmetry (blue) as functions of the network density $\rho$. The optimal density $\rho^*$ is the largest value that minimizes the normalized network asymmetry.

**Performance Metrics**

MANIA can be viewed as a binary classifier: each possible directed edge is classified as *present* or *absent*. We evaluate MANIA based on the following standard metrics for binary classification: the *false positive rate* (or false alarm) $p_f$, and the *false negative rate* (or miss detection) $p_m$. The former is defined as the fraction of absent edges that are incorrectly classified as present, while the latter is defined as the fraction of present edges that are incorrectly classified as absent.

Also, the *Jaccard similarity* between the sets of edges $E(G)$ and $E(\hat{G})$ of the actual network $G$ and the MANIA network $\hat{G}$, respectively, is defined as

$$J(G, \hat{G}) = \frac{|E(G) \cap E(\hat{G})|}{|E(G) \cup E(\hat{G})|}$$

and it varies between zero (no common edges) and one (identical networks).

**Optimal Threshold-Based Network Inference**

We can also compare the network that results from MANIA with the network that would result if we knew the optimal value $\tau^\text{opt}$ of the connectivity threshold, i.e., the value of $\tau$ that maximizes the Jaccard similarity between the inferred network $G_\tau$ and the ground truth network $G$:

$$\tau^\text{opt} = \arg\max_{0 < \tau < 1} J(G_\tau, G)$$

Even though it is not possible to know this optimal threshold value when analyzing real tractography data, we can easily compute its value (or range of values) in experiments with synthetically generated networks, where the ground-truth network $G$ is known.

**Edge Ranking and Confidence Metric in MANIA**

The output of MANIA is an unweighted directed network. We can quantify the level of confidence we have in each edge with the following edge ranking scheme.

Let $\rho_\alpha$ be the minimum network density at which edge $\alpha$ is present. If the edge $\alpha$ is from a source $X$ to a target $Y$, the lower $\rho_\alpha$ is, the higher the fraction of streamlines from $X$ to $Y$. Consequently, we can rank edges so that we are more confident in the presence of edge $\alpha$ than of edge $\beta$ if $\rho_\alpha < \rho_\beta$ (see Figure 3).

We define a *confidence metric* for an edge $\alpha$ that is present in the MANIA network (i.e., $\rho_\alpha < \rho^*$) as follows

$$C(\alpha) = \frac{\rho^* - \rho_\alpha}{\rho^*}$$

$C(\alpha)$ varies from 0 (the edge is only marginally present) to 1 (highest confidence that the edge is present). Similarly, if edge $\alpha$ is absent from the MANIA network (i.e., $\rho_\alpha > \rho^*$), its confidence metric is defined as

$$C(\alpha) = \frac{\rho^* - \rho_\alpha}{1 - \rho^*}$$

and $C(\alpha)$ varies from 0 (the edge is only marginally absent) to $-1$ (highest confidence that the edge is absent).

We also define a confidence metric for a pair of nodes $(X,Y)$, as the arithmetic mean of the confidence metric of the two directed edges between $X$ and $Y$,

$$C(X,Y) = \frac{C(X \to Y) + C(Y \to X)}{2}$$
Note that one of the two edges may be present while the other may be absent. In that case, the confidence of the corresponding node-pair will be less than the confidence of the present edge.

Note that this edge confidence metric is not related to connection “strength” or “quality,” and the resulting network is still meant to be interpreted as an unweighted graph.

**Group Analysis Using MANIA**

If the objective is to create a single “average network” based on data from several subjects, the question is how to best aggregate the tractography data from the given group. One approach is to average the diffusion MRI data, after transforming them in a standard space. Another approach is to average the fraction of streamlines from a given seed voxel to a given target ROI, across all subjects. The previous approaches can be sensitive to outliers, variations in the diffusion MRI process across subjects, tractography errors and mapping/warping into a standard space.

A recently proposed method is to construct a group-level network based on the consistency of the edge weights across subjects (Roberts et al., 2016); this approach preserves edges that appear consistently strong for their length, rather than penalizing long-distance connections. This approach may not be applicable however when the group size is small.

Another method is to construct an individual network for each subject, perhaps using MANIA, and then form an aggregate network by only keeping those edges that appear in a large fraction of subjects. This approach requires a group-level threshold for the minimum fraction of subjects that should have a connection. For instance, de Reus et al. have proposed a statistically rigorous method to compute such a threshold (de Reus and van den Heuvel, 2013a).

Here we propose a different group analysis method, referred to as group-MANIA, that is based on the aggregation of edge-rankings across subjects. The rank-based nature of this method makes it robust to outliers.

As in the previous section, the edges of a subject can be ranked based on the minimum network density at which an edge first appears. We are more confident in the presence of edge \( \alpha \) than of edge \( \beta \) if \( \rho_\alpha < \rho_\beta \) (see Figure 3).

Suppose that we compute an edge rank vector \( R_m \) for each subject \( m \), so that the lowest rank \( R_m(1) \) corresponds to the edge for which we are most confident. The number of possible (not necessarily present) directed edges is the same for all subjects: \( N(N-1) \) where \( N \) is the number of network nodes.

Given a group of size \( M \), we have \( M \) distinct rank vectors \( R_m, m = 1 \cdots M \). The computational problem of rank aggregation (Schalekamp and van Zuylen, 2009) is to compute an optimal permutation \( \hat{R} \) of the \( N(N-1) \) possible edges that captures as well as possible the ordering relations in the \( M \) input rank vectors. Specifically, the *Kemeny distance* between two rank vectors \( R_1 \) and \( R_2 \) is defined as

\[
\sum_{i} \sum_{j} \delta_{ij} (R_1, R_2)
\]

where \( \delta_{ij} (R_1, R_2) = 1 \) if \( R_1 \) and \( R_2 \) disagree in the relative position of elements \( i \) and \( j \), and zero otherwise. Rank aggregation aims to compute a vector \( \hat{R} \) that minimizes the cumulative Kemeny distance between \( \hat{R} \) and all input rank vectors \( R_m \). It is an NP-hard problem, and so it is typically solved heuristically. We use the Quicksort algorithm (Ailon et al., 2008) since it has been shown to provide a good approximation of the optimum solution. QuickSort selects a random edge as pivot at each recursive step, while the remaining edges are separated in a left and right list. The left list includes edges that have a lower rank than the pivot in the majority of the subjects; similarly for the right list. The algorithm proceeds recursively in the left and right lists until all edges are ordered.

After computing the optimal aggregate rank vector, we apply MANIA on \( \hat{R} \) to compute the network with the minimum normalized asymmetry (as in the case of a single subject). Note however that the input to MANIA in this case is an ordered list of edges \( \hat{R} \) rather than the set of connectivity matrices \( T \) (see the MANIA Inputs section). Group-MANIA forms a network with the first \( K = N(N-1) \rho \) edges in \( \hat{R} \), and it computes the normalized asymmetry of that network. It then repeats this step, for all values of \( K \), to identify the network with the minimum value of \( \Phi \). We refer to the resulting network as the rank-aggregated network.

**FiberCup Phantom Dataset**

We utilize the publicly available FiberCup dataset (Poupon et al., 2008, 2010; Fillard et al., 2011) as one of the benchmarks to evaluate MANIA's accuracy. We use a FiberCup acquisition in which the \( b \)-value is 1500. FiberCup resembles a human brain coronal section with 18 ROIs and seven fiber bundles—see Figure 2 of Neher et al. (2014). These fiber bundles are constructed to include sharp turns as well as crossing, kissing and branching points, making it a challenging benchmark for any tractography algorithm. Due to the practical difficulties of generating a realistic phantom, FiberCup does not contain any complex 3D fiber structures.

We apply probtrackx and MANIA on the FiberCup phantom to infer the network of the seven underlying connections. The FDT probabilistic tractography parameters are set to their default values (number of streamlines = 5000, maximum number of
steps = 2000, loop check: set, curvature threshold = 0.2, step length = 0.5 mm, no distance bias correction). Protrackx is seeded at the WGB of the ROIs that are connected to fiber terminals.

**Synthetically Generated Networks**

To evaluate the accuracy and sensitivity of MANIA in a reliable manner we need to rely on synthetic networks rather than actual DTI and tractography data. The benefit of these computational experiments is that we can test MANIA under a wide range of noise conditions and for arbitrary network densities. Unfortunately there are no good statistical models for the noise in the output of tractography data (Jbabdi and Johansen-Berg, 2011). We evaluate MANIA based on a simple noise model that is based on the theory of maximum entropy distributions, as described next.

For simplicity, each ROI of the synthetically generated networks is simply a voxel. Modeling multi-voxel ROIs in these simulation experiments would not add any new insights. Suppose that the directed network between N nodes is represented by the $N \times N$ adjacency matrix $G$. Let $T_{i,j}$ be the fraction of streamlines that originate from node $i$ and terminate at node $j$. Ideally, in the absence of any noise in the DTI data and without any errors in the tractography process, it should be that

$$T_{i,j} = \begin{cases} 1 & \text{if } G_{i,j} = 1 \text{ and } G_{j,i} = 1 \\ 0 & \text{otherwise} \end{cases}$$

(14)

So, if there is an edge between nodes $i$ and $j$ in either direction, the fraction of streamlines from node $i$ to node $j$ should be 100%; otherwise, it should be zero.

In practice, there is significant noise in DTI data and the tractography process can be error-prone, especially when the ROIs are in gray matter and/or when neural fibers cross each other, split or merge, or fan out as they approach their targets. Consequently, the tractography output may show that some streamlines do not reach from node $i$ to node $j$ even when the two nodes are connected, or that some streamlines get from $i$ to $j$ even when there is no connection between the two nodes. We model these errors probabilistically, as follows:

$$T_{i,j} = \begin{cases} 1 - Z_1 & \text{if } G_{i,j} = 1 \text{ and } G_{j,i} = 1 \\ Z_2 & \text{if } G_{i,j} = 0 \text{ and } G_{j,i} = 0 \end{cases}$$

(15)

where $Z_1$ and $Z_2$ are two (generally different) random variables with $[0, 1]$ support. If their probability mass is concentrated close to 0, the results of the tractography process are not significantly affected by noise. On the other hand, if these two random variables are uniformly distributed in $[0, 1]$, the tractography results are completely random and any network inference process is hopeless.

We model the two random variables $Z_1$ and $Z_2$ with the Maximum Entropy distribution with one degree of freedom. In this case, this distribution is the truncated exponential distribution with support $[0, 1]$,

$$f_\zeta(\zeta) = \begin{cases} e^{\frac{-\zeta}{1-\alpha}} - e^{\alpha} & \zeta \in [0, 1] \\ 0 & \text{otherwise} \end{cases}$$

(16)

where $\alpha > 0$ is a parameter that determines the mean and variance of the distribution. Instead of controlling $\alpha$, we control the intensity of noise through the mean of $Z$, $\mu = E[Z] = \frac{1 - (1 + \alpha)e^{-\alpha}}{\alpha (1 - e^{-\alpha})}$

(17)

The two distributions $Z_1$ and $Z_2$ follow this statistical model with means $\mu_1$ and $\mu_2$, respectively. Figure 4 shows the previous distribution for four values of $\mu$. Note that the distribution $Z$ becomes almost “flat” (close to the uniform distribution) when its mean is higher than 0.3, meaning that tractography would be extremely inaccurate when the noise intensity exceeds that level.

In the rest of this paper we limit the range of $\mu_1$ and $\mu_2$ between 0 and 0.3.

**DTI Data, Tractography Parameters, and Selected ROIs**

We also apply MANIA in Diffusion Tensor Imaging (DTI) data collected by an earlier study: “Brain aging in humans, chimpanzees (Pan troglodytes), and rhesus macaques (Macaca mulatta): magnetic resonance imaging studies of macro- and microstructural changes” (Chen et al., 2013). All subjects gave written informed consent, and the study was approved by the Emory University Institutional Review Board. We analyzed the data for 28 of those subjects mostly based on handedness (right-handed females, without a history of psychiatric disorder). Diffusion-weighted images were acquired using a Siemens 3T with a 12-channel parallel imaging phase-array coil. Foam cushions were used to minimize head motion. Diffusion MRI data were collected with a diffusion weighted SF-EPI sequence (Generalized Autocalibrating Partially Parallel Acquisitions [GRAPPA] factor of 2). A dual spin-echo technique combined with bipolar gradients was used to minimize eddy-current effects. The parameters used for diffusion data acquisition...
were as follows: diffusion-weighting gradients applied in 60 directions with a \( b \)-value of 1000 s/mm\(^2\); TR/TE of 8500/95 ms; FOV of 216 \( \times \) 256 mm\(^2\); matrix size of 108 \( \times \) 128; resolution of 2 \( \times \) 2 \( \times \) 2 mm\(^3\); and 64 slices with no gap, covering the whole brain. Averages of 2 sets of diffusion-weighted images with phase-encoding directions of opposite polarity (left–right) were acquired to correct for susceptibility distortion. For each average of diffusion-weighted images, 4 images without diffusion weighting \( (b = 0 \text{ s/mm}^2) \) were also acquired with matching imaging parameters. The total diffusion MRI scan time was approximately 20 min. T1-weighted images were acquired with a 3D MPRAGE sequence (GRAPPA factor of 2) for all participants. The scan protocol, optimized at 3T, used a TR/TE of 2600/900/3.02 ms, flip angle of 8\(^\circ\), volume of view of 224 \( \times \) 256 \( \times \) 176 mm\(^3\), matrix of 224 \( \times \) 256 \( \times \) 176, and resolution of 1 \( \times \) 1 \( \times \) 1 mm\(^3\). Total T1 scan time was approximately 4 min.

The resulting DTI data were processed using the FMRIB’s Diffusion Toolbox (FDT) provided by FSL (FMRIB 4 Software Library) (Behrens et al., 2003). The FDT probabilistic tractography parameters were set to their default values (number of streamlines = 5000, maximum number of steps = 2000, loop check: set, curvature threshold = 0.2, step length = 0.5 mm, no distance bias correction).

We applied MANIA on 18 corticolimbic ROIs. All ROIs are in the left hemisphere and localized in Montreal Neurological Institute (MNI) standard space using the Automated Anatomical Labeling (AAL) (Jenkinson and Smith, 2001) of the WFU PickAtlas toolbox (Lancaster et al., 2000). The ROI acronym as well as the number of voxels in each ROI are shown in Table 1.

The shape of these ROIs are not dilated and we adhere to the standard masks provided in the WFU PickAtlas toolbox. We chose these ROIs because they are known to play a significant role in various psychiatric disorders such as MDD, PTSD, OCD, anxiety and addiction (Mayberg, 1997; Seminowicz et al., 2004; Craddock et al., 2009; James et al., 2009). This sample of ROIs includes cortical, subcortical and limbic regions. Specifically, the cortical ROIs are BA6, BA9, BA10, BA40, BA46, BA47, the limbic are BA24, Th, BS, and the sub-cortical are BA11, BA25, BA32, Acb, Amg, Hp, Ht, Ins, Pcc.

### RESULTS

#### Evaluation with the FiberCup Dataset

**Figure 5A** shows that the MANIA threshold results in almost the same accuracy (Jaccard similarity within 5%) as the optimal threshold (see the Optimal Threshold-Based Network Inference section). However, neither MANIA nor the optimal threshold are able to perfectly reconstruct the correct network. **Figure 5B** shows the network inferred by MANIA. There are two asymmetric false-positive edges (shown in blue) and one false-negative edge (shown in red). However, the confidence metric (shown in **Figure 5C**) of the three miss-classified edges is close to zero, while the confidence metric of the true-positives and true-negatives is significantly different than zero. Further, if we also perform post-symmetrization on MANIA outputs, the false positive edge from ROI-13 to ROI-17 is removed because it has negative confidence.

**Figure 5D** explains the root-cause of the previous miss-classified edges. Seeding Protrackx at ROI-13, most streamlines are incorrectly directed to ROI-17 (instead of ROI-15) because of the overlap with the fiber bundle between ROI-11 and ROI-17. The probability map generated by Protrackx makes its highly unlikely that any threshold-based network inference method would be able to discover the right connection between ROI-13 and ROI-15.

#### Evaluation with Synthetic Data

We evaluate MANIA based on computational experiments with synthetic data and random networks. The “ground-truth” networks \( G \) are constructed as follows. Suppose that \( G \) has \( N \) nodes and density \( \rho \). We place \( \left\lfloor \rho^{\frac{N(N−1)}{2}} \right\rfloor \) undirected edges between randomly selected but distinct pairs of nodes. Note that \( G \) is symmetric by construction because the tractography process cannot infer the true polarity of the underlying neural fibers. Given \( G \), we then create the tractography matrix \( T \) that represents the “noisy” fraction of streamlines between any pair of nodes, as shown in Equation (15). Note that the fraction of streamlines from a node \( X \) to a node \( Y \) is typically different than the fraction of streamlines from \( Y \) to \( X \). In the following experiments, \( N \) is set to 50 nodes, and each experiment is repeated for 1000 networks \( G \).

We first examine the effect of post-symmetrization on the accuracy of both threshold-based network inference and MANIA. In the former, the connections are determined based on a given threshold \( \tau_0 \), as discussed in the Threshold-based Network Inference with Post-Symmetrization section. We denote the Jaccard similarity between the inferred network and the ground-truth network with \( J_{SYM} \) when post-symmetrization is performed, and with \( J_{NO-SYM} \) otherwise. **Figure 6** shows the

<table>
<thead>
<tr>
<th>ROIs</th>
<th>Acronym</th>
<th>Number of voxels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Premotor cortex</td>
<td>BA6</td>
<td>3131</td>
</tr>
<tr>
<td>Insula</td>
<td>Ins</td>
<td>1858</td>
</tr>
<tr>
<td>Ventromedial prefrontal cortex</td>
<td>BA10</td>
<td>1784</td>
</tr>
<tr>
<td>Inferior parietal cortex</td>
<td>BA40</td>
<td>1598</td>
</tr>
<tr>
<td>Dorsolateral prefrontal cortex</td>
<td>BA9</td>
<td>1422</td>
</tr>
<tr>
<td>Mid-brain and pons</td>
<td>BS</td>
<td>1406</td>
</tr>
<tr>
<td>Orbital-frontal cortex</td>
<td>BA11</td>
<td>1243</td>
</tr>
<tr>
<td>Thalamus</td>
<td>Th</td>
<td>1100</td>
</tr>
<tr>
<td>Hippocampus</td>
<td>Hp</td>
<td>932</td>
</tr>
<tr>
<td>Precuneus</td>
<td>Pcc</td>
<td>861</td>
</tr>
<tr>
<td>Inferior prefrontal gyrus</td>
<td>BA47</td>
<td>851</td>
</tr>
<tr>
<td>Ventral anterior cingulate</td>
<td>BA32</td>
<td>721</td>
</tr>
<tr>
<td>Dorsal anterior cingulate cortex</td>
<td>BA24</td>
<td>593</td>
</tr>
<tr>
<td>Dorsolateral prefrontal cortex</td>
<td>BA46</td>
<td>574</td>
</tr>
<tr>
<td>Amygdala</td>
<td>Amg</td>
<td>220</td>
</tr>
<tr>
<td>Subcallosal cingulate</td>
<td>BA25</td>
<td>204</td>
</tr>
<tr>
<td>Nucleus accumbens</td>
<td>Acb</td>
<td>140</td>
</tr>
<tr>
<td>Hypothalamus</td>
<td>Ht</td>
<td>13</td>
</tr>
</tbody>
</table>
difference $\Delta J = J_{SYM} - J_{NO-SYM}$ for several choices of $\tau_0$ as well as for MANIA. Each box-plot is generated from 1000 experiments; in each experiment we generate a random network with density between 0 and 1, while the noise parameters $\mu_1$ and $\mu_2$ are uniformly distributed between 0 and 0.3. The red line corresponds to the median, the box boundaries correspond to the 25 and 75th percentiles, while the dashed lines show the 10 and 90th percentiles. In all cases, $\Delta J > 0$ (one-sided Mann-Whitney U-test—p-values shown next to each box plot), meaning that post-symmetrization helps to improve the accuracy of network inference. This is true for both MANIA and threshold-based inference, even though the improvement is larger for the latter. Because of the positive effect of post-symmetrization, in the rest of the paper we apply it in all network inference experiments.

Figure 7 illustrates the performance of MANIA in the two-dimensional space defined by the noise parameters $\mu_1$ and $\mu_2$ for three values of the network density. Each square in these heat maps is the median across 1000 experiments. The false positive and false negative rates are close to 0 (less than 5%) in the lower-left half of each heat map (i.e., when $\mu_1 + \mu_2 < 0.3$). For higher values of the noise intensity, the accuracy of MANIA depends on the density of the underlying network. In the case of sparse networks, MANIA also infers a sparse network and most errors are false negatives, i.e., MANIA does not detect some of the few existing edges. For dense networks, MANIA also infers a dense network and most errors are false positives, i.e., MANIA detects a few extra edges that do not actually exist. In mid-range densities, the errors are more balanced between false positives and false negatives. In all cases the maximum false positive (or negative) rate when $\mu_1 = \mu_2 = 0.3$ is less than 25%. Recall from Figure 4 that these noise intensity levels should be considered quite high in practice.

Figure 8 compares the MANIA-inferred network with the network that corresponds to the optimal threshold $\tau^{opt}$, as
discussed in the Optimal Threshold-based Network Inference section. Specifically, the heat maps of Figure 8 compare the Jaccard similarity $I_{\text{opt}}$ between MANIA and the ground-truth network, with the Jaccard similarity $I_{\text{opt}}$ between the optimal threshold based network and the ground-truth network. The accuracy of MANIA is typically close to that of the optimal threshold method. Even under the highest noise intensity we consider ($\mu_1 = \mu_2 = 0.3$), $I_{\text{MANIA}}$ is only 10% lower than $I_{\text{opt}}$. These results suggest that MANIA selects automatically a connectivity threshold value that results in almost optimal accuracy, across all possible such threshold values.

Finally, Figure 9 compares MANIA with five given threshold values $\tau_0$. The comparison is in terms of the Jaccard similarity difference $\Delta I = I_{\text{MANIA}} - I_{\text{opt}}$. As in Figure 6, each box-plot is generated from 1000 experiments in which we vary the network density between 0 and 1, and the noise parameters $\mu_1$ and $\mu_2$ between 0 and 0.3. The median $\Delta I$ is always positive and the distribution of $\Delta I$ is skewed toward positive values (one-sided Mann-Whitney U-test – p-values shown next to each box plot) meaning that post-symmetrization helps to improve the accuracy of network inference.

**Case-Study: A Rank-Aggregated Network between 18 ROIs**

We applied MANIA in the DTI data presented in DTI Data, Tractography Parameters, and Selected ROIs, based on the 18 ROIs listed in Table 1. A single rank-aggregated network is constructed, using the group analysis method of the Group Analysis section. Using MANIA, aggregating data from 28 subjects. The rank-aggregated network is shown in the left graph of Figure 10.

Two ROIs (Pc and BA40) appear to not be directly connected with the other 16 ROIs; of course there may be indirect connections through other ROIs that have not been included here (we return to this point in the Discussion section). Every edge in the connected component of Figure 10 has been detected in both directions, i.e., MANIA identifies a completely symmetric network in this case (i.e., no post-symmetrization is needed). The density of the connected component (16 ROIs) is 19%. The color of each edge in the subfigure represents the fraction of the 28 subjects that have the corresponding edge in their individual networks (constructed by MANIA).

We measured the “centrality” of each node in the rank-aggregated network, based on four centrality metrics (degree, closeness, betweenness, PageRank) (Newman, 2010). Different centrality metrics focus on different notions of importance. For instance, the degree centrality metric associates importance with the number of direct connections a node has; BA32 (Ventral anterior cingulate) has the largest number (six) of direct connections in this network (see Table 2). This may be because BA32 is spatially adjacent to both BA10 and BA25, and those ROIs are also of high degree. The betweenness centrality of a node $X$, on the other hand, focuses on the number of shortest paths between any pair of nodes that go through $X$; BA25 (subcallosal cingulate) is the most important node from this perspective because it serves as the “unique bridge” between the 6 red nodes at its left and the 9 blue nodes at its right. BA25 is also the most central node in terms of its average distance to all other nodes (closeness centrality).

Similarly, we measured the edge centrality of all connected node pairs. In terms of edge betweenness centrality, the connection between BA25 and the Nucleus Accumbens (Acb) is by far the most central in this network. It is interesting to note that this edge includes the segment of white matter that is the target of Deep Brain Stimulation (DBS) therapies for the treatment of MDD (Mayberg et al., 2005). In fact, the DBS target is typically the point at which the fibers between (BA25-Acb), (BA25-BA32), and (BA25-BA24) intersect.

Figure 10 also shows some percentiles of the per-subject node-pair confidence metric (median, 25–75th percentiles, 10–90th percentiles, and outliers) for each node-pair that appears connected in at least one of the 28 subjects. The connections between the following node pairs appear in all subjects and have the highest confidence: Hp-Acb, Amg-Acb, BA47-Ins. On the other hand, the following connections appear only in some subjects and their confidence metric varies around zero: Th-BS, BA46-BA9, BA6-Ins. Some connections that appear in 1–2 subjects but have very low confidence are: Pc-BA24, BA11-BA24, Ins-BA25, BA40-BA6.

**DISCUSSION**

Thomas et al. have recently shown that inferring long-range anatomical connections between gray matter ROIs from DWI...
Shadi et al. Symmetry-Based Structural Brain Network Inference

FIGURE 7 | False positive rate and false negative rate of MANIA as a function of $\mu_1$ and $\mu_2$ for sparse networks ($\rho_G = 0.1$), medium density networks ($\rho_G = 0.5$) and dense networks ($\rho_G = 0.9$). Each square is the average of 1000 independent simulations.

FIGURE 8 | The Jaccard similarity difference ($\Delta J = J_{\text{opt}} - J_{\text{MANIA}}$) between MANIA and the optimal threshold-based scheme as a function of $\mu_1$ and $\mu_2$ for sparse networks ($\rho_G = 0.1$), medium density networks ($\rho_G = 0.5$) and dense networks ($\rho_G = 0.9$). Each square is the average of 1000 independent simulations.

data can be significantly inaccurate (Thomas et al., 2014). The authors also note that “(probabilistic tractography methods) are less susceptible to changes in the composition of an ROI but only if an optimized threshold can be derived and used.” More recently Reveley et al. (2015) have investigated the key reasons behind the negative results of Thomas et al. (2014). They showed that the dense system of white matter fibers residing just under the cortical sheet poses severe challenges for long-range tractography, concluding that it is “extremely difficult to determine precisely where small axonal tracts join and leave larger white matter fasciculi.”

In light of the previous results, it is clear that there is a need for new network inference methods. We believe that MANIA is moving in the right direction for the following reasons:

(a) The results of Thomas et al. (2014) suggest that probabilistic tractography can be more accurate than deterministic tractography in terms of sensitivity and specificity as long as its parameters are appropriately optimized. MANIA is indeed mostly applicable to probabilistic tractography, and its main focus is how to “self-configure” its connectivity threshold $\tau$ in an optimized manner, relying on what we expect to be true about the
suggest that it is risky to dilate the given ROIs, which are typically mostly gray matter, so that they also include some white matter voxels. Those voxels may be part of the white matter fiber systems that reside just under the cortical sheet. In other words, if our goal is to understand the connectivity between gray matter ROIs, we should not use tractography seeds that reside in white matter; instead, we need to seed from gray matter even if the diffusion signal is much weaker there. So, we need to expect that some connections may appear as asymmetric, which is what MANIA anticipates.

(b) The results of Reveley et al. (2015) suggest that it is risky to dilate the given ROIs, which are typically mostly gray matter, so that they also include some white matter voxels. Those voxels may be part of the white matter fiber systems that reside just under the cortical sheet. In other words, if our goal is to understand the connectivity between gray matter ROIs, we should not use tractography seeds that reside in white matter; instead, we need to seed from gray matter even if the diffusion signal is much weaker there. So, we need to expect that some connections may appear as asymmetric, which is what MANIA anticipates.

(c) The results of Reveley et al. (2015) can be interpreted as follows: because it is hard for any tractography method to accurately cross the WGB, especially in the case of cortical ROIs, a network inference method should be able to deal somehow with erroneous measurements about specific connections. In other words, just because tractography failed to cross the WGB going from seed X to target Y does not mean that we should conclude that X and Y are not connected. And so, given that the input data about individual connections is quite noisy, we need to examine if there is any additional “hidden structure” in the inference problem that we can exploit. If this is case, we can then look for a solution that satisfies the constraints of that additional structure. In MANIA, this “hidden structure” in the inference problem is that the resulting network should be as symmetric as possible.

(d) Recently, Zalesky et al. (2016) argued that inferring brain networks with both high specificity and sensitivity is beyond the capabilities of current diffusion-weighted MRI, fiber modeling and tractography methods. Further, they showed that specificity is at least twice as important as sensitivity when estimating certain topological properties of brain networks (clustering, efficiency, modularity). As shown in the Evaluation with Synthetic Data section, MANIA tends to emphasize specificity instead of sensitivity (i.e., most errors are false negatives) under sparse connectivity, such as the network between distinct brain modules.

Of course, we do not claim that MANIA addresses every concern about tractography-based network inference. On the contrary, there are more open issues that need to be addressed. Two of them are further discussed next.

Even if the thresholding problem is adequately addressed with MANIA, there is another important problem in structural network inference: the distance bias of the tractography process (Donahue et al., 2016). It is harder to discover connections between distal regions due to the accumulation of uncertainty in long streamlines, causing false negatives for long-range connections (Li et al., 2012a,b). Additionally, it is more likely to incorrectly detect connections between proximal regions, especially in the presence of crossing or turning fibers, causing false positives. The FDT toolbox provides a “distance correction” option by multiplying the number of streamlines that cross a voxel by the average length of those streamlines4—there is no evidence however that this simple form of distance correction is able to improve significantly the accuracy of the network inference process. More sophisticated methods exist however (Morris et al., 2008; Donahue et al., 2016). The method by Morris et al. compares the tractography-generated connectivity probabilities with a null model that gives the corresponding connectivity probabilities with a random tracking process that is dominated by the same distance effects. The method by Donahue et al. regresses out the distance bias from the inter-ROIs connection weights using a log-linear model. We view distance correction methods as an independent processing step that can be applied prior to applying MANIA. For instance, the method by Morris et al. first creates a “null frequency of connection map,” it then filters the “experimental frequency of connection map” that is produced by a probabilistic tractography tool, resulting in the so-called “significance of connection map” (which is supposed to have fewer false positives). MANIA can be then applied on the latter, rather than on the experimental frequency of connection map. It is still not clear if the Morris distance correction method is sufficient to completely address the distance correction bias (Taljan et al., 2011), or whether the method of Donahue et al. is sufficient to completely address the distance bias for cortico-subcortical connections (van den Heuvel et al., 2015). Nevertheless, we anticipate that the combination

4http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FDT/UserGuide.
of MANIA with a distance correction method will improve the accuracy of the resulting networks.

MANIA does not produce a weighted network in which a strength metric is associated with each edge. It provides however a confidence metric for the presence or absence of each possible connection. Interpreting streamline counts in probabilistic tractography as an indication of connection strength is debatable (Jones et al., 2013). It is not yet clear how to differentiate between weak yet probable connections vs. strong but unlikely connections (Zalesky et al., 2016). Nevertheless, a weighted approach like that of Donahue et al. returns an almost fully connected network with weights that span several orders of magnitude (Donahue et al., 2016). In order to achieve good sensitivity and specificity, there is again the need to threshold this weighted connectome (see Figure 4D of the paper by Donahue et al.). For example, in Azadbakht et al. (2015) the authors achieve 74% accuracy but only if the optimum threshold is used. MANIA can address this thresholding problem by first pruning the tractography results to get a set of unweighted but likely connections. Then, one can define the desired weighting scheme for this set, e.g., through fractional scaling (see Donahue et al., 2016).

A network representation consists of both nodes and edges. The clinical and research value of representing a brain as a network depends critically on the selected nodes and on the exact boundaries of the corresponding ROIs (Zalesky et al., 2010). MANIA assumes that the set of given nodes is sufficiently specified, and that their spatial boundaries are accurately defined. In practice, this step of the network inference process is always an "inexact science" given that the functional role of any given ROI is at best only partially understood and the anatomical boundary of each ROI is subject-dependent (Tzourio-Mazoyer et al., 2002).

A voxel-level analysis (van den Heuvel et al., 2008) avoids the selection of functionally specific ROIs but it makes it harder to associate the topological properties of the observed network, which now consists of many thousands of nodes, to any known brain circuits and their function. Again, we view this important issue as orthogonal to MANIA: improved brain parcellation methods, such as data-driven parcellations (Power et al., 2011; Yeo et al., 2011; Glasser et al., 2016b) and decreasing voxel sizes can be used jointly with MANIA to identify structural networks that are consistent, or that can explain well, the observed spatio-temporal correlations in resting-state or task-based fMRI analyses. This coupled exploitation of fMRI and diffusion MRI data has provided valuable insights about the underlying anatomy of the brain structures that result in the...
Default Mode Network (Honey et al., 2009), and they can become more common now that the HCP project provides both functional and diffusion data for hundreds of subjects (Van Essen et al., 2013).

In our 18-ROI case-study, summarized in Figure 10, the use of mostly large ROIs that do not necessarily correspond to distinct functional units, together with the distance bias of the tractography process, may account for the lack of certain expected connections. Two such expected connections are between Pc and BA40 (Greicius et al., 2009), and between BA9 and BA40 (Petrides and Pandya, 2007); the latter is a long-distance connection. Additionally, large cortical ROIs such as BA9 and BA40 are only imprecisely defined, which may also explain the absence of some of their connections. The limbic and subcortical ROIs, on their other hand, are more precisely defined and their connections are mostly running over shorter distances.

These findings suggest that MANIA should be evaluated in the future jointly with, first, advanced distance correction methods, and second, with either more precisely defined ROIs or on a whole-brain parcellation template.

**AUTHOR CONTRIBUTIONS**

KS and CD designed research, KS, SB, DG, HM, and CD performed research, KS and SB analyzed data, and KS, HM, and CD wrote paper.

**ACKNOWLEDGMENTS**

CD and SB are grateful to Prof. Olaf Sporns and his group at Indiana University for hosting SB at the Sporns lab during the summer of 2011 and for discussions about early stages of this work. The authors would like to thank Dr. Jim Rilling for his contribution of the DTI and structural MRI data that served as basis for the analysis presented in this paper, and Dr. Peter Neher for providing us with the FiberCup phantom ROIs.

---

**REFERENCES**


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.

Copyright © 2016 Shadi, Bakhshi, Gutman, Mayberg and Dovrolis. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.