

Visual Analysis of Brain Networks using Sparse Regression Models

LEI SHI, SKLCS, Chinese Academy of Sciences & University of Chinese Academy of Sciences

HANGHANG TONG, Arizona State University

MADELAINE DAIANU, University of Southern California

FENG TIAN, Chinese Academy of Sciences & University of Chinese Academy of Sciences

PAUL M. THOMPSON, University of Southern California

for the Alzheimer's Disease Neuroimaging Initiative*

Studies of the human brain network are becoming increasingly popular in the fields of neuroscience, computer science and neurology. Despite this rapidly growing line of research, gaps remain on the intersection of data analytics, interactive visual representation and the human intelligence - all needed to advance our understanding of human brain networks. This paper tackles this challenge by exploring the design space of visual analytics. We propose an integrated framework to orchestrate computational models with comprehensive data visualizations on the human brain network. The framework targets two fundamental tasks: the visual exploration of multi-label brain networks and the visual comparison among brain networks across different subject groups. During the first task, we propose a novel interactive user interface to visualize sets of labeled brain networks; in our second task, we introduce sparse regression models to select discriminative features from the brain network to facilitate the comparison. Through user studies and quantitative experiments, both methods are shown to greatly improve the visual comparison performance. Finally, real-world case studies with domain experts demonstrate the utility and effectiveness of our framework to analyze reconstructions of human brain connectivity maps. The perceptually optimized visualization design and the feature selection model calibration are shown to be the key to our significant findings.

CCS Concepts: •Human-centered computing → Visualization techniques; •Information systems → Data mining;

Additional Key Words and Phrases: Brain network, connectome, feature selection, visual analysis

ACM Reference Format:

Lei Shi, Hanghang Tong, Madelaine Daianu, Feng Tian, and Paul M. Thompson, 2017. Visual Analysis of Brain Networks using Sparse Regression Models. *ACM Trans. Knowl. Discov. Data.* x, x, Article xx (January

* Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Author's addresses: L. Shi, State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences & University of Chinese Academy of Sciences, Beijing, China; H. Tong, School of Computing, Informatics and Decision Systems Engineering, Arizona State University, Tempe, AZ, USA; M. Daianu, Imaging Genetics Center, Mark & Mary Stevens Institute for Neuroimaging & Informatics, University of Southern California, Marina del Rey, CA, USA; F. Tian, Intelligence Engineering Lab, Institute of Software, Chinese Academy of Sciences & University of Chinese Academy of Sciences, Beijing, China; P. M. Thompson, Departments of Neurology, Psychiatry, Radiology, Engineering, Pediatrics, and Ophthalmology, University of Southern California, Los Angeles, CA, USA.

Part of this paper was published in the Proceedings of IEEE International Conference on Data Mining series (ICDM 2015) entitled "BrainQuest: Perception-Guided Brain Network Comparison".

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2017 ACM. 1556-4681/2017/01-ARTxx \$15.00

DOI: 0000001.00000001

xx:2

L. Shi et al.

2017), 31 pages.
DOI: 0000001.0000001

1. INTRODUCTION

Recent advances in Magnetic Resonance Imaging (MRI) [Huettel et al. 2008] - a non-invasive tool frequently used in medical imaging applications, have revolutionized our understanding of the structure and function of the *in vivo* human brain. Particularly, diffusion MRI [Merboldt et al. 1985], or diffusion weighted imaging (DWI), is widely used to obtain information about the neural architecture of the human brain. Most recent efforts in the field have helped model the human brain as a complex network of connections, also known as the human “connectome”. At microscopic level, the connectome consists of neural pathways, as dictated by the signal from hundreds of billions of neurons (and many more synapses) and their neural connections. At macroscopic level, many findings that focus on the human connectome [Daianu et al. 2013; Daianu et al. 2015; Crossley et al. 2014] report on the structural and functional properties describing the coupling between distinct cortical regions.

In this work, we target the structural human brain network at the macroscopic level reconstructed from neuroimaging data (both DWIs and anatomical MRI scans). The brain network nodes are defined by specialized brain regions parcellated according to neuroanatomical labels or coordinated activities, also known as Region of Interests (ROIs). The network edges are the connections between ROIs inferred from the neuroimaging data [Davidson et al. 2013], where the edge weight represents the strength of the connection, e.g., the number of fibers going through the source and target ROI. In such networks, the number of nodes and ROIs can range from a few tens to a thousand, depending on the granularity of the parcellation. Because a 1000×1000 brain connectivity network is visually complex to be interpreted by users, our preliminary study here focused on the coarser-grained brain networks consisting of 68 or 70 ROIs segmented from the T1-weighted MRI scans by the FreeSurfer software¹ [Fre 2012], and the edges were obtained from the white matter pathways extracted from a whole-brain tractography [Jin et al. 2014].

Much progress has been made with collecting and processing data on the human connectome, e.g., under the topic of brain network analysis in the data mining field [Kong and Yu 2014]. However, it is not yet clear how to accurately illustrate these multivariate and complex networks in order to utilize all existing information and advance our understanding of the human connectome. It is indisputable that the visualization methods are essential in analyzing brain networks and are of interest to various fields of research (e.g., neuroscience researchers on the data analytics side and physicians on the clinical side). In its core, the problem of jointly optimizing the data mining model and the visual representation design on brain networks has barely, if at all, been studied before. Some of the questions that arise relate to the overall brain network representation in healthy and diseased populations, in addition to differentiating factors between healthy and diseased patterns of connectivity, e.g., Alzheimer’s Disease (AD) [Daianu et al. 2013]. In essence, the macroscopic structural brain networks studied here can be interpreted as multi-labeled networks - where multiple labels describe clinical and other properties about the participants and their corresponding brain networks. These labels include the subject’s age, gender, healthy/diseased status and their IQ level, etc. The illustration of network differences correlated with relevant subject labels is the main problem studied in our work.

¹There is a small difference in the number of ROIs across our two brain network data sets because they are processed by different versions of the FreeSurfer software.

In methodology, the motivation behind this work is to initiate and demonstrate the use of effective visual analytics (VA) techniques that combine both visualization and data mining techniques, and then seamlessly integrate them into an interactive visualization system. We envision the necessity of VA techniques on brain network analysis for three reasons. First, the nature and organization of brain networks are highly complex, and at the cortical level, the network is denser than most scale-free social networks previously studied. The graph density² of a 70-region brain network can reach a level of 0.4 in our data set, which means that each region of the brain will connect, on average, to 40% of all other brain regions. In a visualization-only approach, a hairball-like picture is often drawn, making it difficult to analyze and interpret by human users. Second, the brain networks of healthy individuals of similar age may have only small differences. For cognitive tasks such as the network comparison subject to the clinical data (i.e., the declining cognitive scores), users can hardly extract meaningful patterns only from the visualization of raw connectivity data. Third, the data mining models and algorithms for brain network analysis are still in their infancy, VA systems can serve as exploratory tools that aid the development of more accurate models and algorithms.

In this paper, we target two fundamental tasks on the brain network analysis: (1) the visual exploration of the multi-label brain network data set, for example, to visually slice and dice a set of brain networks into multiple groups by their labels; (2) subsequently, the visual comparison among the groups of brain networks categorized by their labels. Recent studies on network visualization proposed elegant algorithms to illustrate and navigate large, dynamic and multivariate networks [Battista et al. 1998; Herman et al. 2000; Kerren et al. 2014], but little has been done on *a set of networks* sorted into multiple diagnostic groups, and how the visualization design can be optimized to support the group-wise visual comparison. On the other hand, network topologies have been previously studied and visualized for comparison purposes [Gleicher et al. 2011], but rarely on weighted networks having nonuniform edge attributes, e.g., the fiber connection strength of brain networks.

Beyond the previous conference publication, the contribution of this extended paper can be summarized as follows:

- *An Integrated Visual Analytics Framework* aimed at the visualization and exploratory analysis of multi-label weighted brain networks. Several interactions customized for the analysis of large complex networks are introduced (Section 3);
- *Visualization Designs for Network Comparison* (Section 5) optimized according to the human perception theory (Section 6.1) and real-life brain network data sets (Section 3.1);
- *Sparse Regression Models* based on a series of regularization methods in comparison to existing network-aware feature selection methods (Section 4), to select discriminative brain network features shown in the visualization interface, and to accelerate the task of visual brain network comparison. The best model can be obtained through the built-in functionality for model selection and performance comparison. Note that in this work, we consider the edge weight on brain networks as features, i.e., the strength of fiber connections. Feature selection (or edge feature selection) means selecting a set of edges on the brain network and using their edge weights/strengths for brain network classification or visual comparison;
- *Multimodal Evaluations*, by the user and quantitative experiments on the JND theory (Section 6.1), visual designs (Section 6.2) and feature selection models (Section 6.3). New case studies are conducted with domain experts from the fields of neu-

² $p/(n^2)$, see Table I for notations.

roimaging and neuromedicine, on the brain network obtained using DWI from distinct groups of subjects (Section 6.4, Section 6.5). All the evaluation results point to the success of the proposed visualization design, computational models, and the integrated visual analytics framework.

2. RELATED WORK

2.1. Brain Network Analysis

Brain network analysis emerges as a compelling topic due to the maturation of non-invasive neuroimaging techniques [Kong and Yu 2014]. The raw neuroimaging data is oftentimes modeled as a high-order function by a three-dimensional image and a time component. Using these complex data, fundamental problems can be formulated on brain connectomics [Davidson et al. 2013]. On the network analysis, researchers define network nodes to detect brain regions with coordinated activities, and define edges to detect binary or weighted relationships between these nodes. Over the brain networks, machine learning algorithms can be used to infer the disease patterns (e.g., Alzheimer's disease [Sun et al. 2009][Huang et al. 2011]) or other variables of interest that correlate with the network difference. In this process, combining neuroimaging data with the additional patient information (e.g., genetic profile and demographic information) can help improve the learning performance [Ye et al. 2008]. In addition, subgraph extraction and analysis is an active research topic on brain networks. New methods have been proposed for frequent and discriminative uncertain graph mining [Zou et al. 2010][Kong et al. 2013][Cao et al. 2015]. Though much progress has been made in this area of study, the problem of jointly optimizing data mining models and visualization designs has barely, if at all, been studied before.

2.2. Feature Selection

Feature selection is a classical problem in the machine learning research. Depending on whether the clustering results are used during the selection process, unsupervised feature selection can be categorized into wrapper models and filter models. Representative wrapping models include FSSEM algorithm [Dy and Brodley 2000], co-selection of rare instance and features [He and Carbonell 2010], mutual-information based criterion [Law et al. 2002] and evolutionary search-based algorithm [Kim et al. 2000]. Different from wrapping models, the selection process in the filtering model is independent with the specific algorithm that generates the clustering result. Representative approaches include Laplacian score [He et al. 2005], maximum information compression index [Mitra et al. 2002], distance-based entropy [Dash and Koot 2009] and many more. On the other hand, supervised feature selections aim to leverage the supervision (e.g., class label, regression output) to guide the selection process for a subset of discriminative features. Representative methods include spectral feature selection [Zhao and Liu 2007], backward-forward algorithm [Zhang 2011], relevance feature vector machines [Cheng et al. 2007]. Readers can refer to [Tang et al. 2014] for a more complete survey. On our problem, feature selection methods have been studied together with the visualization interface [Krause et al. 2014], which can be applied to the brain network case to highlight important edge features. However, it is still unknown that whether the feature selection method can improve the user performance on the visual comparison task, and which feature selection model is the most appropriate on such task.

2.3. Multivariate Brain Network Visual Comparison

Wattenberg pioneered PivotGraph [Wattenberg 2006], an attribute-centric node-link visualization of multivariate networks. PivotGraph leverages a roll-up operation to

pivot the nodes with the same value on one or two attributes into aggregations. The attributes used can be picked manually to generate different PivotGraph views. In the data selection operation, the network can be reduced to only show node aggregations with specified attribute values. Our framework shares the similar idea in abstracting networks by their node attributes. Beyond the aggregation method in PivotGraph, we allow dynamic splitting of network abstractions by additional node attributes, which works as a fundamental operation for the visual comparison task.

Many other multivariate network visualization methods have been proposed since PivotGraph. OntoVis [Shen et al. 2006] abstracted the network based on the ontology graph of social networks. Semantic Substrate [Shneiderman and Aris 2006] proposed a user-defined layout method to place nodes in non-overlapping regions according to their attributes. GraphDice [Bezerianos et al. 2010] applied a scatterplot visual metaphor to the overview of multivariate networks. FacetAtlas [Cao et al. 2010] extracted the multifaceted entities and relationships from a collection of documents. While these methods focus on visually pruning a large multivariate network into smaller and simpler abstractions for the static visualization, our framework extends to support navigation and exploration on these network abstractions.

In the literature, the task of visually comparing weighted (brain) networks with node attributes and edge features is largely unexplored, though the comparison on graph topologies has been studied for a while [Gleicher et al. 2011]. ManyNets [Freire et al. 2010] proposed an interface to analyze multiple attributed networks simultaneously, but it focuses more on the comparison of high-level statistical attributes, and not on their network patterns defined by nodal and edge features as studied in this paper. The recent work by [Alper et al. 2013] might come closest to our work. They evaluated the effectiveness of two visual representations for the weighted graph comparison. However, their work is more on the design side and less is done on the visual analytics loop integrating advanced data analysis models and the comparative visualization optimized for human perception.

While our work targets at the structural brain network and the comparison among multiple subject groups, another class of research visualizes the functional brain network detected by functional MRI (fMRI), where the time dimension can be more important than their subject classification. The recent work of Small Multipiles [Bach et al. 2015] and Time Curves [Bach et al. 2016] develop novel visual metaphors to represent the dynamic brain network over time. Specially, Small Multipiles summarizes a series of brain networks into sequential piles where each pile is displayed by a cover adjacency matrix showing the snapshot of that pile. Time Curves focuses more on the evolution of brain networks over time, which draws a summary curve overlaid with nodes by the Multidimensional Scaling (MDS) projection. The overall pattern of network similarity and dissimilarity can be clearly delineated by Time Curves.

In summary, existing methods for both brain network visualization and analysis mostly consider one aspect of the question at task (i.e., either data mining or visualization) and are suboptimal when both the visual and the analytical constraints are applied simultaneously. The idea of combining visualization, rich user interaction and machine learning algorithms has been pioneered in Apolo [Chau et al. 2011], but their work does not target on the brain network analysis, which is the focus of this paper.

3. VISUAL ANALYTICS FRAMEWORK

In this section, we first give an overview of the human brain network studied in this work, then we propose a modeling of the brain network data and the visual analytics framework over the data model.

xx:6

L. Shi et al.

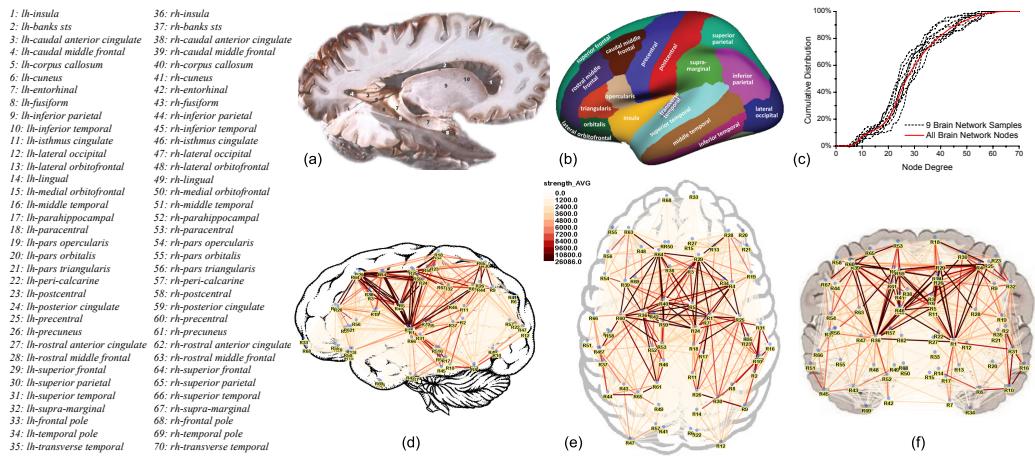


Fig. 1. The human brain network based on 70 cortical ROIs listed in the left column: (a) Grey and white matter compartments; (b) Parcellations defined using the Desikan-Killiany atlas; (c) Nodal degree distributions of individual and all brain networks; (d) Sagittal plane of the aggregated brain network; (e) Axial plane; (f) Coronal plane.

3.1. Human Brain Network and Characteristics

The generation of human brain networks is quite challenging and may be accomplished using various approaches. In this work, we follow the process to generate the brain network as described in these papers [Daianu et al. 2013][Gray et al. 2012]. The raw inputs are the 3D anatomical scans from structural MRI; these images can be segmented into white and grey matter compartments (Figure 1(a)), and further parcellated into 68 or 70 distinct ROIs using the Desikan-Killiany atlas [Desikan et al. 2006] as implemented in the FreeSurfer software [Fre 2012] (Figure 1(b)). These ROIs are listed in the left column of Figure 1. Next, DWIs are used to infer the neural pathways and connectivity patterns of the brain's connectome through a method called tractography [Jones 2008]. DWI is capable of capturing subtle changes in the white matter make-up using measures that are sensitive to white matter fiber integrity and microstructure, otherwise not detectable with structural MRI. From DWI, we define the edges (i.e., fiber bundles) that interconnect the pairs of nodes in the network and allow information transfer among regions of the brain. The edges can be assigned a variety of weights, most commonly, by the function of the edge strength, which is determined by the total number of fiber connections (i.e., fiber density) that pass through a pair of nodes. The resulting brain network describes the overall connectivity pattern of the human brain.

We studied two data sets in this work, each composed of the brain network of multiple subjects together with the demographics (labels) of each subject. The first data set was obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu)³. In total, 202 subjects underwent whole-brain MRI at 16 different sites across North America with ages ranged from 55 to 90 years. Among these 202 subjects, 50 were healthy controls, 72 had early mild cognitive impairment (eMCI), 38 had late mild cognitive impairment (IMCI) and 42 were AD patients. For each subject,

³The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD).

we reconstructed a network with 68 ROIs using the Desikan-Killiany atlas [Desikan et al. 2006]. We also ran whole brain tractography using the Hough transform to define the fiber connections (i.e., network edges) with weights describing the fiber density between a pair of ROIs. To focus on the network-level patterns, we removed the connection from each region to itself (i.e., the diagonal cell in the adjacency matrix), which left 2,278 edge features in each brain network. Each subject is recorded with three labels: gender, age and the diagnosis information.

The second data set included the brain network from 113 subjects released by the OpenConnectome project ⁴. For each subject, a brain network with 70 ROIs (i.e., nodes) according to the Desikan-Killiany atlas was computed. The edge weight between a pair of nodes represented the normalized fiber density between two ROIs. In this OpenConnectome data set, each subject had demographic information on gender, age, Full-Scale IQ (FSIQ) scores, Composite Creativity Index (CCI) and the Big Five personality traits. We classified the value of each measure into several classes to facilitate the network comparison tasks. For example, FSIQ and CCI were categorized into two classes each: the high class with FSIQ or CCI higher than or equal to 100 (the average across population) and the low class with FSIQ or CCI smaller than 100. Notably, all subjects were from young and educated population. Their ages ranged from 18 to 29 years, and 90% of them obtained a FSIQ score above the average. Therefore, on the comparison task for this data set, we focused on the two CCI groups which had more balanced group sizes.

To showcase the necessity for new methods on the brain network visual analytics, we have conducted an empirical study on the characteristics of the 113-subject Open-Connectome data set. In Figure 1(c), the cumulative distributions (CDF) of the nodal degree in the 70-region brain networks are illustrated, including the networks of nine randomly selected subjects and the overall degree distribution of all 113 networks (in red). It is noticeable that the brain network is considerably dense - about a half of nodes have a nodal degree greater than 30 (i.e., connecting more than 40% regions in the whole brain network). Also, the networks are similar among subjects, the degree distributions of all the nine sampled networks vary within a $\pm 15\%$ range from the overall degree distribution. Considering the exponential distribution of the edge strength and the three dimensional node layout, the brain network can be highly complex in a full-scale visualization. In Figure 1, we depicted the aggregated 113-subject brain network in the sagittal plane (left/right) (Figure 1(d)), axial plane (anterior/posterior) (Figure 1(e)), and coronal plane (superior/inferior) (Figure 1(f)). Throughout this study, we will mostly use the axial plane because of its better performance for visual comparison.

3.2. Data Model

We generalized the multi-label brain network data sets into a unified data model to conduct brain network analysis and visualization across multiple subjects. The model is similar to the classical multivariate network model with a set of local attributes attached on each node/edge. The major difference lies in the introduction of global attributes, also known as the network labels, which are defined for each subject at a higher level than the local attributes of each node and edge. For example, subjects can have their demographic information recorded (e.g., age, gender, etc.). To adapt to the use of global attributes, the straightforward change to the data model is to replicate all the labels of a subject onto each node (brain region) of the subject, but this would lead to a significantly larger data set, and slows both the data loading and processing. We designed an extended multi-label graph data model which stores the subject labels in

⁴<http://www.openconnectomeproject.org>

xx:8

L. Shi et al.

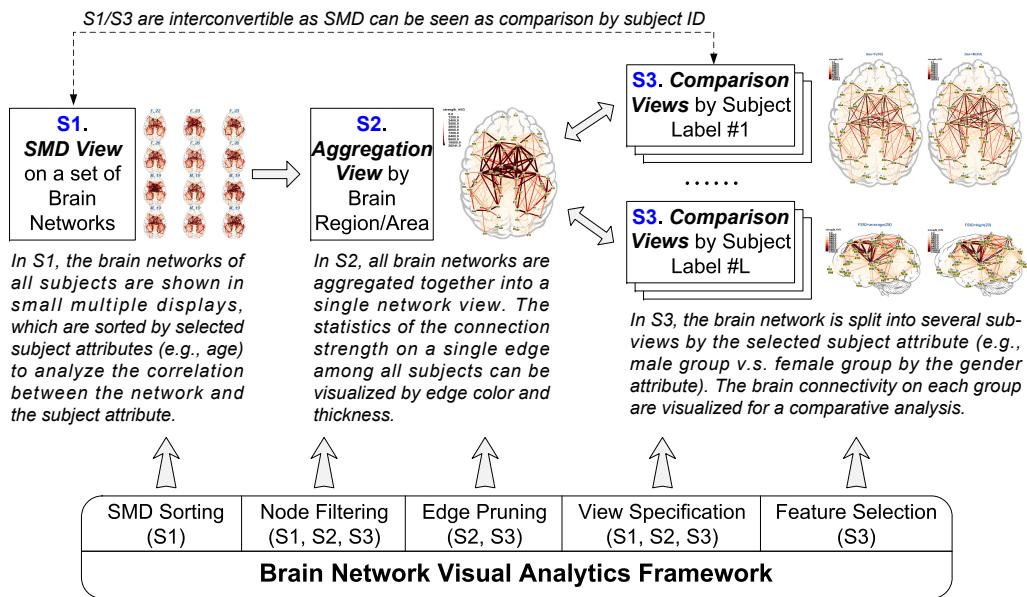


Fig. 2. The visual analytics framework for a set of multi-label brain networks. The initial SMD view (S1) lists the brain networks of all subjects, allowing to sort by subject attributes. The SMD view can be aggregated into a single network (S2) showing the average connectivity of all subjects. The aggregation network can be further split according to certain subject attribute for the comparison purpose.

a global data section. This data model is implemented in the GraphML format⁵ (other lightweight formats such as JSON are also possible). The new content beyond the standard network model includes the subject metadata under `<key>` tags, the subject labels under `<graph>` tags, and the replicated subject index on each node under the `<node>` tag, all without violating the GraphML standard. Upon an attribute query on graph nodes, the graph metadata is first examined to find out whether the query is on the global attribute of the subject or the local attribute of the graph node. In case of global attribute, the subject index is retrieved as the key to query the subject label, while in case of local attribute, a standard GraphML query process is followed.

3.3. A Three-Stage Visual Analytics Framework

Based on the proposed data model, we introduced a visual analytics framework that can not only support the visual comparison among sets of multi-label brain networks, but also conduct explorative analysis over the brain network data set. In Figure 2, the framework is conceptualized into a 3-stage pipeline. In the initial stage (S1), the set of input brain networks are visualized in small-multiple displays (SMD), with each subject's brain network shown as a thumbnail. These thumbnails can be sorted by certain attributes of the subject to reveal the correlation of the brain connectivity with the selected subject attributes. This feature is known as the SMD sorting. The SMD view gives an overall snapshot of the entire data set, but is information-overloaded for any detailed analysis on network connectivity. To proceed to the second-stage analysis, users issue an interaction to aggregate all the brain network nodes by their region indices. In other words, the nodes from all subjects indicating the same brain region are represented by one group node, and the edges are aggregated accordingly. In this

⁵<http://graphml.graphdrawing.org>

Visual Analysis of Brain Networks using Sparse Regression Models

xx:9

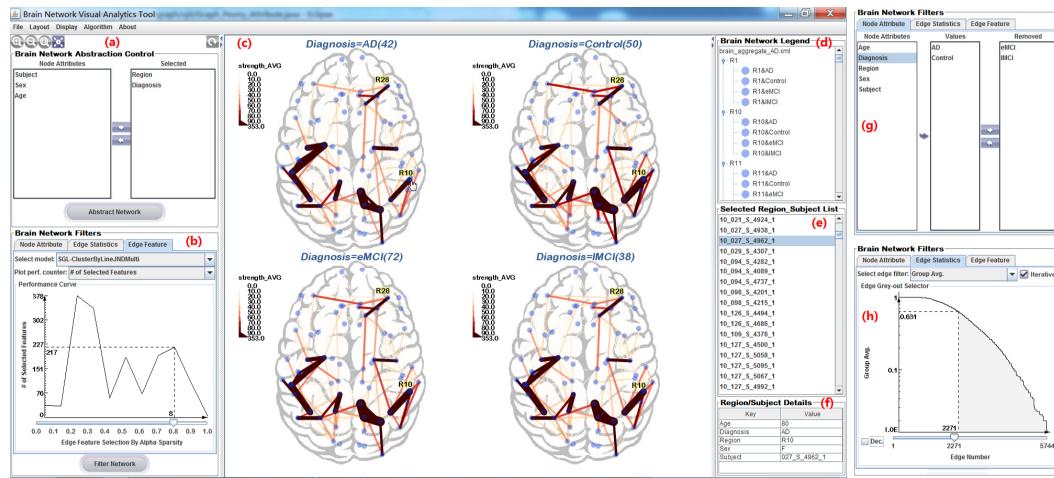


Fig. 3. The user interface of the proposed brain network visual analytics tool ((a)~(f) compose one screen shot of the tool, (g)(h) are alternative tabs of the bottom-left filter panel). The data set is part of the ADNI Consortium and includes 202 subjects: 42 Alzheimer's Disease (AD) patients, 38 patients with late mild cognitive impairment (IMCI), 72 with early mild cognitive impairment (eMCI), and 50 healthy controls. The edge color saturation and line thickness indicate the average fiber strength at each connection. Only discriminative edge features selected by a chosen model are displayed for comparison purposes. Our results are in line with previous findings in the neuroscience literature [Thompson et al. 2001; Daianu et al. 2013], which indicate that AD patients undergo certain breakdown of fiber connections in the frontal, temporal and parietal lobes, predominantly in the left hemisphere (right part in this back-side view, e.g., R10 and R28).

way, a single view representing the brain network of all the subjects is displayed, as shown in the S2 stage of Figure 2. This aggregation view enables the visual analysis of the overall brain connectivity, and also allows pattern discovery on an average brain network structure. In the third stage (S3), the single network view is split into multiple sub-views by adding one or more subject labels into the attribute list for aggregation. The resulting side-by-side visualization can help users complete the visual comparison task and analyze the relationship between the brain network connectivity and the subject's label information. For example, when users select the gender label in our data set, two brain network sub-views aggregated from the network of 50 female subjects and 63 male subjects are generated and displayed in a side-by-side comparison display. Note that the three-stage pipeline presented here is a recommended visual analytics paradigm designed based on our practice on the brain network analysis. Many other analysis trails are also feasible within our framework. For example, from the SMD view (S1) showing the brain network of every single subject, users can directly proceed to the visual comparison view (S3) by configuring an appropriate aggregation attribute list.

We have implemented a brain network visual analytics tool based on the proposed three-stage framework. Figure 3 gives the user interface. Besides the key design feature of aggregation-by-attribute that stitches together adjacent stages in the pipeline, our tool also supports several other user interactions to help customize the visualization for the purpose of brain network comparison, explorative analysis and navigation. These interactions are summarized as below.

SMD sorting. The thumbnail view in SMD can be arranged by selected subject attributes. Two subject attributes can be specified. First, the brain graphs are sorted into multiple groups by their value on the primary subject attribute. Within each subject group, the brain graphs are further sorted by the secondary subject attribute. For

example, in the S1 stage of Figure 2, subject's brain graphs are sorted first by their gender and then by age. The labels on top of each brain graph in the SMD view indicate the selected primary and secondary attribute values on each subject.

Node/Subject filtering. The bottom-left part of the visualization interface (Figure 3(b)) is equipped with a multi-tab network filtering panel. The first tab (Figure 3(g)) allows users select one item from the local node attributes or global subject labels, as defined in the data model. The values of the selected attribute/label from all brain network nodes are then listed. For example, when "diagnosis" is selected, the value list of "AD, eMCI, lMCI, Control" shows up. Users can add binary filters to remove the subjects with undesirable attribute values from the visualization. In Figure 3(g), the filters are configured to filter out subjects with eMCI and lMCI in diagnosis. Note that this filtering mechanism works at all the three stages of the proposed pipeline in an iterative manner that the network display in all the following operations will be affected.

Edge aggregation pruning. The second tab of the network filtering panel, as shown in Figure 3(h), allows users to prune brain network edges by their statistical measures. In the network aggregation/comparison view, each edge aggregation is composed of a set of single-subject edges on which several statistical measures can be defined, such as the average, standard deviation and coefficient of variance of edge strengths. Users can select one of these measures to display its CDF over all edge aggregations. Dynamic queries on both the top and the bottom edge aggregations are provided to guide the edge pruning operation. Note that we support two modes on the edge pruning. In the default mode, the edges are pruned temporarily and will come back upon the next operation. In another iterative mode, the pruned edges are removed in the data model, so that all the following operations will work on the pruning result, which enables the iterative analysis over brain networks.

View specification. The network view in each stage can be customized to deliver more information. First, the statistical measures on the node and edge aggregation can be mapped into different visual channels on the network display, including the color saturation, the line thickness and the node size. Linear feature mapping is used by default, and we also propose non-linear mappings to optimize the visual comparison performance (Section 5). The edge measures are also accessible as textual labels. Second, by using different projections, we support to observe the 3D brain network from three viewpoints: the sagittal view, the axial view and the coronal view. Third, users can apply traditional network interactions, such as zoom and pan, drag and drop, click and range selection, to examine the details of the brain network. By the selection interaction, the information of the selected nodes/aggregations/subjects is shown in the list panel (Figure 3(e)) and the detail panel (Figure 3(f)) respectively.

Edge feature selection. On the third-stage comparison view (S3), the aggregated brain networks for different groups are often indistinguishable in the visualization, due to the similar brain structure among subjects. In our framework, we introduced the computational edge feature selection method to extract fiber connections that are both discriminative and visually salient among comparing subject groups. This is achieved through the interaction on the third tab of the network filtering panel (Figure 3(b)). Users start by designating one feature selection model from the drop down list. This model is applied so that only the edge features selected by the model are displayed in the comparison view. A default model parameter (i.e., the sparsity in this implementation) is used initially, and the parameter can be tuned with the model selection slider (bottom part of Figure 3(b)). Above the slider, the number of selected features under each sparsity parameter is depicted by default as the performance curve. This curve can be switched to report other performance measures of the model. The candidate measures include the cross-validated prediction accuracy, the perception-level visibil-

Table I. Notations.

SYMBOL	DESCRIPTION
N, G_i	# of subjects and their brain graphs
n, p, e_j	# of nodes, # of edges and each edge in the brain graph
X, x_i, x_{ij}	edge weight matrix on all subjects, edge weight vector on G_i , and the component on e_j
Y, y_i	outcome value on all subjects and G_i
K, S_k, V_k	# of groups for the outcome variable, the subset of subjects for each group, and their aggregation views for comparison
\mathcal{R}, r_k, r_{kj}	transfer function on edge aggregations, edge weight on V_k and e_j
γ, γ_j	edge feature selection vector and the component for e_j
$X_\gamma, V_k(\gamma)$	partial edge weight matrix, the view after feature selection

ity by the ratio of visible features, and the clustering coefficient of selected features which indicates the degree of grouping. Users can choose the best model according to their preference on the model performance.

4. SPARSE REGRESSION MODEL

The side-by-side visualization introduced in this work is designed to optimize the human comparison task on edge features (i.e., the brain network connectivity). This is feasible for users when most features in comparison have significant differences. In the experiment of Section 6.1, we injected controlled differences on 50% edge pairs displayed in the visual comparison and varied the difference ratio during the study. Results showed that 50% users can recognize an injected difference larger than 10.9% of the maximal edge feature value, aligning well with the theory of Just Noticeable Difference (JND) as explained in Section 6.1. On the other hand, take the ADNI data set as an example and compare the average brain network of 42 AD patients and 50 healthy controls. Only 4 pairs of edge features have a larger difference than 10.9% of the maximal edge feature value in the aggregated average networks. Applying the feature capping method (Section 5) can increase the number of noticeable edge pairs to 31, but still only covers a 1.36% share of all 2,278 edge pairs displayed in the full-scale comparison. The large amount of similar edge pairs introduce a masking effect that slows or even prohibits users from detecting significantly different edges in the comparative view.

We propose to apply the computational feature selection method to the visual comparison of brain networks. The basic idea is to jointly optimize the utility of the selected features for both statistical analysis and visual comparison. Besides the sparsity requirement in the general feature selection methods, we introduce two additional objectives to optimize on the selected features: (1) the group-level classification and prediction accuracy; (2) the visual difference among groups in the comparison. In the final model-driven visualization, only the selected edge features are displayed. This helps users focus on significant differences, without interruptions by the homogenous background mask. On the other hand, in the scenario of brain network analysis, there can be a much smaller number of subjects (N) than the number of edge features (D). The feature selection method greatly reduces the effect of overfitting in the predictive model.

Before introducing the feature selection method for visual comparison, we first define some necessary notations, which are listed in Table I. The raw input of the method is the brain network of N subjects, denoted by G_1, \dots, G_N . Each network has the same number of nodes (cortical regions), denoted by n ; and p edges (fiber connections) between pairs of nodes, denoted by e_1, \dots, e_p . The edge weight is defined by the strength of the edge, in our scenario, the total number of fiber connections going through the source and target cortical regions. On G_i , the network of the i th subject, the edge

strengths are denoted by the weight vector $X_i = (x_{i1}, \dots, x_{ip})'$. For simplicity, we assume that every brain network has the same number of edges by $p = \frac{n(n-1)}{2}$. For those edges that do not have fiber connection, we set their components in the weight vector to zero.

At the network level, each subject and their brain network is associated with a discrete outcome variable, e.g., the demographics and the diagnosis of a patient. The value of the outcome on N subjects is denoted by the vector $Y = (y_1, \dots, y_N)'$, where y_i has K possible levels. This outcome variable classifies all subjects into K disjoint subsets by S_1, \dots, S_K . In the visual comparison, the brain networks in each subset are aggregated together into one sub-view by the region index, thus generating K sub-views in total, denoted by V_1, \dots, V_K . Due to the homogeneity of brain networks, each sub-view still has n nodes and p edges. The edge weight on each sub-view is determined by a transfer function \mathcal{R} over individual edge weights in the single-subject brain network. By default, we apply the mean function which is used in the standard visualization tool to illustrate the average brain connectivity of a group of subjects. The edge weight vector on the sub-view V_k is denoted by $r_k = (r_{k1}, \dots, r_{kp})'$, where r_{kj} denotes the weight of the j th edge feature. In this work, without loss of generality, we target the pairwise comparison ($K = 2$) between two sub-views (V_1, V_2) aggregating the brain networks of two subject groups split by one binary label. A typical example is the pairwise comparison between the brain networks of AD patients and healthy controls. Note that, the generalization to multi-category or continuous labels is straightforward, and omitted for brevity.

Finally, over the brain networks, the feature selection method chooses a subset of edge features for comparison, which is defined by the feature selection vector $\gamma = \{0, 1\}^p$. The j th edge feature will be selected if $\gamma_j = 1$. According to our design rationale, the problem is defined as selecting a set of features to satisfy two objectives simultaneously.

- D1. **Predictive power** by maximizing the binary prediction accuracy on the outcome label with selected features:

$$\max P(\hat{y}_i = y_i | X_\gamma, y),$$

where X_γ denotes the partial design matrix after the feature selection, \hat{y}_i is the predicted label on graph G_i ;

- D2. **Significant difference in the visual comparison** by enforcing a lower bound on the visibility of difference:

$$P(|r_{1j} - r_{2j}| \geq JND | \gamma_j = 1) \geq \xi,$$

where the visibility of difference is defined by the ratio of features with visible difference in all the selected features. The difference on a feature between sub-views for comparison is considered as visible if the absolute value of difference is no smaller than the perception-level just noticeable difference, denoted as JND . The desired visibility threshold is denoted as ξ (Section 6.1).

In principle, many existing feature selection methods can be applied in our problem: (1) the unsupervised feature selection with a mutual-information based criterion over the feature clustering results [Law et al. 2002]; (2) the statistical hypothesis testing with an one-way analysis of variance (ANOVA, the F-test). The ANOVA test checks whether each feature takes different values among comparing groups in a statistically significant way; (3) the sparse regression model that extends the basic regression analysis with feature norm constraints to impose the sparsity. Finally, we adopt the sparse regression model in this work for the following reasons. First, unsupervised feature selections generate universal results for all outcome variables (labels) and thus are computationally efficient. However, the class labels essential in the visual comparison

are not considered, which degrades the predictive performance. Second, independent feature selection methods such as the hypothesis testing only look at each feature at a time and therefore do not model either the interaction effect or the inherent feature grouping nature on brain networks. For example, on the OpenConnectome data set, the 129 significant edge features selected by one-way ANOVA predict the outcome of CCI group with a 0.523 accuracy, which is very close to the baseline method of random guess. On the contrary, the sparse regression model applies a joint optimization over the prediction accuracy and the overall sparsity for feature selection. The interaction and grouping effect among features can be captured through the feature norm constraints (e.g., the group lasso) to model the underlying brain network structure. The best prediction accuracy with the sparse regression model reaches 90% on some cases of our data set. Third, the sparse regression model can be naturally generalized to deal with different types of outcomes, including the binary (e.g., high CCI v.s. low CCI), multi-category (e.g., the healthy control, early and late mild cognitive impaired, and the AD patient) and continuous variables (e.g., age of subject).

The basic sparse regression model, also known as the lasso [Tibshirani 1996], predicts the binary outcome (the subject label) while selecting important edge features. This model has the following objective function:

$$\text{Minimize} \quad \sum_{i=1}^N \log(1 + e^{-y_i W^T X_i}) + \lambda \|W\|_1 \quad (1)$$

where $W = (w_1, \dots, w_p)'$ denotes the weight vector for all the p edge features. The edge with a larger weight means that it has a higher influence on the outcome variable. The lasso model explicitly combines two terms: the Negative Log Likelihood (NLL) of a logistic regression model, and a L1-norm regularization term. The NLL term stresses the predictive power of the model, while the regularization term shrinks all components in the weight vector W towards zero to achieve the model sparsity, i.e., the effect of feature selection. Here the parameter λ controls the degree of sparsity.

The lasso model achieves good prediction performance, however, when interaction effects among features are strong, it tends to select only one feature from each correlated feature group. This prohibits interpreting the overall comparison picture. A better model is to use both L1-norm and L2-norm as regularization terms, which is called the Elastic Net model [Zou and Hastie 2005] (i.e., stretches a fishing net to retain all big fishes).

$$\text{Minimize} \quad \sum_{i=1}^N \log(1 + e^{-y_i W^T X_i}) + \alpha \lambda \|W\|_1 + (1 - \alpha) \lambda \|W\|_2 \quad (2)$$

where the new parameter $\alpha = [0, 1]$ balances between the sparsity and the feature grouping effect.

Another similar model called the sparse group lasso has been proposed recently [Simon et al. 2013], which also combines the L1 and L2 norms in the regularization and allows the control of both within-group sparsity and group-wise sparsity. This improved model enjoys an additional advantage over the Elastic Net in that it can specify the feature grouping information by exploiting the inherent brain network structure. We apply a variant of this sparse group lasso model in our scenario, which has the following objective function.

$$\text{Minimize} \sum_{i=1}^N \log(1 + e^{-y_i W^T X_i}) + \alpha \lambda \|W\|_1 + (1 - \alpha) \lambda \sum_{m=1}^M \theta_m \|W^{(m)}\|_2 \quad (3)$$

where M denotes the number of feature groups, $W^{(m)}$ is the partial weight vector of the m th feature group.

The major difference of the proposed model from the original sparse group lasso [Simon et al. 2013] lies in the new parameter $\theta_m > 0$ that controls the priority of each group of features. θ_m is therefore called the priority parameter. This parameter is optimized according to the human perception theory in Section 6.1. The heuristics is, when one group of features are relatively unnoticeable in the visual comparison, we set the corresponding θ_m to a larger value than one, so that the probability to select this group is reduced. On the contrary, when there are significant visual differences on a feature group, we apply a θ_m smaller than one to increase the feature selection probability. More details on this prioritized sparse group lasso model are introduced in [Shi et al. 2015].

Note that the prioritized sparse group lasso method considers the feature grouping information on brain networks, but does not explicitly incorporate the graph structure in the regression process. In fact, there are a few recently proposed lasso-based methods that are optimized for the network scenario, such as Network Lasso [Hallac et al. 2015] and graph-guided fused lasso [Chen et al. 2010]. Nevertheless, on the brain network comparison scenario, all these methods can not be directly applied. Network Lasso formulates a generic optimization problem on networks, whose minimization objective adds together two convex cost functions on nodes and edges respectively. The edge cost function is fixed to the square loss, i.e., the L2-norm, so that Network Lasso is more appropriate for node clustering in addition to optimization, but not good for feature selection. This is analogous to the use of L1 norm in lasso for sparse solutions rather than the ridge regression with L2 regularization. On the other hand, the graph-guided fused lasso (GFllasso) models the relationship of multiple outcome variables as a graph and optimizes the feature selection on all outcome variables in a single process. The goal is to obtain a similar set of input features (i.e., to fuse them) for the outcomes related by the graph. Compared to our problem, GFllasso operates and constructs the graph on the output variables for multi-task learning, but does not incorporate the graph structure on input features, such as the brain network. More importantly, all these network/graph-based lasso methods define the input feature/variable on network nodes, while in our scenario the features are the connectivity strength of network edges. The straightforward way to convert to the line graph, i.e., define edges as nodes, can somehow solve this problem, but the resulting line graph will have a quite different structure due to the high edge density of brain graphs, most likely dominated by connected cliques. Applying the network-based lasso on the line graph will lose the structural and group information on the original brain network. A promising future work will be exploring the subset relationship among brain network edges that can be taken into account in a standard fused lasso model [Tibshirani et al. 2005].

To solve the proposed variant model of sparse group lasso, we apply the Moreau-Yosida regularization based algorithm in [Liu and Ye 2010]. The best model parameters are determined in two steps. First, for parameters depending only on the data set, e.g., the overall sparsity λ , we sample over possible value ranges and evaluate each resulting model by a 10-fold cross-validation. The parameter leading to the best prediction performance is applied. On the other hand, for parameters that control the interpretability of the model, e.g., the group-wise sparsity parameter α , we allow users to manually tune the model, with the corresponding performance measures shown as feedbacks. In the real usage of the model, all the edges having nonzero components in the weight vector $W = (w_1, \dots, w_p)'$ are displayed in the side-by-side visual comparison. All the other edges having the zero weight, which are classified as the background mask, are removed in the comparative view.

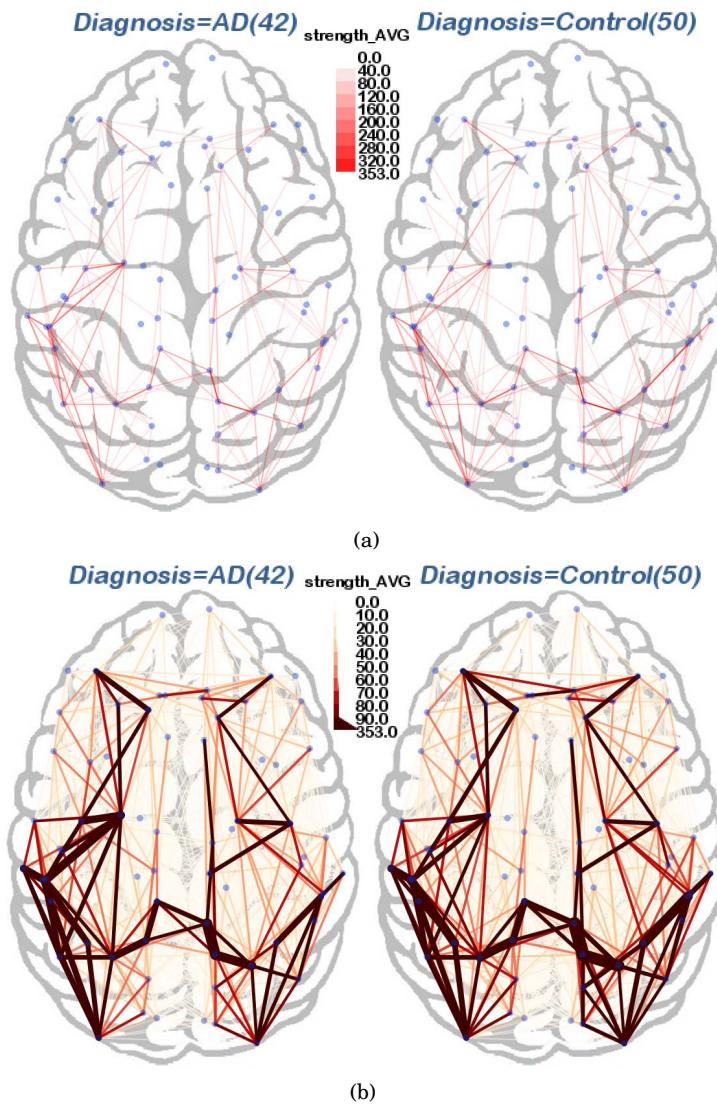


Fig. 4. Alternative designs in mapping the brain connectivity measure to the visual channel: (a) initial design using the linear mapping from the full feature space to the color saturation; (b) optimized design based on the perception theory, applying three new methods: color palette, feature capping, redundant coding.

5. VISUALIZATION

The primary task targeted in this study is the visual comparison among brain networks of different subjects or subject groups. Figure 4 illustrates examples of the design to serve this task on the case of the ADNI data set. Our design juxtaposes the brain networks of AD diagnostic groups in the side-by-side layout for comparison. In each sub-view of the layout, a sketch of the human brain in the axial view is depicted as the background image, having the same aspect ratio with the dimensions of a real human brain. Over the background image, 68~70 light blue nodes are drawn to represent the cortical regions of the human brain by the Desikan-Killiany atlas, with each node placed at the geometric center of a region. Because of the similarity among hu-

man brain networks, the individual/aggregated networks drawn in all sub-views share the same nodal alignment (i.e., the brain region division) and layout, which helps to accelerate the visual comparison task. While the originally measured brain regions are in a three-dimensional space, our visual design supports three alternative projections in the final node layout, i.e., embedding the brain network into YZ/XY/XZ planes, namely the Sagittal/Axial/Coronal views, as shown in Figure 1(d)(e)(f) respectively.

On the other hand, the fiber connections between pairs of regions are drawn as straight lines in the graph. The key design choice is how to visually represent the connectivity measure (aka edge feature) of the individual edge or the edge aggregation to fulfill the network comparison requirement. We propose two separate designs. The initial design follows a traditional linear visual mapping. For example, the average strength of the edge aggregation in a group of brain networks is represented by a line color saturation and/or a line thickness proportional to the value of the strength, as shown in Figure 4(a). More edge statistical measures can be displayed in similar visuals for the comparison, including the number of individual edges, the probability for each edge, and the average, sum, standard deviation, and coefficient of variance of edge strengths on the aggregation. A legend is embedded in the top-left of each sub-view to indicate the edge measure in the display. For example in Figure 4, the legend indicates the color saturation used for different range of average edge strengths and the detailed visual mapping of features. To ensure the visibility of important edges, all the network edges are drawn in the ascending order of the selected edge measure, so that visually stronger edges are shown on top of the weaker ones.

Over the linear visual mapping design on network edges, we introduce an improved nonlinear design as shown in Figure 4(b). According to the experiment results presented later in Section 6.2, this new design can optimize the task performance in comparing the edge features among brain networks. Three guidelines are followed in this nonlinear visual design.

Color Palette. Beyond the continuous mapping from the edge measure (feature value) to the visuals, we apply a discrete visual mapping mechanism. For example, when both the color saturation and the line thickness are selected, a 9-class sequential color palette is designed following the suggestion in ColorBrewer [Harrower and Brewer 2003]. The number of classes of the color palette is determined by the experiment result in Section 6.1 that more than 50% people are sensitive to a visual difference larger than 10.9% of the maximal value in both views. Therefore, the number of classes is selected as $1/0.109 \approx 9$, so that most people will sense the difference between adjacent classes. The edge measures are then fitted into 9 sequential bins by their value. The value in each bin is drawn by the color of the corresponding class in the palette. Using the color palette can effectively enhance stereoscopic depths in the visualization and ease the detection of differences in the visual comparison.

Feature Capping. Through a preliminary analysis of the brain network in our data set (Section 3.2), it was found that only a few edge features have a larger difference between comparative views than the noticeable threshold in human perception (i.e., $> 10.9\%$). We developed the feature capping method to amplify the small features so that the visual differences are stronger and more noticeable in the comparison. For example, in Figure 4(a), the case of ADNI data set comparing 42 AD patients and 50 healthy controls, the maximal edge strength is 353. If we pick a pseudo upper bound of 90 and use the 9-class binning, the visible difference threshold is reduced to $90 \times 10.9\% \approx 10$ from $353 \times 10.9\% \approx 39$. For edges with a strength larger than 90, we use a single capping color to encode the edge measure (dark red by default). Applying this feature capping to compare the aggregated network of AD versus control subject groups, 96 features out of $2,278 \times 2$ edge aggregations are capped and only 6 pairs of the capped features have visible differences larger than the noticeable thresh-

old (10.9% of the maximal edge value). Figure 4(b) presents the design in the resulting 10-color palette with the last class indicating all the feature values above the cap. Comparing Figure 4(b) with Figure 4(a), except for a few individual edges with significant differences found in both figures, we also identified in Figure 4(b) two areas of cortical regions in the middle-left and bottom-right where several edges in each area had noticeable differences between the two sub-views.

Redundant Coding. In the improved design, more than one visual channels on the edge can be selected to display the same edge measure, which is called the redundant coding method. For example, use both color saturation and line thickness to encode the average edge strength. The experiment results in Section 6.1 show that the redundant coding can significantly improve user's performance in the visual comparison. As the feature value is capped in the color coding, we also propose to vary the line thickness to show the difference above the value cap for coloring. To ensure a coherent visual mapping design, we carefully calibrate the slope of the line thickness above the value cap, i.e., to 2.2 times of the slope below the cap, according to the JND profile on the redundant coding and the thickness-only coding ($0.238/0.109 \approx 2.2$). This is illustrated in the legend of Figure 4(b).

6. EVALUATION

6.1. JND theory and perception experiment

As described in previous sections, our visualization and feature selection model design rely on the assumption that human users will only sense relatively large visual differences. In perception theory, this is known as the theory of just-noticeable difference (JND). Quantitatively, JND is defined as the minimal amount of perception magnitude that something must be changed for human to notice the difference. Given a reference stimulus of magnitude I on certain human perception channel, which is the original intensity of the sensible signal, the JND profile, denoted as $JND(I)$, quantifies the minimal magnitude of the increased stimulation $I + JND(I)$, at which just $P\%$ of human users can detect changes from the previous stimulation magnitude. Normally P takes the value of 50, so that a half of the population will sense the change at least as large as $JND(I)$. By Weber's Law, JND profile is approximately proportional to the original intensity in that: $JND(I) = k \cdot I$. The factor k takes a constant value, but varies significantly across different user bases and modalities of human perception (e.g., sound, vision and heaviness).

On data-related perceptions, Chou and Li have studied the pixel-level JND model for understanding image and video [Chou and Li 1995]. In this work, we further extend the JND theory to the perception of node-link graphs. An edge on a subgraph G is said to be (just) noticeable if its visual difference between groups is at least $JND(G)$. To determine the value of $JND(G)$ in our scenario, we conduct a controlled user experiment. Three visual coding methods are used to display the edge features in the visual comparison, including color saturation, line thickness or both. The detail of the experiment is given below.

Design. We recruited 17 subjects for the experiment, 10 were male and 7 were female. All subjects were graduate students in CS or art programs. They were tested before the experiment to avoid the color-blindness. The experiment followed a within-subject design that every subject entered all the tasks (65 in total) and each task was independent of each other. The first 5 tasks were designed for the training purpose and the remaining 60 tasks were the test phase, with 20 tasks testing each of the three visual coding methods: color saturation, line thickness, and the redundant coding in both color and thickness.

xx:18

L. Shi et al.

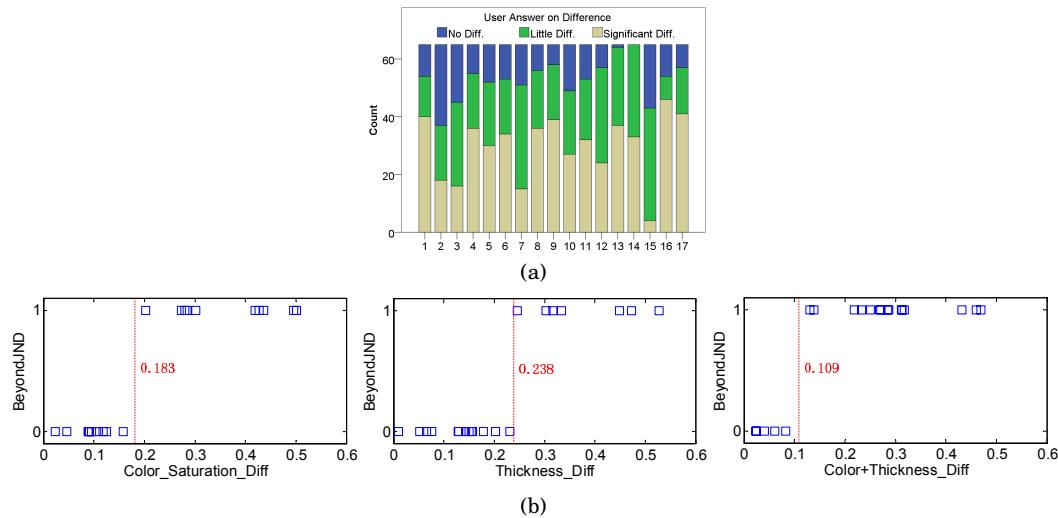


Fig. 5. The user study result to calibrate the subgraph JND model in our scenario: (a) Distribution of user's answers on the visual difference. (b) 50% JND of three visual coding methods: color saturation, line thickness, color+thickness.

Data and task. There were two sub-views to compare in each task of the study: one with the original brain network, and the other with planned differences added on the edges of some subgraph. All original networks were the average brain network of the 113-subject data set from OpenConnectome. The other sub-view with difference was generated as such: first, we randomly picked one nontrivial subgraph from the pre-computed graph clusters; then, each edge in this subgraph was selected with some probability (0.5 by default). All selected edges were increased/decreased in their feature values by a ratio of the original value. The ratio was uniformly controlled between 5% and 100% (20 samples). One of the three visual coding methods was applied to display the difference in each sample, so that we had 60 samples/tasks in total for each subject. We ensured a balanced design so that the full parameter space of each method can be explored. For each task, users were asked to choose from three levels of differences between the two comparative sub-views: 1) no difference; 2) little difference (random noise); and 3) significant difference. We recorded both the user's choice and their completion time.

Results. We collected 1105 user answers in total and analyzed their overall distribution. As shown in Figure 5(a), on average, users indicated 82.7% tasks to have at least little difference and 48.5% with significant difference. In our scenario, users are expected to detect meaningful brain network differences, which is more stringent than the original definition of just-noticeable. Therefore, we decided to use significant v.s. non-significant as our boundary of JND. Also we noticed that one user (#15) had largely skewed answers from the other users, and we excluded his entries from the analysis. For each of the 60 tasks in the testing phase, we checked whether there were at least 50% users answering with the significant difference. This result described whether the task setting was beyond the JND or not. The binary outputs were then used as the outcome in a logistic regression model, and the controlled ratio of difference was used as the input. In Figure 5(b), it is shown that in every visual coding method, the beyond/below JND outcome can be perfectly classified by the ratio of feature difference. A classification boundary of 0.183 for the color saturation coding, 0.238 for the line thickness coding, and 0.109 for the color+line thickness redundant coding were

derived by solving the logistic regression model. These results demonstrate that the color coding has a better capacity than the line thickness coding in visual comparison, while the redundant coding achieves the best performance. Notice that the ratio of difference here is measured as the absolute difference of feature values in comparison divided by the maximal edge feature value in the background subgraph. Therefore, the ratios are not uniformly distributed on [0,1] as we have set them initially.

The experiment result suggests many implications to our proposed methods. First of all, it is shown that the JND notion from the perception theory still holds in the visual comparison of brain networks. Based on this notion, only a small portion of edge features (< 1% in the ADNI data set) between the brain networks of different diagnostic groups are considered to be noticeable by human users. To increase the number of noticeable edge features in the comparison, we have introduced the optimized visualization design with feature capping and redundant coding methods, whose performance is evaluated in the next subsection.

6.2. Design experiment

We conducted a controlled user experiment to compare our visualization and feature selection method with alternative visual designs and feature selection models, by comparing their task performance in the experiment.

Design. We recruited 20 subjects, 12 were male and 8 were female. All subjects were graduate students in CS or art programs. They were tested before the experiment to avoid the color-blindness. The experiment followed a between-subject design and the subjects were randomly divided into two groups for comparison. The subjects in the first group conducted the experiment with the initial linear visualization design (Figure 4(a)) and the subjects in the second group conducted the experiment with the optimized design (Figure 4(b)). Each subject was required to complete 5 tasks. The first task was a sample visual comparison task for training. The other four tasks were the same in data and task design, but applied four different feature selection models: the basic lasso model, the Elastic Net model combining L1 and L2 regularization, the standard sparse group lasso (SGL) model, and the prioritized sparse group lasso model proposed in this work (p-SGL).

Data and task. In this study, the basic task was again to compare two brain network sub-views in the side-by-side visualization, where the network is generated from the 113-subject OpenConnectome data set. The sub-view in the left displayed the average brain network of 60 subjects with high CCI scores (≥ 100), and the sub-view in the right displayed the average brain network of 53 subjects with low CCI scores (< 100). Only edges selected by the current feature selection model were shown in the visual comparison. For each task, the subject was asked to select all the edges that they found to have a significant difference between the sub-views in comparison. A point selection interaction was developed to help users select discriminative edges. All users were instructed to work in a best-effort manner. We recorded all the edges they selected and the time of completion on each task.

Result. The raw experiment result of the user selected edges were processed to compute the visibility of difference measure in the user side, i.e., by the percentage of edges selected by users out of the original model-selected edges. This visibility measure was summarized as the grouped box plot in Figure 6(a). It can be found that the optimized visualization design consistently outperforms the initial design on the visibility of difference, under all feature selection models. The improvements are significant on Elastic Net ($p < 0.001$), SGL ($p < 0.016$), and p-SGL ($p < 0.028$). This result is supported by the JND theory that the optimized design has a smaller JND profile than the initial design, and therefore more differences can be manually discovered. Meanwhile, the p-SGL model achieves the best visibility of difference, which aligns well with

xx:20

L. Shi et al.

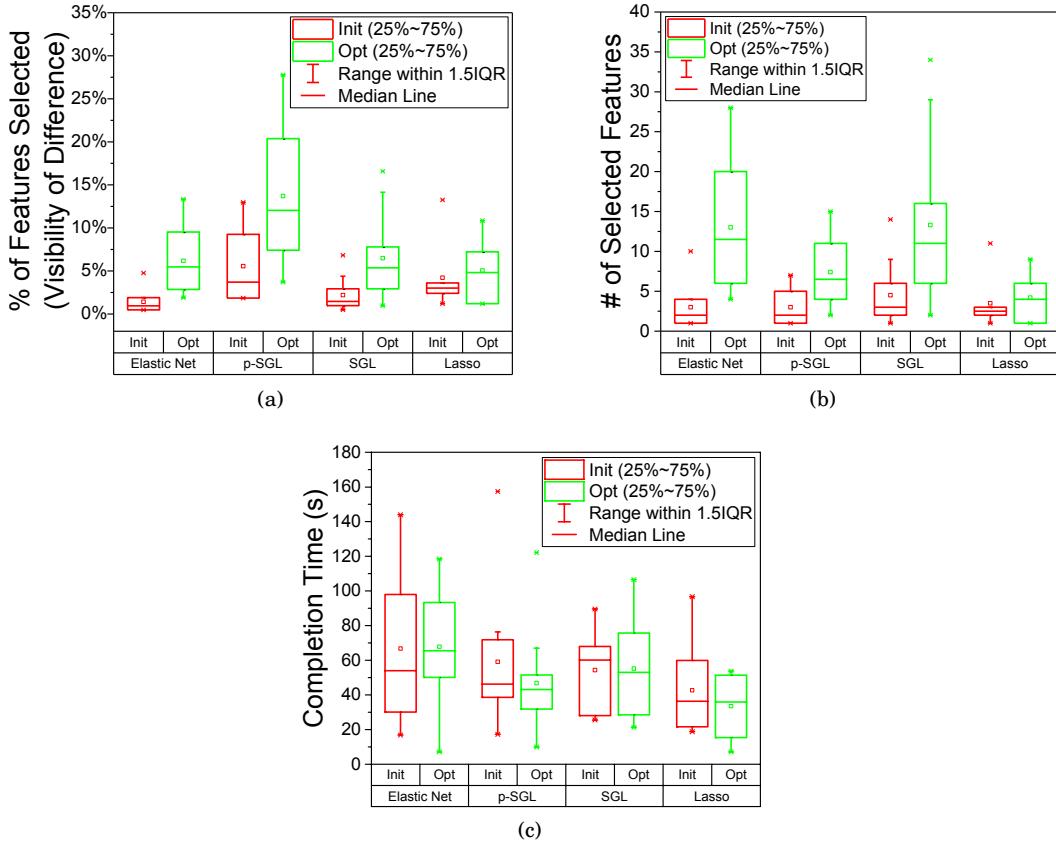


Fig. 6. The user study result comparing the initial visualization design in linear feature coding (Init) with the optimized design applying the nonlinear coding (Opt): (a) Percentage of edge features selected by the user in each experiment, corresponding to the visibility of difference; (b) The actual number of features selected by users; (c) User's completion time.

the proposed objective $D2$ of p-SGL to guarantee the visibility of difference bound. We should note that on the actual number of selected features (Figure 6(b)), p-SGL did not have an advantage. Users selected the highest number of features under Elastic Net and SGL models where their initial number of features in the visual comparison are also the highest. With p-SGL model, the initial model-selected number of features is the smallest, and therefore p-SGL suffers from a highest variance on the visibility of difference. Finally on the completion time measure, as shown in Figure 6(c), the optimized visual design performed comparably as well as the initial design on two models, and had a moderately smaller time cost on the other two models (not significant).

6.3. Quantitative experiment on feature selection models

Besides the user performance in completing the visual comparison task, we also evaluate the effectiveness of feature selection models by their quantitative performance measures.

Data and Task. In this experiment, we focus on the 113-subject OpenConnectome data set. The task is to computationally predict the binary CCI label on the brain network using their fiber connection strengths as edge features. All 2415 edge features are pre-processed before the experiment to remove all trivial brain connectivities for

Table II. The performance of feature selection models in maximizing the prediction accuracy

Measure Model	Accuracy	Visibility	#Feature
No Feature Selection	0.433	0.189	1354
Lasso	0.855	0.232	83
Elastic Net	0.788	0.212	165
SGL	0.847	0.219	118
p-SGL	0.942	0.278	54

Table III. The performance of feature selection models in maximizing the ratio of visible differences (visibility)

Measure Model	Accuracy	Visibility	#Feature
No Feature Selection	0.433	0.189	1354
Lasso	0.855	0.232	83
Elastic Net	0.732	0.251	203
SGL	0.75	0.375	61
p-SGL	0.683	0.528	9

Table IV. The performance of feature selection models on the trade-off of maximizing the prediction accuracy while ensuring a large visibility (> 0.3)

Measure Model	Accuracy	Visibility	#Feature
No Feature Selection	0.433	0.189	1354
Lasso	0.855	0.232	83
Elastic Net	0.732	0.251	203
SGL	0.783	0.322	60
p-SGL	0.863	0.314	71

which the average strength among subjects is smaller than 10 (note that the maximal connectivity strength is 26474 in this data set). This leaves 1354 nontrivial edge features. Over these features, four feature selection models are applied and compared: lasso; Elastic Net; sparse group lasso (SGL); and prioritized SGL (p-SGL) under a ratio of visible difference threshold (ξ) of 0.3. For each model except the basic lasso, we vary the group-wise sparsity parameter α from 0 to 1 to cover a full parameter space. The baseline method is set to the prediction with all edge features, i.e., without feature selection. The underlying prediction algorithm adopts the debiased logistic regression model over the selected edge features.

Two measures are considered in the experiment according to the design objectives defined in Section 4: (1) the prediction *accuracy* of the binary brain network label using the model-selected features ($D1$); (2) the *visibility* of difference in all selected features ($D2$). Note that the prediction accuracy is calculated in a 10-fold cross-validation by a random partition of the data. The visibility of difference is computed by comparing the difference on each selected feature with the JND value determined by the experiment in Section 6.1, i.e., 10.9% of the maximal feature value in the comparison view. The number of selected features (#feature) is also recorded in each experiment setting.

Result. We collect and report experiment results under three different parameter settings. In the first setting, we vary the sparsity parameter α in each model to achieve its best prediction performance. The result is listed in Table II. It is shown that, without the feature selection, the prediction accuracy holds at 0.433, even worse than the null model ($60/113 = 0.531$). With the basic feature selection model of lasso, the accuracy rises to 0.855 using 83 selected features. Applying the Elastic Net and SGL can increase the number of selected features (165 and 118), but the prediction accuracy drops a little to 0.788 and 0.847. The proposed p-SGL model enjoys the best prediction

accuracy of 0.942 because it tends to select the features with large visibility from the SGL model, which helps to reduce the overfitting effect. The visibility measure is also the highest applying the p-SGL model.

In the second parameter setting, we navigate through each model space to maximize the visibility measure. The results on the baseline model (no feature selection) and the lasso model do not change from the first setting because there is no flexibility to tune these models (Table III). Both Elastic Net and SGL models increase the visibility measures to 0.251 and 0.375, while sacrificing a bit in the prediction accuracy. However, it is the proposed p-SGL model that achieves the best visibility of 0.528 by finding an extreme case of only 9 features with large comparative differences. This result validates the design objective of p-SGL to explicitly consider the visibility of feature differences.

In the real usage, we favor the feature selection model that balances between the predictive power and the interpretability (i.e., the visibility of difference in this scenario). The third parameter setting fulfills this requirement by tuning each model to its best prediction accuracy while ensuring a large visibility of difference. In this case, we apply the constraint of $\xi = 0.3$, i.e., expecting a visibility measure larger than 0.3. As shown in Table IV, only SGL and p-SGL can meet this constraint. In comparing the prediction accuracy, p-SGL is preferred ($0.863 > 0.783$). This demonstrates that p-SGL can achieve a better trade-off in terms of the prediction accuracy and the visibility of feature differences.

In summary, the computational feature selection method in complement to the visual comparison design is demonstrated to be effective in both the predictive performance and the user performance in completing the comparison task. By applying a basic Lasso model, the prediction accuracy on the CCI label of the OpenConnectome data set increases to two times of that without the feature selection method. This is because of the significant overfitting effect when the number of raw edge features is much larger than the number of brain network samples. In the user experiment, we are not able to directly compare the performance with and without the feature selection method. With the original brain network visualization, users can not adequately perform the point selection of edge features due to the huge density of brain connectivities. Preliminary implications can be made by comparing the model that selects more edges (Elastic Net, 165 features, 78.8% prediction accuracy) with the model that selects much less edges (p-SGL, 54 features, 94.2% prediction accuracy): with the feature selection model (or the model selecting less but informative edges), a larger percentage of selected edges will be considered as useful by users, and the user's effort in terms of the task completion time is generally smaller. Among the feature selection models, the proposed p-SGL achieves the best trade-off in the prediction accuracy, user utility and time cost on the comparison task. This validates the design rationale of our feature selection model to balance the prediction performance and the significance of visual differences. The purely predictive model, such as Lasso, suffers from the weak user performance; while the statistical hypothesis testing only considering the (visual) significance of individual features leads to a poor prediction performance.

6.4. Case study on the ADNI data set

Following the quantitative experiments, we applied the proposed visualization and computational method in two real-world cases and studied their qualitative user performance, including the findings on the comparison task, the discoveries in the explorative analysis and the user experience and feedback.

The first case focuses on the ADNI data set relating to Alzheimer's Disease. In total, the brain network of 202 subjects were measured and visualized - 50 healthy controls, 72 with early mild cognitive impairment (eMCI), 38 with late mild cognitive impairment (LMCI) and 42 AD patients, see Section 3.2 for more details. We invited one neu-

Visual Analysis of Brain Networks using Sparse Regression Models

xx:23

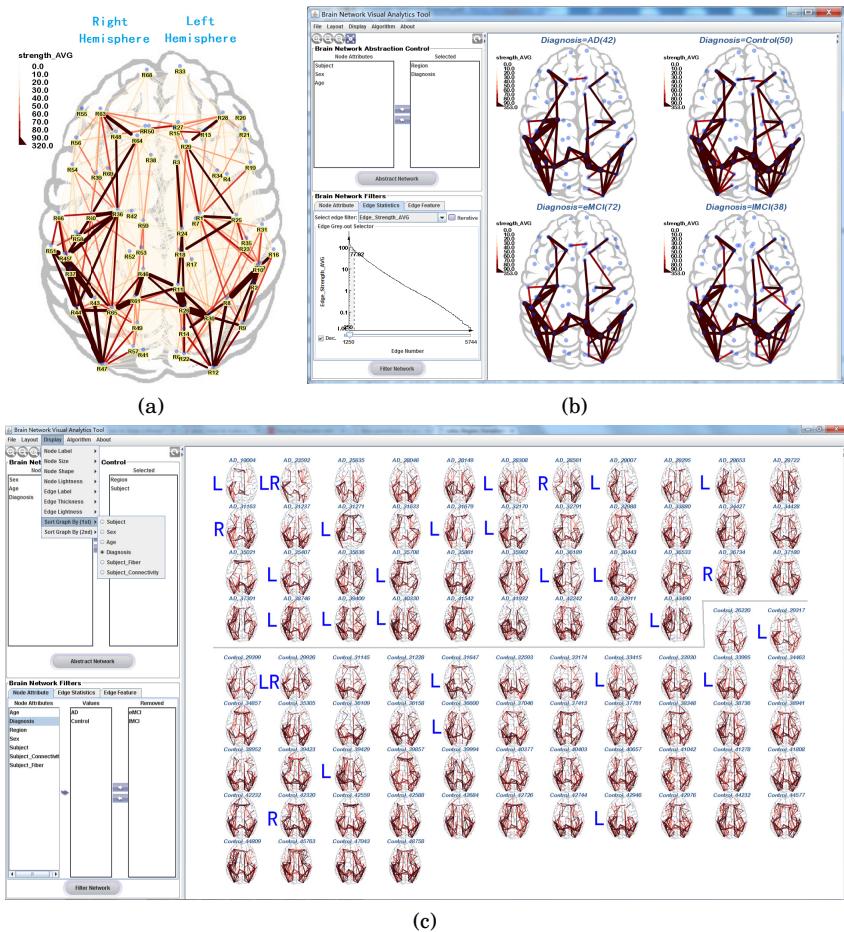


Fig. 7. The asymmetry on AD-related brain networks shown by: (a) the aggregation view of all 202 subjects; (b) the comparison view by diagnostic groups. Only the top connectivity edges are shown; (c) the SMD view listing the brain network of AD patients and normal controls.

roimaging scientist to analyze this brain network data set using our visualization tool. The primary goal was to verify the previous neuroscience findings on AD patients. For example, it was reported that the network asymmetries between the left and the right brain hemispheres of the AD patients were higher than that of the healthy control [Thompson et al. 2001; Daianu et al. 2013].

The scientist initiated the analysis with our tool from the default aggregation view showing the average brain network of 202 subjects (Figure 7(a)). The visualization applied the color+thickness redundant coding design. In this aggregation view, it is noticeable that the left side of the brain network (i.e., the right hemisphere) and the right side (i.e., the left hemisphere) are clearly asymmetric. The right hemisphere has thicker connections than the left hemisphere, which corresponds well to the findings in [Thompson et al. 2001]. In a more recent work [Daianu et al. 2013], researchers have followed up to study the differences in network connectivity patterns between AD patients and the MCI subtypes. Similarly in our tool, the scientist repeated the analysis by splitting all brain networks by the AD diagnostic group (Figure 7(b)). The most significant findings were revealed when the edge features were filtered by the connec-

xx:24

L. Shi et al.

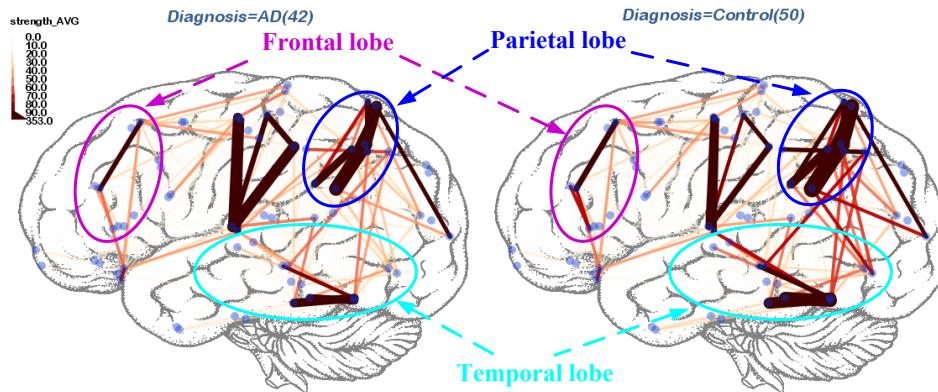


Fig. 8. The sagittal view displaying the same edge features selected by our p-SGL model. Key brain regions relevant to the outcome of AD pathology are annotated.

tion strength to retain the top 250 most interconnected edges. The control group has, on average, thicker edges (i.e., stronger connections) in the left hemisphere than AD and MCI patients. To further understand the distribution of this bilateral asymmetry, the scientist rolled back to the SMD view showing a list of thumbnail displays for AD patients and controls, after filtering out the intermediate MCI patients. As shown in Figure 7(c), the brain networks are sorted first by diagnosis (i.e., AD or control) and then by the overall connectivity strength. It can be found that more brain networks are asymmetrical among AD patients (the upper group) than normal controls (the lower group). We asked a human observer to annotate all the networks in Figure 7(c) that can be visually distinguished as asymmetrical, with a “L” when the left side of the brain has significantly stronger connection than the right side in some part, a “R” for the opposite case, and a “LR” when both sides have stronger connections in some part than the other side. Clearly, the “L” asymmetry dominates both AD and control groups, indicating the destruction of the left hemisphere connectivity. The AD group has two times the “L” asymmetry (16) compared with the control group (8). Moreover, the “L” asymmetry happens relatively irrelevant of the overall connectivity among AD patients, while for the control group, it happens more on the low overall connectivity people, possibly linked to other connectome deficiencies than AD.

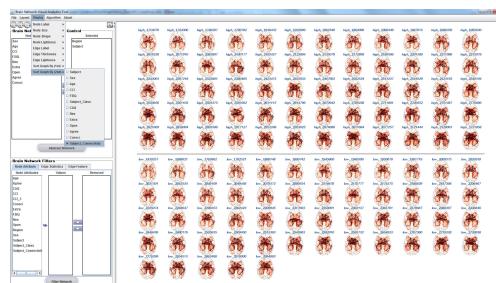
To drill down to the specific brain regions and connectivities leading to the asymmetry, the scientist applied the edge feature selection method in our tool. There were many possible models from the drop-down list of the feature selection tab. The scientist tried with several models and compared them on the model performance, which is displayed as curves on the feature selection tab. Finally the model of SGL-ClusterByLineJNDMulti was selected, with an alpha sparsity of 0.8 in the highest prediction accuracy setting (60.9%). The resulting visualization in Figure 3(c) demonstrated several key regions in the left hemisphere where the AD-afflicted groups have connection breakdown compared with the healthy control group. Notably, R10 (inferior temporal) and R28 (rostral middle frontal) were detected, which fits well the report in [Loewenstein et al. 1989] on the left hemisphere hypometabolism in the frontal, temporal, parietal lobes of AD patients. These findings were more salient in the sagittal view of Figure 8 where relevant lobes were annotated.

6.5. Case study on the OpenConnectom data set

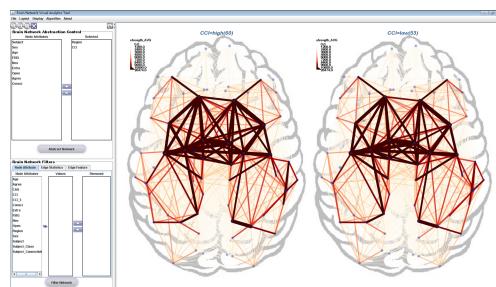
In the second case study, we worked on the 113-subject data set from the OpenConnectome project, as described in Section 3.2. The brain network data model was similar to

Visual Analysis of Brain Networks using Sparse Regression Models

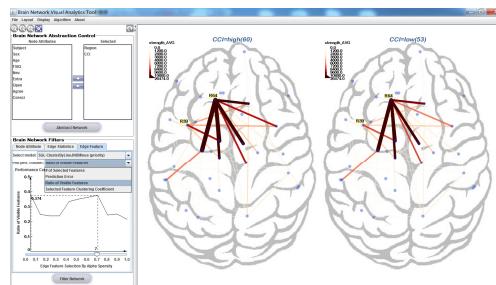
xx:25



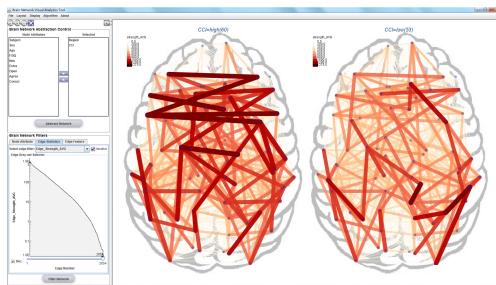
(a)



(b)



(c)



(d)

Fig. 9. The visual comparison on the OpenConnectome data set: (a) the SMD view listing all brain networks by the CCI class and the overall connectivity; (b) comparing between the average brain network of two CCI groups; (c) after applying the p-SGL feature selection model, interesting patterns are located which are in line with findings in the neuroscience literature; (d) on weak edge features - the high CCI group has stronger connections than the low CCI group.

that of the ADNI data set, which consisted of 70 ROIs as nodes, and fiber connections between ROIs as edges. Due to the different measurement methods, the distribution of edge strengths can vary greatly from that of the ADNI data set. We deployed our visualization tool to compare the brain network of the group of 60 high-CCI subjects with that of 53 low-CCI subjects.

In this experiment, we interviewed a senior doctor in the neurology department of a prestigious hospital, who had decades of experience in the field of neuromedicine. Because the doctor was not familiar with the visualization interface, we operated the tool ourselves during the study while receiving verbal feedbacks from the doctor. We started from the SMD view listing the brain network of all 113 subjects as thumbnails, sorted by the CCI class (i.e., high and low) and the overall connectivity strength, as shown in Figure 9(a). The doctor derived a rough pattern that the brain networks of low-CCI subjects tend to have weaker overall brain connectivity than high-CCI subjects. To confirm this finding, we proceeded to the comparison view by the CCI class (Figure 9(b)), however, the doctor found it hard to visually identify differences from the side-by-side comparison.

By integrating computational feature selection models, we drilled down to the fine-grained edge feature comparison view. After navigating among models by comparing their prediction performance, we found that almost all models achieved a better prediction accuracy with fewer edge features selected. The basic lasso model obtained a good prediction accuracy (0.855) with only 83 features, but their selected edge features were scattered out in the graph and there was no interpretable pattern to discover. We then switched to the proposed model optimized with the perception theory (p-SGL), and picked the model parameter to strike a balance between the prediction accuracy and the ratio of visible features (the percentage of features with the visual difference larger than the JND profile). Using this model, noticeable comparative patterns were found, as shown in Figure 9(c). It was inferred that the connections at region #64 (rh-superior frontal) and #39 (rh-caudal middle frontal) were important for the CCI difference among populations. In fact, there is a strong tie between our result and the findings from the neuroscience literature. The superior frontal region is involved in self-awareness [Goldberg et al. 2006], while it has been hypothesized that self-awareness strongly influences human creativity [Silvia and Phillips 2004]. At a higher level, both region #64 and #39 are in the right hemisphere, and in Figure 9(c), the high CCI group have stronger connections than the low CCI group between #64, #39, and several regions in the left hemisphere. It is well-known that the right hemisphere of the brain is in charge of creativity and the left brain focuses on logic, among other functions. This showed that the coordination between the left and right brain can also be important to people's creativity level.

In another analysis scenario, we applied the edge aggregation filters to only display weak edges that have smaller average connection strengths (<1000). The resulting comparison view for different CCI groups revealed a clear pattern, as shown in Figure 9(d). For weak edges, the high CCI group had, on average, stronger connections than those in the low CCI group. On the other hand, we did not detect similar patterns when selecting only strong edge features.

6.6. Expert feedback

Besides controlled experiments, we also deployed the brain network visual analytics tool in the same hospital where we conducted the second case study. Three neurology doctors working closely with MRI/DTI scans and neurological patients were invited to a pilot trial of our tool. After the training and usage sessions, they were asked to give feedbacks on: (1) the functionality/task of aggregating brain networks by subject's labels and the resulting comparison among groups; (2) the functionality/task of

machine-learning based feature selection for an improved visual comparison; (3) any other brain network related tasks that can be supported by the tool in future.

On the first task, all of them were positive to the value of the aggregation view because of its simplicity and intuitiveness in the interaction and visualization. Some representative comments and suggestions include: “Most tools I have used present comparison views between the patient group and the control group. This new tool is novel in that it can not only support such standard comparison, but also allows to compare the networks by any other attributes in the data. This can be helpful for both the neurological disease study and the brain network study of community population”, “Beyond the current comparison on categorical attributes, it will be more helpful if the ordinal attributes can also be supported (e.g., aging), such that the trend in the brain network changes can be detected”.

On the second task, they found the feature selection to be helpful in revealing the significant difference on brain network structure, but they also suggested several improvements. First, after the visualization of network differences, users may prefer to export quantitative findings or analysis reports with the tool. These quantitative indicators (e.g., p-value) can also be displayed in the comparison view. Second, from the clinical point of view, not all ROIs are “hot spot” for specific diseases. These domain knowledge on “hot spot” can be incorporated to improve the feature selection method in the clinical setting.

Finally, the doctors mentioned many other network-related tasks that can be supported by the visual analytics method, mostly from a clinical viewpoint. First, the current subject labels, such as the diagnosis class of AD and control, are identified in a coarse-grained manner. In fact, there are many fine-grained clinical symptoms and investigation results (e.g., cognitive and motor function evaluation) that are linked to the change in brain images. The correlation analysis and visualization of these fine-grained evidences can be more valuable in the clinical study. Second, beyond the group-level comparison, it may also be helpful to visually compare the individual patient’s network with the historical patient/control group, especially during their recovery process from neurological diseases.

7. CONCLUSION

This paper proposes an integrated framework to visually analyze huge amount of *in vivo* human brain networks. In our framework, brain networks can be aggregated, filtered, split, and compared by their subject-level labels (e.g., demographics, diagnostic groups), and statistical measures on cortical regions and fiber connections (e.g., the average fiber strength). To optimize the performance of these visual analytics tasks, we have introduced both elaborate visual comparison designs and the sparse regression models for the discriminative feature selection. Controlled user experiments were conducted to calibrate the visual design and the computational model according to the established perception theory for visual comparison. In the evaluation, we have studied the usefulness of our framework and methods through quantitative experiments and two real-world cases comparing between brain networks of healthy population and the patients affected by the Alzheimer’s disease; and between ordinary people with different creativity levels. The findings discovered with our tool implementing the proposed methods correspond well with the validated results on the neuroscience literature, which demonstrates the success of our approach.

Finally, the current work still has several limitations which we plan to improve in future. First, our comparison design follows the standard side-by-side visualization, while the single view comparison method has not been studied in this scenario, such as the overlaid visualization and the explicit visual coding. Second, we have stayed with the node-link representation of brain networks in this work, while it has been shown

in the previous study that the matrix representation can enjoy more advantages in comparing brain networks [Alper et al. 2013]. It will be interesting to further compare these two visual designs under the real-world brain network analysis scenario and data set. Third, most of this work has focused on the coarse-grained region-level brain connectivity. When we drill down to the fine-grained brain networks, e.g., voxel-level or even neuron-level, how to deal with the scalability issue in both the visualization and the computational model design remains an open challenge.

ACKNOWLEDGMENTS

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. This work is also supported by China National 973 project 2014CB340301, NSFC No. 61379088, 61422212, DTRA under the grant number HDTRA1-16-0017, Army Research Office under the contract number W911NF-16-1-0168, NIH under the grant number R01LM011986 and a Baidu gift.

REFERENCES

2012. FreeSurfer. <http://surfer.nmr.mgh.harvard.edu/>. (2012).
- Basak Alper, Benjamin Bach, Nathalie Henry Riche, Tobias Isenberg, and Jean-Daniel Fekete. 2013. Weighted graph comparison techniques for brain connectivity analysis. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*. 483–492.
- B Bach, Riche N Henry, T Dwyer, T Madhyastha, J-D Fekete, and T. Grabowski. 2015. Small MultiPiles: Piling time to explore temporal patterns in dynamic networks. *Computer Graphics Forum* 34, 3 (2015), 31–40.
- B. Bach, C. Shi, N. Heulot, T. Madhyastha, T. Grabowski, and P. Dragicevic. 2016. Time curves: Folding time to visualize patterns of temporal evolution in data. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2016), 559–568.
- Giuseppe Di Battista, Peter Eades, Roberto Tamassia, and Ioannis G. Tollis. 1998. *Graph Drawing: Algorithms for the Visualization of Graphs*. Prentice Hall PTR.
- Anastasia Bezerianos, Fanny Chevalier, Pierre Dragicevic, Niklas Elmquist, and Jean-Daniel Fekete. 2010. GraphDice: A System for Exploring Multivariate Social Networks. *Computer Graphics Forum* 29, 3 (2010), 863–872.
- Bokai Cao, Xiangnan Kong, Jingyuan Zhang, Philip S. Yu, and Ann B. Ragin. 2015. Mining Brain Networks using Multiple Side Views for Neurological Disorder Identification. In *Proceedings of IEEE International Conference on Data Mining (ICDM)*. 709–714.
- Nan Cao, Jimeng Sun, Yu-Ru Lin, David Gotz, Shixia Liu, and Huamin Qu. 2010. FacetAtlas: Multifaceted Visualization for Rich Text Corpora. *IEEE Transactions on Visualization and Computer Graphics* 16, 6 (2010), 1172–1181.
- Duen Horng Chau, Aniket Kittur, Jason I Hong, and Christos Faloutsos. 2011. Apolo: making sense of large network data by combining rich user interaction and machine learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 167–176.

Visual Analysis of Brain Networks using Sparse Regression Models

xx:29

- Xi Chen, Seyoung Kim, Qihang Lin, Jaime G Carbonell, and Eric P Xing. 2010. Graph-structured multi-task regression and an efficient optimization method for general fused Lasso. *arXiv preprint:1005.3579* (2010).
- Haibin Cheng, Haifeng Chen, Guofei Jiang, and Kenji Yoshihira. 2007. Nonlinear feature selection by relevance feature vector machine. In *Machine learning and data mining in pattern recognition*. Springer, 144–159.
- Chun-Hsien Chou and Yun-Chin Li. 1995. A perceptually tuned subband image coder based on the measure of just-noticeable- distortion profile. *IEEE Transactions on Circuits and Systems for Video Technology* 5, 6 (1995), 467–476.
- Nicolas A Crossley, Andrea Mechelli, Jessica Scott, Francesco Carletti, Peter T Fox, Philip McGuire, and Edward T Bullmore. 2014. The hubs of the human connectome are generally implicated in the anatomy of brain disorders. *Brain* 137, 8 (2014), 2382–2395.
- Madelaine Daianu, Neda Jahanshad, Talia M Nir, Clifford R Jack, Michael W Weiner, Matt A Bernstein, and Paul M Thompson. 2015. Rich club analysis in the Alzheimer’s disease connectome reveals a relatively undisturbed structural core network. *Human brain mapping* 36, 8 (2015), 3087–3103.
- Madelaine Daianu, Neda Jahanshad, Talia M Nir, Arthur W Toga, Clifford R Jack Jr, Michael W Weiner, and Paul M Thompson, for the Alzheimer’s Disease Neuroimaging Initiative. 2013. Breakdown of brain connectivity between normal aging and Alzheimer’s disease: a structural k-core network analysis. *Brain connectivity* 3, 4 (2013), 407–422.
- Manoranjan Dash and Poon Wei Koot. 2009. Feature selection for clustering. In *Encyclopedia of database systems*. Springer, 1119–1125.
- Ian Davidson, Sean Gilpin, Owen Carmichael, and Peter Walker. 2013. Network discovery via constrained tensor analysis of fMRI data. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 194–202.
- Rahul S Desikan, Florent Ségonne, Bruce Fischl, Brian T Quinn, Bradford C Dickerson, Deborah Blacker, Randy L Buckner, Anders M Dale, R Paul Maguire, Bradley T Hyman, and others. 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 31, 3 (2006), 968–980.
- Jennifer G Dy and Carla E Brodley. 2000. Feature subset selection and order identification for unsupervised learning. In *Proceedings of the International Conference on Machine Learning*. 247–254.
- Manuel Freire, Catherine Plaisant, Ben Shneiderman, and Jen Golbeck. 2010. ManyNets: an interface for multiple network analysis and visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 213–222.
- Michael Gleicher, Danielle Albers, Rick Walker, Ilir Jusufi, Charles D Hansen, and Jonathan C Roberts. 2011. Visual comparison for information visualization. *Information Visualization* 10, 4 (2011), 289–309.
- Ilan I Goldberg, Michal Harel, and Rafael Malach. 2006. When the brain loses its self: prefrontal inactivation during sensorimotor processing. *Neuron* 50, 2 (2006), 329–339.
- William R Gray, John A Bogovic, Joshua T Vogelstein, Bennett A Landman, Jerry L Prince, and RJacob Vogelstein. 2012. Magnetic resonance connectome automated pipeline: an overview. *IEEE Pulse* 3, 2 (2012), 42–48.
- David Hallac, Jure Leskovec, and Stephen Boyd. 2015. Network lasso: Clustering and optimization in large graphs. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 387–396.
- Mark Harrower and Cynthia A Brewer. 2003. ColorBrewer.org: an online tool for selecting colour schemes for maps. *The Cartographic Journal* 40, 1 (2003), 27–37.
- Jingrui He and Jaime Carbonell. 2010. Coselection of features and instances for unsupervised rare category analysis. *Statistical Analysis and Data Mining: The ASA Data Science Journal* 3, 6 (2010), 417–430.
- Xiaofei He, Deng Cai, and Partha Niyogi. 2005. Laplacian score for feature selection. In *Advances in neural information processing systems*. 507–514.
- Ivan Herman, Guy Melancon, and M. Scott Marshall. 2000. Graph Visualization and Navigation in Information Visualization: A Survey. *IEEE Transactions on Visualization and Computer Graphics* 6, 1 (2000), 24–43.
- Shuai Huang, Jing Li, Jieping Ye, Adam Fleisher, Kewei Chen, Teresa Wu, and Eric Reiman. 2011. Brain effective connectivity modeling for Alzheimer’s disease by sparse Gaussian Bayesian network. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 931–939.
- Scott A Huettel, Allen W Song, and Gregory McCarthy. 2008. *Functional magnetic resonance imaging* (2 ed.). Sinauer Associates.

xx:30

L. Shi et al.

- Yan Jin, Yonggang Shi, Liang Zhan, Boris A Gutman, Greig I de Zubicaray, Katie L McMahon, Margaret J Wright, Arthur W Toga, and Paul M Thompson. 2014. Automatic clustering of white matter fibers in brain diffusion MRI with an application to genetics. *NeuroImage* 100 (2014), 75–90.
- D.K. Jones. 2008. Tractography Gone Wild: Probabilistic Fibre Tracking Using the Wild Bootstrap With Diffusion Tensor MRI. *IEEE Transactions on Medical Imaging* 27, 9 (2008), 1268–1274.
- Andreas Kerren, Helen Purchase, and Matthew O. Ward. 2014. *Multivariate Network Visualization (Proc. Dagstuhl Seminar 13201)*. Springer.
- YeongSeog Kim, W Nick Street, and Filippo Menczer. 2000. Feature selection in unsupervised learning via evolutionary search. In *Proceedings of the ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 365–369.
- Xiangnan Kong and Philip S Yu. 2014. Brain network analysis: a data mining perspective. *ACM SIGKDD Explorations Newsletter* 15, 2 (2014), 30–38.
- Xiangnan Kong, Philip S Yu, Xue Wang, and Ann B Ragin. 2013. Discriminative feature selection for uncertain graph classification. In *Proceedings of the SIAM International Conference on Data Mining*. 82–93.
- Josua Krause, Adam Perer, and Enrico Bertini. 2014. INFUSE: Interactive Feature Selection for Predictive Modeling of High Dimensional Data. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 1614–1623.
- Martin H Law, Anil K Jain, and Mário Figueiredo. 2002. Feature selection in mixture-based clustering. In *Advances in Neural Information Processing Systems*. 625–632.
- Jun Liu and Jieping Ye. 2010. Moreau-Yosida regularization for grouped tree structure learning. In *Advances in Neural Information Processing Systems*. 1459–1467.
- David A Loewenstein, Warren W Barker, Jen-Yueh Chang, Anthony Apicella, Fumihito Yoshii, Paresh Kothari, Bonnie Levin, and Ranjan Duara. 1989. Predominant left hemisphere metabolic dysfunction in dementia. *Archives of neurology* 46, 2 (1989), 146–152.
- Klaus-Dietmar Merboldt, Wolfgang Hanicke, and Jens Frahm. 1985. Self-diffusion NMR imaging using stimulated echoes. *Journal of Magnetic Resonance* 64, 3 (1985), 479–486.
- Pabitra Mitra, CA Murthy, and Sankar K. Pal. 2002. Unsupervised feature selection using feature similarity. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24, 3 (2002), 301–312.
- Zeqian Shen, Kwan-Liu Ma, and Tina Eliassi-Rad. 2006. Visual Analysis of Large Heterogeneous Social Networks by Semantic and Structural Abstraction. *IEEE Transactions on Visualization and Computer Graphics* 12, 6 (2006), 1427–1439.
- Lei Shi, Hanghang Tong, and Xinzheng Mu. 2015. BrainQuest: Perception-Guided Brain Network Comparison. In *Proceedings of IEEE International Conference on Data Mining (ICDM)*. 379–388.
- Ben Shneiderman and Aleks Aris. 2006. Network Visualization by Semantic Substrates. *IEEE Transactions on Visualization and Computer Graphics* 12, 5 (2006), 733–740.
- Paul J Silvia and Ann G Phillips. 2004. Self-awareness, self-evaluation, and creativity. *Personality and Social Psychology Bulletin* 30, 8 (2004), 1009–1017.
- Noah Simon, Jerome Friedman, Trevor Hastie, and Robert Tibshirani. 2013. A sparse-group lasso. *Journal of Computational and Graphical Statistics* 22, 2 (2013), 231–245.
- Liang Sun, Rinkal Patel, Jun Liu, Kewei Chen, Teresa Wu, Jing Li, Eric Reiman, and Jieping Ye. 2009. Mining brain region connectivity for alzheimer's disease study via sparse inverse covariance estimation. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 1335–1344.
- Jiliang Tang, Salem Aleyani, and Huan Liu. 2014. Feature selection for classification: A review. *Data Classification: Algorithms and Applications*. Editor: Charu Aggarwal, CRC Press In Chapman & Hall/CRC Data Mining and Knowledge Discovery Series (2014).
- Paul M Thompson, Michael S Mega, Roger P Woods, Chris I Zoumalan, Chris J Lindshield, Rebecca E Blanton, Jacob Moussai, Colin J Holmes, Jeffrey L Cummings, and Arthur W Toga. 2001. Cortical change in Alzheimer's disease detected with a disease-specific population-based brain atlas. *Cerebral Cortex* 11, 1 (2001), 1–16.
- Robert Tibshirani. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B* (1996), 267–288.
- R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight. 2005. Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67, 1 (2005), 91–108.
- Martin Wattenberg. 2006. Visual exploration of multivariate graphs. In *SIGCHI conference on Human Factors in computing systems*. 811–819.
- Jieping Ye, Kewei Chen, Teresa Wu, Jing Li, Zheng Zhao, Rinkal Patel, Min Bae, Ravi Janardan, Huan Liu, Gene Alexander, and others. 2008. Heterogeneous data fusion for alzheimer's disease study. In

Visual Analysis of Brain Networks using Sparse Regression Models

xx:31

- Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* 1025–1033.
- Tong Zhang. 2011. Adaptive forward-backward greedy algorithm for learning sparse representations. *IEEE Transactions on Information Theory* 57, 7 (2011), 4689–4708.
- Zheng Zhao and Huan Liu. 2007. Spectral feature selection for supervised and unsupervised learning. In *Proceedings of the international conference on Machine learning*. 1151–1157.
- Hui Zou and Trevor Hastie. 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67, 2 (2005), 301–320.
- Zhaonian Zou, Hong Gao, and Jianzhong Li. 2010. Discovering frequent subgraphs over uncertain graph databases under probabilistic semantics. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 633–642.