Multigraph Visualization for Feature Classification of Brain Network Data

Jiachen Wang[1], Shiaofen Fang[1], Huang Li[1], Joaquín Goñi[2], Andrew J. Saykin[3] and Li Shen[3]

[1]Department of Computer & Information Science, Indiana University Purdue University Indianapolis
[2]College of Engineering, Purdue University
[3]Indiana University, School of Medicine

Abstract

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 paper, we introduce a multigraph application in brain network data analysis that has a strong need for multigraph visualization. In this application, multigraph is used to represent brain connectome networks of multiple human subjects. A volumetric data set is constructed from the matrix representation of the multigraph. A volume visualization tool is then developed to assist the user to interactively and iteratively detect network features that may contribute to certain neurological conditions. We apply this technique to a brain connectome dataset for feature detection in the classification of Alzheimer's Disease (AD) patients. Preliminary results show significant improvements when interactively selected features are used.

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

Categories and Subject Descriptors (according to ACM CCS): Visualization [Human-centered computing]: Visualization application domains—Visual analytics

1. Introduction

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 [HWYS14] is a set of graphs that have a common set of nodes but different sets of edges. 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 classi-

fication of neurological diseases. This set of brain networks from mul-tiple subjects is a multigraph as humans have similar brain structures. In this paper, we will show that proper multi-graph visualization can help the analysis process by generat-ing more salient visual features.

Human connectomics [BS09] studies how the human brain is wired and how its function is affected by the con-nectivity pattern using multi-modal neuroimaging data. The human brain is a complex network of approximately 1010 neurons linked by 1014 synaptic connections [Wik16]. Given such an unprecedented complexity, 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 [CGM*12] [Spo11]. The visualization and related visual analytics of this network has not been well studied.

In this paper, we 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 classifi-

cation. By visualizing and comparing the multigraphs of subjects in different classes, we are able to detect and extract network features that are more salient in differentiating data groups for diagnosis and analysis.

In the rest of this paper, related work will first be discussed in Section 2. The application dataset will be described in Section 3. In Section 4, we will describe our 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.

2. Related Work

Human brain connectomics involves several different imaging modalities that require different visualization techniques. Margulies, et al [MBWG13] provided an excellent overview of the various available visualization tools for brain anatomical and functional connectivity data. A tool called ConnectomeExplorer [BAAK*13] can also provide visual exploration of brain connectome data from microscopic images. In brain connectomics, connectome network's connectivity data are usually visualized as graphs. Graph visualization has been extensively studied in information visualization. For connec-tomics application, the networks can be either visualized as separate graphs, away from the anatomical context [SSSB05] or embedded into the brain anatomical context [ABHR*13] [ZEM*12] [XWH13].

As a subfield of information visualization, graph visualization has been extensively studied. Surveys of classic graph visualization techniques can be found in [HMM00] and [Tam07]. In addition, there are also various types of deviations from the classic network structure in graph visualization such as node clustering, multivariate attributes and spatial constraints. Detailed literature reviews of these types of multifaceted techniques is given in [HSS15]. However, few techniques have been developed for multigraphs. Although multivariate graphs [RMM15] [HSS15] can contain multiple edges for pairs of nodes in the same graph, they do not pro-vide 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 [HWYS14]. But it is really focusing on only multi-edge graphs, i.e. the same graph with multiple edges.

A somewhat similar problem is the visualization of dynamic graphs [BBDW14] [KKC14]. 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 [BPF14], 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. 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 visu-alization. While feature visualization has been an active topic in scientific visualization [RPS], using interactive visuali-zation 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 [PVH*03], spatiotemporal GIS [WSH13], bioinformatics [MN06], and neuroimaging [STM15].

3. Brain Network Data Set

Our application is the human brain network data analysis for Alzheimer's Disease classification using an MRI and diffu-sion tensor imaging (DTI) dataset from the Alzheimers Dis-ease 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 1. More details of this process, including the parcellation and tractography algorithms, can be found in [LFG*15]. The pipeline is divided into three steps: (1) Gen-eration of regions of interest (ROIs), (2) DTI tractography, and (3) connectivity network construction.

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.
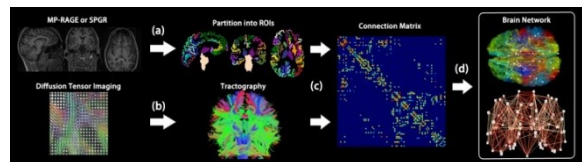
Figure 1: *Creation of structural connectivity networks.*

In our 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.

4. 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.

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. We then generate volume renderings of these volumes side-by-side to detect salient features which are both common within groups and different across groups.

To visualize a multigraph volume, we first need 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. We apply Gaussian filters to "splat" each point to nearby voxels. Figure 2 shows an example of a graph matrix before and after the Gaussian filtering. VTK is then used to carry out the volume rendering. Opacity and color transfer functions are designed for the users to interactively adjust the visual effects to look for salient features. Figure 3 shows the interface of this visualization tool. The right column shows all the individual graph matrices. The left column shows the averages of all the slices (matrices) within each category. The middle column shows all volume rendering results of
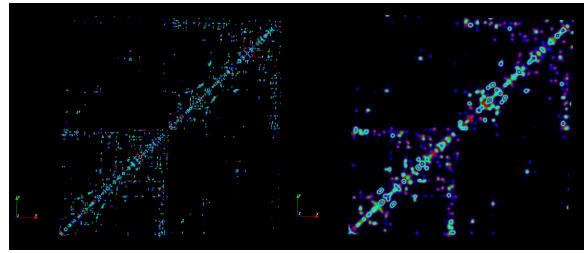

Figure 2: *A graph matrix before & after Gaussian filtering.*

the three multigraph volumes. The trans-formations of the three volume renderings are synchronized to facilitate interactive visual comparisons. 3D interactions such as rotation and scaling, can be applied to the volume rendered images to compare these 3 groups of subjects. 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 4, such regions will be selected on the interface as submatrices. Since this is an interactive process, it is primari-ly 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.
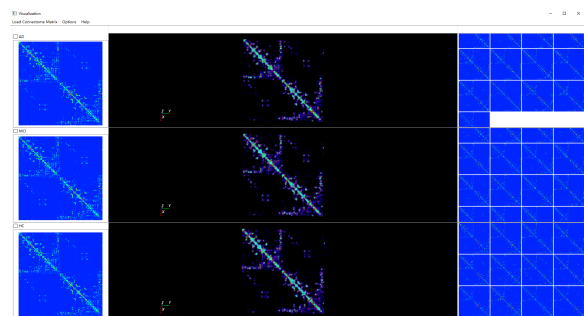

Figure 3: *The interactive feature selection interface.*

Results from Figure 4 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.

5. Feature Analysis

In order to compute a set of features for analysis, the subma-trices 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.
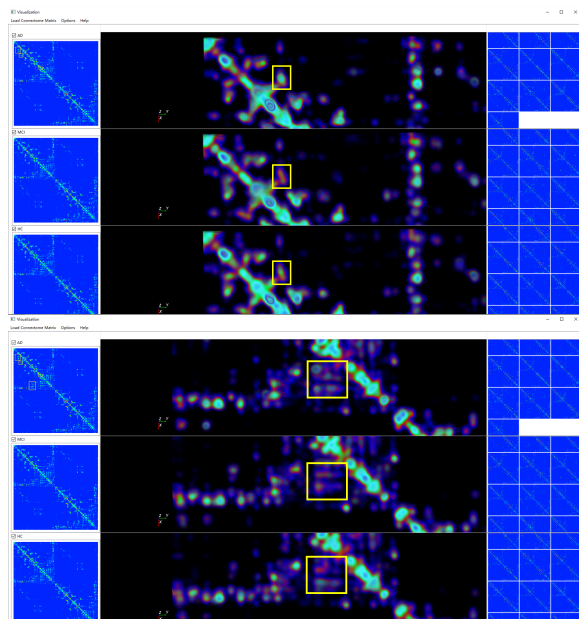
Figure 4: *The interactive feature selection interface.*

Table 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%	

Support Vector Machine (SVM) is used to construct classi-fiers. To do so, a supervised classification algorithm is im-plemented. 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, we limit the feature vector to have about 25 features.

In order to compare with results without the visual feature selection, we also apply PCA on the entire 234x234 multigraph matrices. The resulting PCs will then be selected using a standard best first feature selection algorithm [XYC88]. 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 1 shows the comparisons of the three pairwise classification results using both visually de-tected features and automatically detected features. There are significant improvements in all three tests using the visually detected features.

6. Conclusions

We 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. We believe this a more powerful and efficient paradigm for visual analytics than pure user-centered visual data manipulation.

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

References

[ABHR*13] ALPER B., BACH B., HENRY RICHE N., ISENBERG T., FEKETE J.-D.: Weighted graph comparison techniques for brain connectivity analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2013), CHI '13, ACM, pp. 483–492. URL: http://doi.acm.org/10.1145/2470654.2470724, doi:10.1145/2470654.2470724. 2

[BAAK*13] BEYER J., AL-AWAMI A., KASTHURI N., LICHTMAN J. W., PFISTER H., HADWIGER M.: Connectomeexplorer: Query-guided visual analysis of large volumetric neuroscience data. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (Dec 2013), 2868–2877. doi:10.1109/TVCG.2013.142. 2

[BBDW14] BECK F., BURCH M., DIEHL S., WEISKOPF D.: The State of the Art in Visualizing Dynamic Graphs. In *EuroVis - STARs* (2014), Borgo R., Maciejewski R., Viola I., (Eds.), The Eurographics Association. doi:10.2312/eurovisstar.20141174. 2

[BPF14] 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, pp. 877–886. URL: http://doi.acm.org/10.1145/2556288.2557010, doi:10.1145/2556288.2557010. 2

[BS09] BULLMORE E., SPORNS O.: Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience 10*, 3 (2009), 186–198. 1

[CGM*12] CAMMOUN 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. URL: http://www.sciencedirect.com/science/article/pii/S0165027011005991, doi:http://dx.doi.org/10.1016/j.jneumeth.2011.09.031. 1

[HMM00] HERMAN I., MELANCON 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. doi:10.1109/2945.841119. 2

[HSS15] HADLAK S., SCHUMANN H., SCHULZ H.-J.: A survey of multi-faceted graph visualization. In *Eurographics Conference on Visualization (EuroVis)* (2015), pp. 1–20. 2

[HWYS14] HU H., WU L., YANG C., SONG H.: Interactive multigraph visualization and exploration with a two-phase strategy. *Journal of Systems Engineering and Electronics 25*, 5 (Oct 2014), 886–894. doi:10.1109/JSEE.2014.00102. 1, 2

[KKC14] KERRACHER N., KENNEDY J., CHALMERS K.: The design space of temporal graph visualisation. In *Proceedings of the Eurographics Conference on Visualization (EuroVis' 14, Short Papers Track)* (2014). 2

[LFG*15] 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., ADNI A. T.: *Brain Informatics and Health: 8th International Conference, BIH 2015, London, UK, August 30 - September 2, 2015. Proceedings*. Springer International Publishing, Cham, 2015, ch. Integrated Visualization of Human Brain Connectome Data, pp. 295–305. URL: http://dx.doi.org/10.1007/978-3-319-23344-4_29, doi:10.1007/978-3-319-23344-4_29. 2

[MBWG13] MARGULIES D. S., BÖTTGER J., WATANABE A., GORGOLEWSKI K. J.: Visualizing the human connectome. *NeuroImage 80* (2013), 445 – 461. Mapping the Connectome. URL: http://www.sciencedirect.com/science/article/pii/S1053811913004709, doi:http://dx.doi.org/10.1016/j.neuroimage.2013.04.111. 2

[MN06] MANIYAR D. M., NABNEY I. T.: Data visualization with simultaneous feature selection. In *2006 IEEE Symposium on Computational Intelligence and Bioinformatics and Computational Biology* (Sept 2006), pp. 1–8. doi:10.1109/CIBCB.2006.330985. 2

[PVH*03] POST F. H., VROLIJK B., HAUSER H., LARAMEE R. S., DOLEISCH H.: The state of the art in flow visualisation: Feature extraction and tracking, 2003. 2

[RMM15] RENOUST B., MELANÇON G., MUNZNER T.: Detangler: Visual analytics for multiplex networks. *Computer Graphics Forum 34*, 3 (2015), 321–330. URL: http://dx.doi.org/10.1111/cgf.12644, doi:10.1111/cgf.12644. 2

[RPS] REINDERS F., POST H. F., SPOELDER J. H.: Visualization of time-dependent data with feature tracking and event detection. *The Visual Computer 17*, 1, 55–71. URL: http://dx.doi.org/10.1007/PL00013399, doi:10.1007/PL00013399. 2

[Spo11] SPORNS O.: The human connectome: a complex network. *Annals of the New York Academy of Sciences 1224*, 1 (2011), 109–125. URL: http://dx.doi.org/10.1111/j.1749-6632.2010.05888.x, doi:10.1111/j.1749-6632.2010.05888.x. 1

[SSSB05] SALVADOR R., SUCKLING J., SCHWARZBAUER C., BULLMORE E.: Undirected graphs of frequency-dependent functional connectivity in whole brain networks. *Philosophical Transactions of the Royal Society of London B: Biological Sciences 360*, 1457 (2005), 937–946. URL: http://rstb.royalsocietypublishing.org/content/360/1457/937, arXiv:http://rstb.royalsocietypublishing.org/content/360/1457/937.full.pdf, doi:10.1098/rstb.2005.1645. 2

[STM15] SHI L., TONG H., MU X.: Brainquest: Perception-guided brain network comparison. In *Data Mining (ICDM), 2015 IEEE International Conference on* (Nov 2015), pp. 379–388. doi:10.1109/ICDM.2015.135. 2

[Tam07] TAMASSIA R.: *Handbook of Graph Drawing and Visualization (Discrete Mathematics and Its Applications)*. Chapman & Hall/CRC, 2007. 2

[Wik16] WIKIPEDIA: Connectome — wikipedia, the free encyclopedia, 2016. [Online; accessed 29-April-2016]. URL: https://en.wikipedia.org/w/index.php?title=Connectome&oldid=708781560. 1

[WSH13] WANG J., SISNEROS R., HUANG J.: Interactive selection of multivariate features in large spatiotemporal data. In *2013 IEEE Pacific Visualization Symposium (PacificVis)* (Feb 2013), pp. 145–152. doi:10.1109/PacificVis.2013.6596139. 2

[XWH13] XIA M., WANG J., HE Y.: Brainnet viewer: a network visualization tool for human brain connectomics. *PloS one 8*, 7 (2013), e68910. 2

[XYC88] XU L., YAN P., CHANG T.: Best first strategy for feature selection. In *Pattern Recognition, 1988., 9th International Conference on* (Nov 1988), pp. 706–708 vol.2. doi:10.1109/ICPR.1988.28334. 4

[ZEM*12] ZUO X.-N., EHMKE R., MENNES M., IMPERATI D., CASTELLANOS F. X., SPORNS O., MILHAM M. P.: Network centrality in the human functional connectome. *Cerebral Cortex 22*, 8 (2012), 1862–1875. URL: http://cercor.oxfordjournals.org/content/22/8/1862.abstract, arXiv:http://cercor.oxfordjournals.org/content/22/8/1862.full.pdf+html, doi:10.1093/cercor/bhr269. 2