PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Jiachen Wang

Entitled

Multigraph Visualization for Feature Classification of Brain Network Data

For the degree of Master of Science

Is approved by the final examining committee:

Shiaofen Fang
Chair

Mihran Tuceryan

Shen Li

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy of Integrity in Research" and the use of copyright material.

Approved by Major Professor(s): Shiaofen Fang

Approved by: Shiaofen Fang 12/5/2016

Head of the Departmental Graduate Program Date

MULTIGRAPH VISUALIZATION FOR FEATURE CLASSIFICATION OF BRAIN NETWORK DATA

A Thesis

Submitted to the Faculty

of

Purdue University

by

Jiachen Wang

In Partial Fulfillment of the

Requirements for the Degree

of

Master of Science

December 2016

Purdue University

Indianapolis, Indiana

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor, Dr. Fang, for his advice, patience, and providing me with an excellent atmosphere for doing research. Without his support and persistent help this dissertation would not have been possible. I wish to thank Dr. Shen and Dr. Tuceryan, for their guidance and suggestions as members of my Thesis Committee. Additionally, I would like to thank my friends from Fang group: Huang Li and Shenghui Jiang.

Finally, I would like to thank my family for their love, care and support throughout my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Wang, Jiachen. M.S., Purdue University, December 2016. Multigraph Visualization for Feature Classification of Brain Network Data. Major Professor: Shiaofen Fang.

A Multigraph is a set of graphs with a common set of nodes but different sets of edges. Multigraph visualization has not received much attention so far. In this thesis, I will introduce an interactive application in brain network data analysis that has a strong need for multigraph visualization. For this application, multigraph was used to represent brain connectome networks of multiple human subjects. A volumetric data set was constructed from the matrix representation of the multigraph. A volume visualization tool was then developed to assist the user to interactively and iteratively detect network features that may contribute to certain neurological conditions. I applied this technique to a brain connectome dataset for feature detection in the classification of Alzheimer's Disease (AD) patients. Preliminary results showed significant improvements when interactively selected features are used.

Keywords: graph visualization, multigraph, volume rendering, brain imaging, feature detection.

CHAPTER 1. INTRODUCTION

Data visualization leverages the power of human perception to process complex information and develop novel insights through interaction. Today, data visualization has been playing an increasingly important role in our life along with the ongoing big data revolution. Big data are of high volume, high velocity and high variety, and thus bring in big challenges in data processing, representation and analysis. Visualization not only allows us to represent data but more importantly to interact with the visual representations to drill into lower level of detail, and highlight the key features across large-scale datasets. This is a great advantage of visualization approaches, which collaborate our eyes and brains to gain insights that cannot be achieved by traditional statistical methods.

With the proliferation of network applications in all aspects of modern society (e.g. WWW, social networks, transportation networks, etc.), graph or network visualization becomes increasingly important. While there have been a large number literature in information visualization dedicated to the theory and practices of graph/network visualization, the visualization of multigraphs has received very little attention. This is a somewhat curious phenomenon as multigraph is a very common type of data sets in many network related data analysis applications.

A multigraph [1] is a set of graphs that have a common set of nodes but different sets of edges. Each edge can stand for one type of relationships which could be a mode of communication, a measure of time or any other relative value. Multigraph provides a way to visually perceive the various relationships of the networks and allows deepening the understanding of the internal details, not just the apparent features. It is essential to create a clear layout of multigraph to represent a global structure. Also, multigraph emphasizes interactive visual analysis which allows adjusting the edges and vertices in appropriate ways for detailed presentation.

Many network problems can be modeled as multigraphs. For example, in social networks, the study of network communities may require the understanding and analysis of the different types of connectivity (e.g. different time periods, or different modes of communications). Another important application is in medical and clinical research. For example, when studying human brain networks, network data from multiple human subjects may be collected for analysis or classification of neurological diseases. This set of brain networks from multiple subjects is a multigraph as humans have similar brain structures. In this thesis, I will show that proper multigraph visualization can help the analysis process by generating more salient visual features.

Graph theory defines a network as a set of nodes and the edges between them. Graph theory was originally emerged as a new branch of mathematics back in the 17th century. Not until the middle of the 19th century, graph theoretic concepts started to pervade many areas of science. The study of social network has driven graph theory to make significant progress and to enter a modern era. In random graph, each pair of

nodes in random graphs has equal probability to connect with each other. However, most real-world graphs possess very different properties as random graphs and this deviation from randomness reflect their specific functionality. Some networks, such as the Internet, have degree distributions following power law, which a node has degree k is given as $Prob(k) \sim k^{-\lambda}$. Some networks from biological systems lack a characteristic scale and thus are marked as scale-free. Scale-free networks have been demonstrated that each added node, as it grows, tends to connects with nodes that already have high degree. Many networks have been reported to have small-world attribute, from a wide range of studies of genetic, signaling, communications and neural networks. This indicates that all nodes of a network are linked through relatively few intermediate steps, even they only maintain very few direct communications.

Graph topology can be described by a wide variety of measures. The most fundamental measure of network is degree, which measures the number of connections that one node link to the rest of the network. The degrees of the overall nodes of a network form a degree distribution. Random network has equal possibility for node connections, resulting in a Gaussian and symmetrically centered degree distribution. Non-random network has non-Gaussian degree distributions with a long tail towards high degrees. If the neighbor nodes are directly connected to each other, they form a cluster. Clustering coefficient is another important measure which quantifies the ratio of the number of connections between nearest neighbors of a node to the maximum number of possible connections. Random network has relatively low average clustering, whereas complex network tend to have high clustering. The minimum number of edges

that can form the path from one node to another is called path length. Both random and complex networks have short mean path lengths, on the other hand, regular lattice has long mean path length. Hubs are nodes with high centrality, which measures how many of the shortest paths between other pair nodes pass through a certain node. Therefore a node with high centrality always play significant roles in efficient communication.

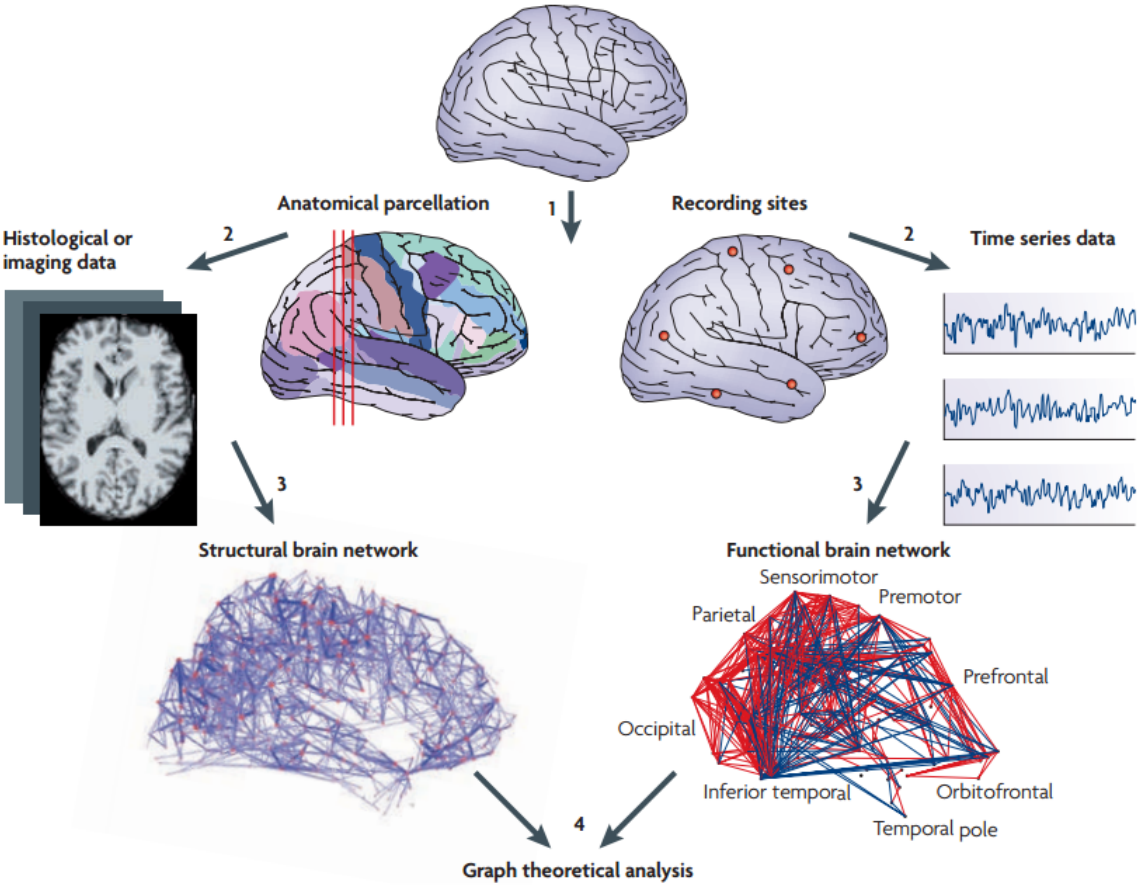

Figure 1.1. Structural and functional brain networks [2].

Human connectomics [2] studies how the human brain is wired and how its function is affected by the connectivity pattern using multi-modal neuroimaging data, as shown in Figure 1.1. The human brain is a complex network of approximately 10^{10}

neurons linked by 10^{14} synaptic connections [3]. It has been widely appreciated that brain network provides physiological basis for information processing and mental representations. Studying this complex network through analyzing its communications among anatomical substrates could provide a novel approach of quantifying the brain's structural and functional systems.

Several attempts have been reported recently to map structural networks of human connectome. He et al first reported large-scale anatomical connection patterns of human cerebral cortex by measuring cortical thickness using Magnetic resonance imaging (MRI) [4]. Graph analysis revealed significant short and long range anatomical connections in brain sub regions. It has also been demonstrated that the probability of finding a connection between two regions with a certain distance followed an exponentially truncated power law model. A following up study further studied cortical morphormetric database and revealed the close relationships between structural network and functional pathways in the cortex [5]. Structure modules were defined as sub regions that are connected morphologically to achieve the maximum network modularity. It has been proved that brain topological modules tightly coupled with known functional domains such as language, strategic, sensorimotor, visual and mnemonic processing. Another approach that has been used to map human brain structural network is diffusion imaging and tractography. A study of 90 cortical and subcortical gray matter regions were conducted to analyze basic statistical aspects of the human brain anatomical network [6]. The weighted network was constructed and factors such as small world attributes, efficiency, degree distribution, vulnerability,

betweenness centrality and motifs composition properties were measured. Several regions, including the precuneus, the insula, the superior parietal cortex and the superior frontal cortex and the superior frontal gyrus were identified as putative bubs of brain network. DTI has limitations detecting crossing fiber Bundles, due to its incompatibility to generate fiber bundle orientation. Diffusion spectrum imaging (DSI) can be applied to improve this difficulty by reconstructing multiple diffusion directions in each voxel.

How do brain structural connections affect or shape functional brain networks? Structural network indicates that neural nodes are connected with each other through specific structural patterns. The functionality of neural node is determined through interacting with other nodes within the same network. Additional functionality linkages resulted by indirect interactions make the network even more powerful and complicated. The local functions of individual node are always resulted by the actions of the whole neural network. Though it is widely recognized that the structural and functional connectivity of brain network is highly complex and dynamic, the mechanism of how neural connectivity is controlled and modified is still not clear. Some evidences indicate that neural network is remodeled constantly, while others suggest that most synaptic spines are stable. How relatively slow structural modifications achieve faster functional changes among neural nodes remains a hot topic among researchers.

Computational approaches offer a complementary approach to discover structure and function relationships of the brain. Given the unprecedented complexity of brain network, we are facing critical computational challenges for comprehensive

mapping and analysis of brain connectivity, across all scales. Research in this area has largely focused on extracting brain networks from structural, functional and diffusion magnetic resonance imaging (MRI) data [7, 8]. The visualization and related visual analytics of this network has not been well studied.

In this thesis, I focus on a specific visual connectomic analysis application: feature classification for brain diseases. Visual feature classification applies feature visualization techniques to provide discriminating power for data classification. By visualizing and comparing the multigraphs of subjects in different classes, I am able to detect and extract network features that are more salient in differentiating data groups for diagnosis and analysis. In the rest of the thesis, related work will first be discussed in Section 2. The application dataset will be described in Section 3. In Section 4, I will describe the multigraph visualization approach and its visual interface for feature detection. The result of this technique applied to a testing data set for Alzheimer's Disease patients will be presented in Section 5. Conclusions and future work will be given in Section 6.

CHAPTER 2. RELATED WORK

2.1. Anatomical and functional brain connectome

Human connectome research is greatly advanced by the innovations in data visualization, from tensor glyphs for diffusion-weighted imaging, to advanced rendering of anatomical tracts, to more recent graph-based representations of functional connectivity data [9]. With brain network's innumerable data dimensions, how to present the mapping of connections in intuitive, informative and candid ways becomes even more challenging. Give the high information content of brain connectome, how to accurately render the content of available data while prioritize the uncertainty of the data is another formidable challenge. Furthermore, to transfer large-scale, complex network data into intuitive, easy to understand visualization requires balance of complexity and simplicity. How to achieve visual simplicity while preserve genuine information is another question we need to think about. It is essential to address the connectome visualization on its progress, available tools, challenges and novel techniques.

Research into anatomical connectivity of the brain has been greatly advance by innovative visualization. Diffusion tensor imaging (DTI) enables displaying more dimensions of data. It models diffusion as a tensor of rank two with six degrees of

freedom at each voxel. Basser et al reported to use glyphs, in the shape of ellipsoids, to define localized visual unit of information [10]. Considering more complex configurations, such as crossing fibers, tensors of rank two are not sufficient to describe the voxel directionality. Researchers developed novel approaches, such as high angular resolution diffusion imaging (HARDI) and diffusion spectrum imaging (DSI), for three dimensional probability distributions [11]. Visualization of HARDI and DSI data can be achieved by constructing the image using spherical polar plots or two tensors rendered cuboids as basic unit. In glyph-based methods, each discrete glyph represents one diffusion tensor. Tractography, on the other hand, represents continuous curves. Various rendering approaches have been reported to produce high quality tractograms, such as tuboids, hair-like structures, triangle strips and point sprites, streamtubes and surfaces, and stylized line primitives. Tractography risks losing information content due to its absolute certainty for fiber rendering, when it comes to representing crossing fibers, splits and termination points. The obvious disadvantage of streamline fiber tracking methods have motivated the development of probabilistic approaches. Probabilistic tracking approaches generate the tract based on the probability of the tract passing through each voxel for numerous path iterations. One classic way to represent probability values is using symbolic color scale. Volume rendering by applying opacity of different levels to generate semi-transparent cloud is also widely used [12].

Besides anatomical connectivity, it is also very important to understand functional connectivity of brain network. One of the classic methods to represent functional connectome is two-dimensional matrix, for which each point describes the

strength of the connection between two regions of interests (ROIs). ROI allows brain connectome to be visually digestible and overlooks the local movements of the neurons within individual ROI. This is a balance of information loss and achievable visualization. Many approaches have been reported for constructing functional brain network. Salvador et al represented brain functional connectivity using undirected graphs. These are the first maps to describe frequency dependence of entire human brain functional networks. In their paper, a two dimensional coordinate system was applied to show each node's anatomical position and the connectional relationships among the nodes are described as lines [13]. It has been suggested that bilaterally homologous regions of the brain tend to be strongly and symmetrically connected. Also, low-frequency components usually are involve in stronger functional connections, in comparing to high-frequency components. Moreover, functional brain network represents small world properties that local connectivity is generally stronger than long-distance connectivity [14].

There has been an increasing interest in understanding the relationship between structural and functional brain networks. Many studies have revealed an overlap between structural and functional connections. Van den Heuvel et al examined both the functional and structural connections of the human brain in a group of 26 healthy subjects using resting-stage magnetic resonance imaging (MRI) and diffusion tensor imaging scans (DTI). Among the nine resting-state functional networks they looked at, eight of them were supported by anatomical white matter tracts [15]. Skudlarski et al applied DTI and state temporal correlations (RSTC) to create global connectivity

matrices covering the whole brain grain matter. This allowed for direct comparisons between functional connectivity measured by RTSC and structural connectivity measured by DTI. The connectivity maps generated by both approaches strongly agreed with each other [16]. With the strong correlations in between structural and functional networks, several properties of structural network have been proved to play essential role in understanding brain functions. One important property is Euclidian distance between two brain regions. To better shaping functional connections, the degree of two nodes in the structural network is another key property. Moreover, shortest paths are also considered to be strong predictors for functional connections as those paths are favorable because of metabolic efficiency and fast communication.

2.2 Graph visualization

As a subfield of information visualization, graph visualization has been extensively studied. Graph visualization has been widely applied to many areas. Web site maps are good examples of graph application. In biology, protein functions, genetic maps, evolutionary trees all can be presented as graphs. Other examples include data structures, data flow diagrams, entity relationship diagrams, logic programming such as SLD trees etc. The key challenge for graph visualization lies in the balance between the size of the displayed graph and the usability of the graph. Generally, displaying every detail of large graph may makes it difficult to view and comprehend [17]. Another challenge is human factor in the visualization. Very few tools for graph visualization

have practically adapted cognitive factors into the applications. This is no doubt is a hot subject for future research.

Traditional graph layouts usually assume that the layout is determined only by the nodes and the edges without additional constraints. Classical tree layout with children nodes deriving from their parent node reflects the intrinsic hierarchy of the data. Reingold and Tilford algorithm is one of the most widely used tree layout algorithm. Radial layout is another classical tree layout algorithm. This approach places nodes on concentric circles according to their depths. Balloon layout, which is mapped according to cone tree algorithm, attaches sibling subtrees to parent node in a circle. To improve the "space problem" associated with traditional tree layout approaches, 3D layout becomes popular as it allows an extra dimension and more space to display large data. 3D version also allows users to interactively navigate the structure and visualize it from multiple angles. Moreover, additional visualization skills, such as opacity, volume, depth queuing, can be applied to achieve better visualization results. Most force-directed approaches are dimension independent, thus allows 2D displays to be generalized to 3D directly. One of the well-known 3D information structures is the cone tree, which is generated by placing children nodes evenly spaced along the base of their parent node at the apex of the cone shape. The distributions of the nodes are based on their hierarchy orders. The body of each cone is transparent, thus allowing the cones behind it to be easily visualized. Hyperbolic layout is another graph visualization technique, which provides a distorted view and generates better visualization results with potentially large trees [18].

As we discussed earlier, visualization is facing the problem raised by large sizes of the graphs. Therefore, navigation and interaction are essential steps for information visualization. Zoom to scale is one of the traditional approaches for displaying detailed information of large graph structures. Geometric zooming simply blow up the graph content. A more applicable approach is semantic zooming, which changes the graph content and allows more details to be revealed. Zooming requires assigning appropriate subgraph levels. When zooming in on certain details, losing contextual information becomes a big problem. Focus+context techniques are applied to improve on such situations. One well-known example is fisheye distortion. This approach achieves the fish-eye lens effect by zooming in the interest areas while still be able to show other portions, though with less details. Fisheye technique can be implemented as a separate step after the layout algorithm module and before the final display. The visual graph generated by the layout algorithm may be affected negatively by the following fisheye step. For example, it may add unwanted edge crossings. Alternatively, focus+context methods can be included as internal modules of the layout algorithm. The distorted view therefore is generated by interacting with the parameters governing the layout algorithm. 3D techniques are also widely used in achieving focus+context distortion. A simple example is mapping the view created by 2D techniques, such as perspective or parallel projections, directly onto 3D surfaces. We have only discussed some simple versions here, much remains to be done for future research in this topic. For example, how to apply multifocal focus+context methods to check different important areas of the graph simultaneously. There are cases that the size of the graph is so huge that it is

impossible to display the whole graph, incremental exploration approaches are usually applied for such situations. Only subgraph is displayed by the system and other portions are displayed as needed. The keys steps for incremental exploration include generating visual window to show the subgraph and reconstructing the content of the window after change.

We've also discussed briefly earlier that in order to achieve a better visualization in both performance and clarity, it is usually essential to reduce the amount of basic elements being viewed. One classic technique is clustering, which is the process of grouping data based on chosen semantics. Using structural information for clustering is referred as structure-based clustering. On the other hand, using semantic data to cluster the elements is referred as content-based clustering. Content-based clustering is not as widely applied as structure-based clustering for graph visualization, as it sometimes is more application-specific and is not general enough to be used in other application areas.

Clustering the data allows us to reduce the visual complexity while preserve the detailed information to a necessary level. The key issue is to look for a balance between the number of clusters and the visible information content. In order to cluster the nodes, node metric is a measure to rank nodes according to the features associated with each one of them. The nodes can be assigned to different groups based on their metric value. One widely applied node feature is degree, which measures the number of edges connected to a node. Since node degree only carries structure information of the graph, it is described as structural-based. Other than node clustering, there are other types of

deviations from the classic network structure, such as multivariate attributes [19]. Multivariate graph visualization shows the graph structure together with node attributes. The attributes can be part of the original data or be computed from the nodes or edges information. Multiple attributes can be incorporated into the visual representation through spatial compositions or temporal compositions. The most commonly seen approach for spatial composition is representation with graph structure as base, which embedded attributes within the nodes. Balanced representation is a simple version of spatial composition, which shows a layout of the graph structure and a multivariate attributes visualization. Attributes can also server as base representation for spatial composition. Each node can be positioned according to its attributes. Temporal compositions can be generated mainly by two different approaches. One is using graph structure as base representation. For example, the visual display can be adjusted based on node attributes to only show the nodes of the hierarchy whose attribute values are within the asked range. Attributes can also be applied as base representation for temporal composition. One example is to apply node attributes as parallel coordinates for visualization.

2.3 Brain connectome visualization

Brain connectome visualization has largely relied on graph theory, which concerns more about describing nodes and edges, rather than the spatial paths connecting nodes. One category of brain connectome visualization techniques tries to maintain the anatomical location and the connections are articulated using lines. Achard

et al created anatomical connectome by using 2D coordinate systems to reflect each node position from a single view [20]. The distance between each pair of regions approximates the path length between them. Anatomical connectome can also be described in 3D, through rendering a brain volume and applying unique curved lines across the space to represent the connections [21]. However, visualization result could be negatively affected as the number of connections increases. This problem can be eased by using functional space that represents functional connectivity distance, rather than anatomical distance. Brain connectome is also widely represented by emphasizing connectivity itself. For example, Zuo et al described brain network highlighting the overall layout. The voxel-wise connections were visualized with the same colors and the overall anatomical and functional details were removed [20]. Due to the complexity of brain network, creating an efficient brain network visualization is usually at the expense of representing the probability of connections through thresholding. Also, as brain connections change dynamically, it is always a challenge as how to integrate temporal domain into the graph along with the spatial. Currently approaches include using color coded ROIs with representative time series, presenting brain shifting along time series to reveal the changes, etc.

2.4 Multigraph

As a subfield of information visualization, graph visualization has been extensively studied. However, few techniques have been developed for multigraphs. Although multivariate graphs [22, 23] can contain multiple edges for pairs of nodes in

the same graph, they do not provide properties that defined by the structures of a complete graph. For example, a node-link visualization method for multigraph and related interactions is proposed in [23]. But it is really focusing on only multi-edge graphs, i.e. the same graph with multiple edges. They proposed a multigraph ranking model that is able to identify nearest neighbor users in multirelational social networks. The whole multigraph is constructed by overlapping all given relationships. The size and number of the node represents how popular the individual is among all relationships. Color was applied to the edges to distinguish different relationships, as shown in Figure 2, upper graph. A somewhat similar problem is the visualization of dynamic graphs [24]. But since the graph structures often change with time, the focus of the visualization algorithms is different. In the case of static graph structures in dynamic graphs, the problem becomes similar to a multigraph problem. In [25], a matrix cube technique was proposed to stack multiple graphs together to form a cube for the visualization of the time-varying changes of the graph edges using information visualization tools such as slicing, small multiples, and color coded projection, as shown in Figure 2.1, bottom graph. Although the graph stacking strategy is similar to our approach, the visualization method for the stacked graphs are very different as we further blur the cube into a volume data set and apply interactive volume rendering for feature detection. Another type of related work is feature extraction by visualization. While feature visualization has been an active topic in scientific visualization, using interactive visualization for feature selection to support data analysis has not been widely studied. Some preliminary work has shown promises in many science and engineering applications

such as flow dynamics [26], spatiotemporal GIS [27], bioinformatics [28] and neuroimaging [29].

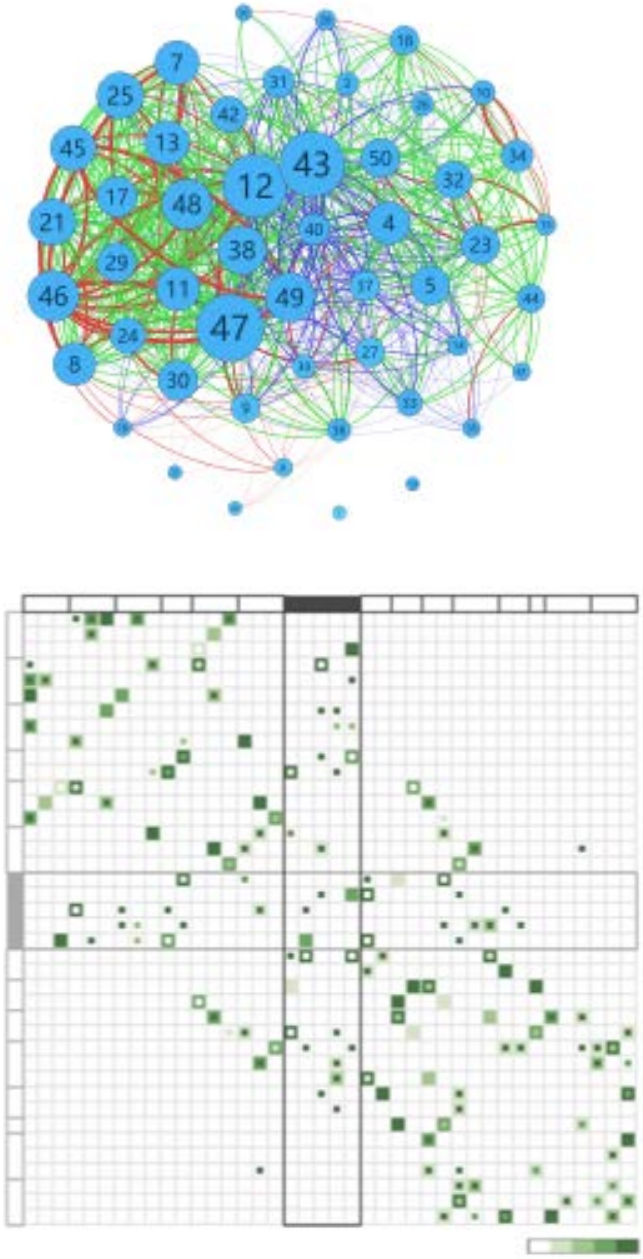

Figure 2.1. Existing multigraph visualization techniques. Upper: Multi-edged graph. Bottom: Stacked adjacency matrices.

CHAPTER 3. BRAIN NETWORK DATASETS

The application I designed and developed is for analyzing human brain network for Alzheimer's Disease classification. The dataset I applied is an MRI and diffusion tensor imaging (DTI) dataset from the Alzheimers Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). An overview of the brain connectome network construction process based on MRI and DTI data is shown in Figure 3. More details of this process, including the parcellation and tractography algorithms, can be found in [30]. The pipeline is divided into three steps: (1) Generation of regions of interest (ROIs), (2) DTI tractography, and (3) connectivity network construction, as shown in Figure 3.1.

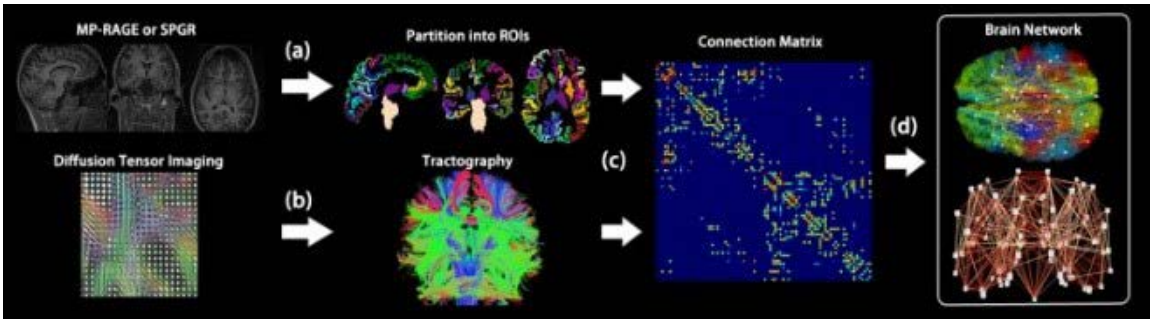

Figure 3.1. Creation of structural connectivity networks.

ROI Generation: Anatomical parcellation is performed on the high-resolution anatomical MRI scan of each subject to obtain 68 gyral-based ROIs, with 34 cortical ROIs

in each hemisphere. These ROIs can be further subdivided so that brain networks at different scales can be constructed.

DTI Tractography: The DTI data are analyzed and processed for fiber tracking using FACT (fiber assignment by continuous tracking). A spline filtering is applied to smooth the tracks.

Network Construction: Nodes and edges are defined in constructing the weighted, undirected network. The weight of the edge is defined as the density of the fibers connecting the pair.

In this study, the brain connectome data were collected for 104 subjects in 3 categories: HC (Healthy Control, 43 subjects); MCI (Mild Cognitive Impaired, 42 subjects) and AD (Alzheimer's Disease, 19 subjects). For each subject, anatomical parcellation is performed on the high-resolution T1-weighted anatomical MRI scan. 234 gyral-based ROIs were obtained for each subject. These ROIs can be further divided into smaller ROIs so that brain networks at different scale can be constructed. Nodes and edges are defined to construct the weighted, undirected network. The weight between each pair of nodes is defined as fiber density, which is number of fiber tracks divided by mean volume of the two ROIs. Connection of two ROIs is defined as the end points of the fiber falling in both the two ROIs. The brain connectome network data from the 104 subjects are 104 graphs. Since all the subjects share the same parcellation, they have the same set of node labels (the brain ROIs) but different connectivity (edges) between the nodes as the fiber densities are different for different subjects between the same pairs of ROI labels. Thus, it is a typical multigraph problem.

CHAPTER 4. METHODS

4.1. Brain parcellation

Brain parcellation divides the brain into a set of non-overlapping regions according to information based on anatomical connectivity, functional connectivity, stimuli based activation, etc [31]. The parcels are defined as clusters to connect anatomical voxels of brain, as shown in Figure 4.1, Main steps of parcellation are as follows. The first step is to define global module and resolution for parcellation. One common strategy is the use of anatomical or functional regions of interest (ROIs). The choice of the regions can be based on the experience of previous experiments. Second step is to modify the solution to yield homogeneous parcels, which should have similar volumes and be similarly compact. K-means algorithm can efficiently achieve this goal by minimizing the intra-class variance in the classification context. The final step is to apply more constraints to improve the parcellation. For example, penalizing the crossing of sulci can be applied as an anatomical constraint.

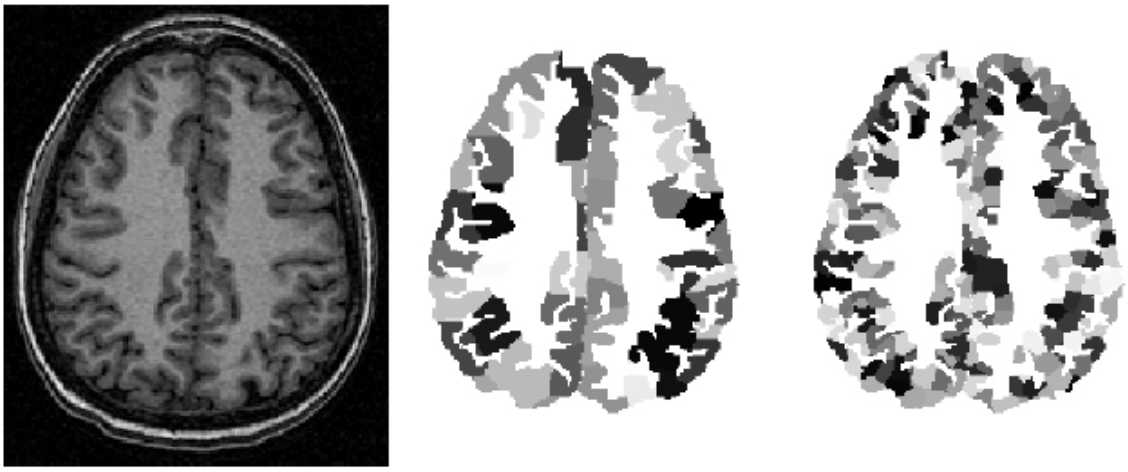

Figure 4.1. Examples of 3D parcellation of brain cortex. Original image (left). Result with 50 parcels per hemisphere (middle). Result with 500 parcels per hemisphere (right).

4.2. Volume rendering

"Volume visualization is a method of extracting meaningful information from volumetric datasets through the use of interactive graphics and imaging, and is concerned with the representation, manipulation, and rendering of volumetric datasets" [by Kaufman, A.; Volume Visualization; IEEE Computer Society Press Tutorial; 1990.] Volume rendering can apply properties such as color and opacity to describe visualization to gain insights [33]. Color and opacity are parameters of light emission and absorption individually. Each voxel of the data can emit or absorb light. The emission model can work by allowing the voxels to only emit their own light or by adding multiple scattering. The absorption modifies the amount of light passes the voxels and can be referred as opacity or transparency of voxels. The properties, such as color and opacity, are assigned to the volumetric data and enable users to visualize not only the surface

but also the inside structure of the data. This step is referred as classification. Sometimes, volumetric data also need to be segmented. For example, when scanning human tissues, different parts may show the same density value and cannot be separated. To solve this problem, segmentation should be applied first for differentiation before classification approached be assigned. Shading technique is usually used to enhance the appearance of the rendering objects. Shading can be defined as global, direct and local methods. Global shading consider the light being exchanged between all objects. Direct shading on the other hand only takes the light directly falls onto the objects. Local shading is computational cheaper comparing to global and direct approaches. It contains ambient, diffuse and specular components. To create the final object, all samples need to be visualized together through a combination of the properties from all the samples for each pixel. Usually the involved data are semi-transparent, therefore the composition step need to be completed in certain order: front to back or back to front. Front to back method allows the process to be terminated when reaching a certain threshold. Back to front method does not allow early termination, however, it requires less computation since there is no need to keep tracking the remaining transparency.

Many algorithms can be taken in volume visualization and those algorithms generally fall into two different categories, as shown in Figure 4.2. One is indirect volume rendering which transfer the data into polygonal iso-surfaces and then perform rendering. The data can also be rendered directly and this is referred as direct volume rendering. Of all available algorithms for volume rendering, splatting is one that has

been mostly widely used. This technique splats each elements as a snow ball according to certain functions such as Gaussian kernels with amplitude scaled by the voxel values. The functions are then projected on to the screen. Splatting can only take the relevant voxels that are stored for later processing. This can greatly improve the computation efficiency for projection and rasterization steps. The shear warp approach is another powerful approach for volume rendering. It was developed by Cameron and Undrill, popularized by Philippe Lacroute and Marc Levoy. The approach takes advantage of a volume and image encoding scheme. The shear warp factorization can achieve low computational overhead because it allows simultaneous traversal of volume and image that skips opaque image regions and transparent voxels [34]. Shear warp algorithm works relatively fast, but at the loss of image accuracy and quality. Another popular rendering technique is 3D texture mapping. Volume rendering applications adapt 3D texture concepts with respect to arbitrary clipping geometries. The idea is to present the 3D text mapping as the trilinear interpolation of the data at an arbitrary point within the domain. The result volume has relatively better quality but may show a transition when rotating the object.

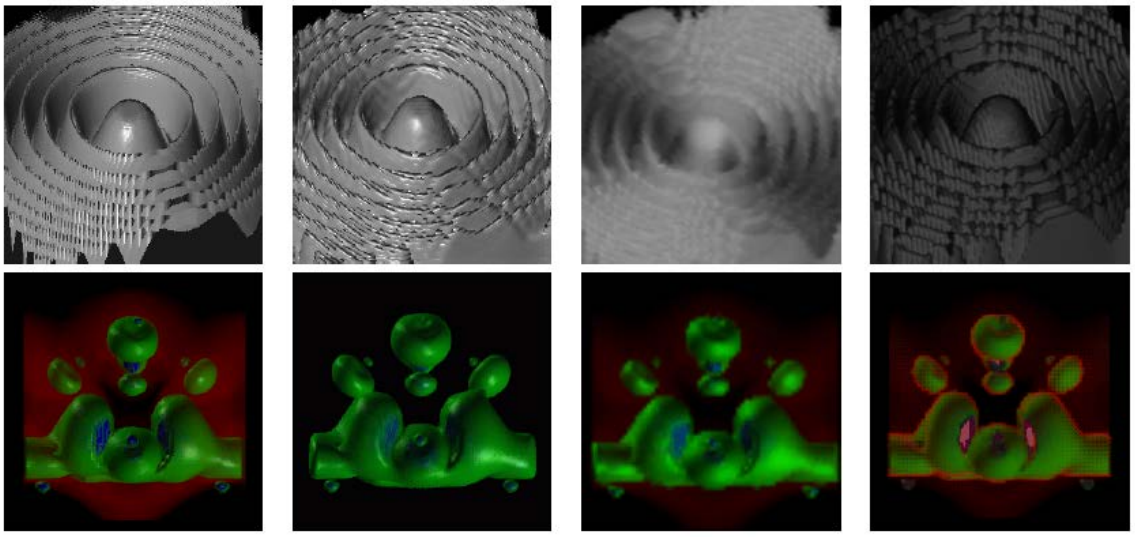

Figure 4.2. Comparison of ray casting, splatting, shear-warp, and 3D texture mapping algorithms. Left to right: Marschner-Lobb function containing high frequencies and simulation of the potential distribution of electrons around atoms (top to bottom) [33].

4.3 Principal component analysis

Principal component analysis (PCA) is one of the most powerful application of linear algebra and it is widely used in all forms of analysis in different areas [34]. PCA provides an efficient way for how to reduce complex datasets to a lower dimension and to extract relevant information that were hidden in the noisy data. The components are uncorrelated with each other and they each contributes to variance of the data. The first principal component has the closest fit in explaining the relationship of the dataset. Later components are serially generated by using the residuals from previous principal component analysis. The whole set of components together account for the total variance of the variables. The first few variables hold most of the variance, therefore

most of the data exist in the first few dimensions and later components are usually unimportant. PCA can be summarized into three steps:

1. Organize the dataset into n x p matrix. n is the number of measurement types and p is the number of trials.

$$\begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1p} \\ y_{21} & y_{22} & \cdots & y_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ y_{n1} & y_{n2} & \cdots & y_{np} \end{pmatrix}$$

2. Subtract mean off each measurement type.

Calculate mean:

$$\overline{Y}_i = \frac{1}{n}\sum_{k=1}^{n} Y_{ki}$$

Calculate standard deviation:

$$S_i = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}\left(Y_{ki} - \overline{Y}_i\right)^2}$$

Normalization:

$$X_{ij} = \frac{Y_{ij} - \bar{Y}_j}{S_j}$$

Normalized matrix:

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}$$

3. Calculate eigenvectors and eigenvalues from the covariance matrix.

$$XX' = \begin{pmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{pmatrix}$$

Where:

$$r_{ij} = \frac{1}{n-1} \sum_{k=1}^{n} X_{ki} X_{kj}$$

Calculate eigenvalue λ with A = XX’ and I is unit matrix:

$$|\mathbf{A} - \lambda\mathbf{I}| = 0$$

Calculate eigenvector X:

$$(\mathbf{A} - \lambda_i\mathbf{I})\mathbf{x} = 0$$

Eigenvalues can be used to make decisions about which eigenvectors should we drop without losing too much information. The eigenvectors with smaller eigenvalues carry less information about data distribution and are of less importance.

CHAPTER 5. RESULTS

5.1 Multigraph Visualization and Interaction

A key step in the data analytics process is feature selection, which is usually a highly technical process hidden from the end user. There is, however, a need for transparency in the feature selection and model creation process, not only for feature interpretability, but also because end user input is highly valuable. Human intuition, knowledge, and perceptual ability to identify patterns are the primary driving force of visual analytics. To this end, interactive data visualization plays a central role.

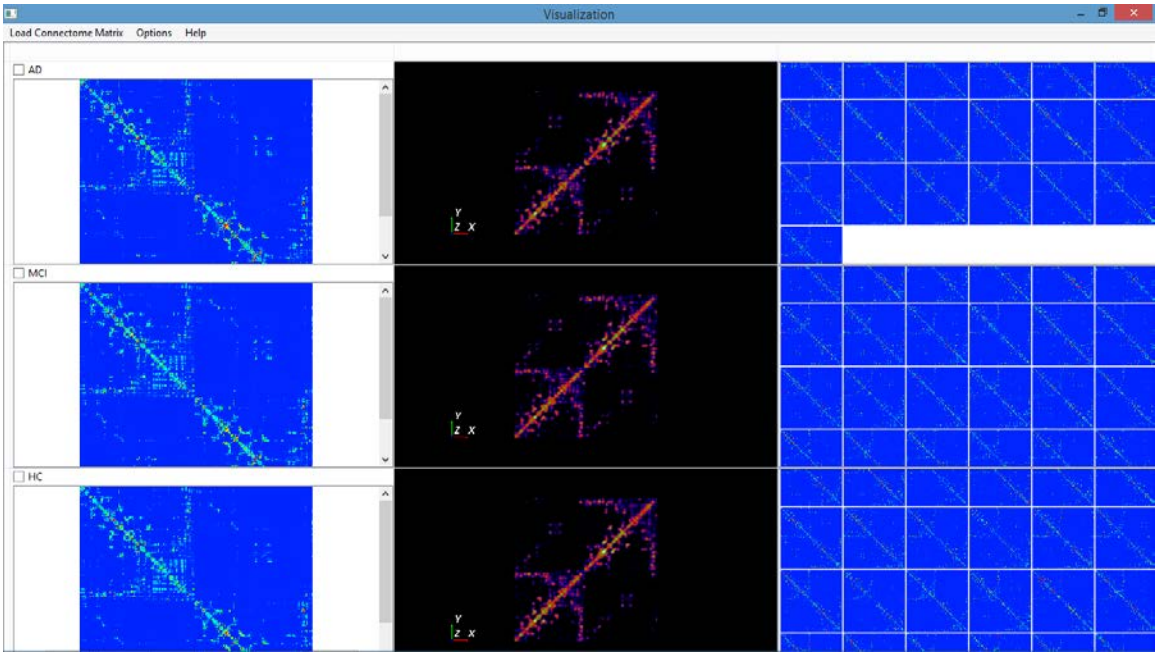

Figure 5.1. User interface tool for brain network visualization.

The goal of the visualization here is to show salient network features that are the most effective in differentiating subjects in these three categories: HC, MCI and AD. Each connectome network can be represented as an adjacency matrix. If we stack together the matrices of a group of subjects, it forms a volume dataset. Naturally, the subjects in the three categories can be grouped together to form three volume datasets. I then generate volume renderings of these volumes side-by-side to detect salient features which are both common within groups and different across groups. Figure 5.1 shows the whole look of the user interface tool that we generated. At the left column of the interface is the 2D maps of the averages brain network for different group, including AD, MCI, HC from top to bottom, as shown in Figure 5.2 There is not significant

difference among the three groups, by comparing the 2D average maps. This suggests 2D average graph is probably not the best way for detecting salient network features in brain disease data.

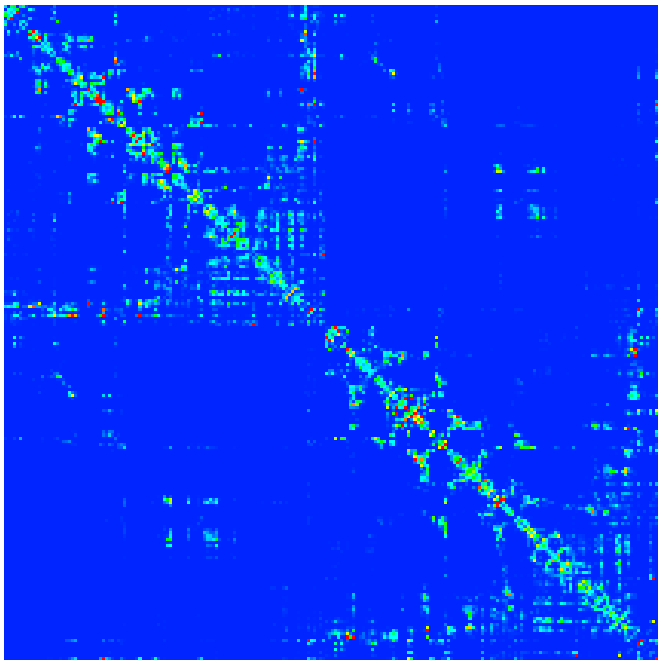

Figure 5.2. Average brain network visualized in 2D. AD, MCI, HC from top to bottom, cont'd.

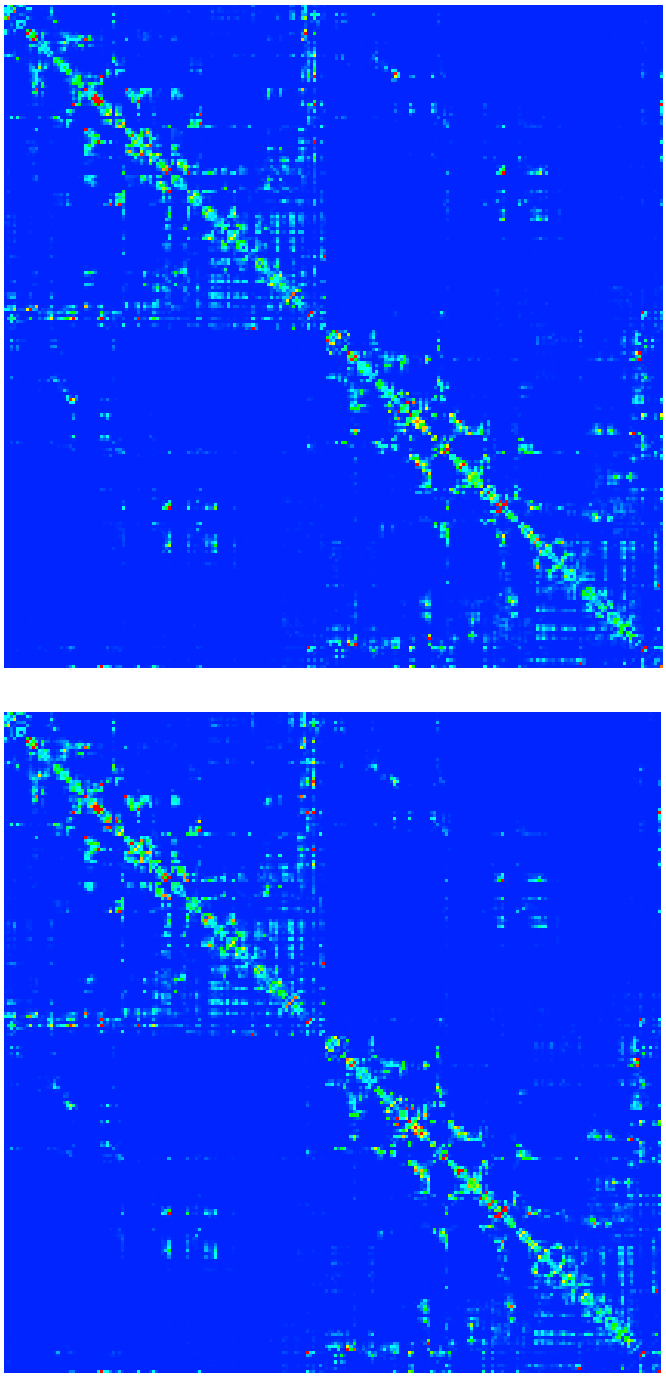

Figure 5.2. Average brain network visualized in 2D. AD, MCI, HC from top to bottom.

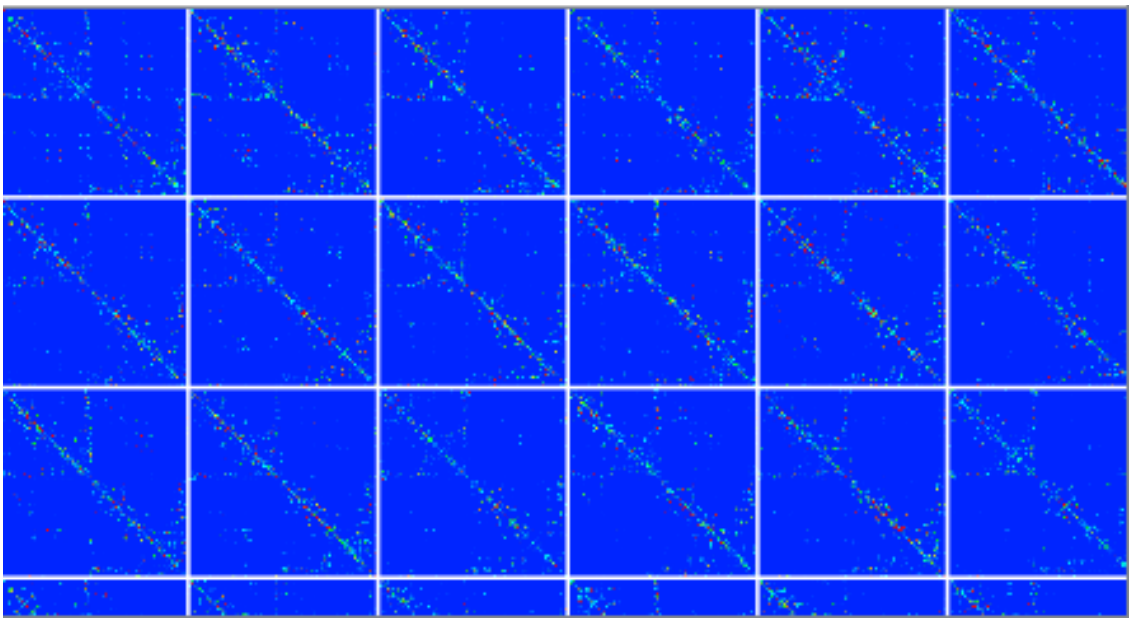

Figure 5.3. Individual brain network visualized in 2D.

Figure 5.3 presents all the individual graph matrices, which are located at the right side column of the user interface. Color is used as the key parameter to show the weight of the edges, which stand for the connection between the ROIs.

To visualize a multigraph volume, the first step is to blur the sparse matrices to generate a cloud-like volume data. This is to enhance the influence of the discrete edges in the adjacency matrices so that volume rendering can be visually more effective. I applied Gaussian filters to “splat” each point to nearby voxels. The data is distributed using an elliptical Gaussian distribution function. Figure 5.4 shows an example of a graph matrix before and after the Gaussian filtering. This approach has greatly improved the volume visualization result. The original sparse graph has been transferred to a connected graph. This technique allows users to better visualize the patterns and

connections within the brain network. VTK is then used to carry out the volume rendering.

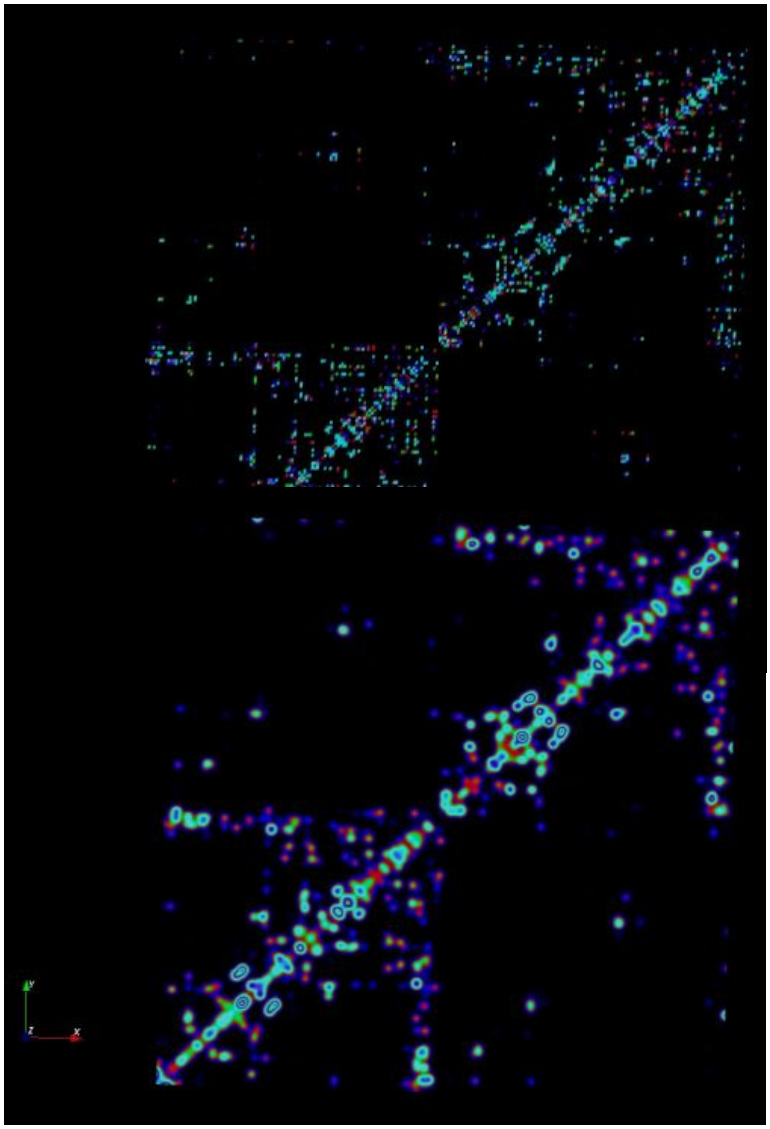

Figure 5.4. A graph matrix before (upper) & after (lower) Gaussian filtering.

The middle column of the user interface shows all volume rendering results of the three multigraph volumes. The transformations of the three volume renderings are synchronized to facilitate interactive visual comparisons. 3D interactions can be applied to the volume rendered images to compare these 3 groups of subjects. Navigation allows users to examine the graphs in different angle, as show in Figure 5.5. Front view provides a better visualization of the patterns and connections of the networks. This angle works better for feature selection. It is also very important to check the side view, which allows us to identify whether the selected feature is consistent among all subjects within the same group. Figure 5.6 shows the zoom in and zoom out functions. These are essential functions that provides a closer look of the selected features. Another powerful function of this interactive tool is the color function, as shown in Figure 5.7 Users can interactively draw different color to different range of data or only draw color to the range of the data that we are interested in. Opacity function offers the user another option for visualizing the network at different iso-surface, as shown in Figure 5.8. High transparency is a better choice when we want to examine all the subjects at the same time and to check if they show the same feature. Low transparency works better to represent the structure of individual iso-surface. These interactive functions are designed to provide the users flexibility when using this tool and to emphasize users' experience on visualization and feature selection.

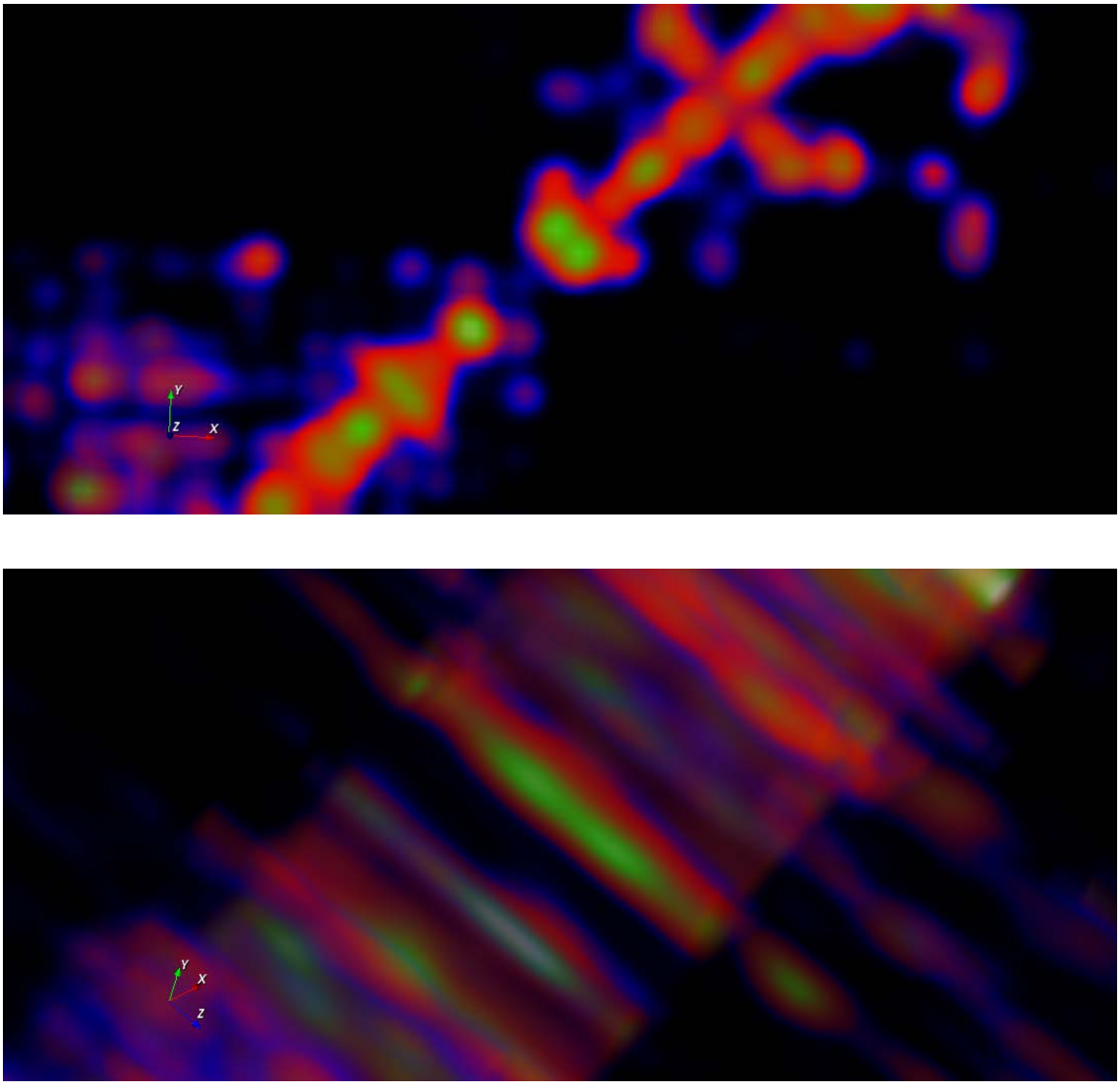

Figure 5.5. Interactive visualization – navigation.

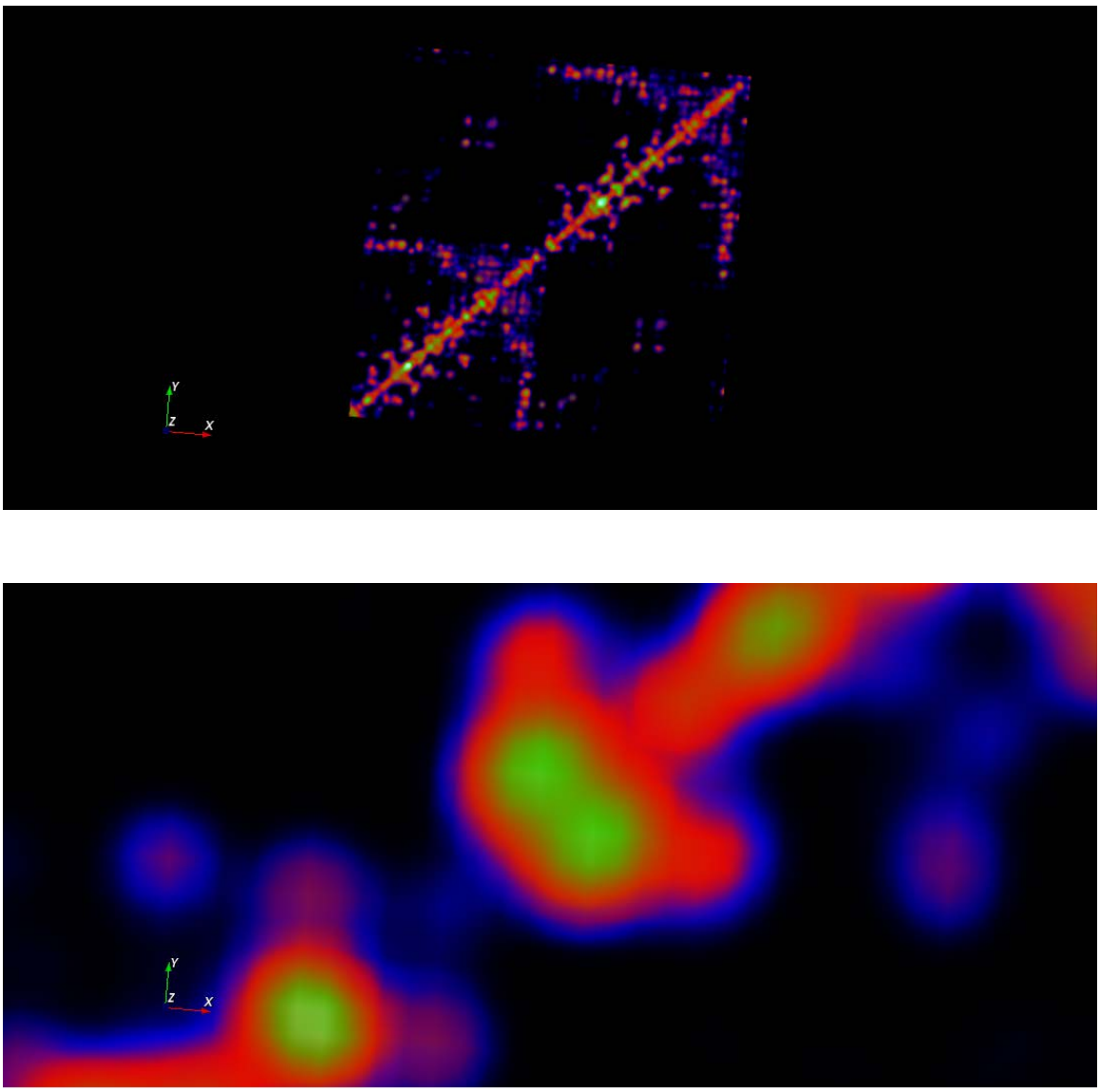

Figure 5.6. Interactive visualization – zoom in and zoom out.

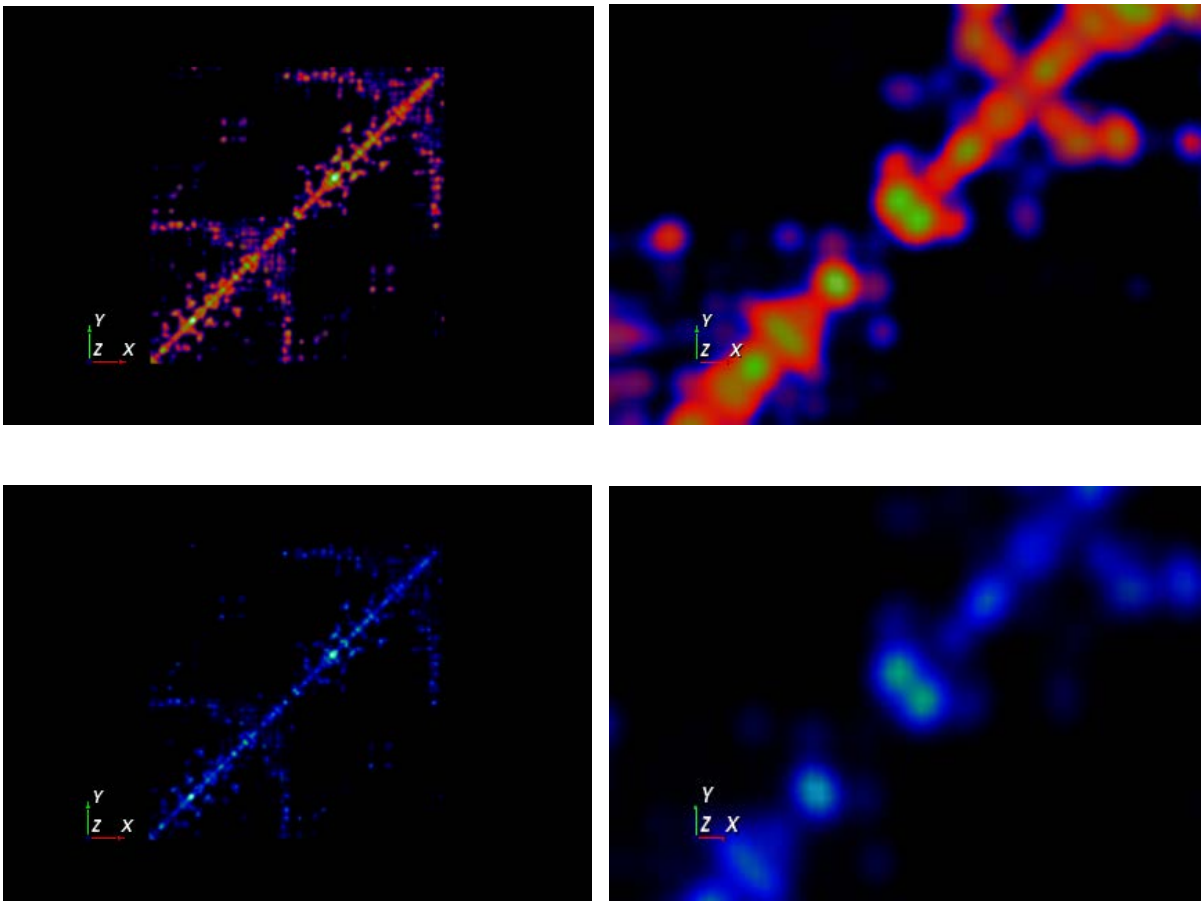

Figure 5.7. Interactive visualization – color function.

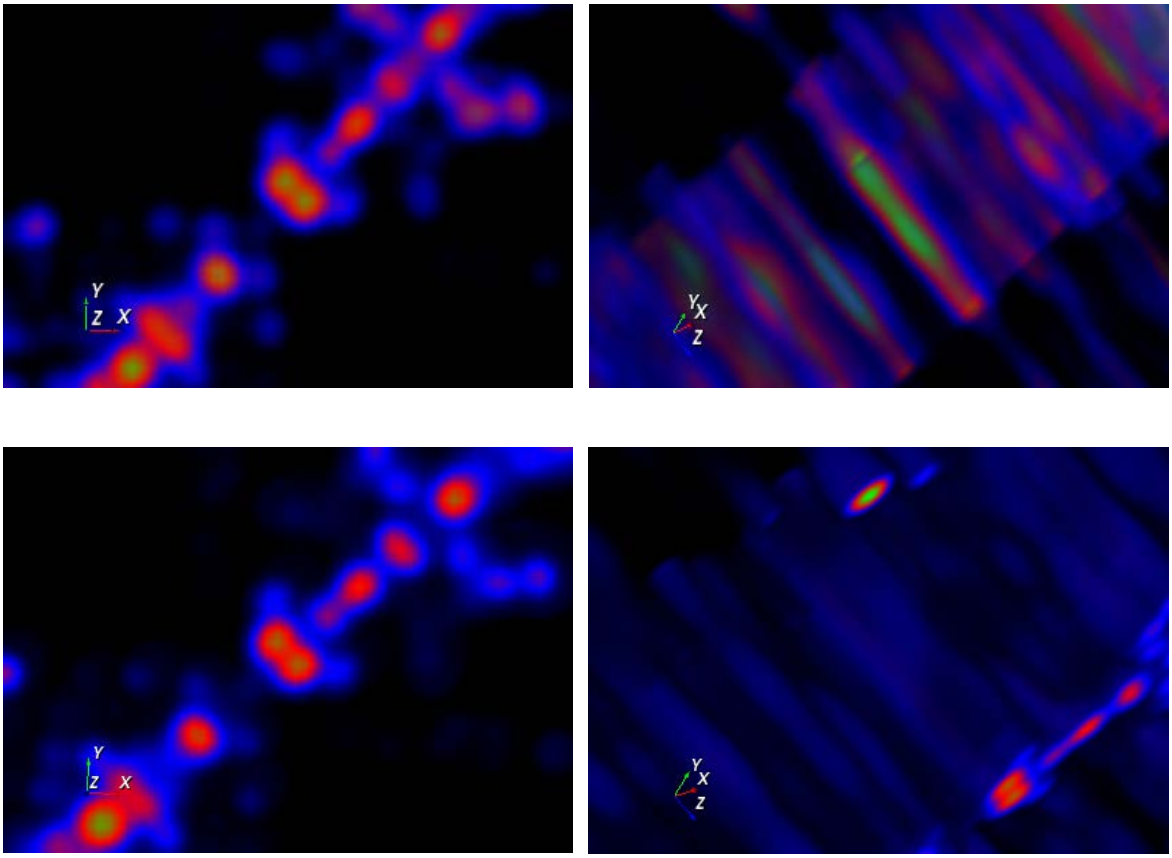

Figure 5.8. Interactive visualization – opacity function.

Users can then interactively identify regions that show the most differences between the three categories, as well as consistencies within their individual groups. As shown in Figure 5.9, such regions will be selected on the interface as submatrices. Since this is an interactive process, it is primarily the users' subjective decisions to identify places where they think they see significant differences. This process may also include change of transfer function for better visual clarities. Results from Figure 5.9 indicate that differences between the three types of networks cannot be easily discovered from statistical averages, as they do not seem to show on the averaged images. The selected submatrices will then be further calculated to form the feature vector for classification analysis.

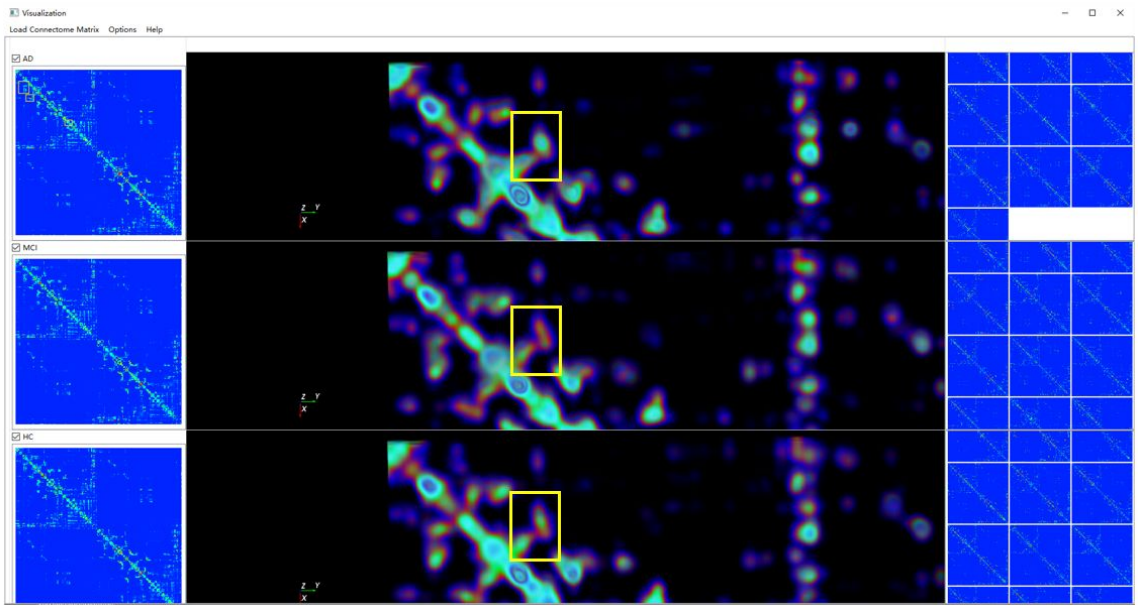

Figure 5.9. Features selected as submatrices, cont'd.

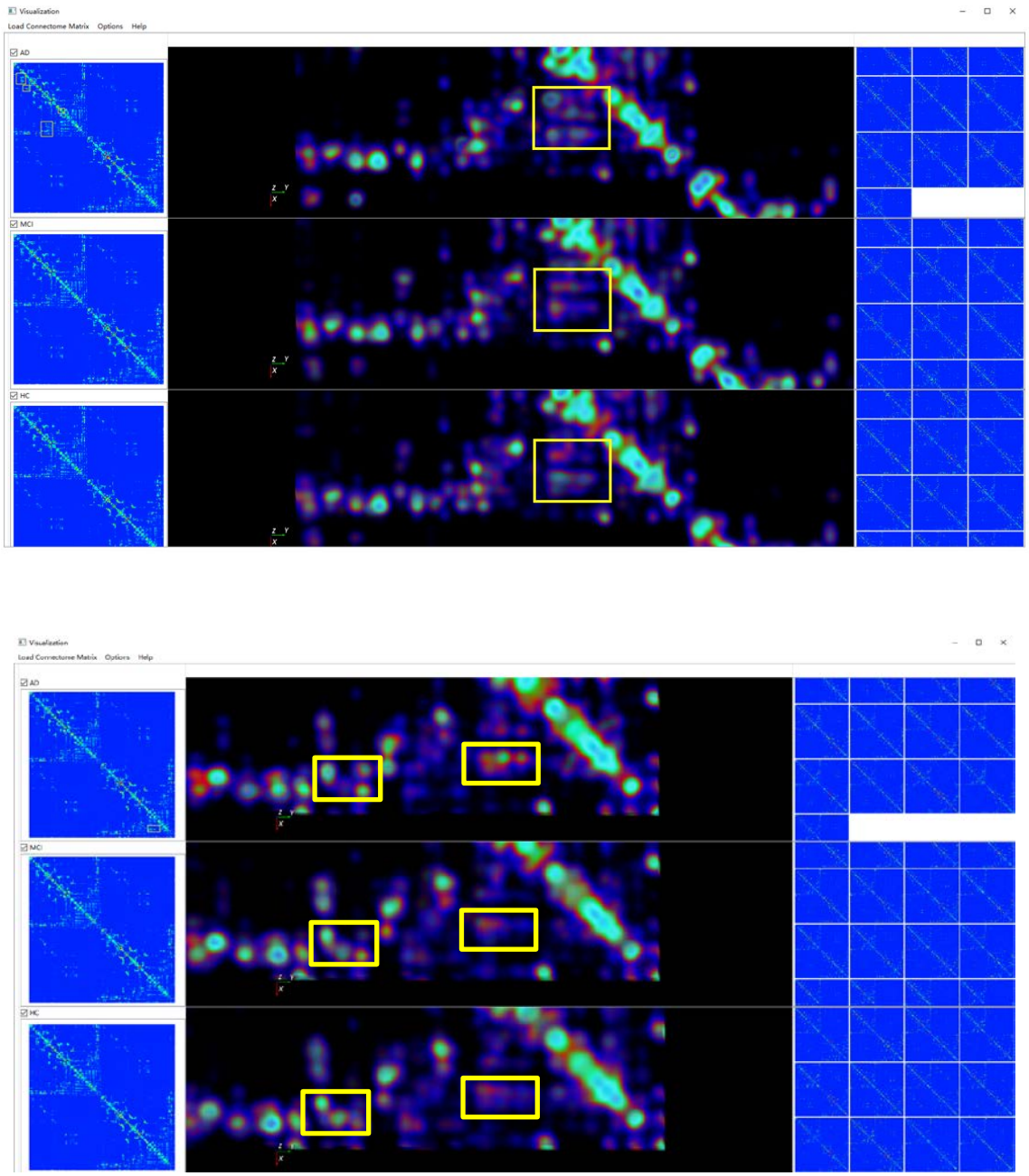

Figure 5.9. Features selected as submatrices.

5.2 Feature Analysis

In order to compute a set of features for analysis, the submatrices identified interactively are processed by Principal component analysis (PCA). PCA is applied to the features defined by the elements in each submatrix individually. The set of all such principal components (PCs) form a collective feature vector for further classification analysis.

Support Vector Machine (SVM) is used to construct classifiers. To do so, a supervised classification algorithm is implemented. A 3-fold cross validation approach is applied. For each category, 2/3 of the total subjects are randomly selected and used as training data and the remaining 1/3 are kept for testing. This process will then be repeated 3 times with different randomly selected subsets, and the results are averaged over the 3 rounds. Our training set has 13 subjects from AD, 29 subjects from HC and 29 subjects from MCI. The test set has 6 subjects from AD, 14 subjects from HC and 13 subjects from MCI. To avoid overfitting, the feature vector was limited to have about 25 features. In order to compare with results without the visual feature selection, PCA was also applied on the entire 234x234 multigraph matrices. The resulting PCs will then be selected using a standard best first feature selection algorithm [35]. This will reduce the feature set to 25 features (same as the visually selected features). The training results from the SVM are then applied to the test set for validation. Table 5.1 shows the comparisons of the three pairwise classification results using both visually detected features and automatically detected features. There are significant improvements in all three tests using the visually detected features.

Table 5.1. Test results of three classifiers: AD vs HC; AD vs MCI; HC vs MCI.

Clinical diagnosis	Visual-features		Auto-features	
	AD(+)	AD(-)	AD(+)	AD(-)
AD	5.3	0.7	4	2
HC	2	12	5	9
	Overall: 87.0%		Overall:65.5%	

Clinical diagnosis	Visual-features		Auto-features	
	MCI(+)	MCI(-)	MCI(+)	MCI(-)
MCI	9.7	3.3	8	5
HC	2.3	11.7	5	9
	Overall: 79.1%		Overall: 62.9%	

Clinical diagnosis	Visual-features		Auto-features	
	AD(+)	AD(-)	AD(+)	AD(-)
AD	4.7	1.3	4	2
MCI	3	10	5	8
	Overall: 77.6%		Overall: 64.1%	

5. 3 Evaluations

The tool was evaluated by biomedical scientists for their comments and suggestions.

[Review by Changhyun Gil. Ph. D. Postdoc at Indiana University Department of Pediatrics]

The authors developed an interactive/interactive volume visualization tool from health and Alzheimer's disease patients' brain connectome networks multigraph data sets. This multigraph visualization tool may help users to better detect the features of Alzheimer's disease. It is a good choice to use multigraph visualization approaches to solve neurobiology problems due to the complexity of brain structure and the dynamic of neuron connectivity. In addition, this method treated multigraph as a volume for disease feature detection and then combined machine learning and data visualization approaches thus is more advantageous than solely user-centered or statistic based tactics. Indeed, the authors showed that the differences between Healthy Control (HC), Mild Cognitive Impaired (MCI) and Alzheimer's disease (AD) could not be distinguished simply by statistic average. However in Table 1, it was shown that after being trained with training data sets using a supervised classification algorithm, visually detected features could classify each pair of group of data better than automatically detected features.

I think this interactive feature detection method using multigraph visualization framework is very promising for the detection of AD features and thus can potentially facilitate researchers to target AD "hot spots" and search for corresponding treatment. In addition, since the preliminary data showed that this method could distinguish between HC and AD, as well as MCI and AD better than auto-features, this method is

also promising for the early diagnosis of AD patients. Furthermore, it will be very interesting to apply this method to the feature detection of other neurological conditions, for example, Parkinson Disease. In the future, similar methods can even be applied to other complex biomedical research questions, like the study of intra-cellular signaling pathways and metabolic pathways.

Though the preliminary data is very encouraging, some additional works are suggested to further validate this method: First, the author suggested that visually detected features performed better than auto-features in the classification between 3 pairwise groups (HC vs. MCI, MCI vs. AD and HC vs. AD). However, it would be interesting to see if it can also better classify among all 3 groups (HC vs MCI vs AD). Second, the brain connectome data included 43 HC, 42 MCI 19 AD subjects. A different database, or a bigger pool of patients is needed to further validate this method.

CHAPTER 6. CONCLUSIONS

I have presented a multigraph visualization framework for interactive feature detection using brain image data. Treating multigraph as a volume for interactive feature detection is a novel approach, and the results look promising. Interactive feature detection through data visualization effectively bridge visualization and data mining, and is able to take advantages of both human perceptual abilities and the power of data mining algorithms. I believe this a more powerful and efficient paradigm for visual analytics than pure user-centered visual data manipulation.

In the future, I would like to develop a richer set of interactive operations with the multigraph visualization platform, including perhaps the interactive visualization of analysis results to evaluate each news selected feature vector.

REFERENCES

REFERENCES

[1] Hu H., Wu L., Yang C., Song H.: Interactive multigraph visualization and exploration with a two-phase strategy. Journal of Systems Engineering and Electronics, 2014, 25(5): 886-894.

[2] Bullmore E., Sporns O.: Complex brain networks:graph theoretical analysis of structural and functional systems. Nat Rev Neurosci 10, 3, 2009, 186–198.

[3] Wikipedia C.: Connectome, 25 September 2011 20:35 UTC.

[4] He Y., Chen Z. J. & Evans A. C.: Small-world anatomical networks in the human brain revealed by cortical thickness from MRI. Cereb. Cortex 17, 2007, 2407–2419.

[5] Chen ZJ, He, Y, Rosa-Neto P, Germann J, Evans AC.: Revealing modular architecture of human brain structural networks by using cortical thickness from MRI. Cereb Cortex 2008 Oct 10;18(10): 2374-81. Epub 2008 Dec 10.

[6] Iturria-Medina Y., Sotero R. C., Canales-Rodriguez E. J., Aleman-Gomez Y. & Melie-Garcia L. Studying the human brain anatomical network via diffusionweighted MRI and graph theory. Neuroimage 40, 2008, 1064–1076.

[7] Cammou L., Gigandet X., Meskaldji D., Thiran J. P., Sporns O., Do K. Q., Maeder P., Meuli R., Hagmann P.: Mapping the human connectome at multiple scales with diffusion spectrum mri. Journal of Neuroscience Methods 203, 2, 2012, 386–397.

[8] Sporns O.: The human connectome: a complex network. Annals of the New York Academy of Sciences 1224, 1, 2011, 109–125.

[9] Margulies D. S., Bottger J., Watanabe A., Gorgolewski K. J.: Visualizing the human connectome. NeuroImage 80, 2013, 445 – 461.

[10] Basser P.J., Mattiello J., LeBihan D.: MR diffusion tensor spectroscopy and imaging. Biophys. J. 66, 1994, 259–267.

[11] Tuch D.S., Reese T.G., Wiegell M.R., Makris N., Belliveau J.W., Wedeen V.J.: High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. Magn. Reson. Med. 48, 2002, 577–582.

[12] Von Kapri A., Rick T., Caspers S., Eickhoff S.B., Zilles K., Kuhlen T.: Evaluating a visualization of uncertainty in probabilistic tractography. Proc. SPIE Medical Imaging 2010: Visualization, Image-Guided Procedures, and Modeling, 7625.

[13] Salvador R., Suckling J., Schwarzbauer C., Bullmore E.: Undirected graphs of frequency-dependent functional connectivity in whole brain networks. Philos. Trans. R. Soc. Lond. B Biol. Sci. 360, 2005, 937–946.

[14] Achard S., Salvador R., Whitcher B., Suckling J., Bullmore E.: A resilient, lowfrequency, small-world human brain functional network with highly connected association cortical hubs. J. Neurosci. 26, 2006, 63–72.

[15] Van Den Heuvel M. P., Mandl R. C., Kahn R. S., Pol H., and Hilleke E.: Functionally linked resting-state networks reflect the underlying structural connectivity architecture of the human brain. Human Brain Mapping, 30(10), 2009, 3127-3141.

[16] Skudlarski P., Jagannathan K., Calhoun V. D., Hampson M., Skudlarska B. A. and Pearlson G.: Measuring brain connectivity: diffusion tensor imaging validates resting state temporal correlations. NeuroImage, 43(3), 2008:554-561.

[17] Herman I., Melanconm G., Marshall M.S.: Graph visualization and navigation in information visualization: A survey. IEEE Transactions on Visualization and Computer Graphics 6, 1, Jan, 2000, 24–43.

[18] Lamping J., Rao R., Pirolli P.: A Focus+context Technique Based on Hyperbolic Geometry for Visualizing Large Hierarchies, Human Factors in Computing Systems, CHI '95 Conf. Proc.,1995.

[19] Hadlak S., Schumann H., Schulz H. J.: A survey of multi-faceted graph visualization. In Eurographics Conference on Visualization (EuroVis), 2015, 1–20. 2.

[20] Zuo X. N., Ehmke R., Mennes M., Imperati D., Castellanos F.X., Sporns O., Milham M.P.: Network centrality in the human functional connectome. Cereb. Cortex 22, 2012, 1862–1875.

[21] Foucher J.R., Vidailhet P., Chanraud S., Gounot D., Grucker D., Pins D., Damsa C., Danion J.-M.: Functional integration in schizophrenia: too little or too much? Preliminary results on fMRI data. NeuroImage 26, 2005, 374–388.

[22] Renoust B., Melancon G., Munzner T.: Detangler: Visual analytics for multiplex networks. Computer Graphics Forum 34, 3 2015, 321–330.

[23] Mao M., Lu J., Zhang G., Zhang J.: Multirelational Social Recommendations via Multigraph Ranking. IEEE Transactions on Cybernetics, vol. PP, 99, 1-13.

[24] Kerracher N., Kennedy J., Charmers K.: The design space of temporal graph visualization. In Proceedings of the Eurographics Conference on Visualization (EuroVis' 14, Short Papers Track), 2014, 2.

[25] Bach B., Pietriga E., Fekete J. D.: Visualizing dynamic networks with matrix cubes. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, New York, NY, USA, 2014, CHI '14, ACM, 877– 886.

[26] Post F. H., Vrolijk B., Hauser H., Laramee R. S., Doleisch H.: The state of the art in flow visualization: Feature extraction and tracking, 2003, 2.

[27] Wang J., Sisneros R., Huang J.: Interactive selection of multivariate features in large spatiotemporal data. In 2013 IEEE Pacific Visualization Symposium (PacificVis), Feb 2013, 145–152.

[28] Maniyar D. M., Nabney I. T.: Data visualization with simultaneous feature selection, IEEE Symposium on Computational Intelligence and Bioinformatics and Computational Biology, Sept 2006, 1–8.

[29] Shi L., Tong H., Mu X.: Brainquest: Perceptionguided brain network comparison. In Data Mining (ICDM), IEEE International Conference, Nov 2015, 379–388.

[30] Li H., Fang S., Goni J., Contreras J. A., Liang Y., Cai C., West J. D., Risacher S. L., Wang Y., Sporns O., Saykin A. J., Shen L., Andi A. T.: Brain Informatics and Health: 8th International Conference, BIH 2015, London, UK, August 30 - September 2, 2015.

[31] Flandin G., Kherif F., Pennec X., Rivi`ere D., Ayache N., Poline J.B.: Parcellation of brain images with anatomical and functional constraints for fMRI data analysis. Projet Epidaure, 2004 Route des Lucioles, 06902 Sophia Antipolis, France Service Hospitalier Fr´ed´eric Joliot, CEA, 4 Place du G´en´eral Leclerc, 91406 Orsay, France, IFR 49, Institut Imagerie Neurofonctionnelle, Paris, France.

[32] Meißner M., Pfister H., Westermann R., Wittenbrin C.M.: Volume Visualization and Volume Rendering Techniques. EUROGRAPHICS '2000.

[33] Lacroute P. and Levoy M.: Fast volume rendering using a shear-warp factorization of the viewing transform. In Computer Graphics, Proceedings of SIGGRAPH 94, pages 451–457, July 1994.

[34] Turk M. and Pentland A.: Eigenfaces for Recognition, Journal of Cognitive Neuroscience, vol. 3, no. 1, pp. 71-86, 1991.

[35] Xu, L, Yan, P, Chang, T. Best first strategy for feature selection. 9th International Conference on Pattern Recognition, 706-708, 1988.