2012 IEEE 12th International Conference on Data Mining Workshops

Discovering Aberrant Patterns of Human Connectome in Alzheimer's Disease via Subgraph Mining

Junming Shao*†, Qinli Yang‡, Afra Wohlschläger†, Christian Sorg†
*Department of Neuroradiology, Technical University of Munich, Germany
†Institute for Computer Science, University of Mainz, Germany
‡Institute for Computer Science, University of Munich, Germany

***Abstract*—Alzheimer's disease (AD) is the most common cause of age-related dementia, which prominently affects the human connectome. Diffusion weighted imaging (DWI) provides a pro-mising way to explore the organization of white matter fiber tracts in the human brain in a non-invasive way. However, the immense amount of data from millions of voxels of a raw diffusion map prevent an easy way to utilizable knowledge. In this paper, we focus on the question how we can identify disrupted spatial patterns of the human connectome in AD based on a data mining framework. Using diffusion tractography, the human connectomes for each individual subject were constructed based on two diffusion derived attributes: fiber density and fractional anisotropy, to represent the structural brain connectivity patterns. Then, these human connectomes were further mapped into a series of unweighted graphs by discretization. After frequent subgraph mining, the abnormal score was finally defined to identify disrupted subgraph patterns in patients. Experiments demonstrated that our data-driven approach, for the first time, allows identifying selective spatial pattern changes of the human connectome in AD that perfectly matched grey matter changes of the disease. Our findings further bring new insights into how AD propagates and disrupts the regional integrity of large-scale structural brain networks in a fiber connectivity-based way.**

***Keywords*-Human Connectome, Diffusion Tensor Imaging, Alzheimer's Disease, Subgraph Mining**

I. Introduction

Unraveling the structural connectivity in the human brain (i.e. the human connectome [17]) is challenging for neuroscientists for over a century. Brain's wiring captures the basic feature of nervous system organization and is fundamental to understand the brain's functions, which depend critically on the integration of regionally remote neural activity. In recent years, the advances of imaging technologies have provided a promising avenue of exploring structural brain connectivity in-vivo. The study of the human connectome has therefore attracted a huge of attention, especially for the research on various pathologies related to white matter changes such as Alzheimer's disease, Schizophrenia, and Multiple Sclerosis. Discovering the disrupted wiring of the brain is essential for a better understanding of the pathophysiology of these diseases, finally to target potential treatments more specifically. Thereby a critical question is how to map aberrant connectivity spatially as specific as possible in these diseases. In this study, we aim to answer this question in patients with Alzheimer's disease based on a data mining framework.

Alzheimer's disease (AD), a neurodegenerative disorder characterized by increasing cognitive and behavioral deficits [1], is neuropathologically defined by amyloid plaques, neurofibrillary tangles and the loss of neurons (i.e. atrophy), with changes starting regionally and spreading out gradually across brain's grey matter [2], [19]. Meanwhile, post-mortem histological and in-vivo imaging studies demonstrated widespread alterations of patients' white matter, involving frontal, occipital, and temporal lobes [7], [3] and selected tracts such as the corpus callosum or cingulum [15]. By far, AD is the most common cause of age-related dementia. Other causes are vascular diseases such as stroke or neurodegenerative diseases such as Lewy body disease or frontotemporal lobe degeneration. Due to the "aging society", more and more people are diagnosed with dementia. Therefore, there is a strong interest in expanding our knowledge of diseases causing dementia, especially for AD. Recent findings demonstrated the selective disruption of large-scale brain functional networks. However, the aberrant patterns in structural brain networks in AD are still poorly understood.

Diffusion weighted magnetic resonance imaging (DWI) is a technique that can be used to explore white matter microstructure in-vivo, using water diffusion properties as a probe. The DWI signal is sensitive for the diffusion of water molecules that it is along the direction of axons and restricted in the direction perpendicular to them. During DWI, multiple brain images are acquired where each is sensitive for a distinct direction. At each voxel, measured data are fitted to a mathematical diffusion tensor model describing diffusion as an ellipsoid or tensor. Both local properties of water diffusion (such as fractional anisotropy (FA) or mean diffusivity (MD)) can be derived from the voxel-wise diffusion tensor [12]. However, after diffusion tensor calculation, we obtain the diffusion-derived FA or MD maps which include millions of FA or MD values to capture the white matter microstructure. The remaining problem in clinical research is how can we make the information of millions of voxels to utilizable knowledge. In particular, how to find the disrupted structural connectivity

978-0-7695-4925-5/12 $26.00 © 2012 IEEE
DOI 10.1109/ICDMW.2012.9

in AD relying on these vast amount of data. To date, the most frequently used method is to perform the voxel-wise comparison between the group of AD and healthy controls (HC) using mass-univariate statistical analysis [3], [15]. In principle, this type of methods tend to lose the individual information and is restricted to data with Gaussian distribution. Moreover, selectively disrupted spatial patterns of the structural connectivity in the human brain cannot be explored. Therefore, a more promising approach is to construct the individual structural connectivity networks for each subject and then investigate the aberrant connectivity patterns in AD by introducing the subgraph mining. To the best of our knowledge, it is the first time to relate the subgraph mining to large-scale structural brain networks.

The remainder of the paper is organized as follows. We start with a review of related work in Section II. Section III provides a description of the data set including the preprocessing steps. In Section IV we subsequently explain our framework and how we construct the human connectome, map the weighted networks into multiple unweighted networks and discover the aberrant patterns of human connectome based on frequent subgraph mining. Section V presents our results and the interpretations by medical experts are further given in Section VI. The conclusion is presented in Section VII.

II. Related work

During the past several decades, the study on dementia and AD has attracted a large amount of attention in diverse fields. Currently, one of the most common used approaches to study AD by neuroimaging is called statistical parametric mapping (SPM), which investigates the change of each voxel independently, i.e. a mass-univariate statistical analysis, such as voxel-wise Student's t-test or ANOVA. Here, only a brief survey of AD studies in the context of data mining is provided, which can be mainly classified into two categories.

Graph-based Analysis: Unlike the voxel-besed analysis such as SPM, the graph theoretical analysis focuses on the large-scale brain networks based on graph properties, such as clustering coefficient, shortest path, centrality and efficiency. Stam et al. [18] applied graph-based analysis on the functional connectivity, i.e. synchrony of time series of neural signals, of beta band-filtered electroencephalography (EEG) channels and found that AD was characterized by a loss of small-world network characteristics. He et al. [9] investigated the aberrant large-scale structural networks derived from cortical thickness measurements in AD by exploring the small world property, node centrality and network robustness. Recently, based on the DWI-based tractography and graph theory, Lo et al. [13] has indicated that AD dementia leads to changes in the topological organization of individual, fiber-based, structural connectivity networks, and that these changes correlate with cognitive deficits. However, all these approaches investigate the aberrant patterns in AD from a global perspective, the local disrupted spatial patterns are not considered.

Patterns-based AD Prediction: The quick, reliable and early detection of AD in predementia stages is a challenging problem in clinical diagnosis as it is largely based on symptom history and examination supported by neuropsychological evidence of the profile of cognitive impairment. Biological particularly neuroimaging-based score are in most cases only used to exclude other potential causes of cognitive impairments. To deal with such problem, the machine learning techniques were applied for the improvement of AD diagnosis. Klöppel et al. [10] used T1-weighted grey matter images and SVM to separate pathologically verified AD patients from healthy controls with accuracy of 96%. Similarly, Davatzikos et al. [4] first identified the grey matter regions in which the tissue density correlated well with the clinical variable via a voxel-by-voxel calculation of the Pearson correlation coefficient and then classified AD subjects via a pattern classification technique. Recently, Shao et al. [16] constructed the individual human connectome and used the whole connectivity patterns to predict AD based on feature selection and classification. This finding suggested that individual human connectome may has the potential of providing an imaging- and white matter-based biomarker for the distinction between healthy aging and very early AD. All these approaches allow deriving a feature vector representing the specific pattern of atrophy in a single individual and then predict AD based on the pattern classification scheme.

However, both graph-based analysis and patterns-based AD prediction do not allow investigating regionally specific connectivity patterns that represent AD's effect. I.e. these approaches can not answer the question which regional connectivity-pattern is affected by AD. The study presented in this paper differs from previous work by combining DWI-based tractography and subgraph pattern mining. In contrast to voxel-based analysis, we reconstructed the human connectome which naturally reflect the density and integrity of long-range connections between cortical regions. The aberrant patterns of human connectome were detected by subgraph mining, which is able to provide new insight into selectively disrupted spatial connectivity in AD.

III. Data and Preprocessing

In this study, 17 patients with AD (range 55 to 83 years with an average of 68.9 years +/- 8.1, 7 females) and 21 healthy controls (range 56 to 85 years with an average of 66.3 +/-7.4, 13 females) have participated. All participants provided informed consent in accordance with the Human Research Committee guidelines of the Klinikum Rechts der Isar, Technische Universität, München. Examination of every participant included medical history, neurological examination, neuropsychological assessment, structural MRI, and (for patients only) informant interview (Clinical Dementia Rating) as well as blood tests.

For each subject, we acquired diffusion weighted MRI images on a 3T-MRI scanner (Achieva, Philips) using a pulsed gradient spin-echo echo planar imaging sequence with a parallel imaging (SENSE) factor of 2.5, TE = 60 ms, and TR = 6516 ms. Images were measured for 112×112 matrix size of slice and subsequently reconstructed for a 128×128 matrix size, with a resolution of 1.75 mm in plane and a slice thickness of 2 mm. A total of 60 contiguous slices were acquired to give complete brain coverage containing 128×128×60 voxels with size 1.75 × 1.75×2 mm^3. Diffusion gradients were applied in 15 non-collinear directions with b = 800 s/mm^2. B0 image without diffusion weighting, b = 0 s/mm^2, was additionally acquired. Distortion induced by motion for diffusion weighted MRI images was first corrected by aligning all diffusion-weighted images to the non-diffusion (B0) image using a 2-D linear registration algorithm (AIR, Automated Image Registration) [20]. After that, all diffusion-weighted images were fitted to a mathematical diffusion tensor model for each voxel. The local diffusion measures like FA and MD, reflected the local density and integrity of white matter, were calculated voxel-by-voxel.

IV. Method

In this section, we proposed a data mining framework to discover the aberrant patterns of human connectcome in AD. In the first step, we constructed the human connectomes to represent the connectivity patterns of fiber pathways for each subject. After that, these human connectomes were further transformed into a set of unweighted graphs by discretization. We then applied frequent subgraph mining algorithm FSG on the data set of all unweighted graphs. In our case, the goal was to find the most aberrant subgraphs which are different between healthy controls and AD patients. Therefore, we further defined the abnormal score to measure the degree of abnormality of these subgraph patterns. In the following, we start with some preliminaries.

A. Preliminaries

Definition 1 (**Labeled Undirected Graph**) Let $g = (V, E, L, l)$ be a labeled undirected graph, where V is a set of nodes and $E \subseteq V \times V$ is a set of edges. $e = \{u, v\} \in E$ indicates a connection between the nodes u and v. L is a set of labels, and l is a mapping function that assigns labels to vertices V and edges E. In this study, only the nodes are associated with labels.

Definition 2 (**Subgraph**) For two labeled undirected graphs $g_s = (V_s, E_s, L_s, l_s)$ and $g = (V, E, L, l)$, we say g_s is a subgraph of g if $V_s \subseteq V, E_s \subseteq E, L_s \subseteq L$ and $l_s = l$.

Definition 3 (**Isomorphism**) For two labeled undirected graphs $g' = (V', E', L', l')$ and $g = (V, E, L, l)$, if there is a bijection f between the nodes of graph g' and g, $f : V(g' \rightarrow V(g))$ such that $e' = \{u', v'\} \in E'$ iff $e = \{u, v\} \in E$ where $u = f(u')$ and $v = f(v')$, then g' and g are isomorphic graphs.

Definition 4 (**Subgraph Frequency**) Given a set of graph G, the frequency of a subgraph g_s is defined as:

$$fq(g_s|G) = \frac{|g_s\text{is a subgraph of } g \text{ and } g \in G|}{|G|}$$

B. Human Connectome Construction

Here we constructed the human connectome for each subject. Figure 1 illustrates the construction of human connectome in our approach, which was mainly involved the following three steps.

Cortical Parcellation: For cortical segmentation, the whole brain of each participant was segmented into 96 cortical regions via Harvard-Oxford cortical structural atlas (http:// www.fmrib.ox.ac.uk/fsl). Each individual's non-diffusion weighted (B0) image was first affine-registered to the ICBM 152 template of Montreal Neurological Institute space (MNI, http://www.bic.mni. mcgill.ca), which represents a standardized brain template of the same stereotactic space as Harvard-Oxford atlas, to obtain the transformation matrix (T). The inverse transformation matrix (T^{-1}) was then applied to the Harvard-Oxford atlas and (B0) image to generate corresponding cortical regions in each individual's diffusion-weighted image native space. Table I shows the name of cortical regions in this Harvard-Oxford atlas.

Diffusion Tractography: To determine the connectivity between pair-wise regions, diffusion tractography was used. Here, the deterministic fiber tracking algorithm TEND [14] was applied to investigate brain's white matter for each subject. Since voxels with high FA are more likely to contain a high proportion of white matter, all voxels with FA¿0.3 were selected as seed points of fiber tracking [14]. Tracking started from these seed points, using the major eigenvectors of seed points as the original propagation directions. Then tracts were propagated by using the entire diffusion tensor to deflect the estimated fiber trajectory in both directions. Tracking stopped in voxels with $FA < 0.2$ or physiologically implausible curvature of the track (> 60 degrees) [14].

Human Connectomes: The output of both cortical parcellation and diffusion tractography was combined to construct individual human connectome for each subject. Each atlas-based region was regarded as a network node. Connectivity/edge of each pair of nodes was measured by fibers across two regions. If there existed at least one fiber with endpoints in one pair of regions (e.g. region i and region j), the two cortical regions were assumed to be connected [8]. For each connection two attributes were calculated: (a) fiber density FD_{ij} of a connection was defined as proportion of all fibers connecting the two regions (n_{ij}) over the total number of fibers of the subject (n_{all}), i.e. $FD_{ij} = n_{ij}/n_{all}$. (b) FA_{ij} of a connection was defined as the mean value of FA across all voxels of all connection fibers. Each attribute

Table I
THE ANATOMICAL LABELS OF HARVARD-OXFORD ATLAS CORTICAL REGIONS.

ID	Cortical Regions (Abbreviation)	ID	Cortical Regions (Abbreviation)
1	Frontal Pole (FP)	25	Frontal Medial Cortex (FMC)
2	Insular Cortex (IC)	26	Supplementary Motor Cortex (JLC)
3	Superior Frontal Gyrus (SFG)	27	Subcallosal Cortex (SC)
4	Middle Frontal Gyrus (MFG)	28	Paracingulate Gyrus (PG)
5	Inferior Frontal Gyrus, pars triangularis (IFGT)	29	Cingulate Gyrus, anterior division (CGA)
6	Inferior Frontal Gyrus, pars opercularis (IFGO)	30	Cingulate Gyrus, posterior division (CGP)
7	Precentral Gyrus (PG)	31	Precuneous Cortex (PC)
8	Temporal Pole (TP)	32	Cuneal Cortex (CC)
9	Superior Temporal Gyrus, anterior division (STGA)	33	Frontal Orbital Cortex (FOC)
10	Superior Temporal Gyrus, posterior division (STGP)	34	Parahippocampal Gyrus, anterior division (PGA)
11	Middle Temporal Gyrus, anterior division (MTGA)	35	Parahippocampal Gyrus, posterior division (PGP)
12	Middle Temporal Gyrus, posterior division (MTGP)	36	Lingual Gyrus (LG)
13	Middle Temporal Gyrus, temporooccipital part (MTGT)	37	Temporal Fusiform Cortex, anterior division (TFCA)
14	Inferior Temporal Gyrus, anterior division (ITGA)	38	Temporal Fusiform Cortex, posterior division (TFCP)
15	Inferior Temporal Gyrus, posterior division (ITGP)	39	Temporal Occipital Fusiform Cortex (TOFCA)
16	Inferior Temporal Gyrus, temporooccipital part (ITGT)	40	Occipital Fusiform Gyrus (TOFCP)
17	Postcentral Gyrus (POG)	41	Frontal Operculum Cortex (FOC)
18	Superior Parietal Lobule (SPL)	42	Central Opercular Cortex (COC)
19	Supramarginal Gyrus, anterior division (SGA)	43	Parietal Operculum Cortex (POC)
20	Supramarginal Gyrus, posterior division (SGP)	44	Planum Polare (PP)
21	Angular Gyrus (AG)	45	Heschl's Gyrus, includes H1 and H2 (HG)
22	Lateral Occipital Cortex, superior division (LOCS)	46	Planum Temporale (PT)
23	Lateral Occipital Cortex, inferior division (LOCI)	47	Supracalcarine Cortex (SC)
24	Intracalcarine Cortex (IC)	48	Occipital Pole (OP)

Note: As the left and right hemispheres are symmetrical, only the anatomical names of 48 cortical regions are displayed. *.L and *.R indicate the left and right cortical regions in the study.

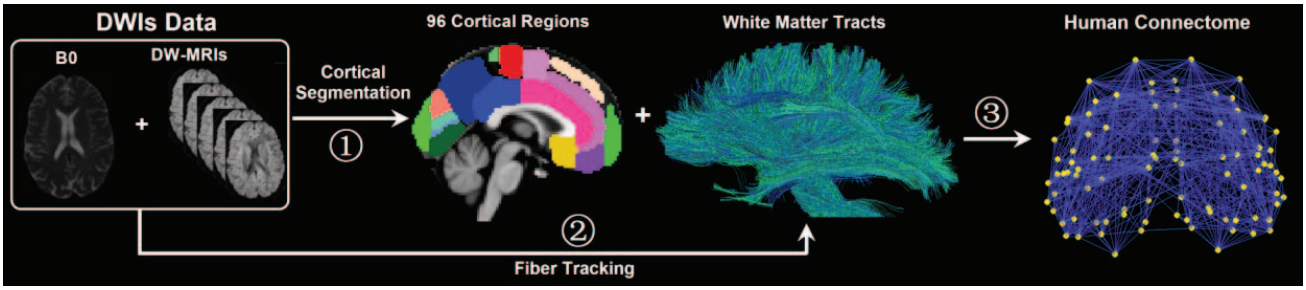

Figure 1. The framework of our approach to construct human connectome based on cortical segmentation and white matter tractography. (1) Harvard-Oxford cortical structural atlas were applied to generate corresponding cortical regions in each individual's DWI native space. (2) Whole brain diffusion tractography was performed to construct axonal trajectories across the entire white matter. (3) Individual human connectome were constructed by combining the output of both cortical parcellation and diffusion tractography.

characterizes the different properties of the connection, see discussion in Section VI and Fig. 5(A1, A2). Finally two human connectomes based on attributes of FD and FA were obtained for each subject.

C. Converting Weighted Networks into a Set of Unweighted Networks

With cortical parcellation and diffusion tractography, we finally established two human connectomes for each subject based on the attributes of FD and FA. However, it is a nontrivial task to quantify the regional connectivity patterns in the weighted networks directly. To better capture the disrupted spatial patterns of structural brain connectivity, the weighted networks were simply mapped into a set of unweighted graphs by thresholding with different weights. In this way, the discovery of the disrupted spatial patterns in weighted networks can be easily explored on multiple unweighted graphs.

In our study, specifically, for each connection, we first calculated the minimal and maximal weights w_{ij}^{min} and w_{ij}^{max} across all subjects. Then we divided the interval $[w_{ij}^{min}\ w_{ij}^{max}]$ into K bins (K = 30 in this study). In principle, the larger K, the less information is lost due to the discretization. Then we used different $K-1$ cutoff values to map the weighted networks into a series of unweighted networks. Figure 2 gives a simple example of mapping weighted network into unweighted networks. Take the connection between nodes A and B for illustration.

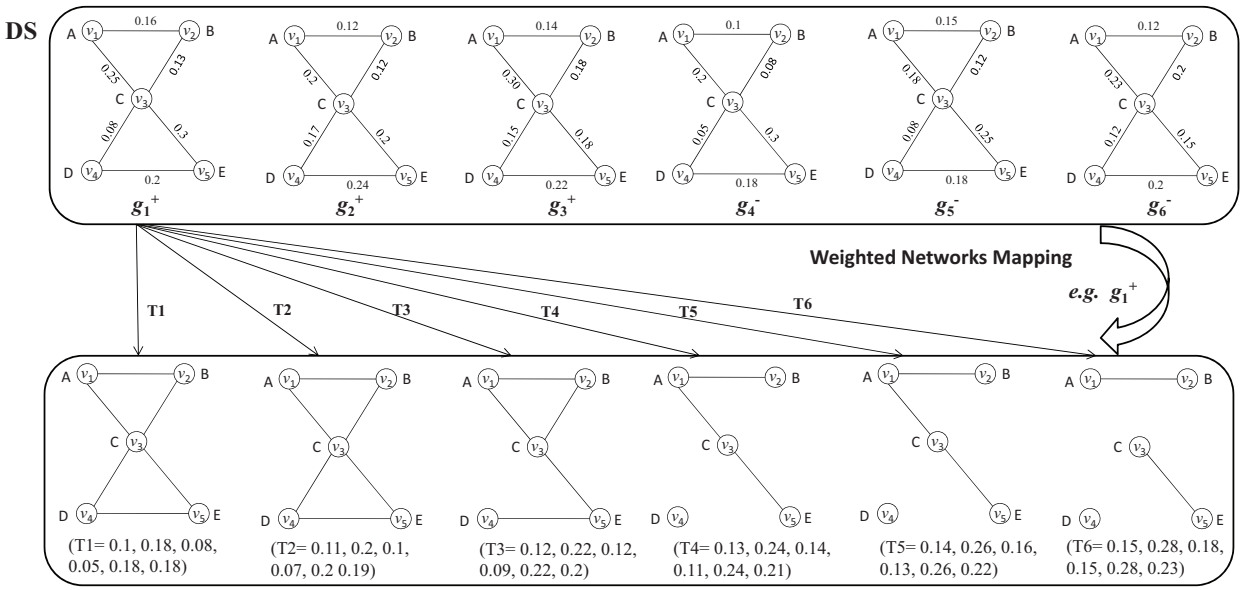

Figure 2. Illustration of the strategy of mapping the weighted network into a set of unweighted networks. Here, DS is the data set. g_1^+, g_2^+, g_3^+ and g_4^-, g_5^-, g_6^- indicate the positive samples (healthy subjects) and negative samples (AD patients) respectively. $K = 6$ and $T1 = (e_1^1, e_2^1, e_3^1, e_4^1, e_5^1, e_6^1)$ represent the first thresholding with cutoff values for edges AB, AC, BC, CD, CE and DE respectively. $(T1, \cdots, T6)$ indicate the different thresholds of cutoff values for graph g_1^+.

The range of weights of this connection in this data set is $[0.1\ 0.16]$ and then is divided into $K = 6$ bins. The cutoff values of these bins were $(0.1, 0.11, 0.12, 0.13, 0.14, 0.15)$ respectively. Using these cutoff values, we can convert this weighted connection into several unweighted connections. This procedure was repeated for each connection, then each weighted graph (e.g. g_1^+) can be transformed into a set of unweighted graphs as indicated in the bottom row of Figure 2.

D. Frequent Subgraph Mining on Brain Structural Connectivity Networks

After weighted networks mapping, in this section we applied the FSG algorithm [11] to find all connected subgraphs that appear frequently in our brain structural networks database. The FSG algorithm is actually an extension of traditional frequent itemset discovery where edges in subgraphs items correspond to the items in a data set of transactions. Before the description of FSG, we first give the definition of frequent subgraph mining.

DEFINITION 5 (FREQUENT SUBGRAPH MINING) Given a set of undirected labeled graphs G and a parameter s such that $0 < s \leq 1$, find all connected undirected graphs that are subgraphs in at least $s \cdot |G|$ of the input graphs.

FSG first searched for all frequent single and double edge subgraphs. Afterwards, the algorithm discovered the larger frequent subgraphs by iteratively applying the following two steps:

Candidate Generation: In this phase, all candidate subgraphs of size $(k+1)$ were generated by joining two frequent size-k subgraphs that they have a common size-$(k-1)$ subgraph and the interaction of the two size-$(k-1)$ connected subgraphs with the smallest and the second smallest canonical label of the two frequent subgraphs was $\emptyset$. Canonical labeling is a unique string code used to represent a given graph, which can efficiently perform a number of operations such as checking whether or not a candidate pattern satisfies the downward closure property of the support condition, or frequent subgraph ordering. By the use of this strategy of candidate generation, the number of redundant and non-downward closed patterns can be effectively pruned.

Frequency Counting: Once candidate subgraphs were generated, their frequencies were computed using an Apriori style strategy. In addition, unlike building the hash-tree for efficient frequent itemset discovery, FSG used Transaction identifier (TID) lists to speed up frequency counting. With this TID, the algorithm can prune the candidate subgraph without performing any subgraph isomorphism computations if the intersection of the TID lists was bellow the minimum support level.

E. Detection of Patterns in Brain Connectome

After performing frequent subgraph mining, possible existing subgraphs of the structural network were obtained. However, what we were interested in was to find the aberrant subgraph patterns in AD. Formally, given a positive set and a negative set of graphs, the goal was to find the

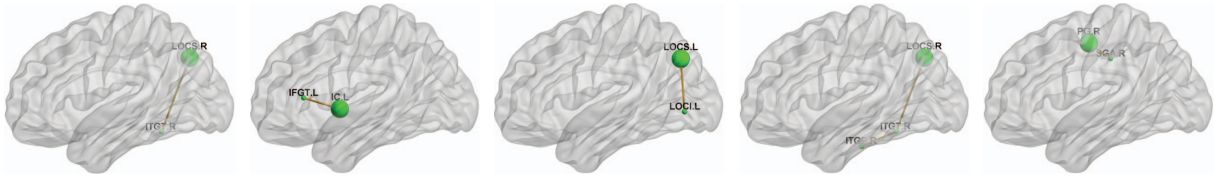

Figure 3. The top 5 aberrant patterns of brain connectome in AD based on the attribute of Fiber Density.

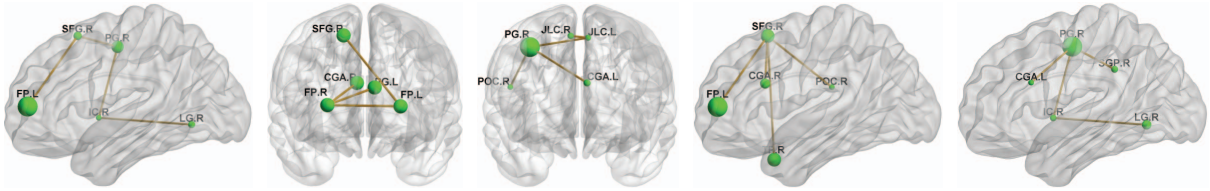

Figure 4. The top 5 aberrant patterns of brain connectome in AD based on the attribute of FA.

discriminative subgraph patterns with the highest difference between two sets of graphs. As the integrity of structural brain connectivity decreased in AD, in this study, we only needed to consider aberrant subgraph patterns with high frequencies in group of healthy subjects and low frequencies in the AD group. Therefore, we defined the *abnormal score*, to measure the degree of abnormality of frequent subgraph patterns between the groups of healthy controls and AD.

DEFINITION 5 (**ABNORMAL SCORE**) Intuitively, the abnormal score of a subgraph pattern g_s is simply defined as the difference between its positive and negative frequency.

$$Score(g_s) = fq(g_s|G^+) - fq(g_s|G^-)$$

The larger the abnormal score, the larger the difference of the pattern between two groups. $Score(g_s) = 1$ means the subgraph g_s exists in all graphs of the group of healthy subjects and there is no such pattern in any graphs of the group of AD.

V. EXPERIMENTS

In this section, we reported a series of experiments to evaluate the performance of our approach. After the construction of human connectomes and weighted networks mapping, we applied the FSG algorithm to find frequent subgraphs in all unweighted networks with support $s = 40\%$ and the subgraphs with maximal edges of 4 in all experiments.

A. AD-related Abnormal Patterns

Based on the constructed human connectomces for each subject, we investigated aberrant spatial sub-patterns in AD by frequent subgraph mining and abnormal score. Figure 3 and 4 plot most important disrupted subgraph patterns in AD based on the attributes of FD and FA, respectively. Due to space limitation, only the top 8 patterns were displayed. In Figure 3, most disrupted sub-patterns based on the attribute of FD have only one edge. Importantly, the disrupted wiring was focused on the parietal, temporal, and occipital lobes, which were primarily affected by cell loss and atrophy in early stages of AD dementia.. Compared to fiber density, the disrupted patterns based on FA were of a relatively large scale, where patterns spanned several lobes but consistently including frontal ones. For more detailed interpretation of these distinctive abnormal patterns concerning FD and FA, see the discussion in Section VI. In order to evaluate these abnormal patterns, we further analyzed the significance of these patterns to separate two groups. In detail, for each abnormal pattern, we computed the average value of FA and FD across edges. Then the comparison of this pattern across the two groups was performed by the use of two sample t-tests. Table II shows abnormal scores and corresponding P-values of top 10 disrupted patterns based on the attribute of FD and FA, respectively. All subgraph patterns were significantly different across groups for both fiber density and fiber fractional anisotropy ($P - value < 0.05$).

Table II
EVALUATION OF THE ABNORMAL SUBGRAPH PATTERNS WITH STATISTICAL T-TEST. AS: ABNORMAL SCORE.

Order	Fiber Density		Fractional Anisotropy	
	AS	P-Value	AS	P-Value
1	0.289	3.99e-04	0.564	2.13e-05
2	0.275	6.71e-04	0.564	3.38e-05
3	0.263	1.30e-03	0.560	1.39e-05
4	0.243	3.30e-04	0.555	1.51e-04
5	0.236	0.013	0.551	1.10e-3
6	0.228	0.005	0.551	7.06e-04
7	0.225	0.009	0.549	3.84e-04
8	0.216	0.022	0.548	1.27e-05
9	0.214	0.007	0.546	1.29e-04
10	0.194	0.004	0.545	2.21e-04

B. Hotspot Regions

Beyond the mining of disrupted connectivity patterns, we were also interested in most prominently affected cortical regions in AD concerning fiber density and integrity. To identify these hotspot regions, we examined all detected top

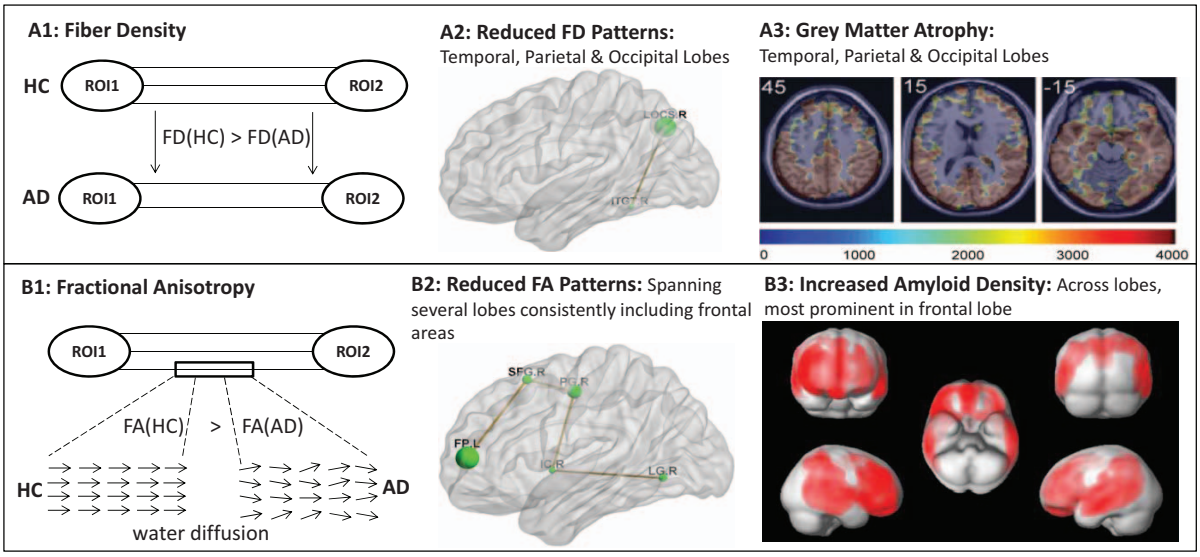

Figure 5. The interpretation of disrupted spatial patterns. A1: Illustration of atrophy of white matter characterized by fiber density; A2: Disrupted spatial patterns derived by fiber density mainly located in temporal, parietal and occipital lobes; A3: Corresponding finding of grey matter atrophy in temporal, parietal and occipital lobes in [5]. B1: Illustration of decreased integrity of white matter characterized by FA; B2: Disrupted spatial patterns derived by FA spanning several lobes consistently including frontal areas; B3: Corresponding increased amyloid density across lobes, but also prominent in frontal lobes [6].

Table III
THE HOTSPOT CORTICAL REGIONS OF AD DERIVED FROM ABNORMAL SUBGRAPH PATTERNS.

Order	Fiber Density	Fractional Anisotropy
1	IC.L	SFG.R
2	ITGT.R	FP.L
3	LOCS.R	PC.R
4	FP.L	FP.R
5	PG.L	PG.R
6	LOCS.L	IC.L
7	AG.R	CGA.L
8	TP.L	CGA.R
9	STGA.L	IC.R
10	STGP.L	TP.L

100 abnormal patterns for each attribute and counted the appearance of each cortical regions. I.e. we computed the summation of the degree for each cortical region in the top 100 abnormal patterns. Table III shows the top 10 hotspot cortical regions of AD derived from abnormal subgraph patterns. In the attribute of fiber density, these identified regions were predominately located in the temporal, parietal occipital lobes. For the attribute of FA, most hotspot regions were brain hubs that characterized by a high degree of brain connectivity [8].

VI. DISCUSSION

In the current study, diffusion tractography-based human connectomes and subgraph mining were used to study the alternations of regional specific connectivity patterns in AD. According to the experimental results, the detected aberrant subgraph patterns based on the attributes of fiber density and FA including disrupted hotspot brain areas are perfectly in line with previous findings demonstrating AD's impact on these areas by increased amyloid density, brain atrophy or altered white matter properties [1], [3], [7], [5], [6].

Beyond that, the most intriguing point of our result is the distinct distribution of aberrant subgraph patterns of human connectome in AD concerning fractional anisotropy (FA) and fiber density (FD) (Figure 5). FD characterizes the white matter fibers' atrophy or loss (Fig. 5(A1)) and FA represents white matter fibers' integrity (Fig. 5(B1)). White matter fiber loss concerns mainly the parietal, temporal and occipital lobes (Fig. 3, Fig 5(A2)) perfectly matching grey matter atrophy in AD ([5], Fig. 5(A3)). Aberrant fiber integrity concerned fiber patterns span several lobes but consistently including frontal areas. These patterns nicely match grey matter patterns of increased amyloid plaque density in AD ([6], Fig. 5(B3)) with strongest increases in frontal lobes, that amyloid plaques are the hallmark of AD and reflect grey matter integrity [1]. Our findings support and extend previous grey matter findings of AD for white matter fiber-based aspects, and provide first evidence that AD impacts grey and white matter in parallel and regionally selective.

More generally (and from a data analysis point of view), existing analysis of brain connectivity (independently whether it concerns structural, neurochemical, and activity-related interaction) apply univariate element-wise or multivariate global approaches (graph-based analysis, pattern classification, see Section II) to analyze underlying brain connectivity patterns. Since brain disorders such as AD or major depression selectively alter the organization of

fiber pathways in the human brain, the key question is which parts of the whole organization are affected. Our data-driven framework provides a potential way to identify these interesting disrupted regional patterns of connections directly to gain deeper insight of the dysfunctions of AD.

VII. Conclusion

In this paper, we presented a data mining framework to discover abnormal patterns of the human connectome in AD. The goal of the study was not to develop new approaches, but to apply current data mining methods like FSG to gain a better understanding of various forms of dementia. To this end, after construction of the human connectomes for each subject, the weighted human connectomes were mapped into a set of unweighted graphs. Frequent subgraph mining algorithm was used to reveal existing subgraphs in human connectome. The most disrupted subgraphs in AD were identified by frequent subgraph mining and abnormal score. Experiments demonstrated that the identified disrupted spatial patterns of human connectome perfectly matched previous patterns of grey matter changes in AD. Experiments provided the first evidence that AD affects white matter in a way that ensembles spatially grey matter changes. The findings further gained new insights into how AD propagates and disrupts the density and integrity of large-scale human connectome. In the future, we plan to extend our data-driven approach to subjects at increased risk for AD (Mild Cognitive Impairments, APOE4 genotype) and to relate fiber-based white matter reductions with distinct pathological grey matter properties such as amyloid density.

References

[1] K. Blennow, M. J. de Leon, and H. Zetterberg. Alzheimer's disease. *Lancet*, 368:387–403, 2006.

[2] H. Braak and E. Braak. Neuropathological stageing of alzheimer-related changes. *Acta Neuropathol (Berl)*, 82:239–259, 1991.

[3] T. Chua, W. Wen, M. Slavin, and P. Sachdev. Diffusion tensor imaging in mild cognitive impairment and alzheimer's disease: a review. *Curr. Opin. Neurol.*, 21:83–92, 2008.

[4] C. Davatzikos, Y. Fan, X. Wu, D. Shen, and S. M. Resnick. Detection of prodromal alzheimer's disease via pattern classification of magnetic resonance imaging. *Neurobiol. Aging*, 29:514–523, 2008.

[5] S. Derflinger, C. Sorg, C. Gaser, N. Myers, M. Arsic, A. Kurz, C. Zimmer, A. Wohlschläger, and M. Mühlau. Grey-matter atrophy in alzheimer's disease is asymmetric but not lateralized. *J. Alzheimer's Dis.*, 25(2):347–357, 2011.

[6] A. Drzezga, T. Grimmer, G. Henriksen, M. Mühlau, R. Perneczky, I. Miederer, C. Praus, C. Sorg, A. Wohlschläger, M. Riemenschneider, H. J. Wester, H. Foerstl, M. Schwaiger, and A. Kurz. Effect of apoe genotype on amyloid plaque load and gray matter volume in alzheimer disease. *Neurology*, 72(17):1487–1494, 2009.

[7] E. Englund. Neuropathology of white matter changes in alzheimer's disease and vascular dementia. *Dement. Geniatr. Cogn.*, 9:6–12, 1998.

[8] P. Hagmann, L. Cammoun, X. Gigandet, R. Meuli, C. J. Honey, V. J. Wedeen, and O. Sporns. Mapping the structural core of human cerebral cortex. *PLoS Biol.*, 6(e159):1479–1493, 2008.

[9] Y. He and A. E. Z. Chen. Structural insights into aberrant topological patterns of large-scale cortical networks in alzheimer's disease. *J. Neurosci.*, 28:4756–4766, 2008.

[10] S. Klöppel, C. M. Stonnington, C. Chu, B. Draganski, R. Scahill, J. D. Rohrer, N. C. Fox, C. R. J. Jr, J. Ashburner, and R. S. J. Frackowiak. Automatic classification of mr scans in alzheimer's disease. *Brain*, 131:681–689, 2008.

[11] M. Kuramochi and G. Karypis. An efficient algorithm for discovering frequent subgraphs. *IEEE Trans. Knowl. Data Eng.*, 16(9):1038–1051, 2004.

[12] L. Lamport. *Introduction to Diffusion Tensor Imaging*. Elsevier, New York, 2007.

[13] C. Y. Lo, P. N. Wang, K. H. Chou, J. Wang, Y. He, and C. P. Lin. Diffusion tensor tractography reveals abnormal topological organization in structural cortical networks in alzheimer's disease. *J. Neurosci.*, 30:16876–16885, 2012.

[14] M. L. M, D. M. Weinstein, J. S. Tsuruda, K. M. Hasan, K. Arfanakis, M. E. Meyerand, B. Badie, H. A. Rowley, V. H. amd A. Field, and A. L. Alexander. White matter tractography using diffusion tensor deflection. *Human Brain Mapping*, 18:306–321, 2003.

[15] S. E. Rose, F. Chen, J. B. Chalk, F. O. Zelaya, W. E. Strugnell, M. Benson, J. Semple, and D. M. Doddrell. Loss of connectivity in alzheimer's disease: an evaluation of white matter tract integrity with colour coded mr diffusion tensor imaging. *J. Neurol. Neurosur. PS.*, 69:528–530, 2000.

[16] J. Shao, N. Myers, Q. Yang, J. Feng, C. Plant, C. Böhm, H. Förstl, A. Kurz, C. Zimmer, C. Meng, V. Riedl, A. Wohlschlägera, and C. Sorg. Prediction of alzheimer's disease using individual structural connectivity networks. *Neurobiol. Aging*, 33(12):2756-2765, 2012.

[17] O. Sporns, G. Tononi, and R. Kötter. The human connectome: A structural description of the human brain. *PLoS Comput. Biol.*, 1(4):e42, 2005.

[18] C. J. Stam, B. F. Jones, G. Nolte, M. Breakspear, and P. Scheltens. Small-world networks and functional connectivity in alzheimer's disease. *Cereb. Cortex*, 17:92–99, 2007.

[19] D. R. Thal, U. Rueb, M. Orantes, and H. Braak. Phases of ab-deposition in the human brain and its relevance for the development of ad. *Neurology*, 58:1791–1800, 2002.

[20] R. P. Woods, S. T. Grafton, J. D. Watson, N. L. Sicotte, and J. C. Mazziotta. Automated image registration: Ii. intersubject validation of linear and nonlinear models. *J. Comput. Assist. Tomogr.*, 22:153–165, 1998.