A Visual Analytics System for Brain Functional Connectivity Comparison across Individuals, Groups, and Time Points

Takanori Fujiwara*   Jia-Kai Chou†   Andrew M. McCullough‡   Charan Ranganath§   Kwan-Liu Ma¶

University of California, Davis

ABSTRACT

Neuroscientists study brain functional connectivity in order to obtain a deeper understanding of how the brain functions. Current studies are mainly based on analyzing the averaged brain connectivity of a group (or groups) due to the high complexity of the collected data in terms of dimensionality, variability, and volume. While it is more desirable for the researchers to explore the potential variability between individual subjects or groups, a data analysis solution meeting this need is absent. In this paper, we present the design and capabilities of such a visual analytics system, which enables neuroscientists to visually compare the differences of brain networks between individual subjects as well as group averages, to explore a large dataset and examine sub-groups of participants that may not have been expected a priori to be of interest, to review detailed information as needed, and to manipulate the data and views to fit their analytical needs with easy interactions. We demonstrate the utility and strengths of this system with case studies using a representative functional connectivity dataset.

Index Terms: I.3.8 [Computer Graphics]: Applications;

1 INTRODUCTION

In the field of cognitive neuroscience, resting-state functional connectivity (RSFC) data are commonly adopted to draw inferences about what type of cognitive processing is occurring as a result of some recent experimental manipulation. RSFC data are obtained by continuously acquiring blood-oxygen level dependent (BOLD) images using functional magnetic resonance imaging (fMRI) technology while the subjects are instructed to do nothing except lie in the machine with their eyes closed. Traditionally BOLD fMRI images are used to infer levels of neural activity in specific regions of the brain during a controlled task, but these levels of neural activity can also be used to measure the connectivity of different brain regions during rest. Understanding what the brain is doing during rest can re-shape our understanding of how the brain functions.

Studying RSFC of the brain helps neuroscientists learn how activity in the brain results in complex cognition and behavior. In general, neuroscience researchers design experimental conditions to examine how the different conditions or tasks would influence RSFC at a group-level (i.e., RSFC or other network connectivity measures averaged across subjects within each group).

In a typical RSFC data analysis conducted by neuroscientists, a set of commonly used statistical models, which are usually chosen *a priori*, is applied to find correlation between different aspects of the data. Due to the high complexity and dimensionality of the data, correlation found at the group-level may not necessarily reflect

---
*e-mail:tfujiwara@ucdavis.edu
†e-mail:jkchou@ucdavis.edu
‡e-mail:amccullough@ucdavis.edu
§e-mail:cranganath@ucdavis.edu
¶e-mail:ma@cs.ucdavis.edu

patterns at the individual-level. Research [12] has revealed that there exists immense variation in RSFC across individuals (and even within individuals, across different scan sessions). It is oftentimes an interest of researchers' to examine exactly what factor(s) alter RSFC.

As a result, it is essential that neuroscientists have the tools available to conduct a detailed and flexible examination of RSFC datasets, including examination of data at the level of the individual. The major requirement of such a tool is to let users quickly identify the desired subsets of a large dataset, examine the average RSFC from that subset, and compare the RSFC data from any individual to the average RSFC of any desired sub-group. However, an effective tool that serves such purposes in the field is still lacking.

In this paper, we present the design of an interactive visual analytics system which aims to address the issues mentioned above. Case studies were conducted to demonstrate the utility of different functions of the system. We summarize the contributions of the system in terms of how it can advance the way neuroscientists perform data analysis and visualization in brain research as follows:

- The system enables comparison of patterns of brain region activity correlations between a large collection of subjects and between multiple scan sessions. In addition, the ability to examine large RSFC datasets at multiple levels of analysis (i.e., individuals, small or large groups) is available.

- The system integrates necessary visualization and analytical methods to help users explore and find differences/similarities of data based on information obtained from multiple aspects.

- The system provides a fully interactive interface to fulfill the users' often-changing analysis needs, such as flexibly incorporating or excluding additional factors from the analysis.

2 RELATED WORK

In the literature of brain functional connectivity, there are two general directions of research. One is focused more on comparing the differences of brain networks between individuals [11]. Finn et al. described how the functional connectivity is unique by individual and can be used to distinguish an individual subject from others [12]. Geerligs et al. compared the similarity of the subjects' correlation matrices and claimed that age could be a factor that introduces differences of functional connectivity between groups of individuals [13]. Arian et al. evaluated the effect of acquisition parameters of fMRI and analysis methods on individual differences by analyzing similarities between each subject's correlation matrix [1].

The other direction is to analyze the time varying aspect of the brain functional connectivity. Allen et al. addressed how this aspect can unveil the flexibility in the functional coordination between different neural systems [2]. Chalhoun et al. used the term of "chronnectome" to indicate the time varying connectivity of a brain and reviewed several multivariate approaches for their effects on characterizing brain function [9]. Hutchison et al. suggested that the dynamic functional connectivity metrics may change in macroscopic neural activity patterns underlying critical aspects of cognition and behavior [20]. Deco et al. analyzed the variations in RSFC across

time [10]. They found that even at rest, the dynamic capabilities of the brain are reflected in changes in connectivity between regions.

Given the large complexity and dimensionality of brain network data, effective (visual) analytics tools and methods are essential to provide scientists the ability to explore different aspects of their data and to formulate new hypotheses that might lead to future research possibilities. Margulies et al. [24] and Pfister et al. [26] provided an overview of various visualization methods that have evolved for anatomical and functional connectivity data.

Several tools have been developed to analyze and visualize the functional connectivity data. *Graph-Analysis Toolbox* (GAT) provides graph analysis and visualization tools for structural and functional brain networks [19]. LaPlante et al. developed *Connectome Visualization Utility* (CVU) that can visualize the connectivity data as a matrix and a node-link diagram on a circular layout or an anatomical brain image [22]. Hassan et al. created a software called *EEGNET* which can be used to analyze and visualize EEG and MEG recordings in both 2D and 3D network views [16]. *eConnectome* also provides the similar methods for EEG and MEG recordings [17]. Xia et. al. built *BrainNet Viewer* [32] which can display the 3D brain network with some user-defined properties by coloring the brain surface or nodes and links. Kruschwitz et al. presented *GraphVar* [21], which integrates multiple softwares including *BrainNet Viewer* [32] and the *Brain Connectivity Toolbox* [27], with a user-friendly GUI. In *Connectome Viewer Toolkit*, Gerhard et al. proposed the *Connectome File Format* which can be used to share the structure and meta data of the connectome, so that the data can be easily handled and visualized by their Python script editor [14]. The tools mentioned above are mainly designed to support statistical analysis of functional brain connectivity data, or to visualize data from one subject or from a pre-determined group of subjects, instead of facilitating the task of examining a large collection of functional connectivity datasets in an efficient and user-friendly way. Moreover, the above tools are generally not capable of easily incorporating additional information about the subjects or datasets, such as participant variables (e.g., age), behavioral measures (e.g., task performance measures), or physiological variables (e.g., hormone levels).

Previous work has attempted to use visualization methods to compare the brain functional connectivity between subjects. Alper et al. assessed the usability of their proposed augmented adjacency matrix and node-link visualizations on the task of comparing brain connectivity patterns [3]. Shi et al. took advantage of some feature selection algorithms in the data mining field to extract the most discriminative and perceptually interpretable subgraph from the selected brain networks [29]. Yang et al. proposed the blockwise brain networks comparison [33]. They separated ROIs into six lobes and applied their developed clustering algorithms within each lobe. These classified and clustered ROIs were visualized with *NodeTrix* [18] which is a hybrid visualization combining node-link diagrams and matrix representations. In these ways developed by Shi et al. [29] and Yang et al. [33], the complexity of making visual comparisons between brain networks can be greatly reduced. Bach et al. proposed visualizing time varying brain networks with a series of matrices as each matrix represents a temporal snapshot of the brain [5]. Matrices within neighboring time points are then piled together if their variance is smaller than a certain threshold. Bach et al. introduced *Time Curves* [6] which employs multidimensional scaling (MDS) [31] to embed both temporal and self-similarity information of a time varying brain network into the 2D space. However, it is still difficult for the users to compare multiple subjects and explore possible reasons for differences and similarities in network connectivity patterns.

To summarize the principle absence in existing software, the existing tools [16, 17, 19, 21, 22, 32] used by neuroscientists are not designed for simultaneous comparison or visualization of data at the level of individuals and groups. Several existing works focused on comparing between a few groups or examining time-varying
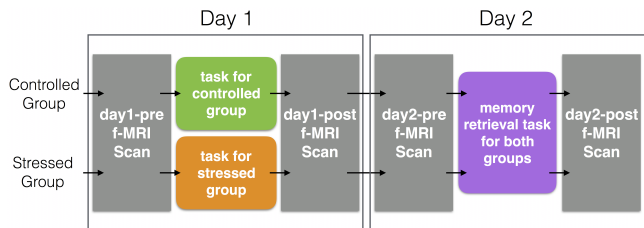

Figure 1: A brief overview of the dataset used in our study, including the information of participant grouping, experimental tasks, and scan sessions.

signals [3, 5, 6, 29, 33]. Compared with them, our system focuses on enabling efficient comparison of a large collection of subjects and multiple scan sessions while also enabling an efficient means to examine subsets of the data as desired, and incorporating a variety of subject variables along with a user-friendly interface.

3 DATASET AND DESIGN GOALS

In this section, we introduce the driving dataset and the design goals of the system. We should note that the possibility to extend the system to be used for analyzing brain connectivity data beyond the data described here has also been considered when designing the system.

3.1 Dataset Description

The dataset we used contains correlation matrices representing the functional connectivity between brain regions during resting-state fMRI (rs-fMRI) scans. Figure 1 provides a brief overview of the dataset. The fMRI images were acquired from 50 subjects during multiple phases of an experimental study of the effects of stress on brain activity related to memory. After completing a memory encoding task, the subjects were separated into two groups, a stressed group and a control group. Before and after the stress (or control) task, the subjects completed an rs-fMRI scan of 7 minutes and 10 seconds. During this time, the subjects were instructed to lie silently in the machine with their eyes closed but were given no additional instructions. On the following day, the subjects completed two additional rs-fMRI scans, one before and one after a memory retrieval task. Thus, there are 4 rs-fMRI scan sessions for each subject: pre- and post-stress (or control) and pre- and post-retrieval. More detailed information about the data is described as follows.

Brain regions — Each subject's data was first pre-processed and warped to fit into a standardized brain template. Then, activity estimates from 137 anatomical regions of interest (ROI) were extracted for each time point during the scan and the pairs of ROI activity estimates were correlated across time points for further analysis.

Matrices — Each rs-fMRI scan is represented by a matrix containing $137 \times 137 = 18769$ correlation values. Because the correlation values are duplicated within the matrix (i.e., matrices are symmetrical about the diagonal), the actual number of unique correlation values is: $(18769 - 137)/2 = 9316$. In addition, correlation values are between -1.0 and 1.0.

Subjects — As is common practice in studies using fMRI, data were excluded from individual subjects who exhibited more than $3\,\text{mm}$ total movement in any direction during any scan session. Exclusion of those data is necessary because it is impossible to infer the spatial origin of any signal measured after the subject moved. These criteria resulted in exclusion of all rs-fMRI data for four subjects, and of an additional three subjects from day 2, leaving the complete dataset from 43 subjects, and day 1 data from 46 subjects.

Meta info — Additional data collected from each subject were incorporated into the software package to examine if those variables influence RSFC. The information includes experimental group

(stress or control), hours slept, age, and cortisol response (i.e., a physiological measure of stress).

The main research question the neuroscientists wish to answer with this dataset is: How does acute stress alter the functional connectivity of ROIs known to be involved in solidifying memory for recent experience. The most direct way is to explore differences in patterns of RSFC between the stress and control groups immediately after the stress manipulation (day1-post scan). The neuroscientists also wish to examine whether additional factors beyond the stress manipulation contribute to (dis)similarities in patterns of RSFC across all ROIs in the brain. Finally, the neuroscientists wish to investigate how functional connectivity across all ROIs changes as a function of the scan sessions, and whether the stress/control task or other factors influence those differences in consistent ways.

3.2 Typical Analytics Flow and Design Goals

We next describe a typical analytics flow that researchers follow when conducting brain connectivity analysis. We then present the design goals of our system.

3.2.1 Typical Analytics Flow

Studies of RSFC typically involve a statistical comparison of connectivity patterns between two groups of subjects, or comparison of connectivity before and after an experimental manipulation. Researchers normally study RSFC data at the group level while not thoroughly examining data from individuals mainly because of the high cost of time and the lack of effective comparison methods/tools. Moreover, RSFC studies are typically conducted with fairly specific hypotheses regarding how connectivity will differ between groups of subjects or as a function of the manipulation. After pre-processing rs-fMRI data, many researchers move directly to running statistical comparisons to test hypotheses about certain brain regions. In order to communicate the results, researchers then typically create a visualization such as a node-link diagram on a transparent brain or a correlation matrix representing the group-level differences in connectivity. However, a great deal of information is lost in such a process — information regarding any individual subject's RSFC can be lost in group-level analyses, and examining such high dimensional data solely with respect to one's a priori hypotheses negates the opportunity to thoroughly explore alternative reasons for RSFC differences amongst the subjects.

3.2.2 Design Goals (DG)

Each design goal (DG) was determined based on iterative discussions with the collaborating neuroscientists about how an effective visual analytics system can serve or facilitate their analytics tasks and needs.

DG1: The system should allow users to compare the differences in brain networks between several dozens of subjects. In a typical neuroscience study, several dozens of subjects are recruited. However, most of the time neuroscience researchers only consider the averaged brain networks of different groups of subjects for comparison. The differences between each individual within the group or across time are oftentimes neglected, and rarely examined thoroughly such as by incorporating other, non-MRI data to examine other potential causes for network (dis)similarities. This is mainly because the brain networks typically have high complexity and dimensionality and there existed no well-defined approach to support such a task. The first goal is then to enable the capability for researchers to compare a collection of individuals' brain networks with an expressive and intuitive manner.

DG2: The system should enable users to analyze the variations of the subjects' brain networks over multiple sessions. In general cases, the recruited subjects are asked to complete multiple brain scans (i.e., sessions) at multiple time points so that researchers can observe or study the change of functional brain networks after

certain tasks are performed. Therefore, it is important for the system to support such a functionality.

DG3: The system should provide some visualization methods that the neuroscientists are already familiar with. Correlation matrices and node-link graphs are two visualization methods commonly used in neuroscience for presenting detailed functional connectivity. A correlation matrix shows all pairs of correlations between any two brain subregions without occlusion while a node-link diagram presents the relationships between brain subregions along with the information of their actual positions. Including these two visualizations in the system will allow the neuroscientists to examine detailed information about functional brain networks, to explore functional connectivity data much more thoroughly and intuitively, to verify research findings, and to communicate those findings with visualizations.

DG4: The system should allow for the incorporation of other data collected from participants into the analytical process. In many cases, vast amounts of data beyond the subjects' brain activities are also collected, such as demographics info (e.g., age, gender), behavioral measures (e.g., task performance preceding or following rs-fMRI data acquisition), or physiological measures (e.g., hormone levels). Examining the influence of these variables on functional connectivity may lead to new research findings or suggest new directions for future research. As a result, the system should provide an analytics method that allows the user to easily incorporate this information.

DG5: The system should help users easily interact with the data. Due to the complexity and dimensionality of RSFC data, there is a wide variety of aspects that one could incorporate in the visualization and analysis of the data. The analytical process may proceed and change according to different patterns that emerge or knowledge that is discovered when incorporating the variety of subject information described above (DG4). It is essential to provide an intuitive and fully interactive user interface so as to serve the neuroscientists' broad and ever-changing analytics needs.

4 SYSTEM DESCRIPTION

In order to meet the design goals, the system is developed to support three major functionalities: 1) visualizing the differences between subjects and sessions, 2) reviewing detailed information, and 3) linked interactions between views. Details of each functionality will be explained in the ensuing sections. The user interface of the system is displayed in Figure 2, which consists of five components: (a) settings panel, (b) MDS view, (c) correlation matrix view, (d) 3D graph view, and (e) information view. The analytical flow of using the system is summarized in Figure 3 as each step involves one or more views/panel in the user interface. The three functionalities are each associated with some of the blocks or texts shown in Figure 3 and are color-coded in blue, green, and red, respectively.

4.1 Visualizing the Differences between Individuals and Sessions (DG1, DG2, DG4)

One of the major contributions of the system is to let users visually compare the differences between individuals and sessions. To do so, the user can first configure the parameters in the settings panel, and then the visualized results will be displayed in the MDS view.

4.1.1 Settings Panel (Figure 2A)

In the settings panel, the user can first decide whether to start with the "original" dataset or to begin with one of the "post-processed" datasets. The "original" data includes correlation values of each scan for each individual as a matrix described in Section 3.1. Using the "original" dataset, we obtain two "post-processed" datasets by calculating the difference of each individual's data across multiple sessions. Specifically, the "Diff between days" dataset is derived by calculating the difference of each individual's pre- and post-task
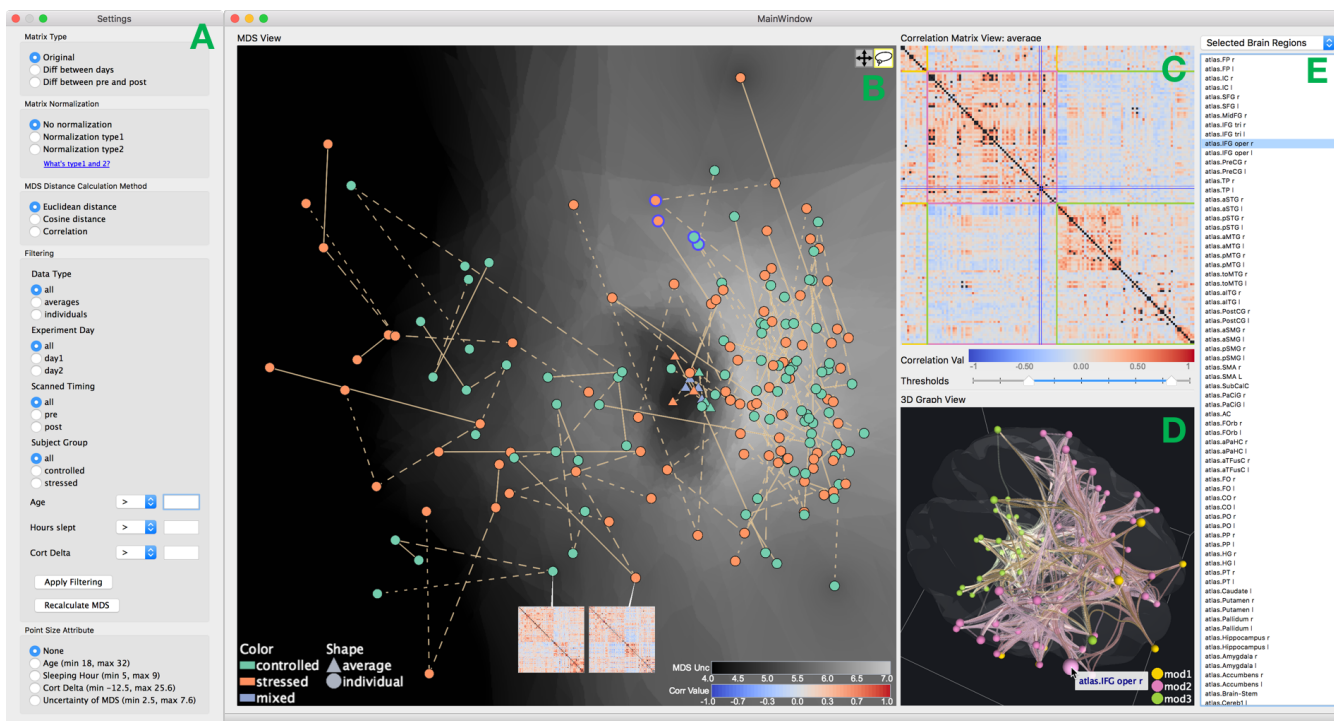
Figure 2: The user interface of the system, which contains five components: (a) settings panel, (b) MDS view, (c) correlation matrix view, (d) 3D graph view, and (e) information view.
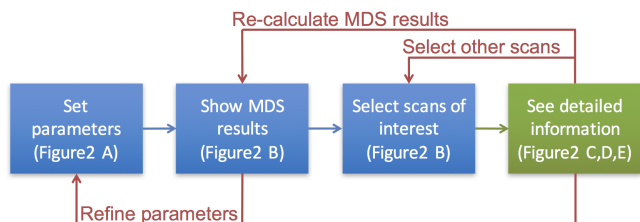

Figure 3: The analytical flow of using the system. Each step involves one or more views/panel in the user interface.

scans between day 1 and day 2 while the "Diff between pre and post" dataset is obtained by computing the difference between each individual's pre- and post-task scans both in day 1 and day 2.

Then, the user can determine if he/she wants the data to be normalized before analysis. The normalization can be used to serve the purpose of avoiding excessive individual differences at the time of scanning caused by fMRI. We provide two pre-defined normalization methods in the system. Next, the user can choose from one of the three distance metrics for the multidimensional scaling (MDS) calculation. After that, several data filtering options are provided so that the user can easily focus the analysis scope on a certain subset of the data. Finally, the size of the nodes can be configured to represent either some meta-information of the subjects or the uncertainty value introduced by the MDS method.

4.1.2 MDS View (Figure 2B)

After setting up the parameters, the differences of brain scans between individuals, between different sessions, and between groups are visualized in the MDS view. MDS is a popular method in mul-

tivariate analysis to show the (dis)similarities between objects in low dimensional space. Our system applies classical MDS [31] in 2D target space. Classical MDS takes a distance matrix as an input and calculates a desired position in 2D space for each object such that between-object distances are preserved as much as possible. In the MDS view, the brain scan of each subject in a scan session is stored as a 18769×1 vector of correlation values and represented as a node. To derive the distance between any two nodes, our system supports three distance metrics: Euclidean, Pearson's correlation, and cosine distance. Although Euclidean distance is the most popular scale, it is easily affected by differences in the magnitudes of the values. To better support the case of comparing (dis)similarity of correlation matrices between different individuals, options to use Pearson's correlation and cosine distances are also included.

The position of a node is then determined by the MDS algorithm. The attributes of a node are used to present other information of a subject. The color of the nodes is used to show the group information of the subject, which is normally collected in brain studies for researchers to study the correlation or the disassociation within or between certain groups of subjects. In our driving application, the subjects are categorized into "controlled" and "stressed" groups and are represented by green nodes and orange nodes, respectively. The colors for categorical data in the MDS view and the 3D graph view were selected from ColorBrewer [15], which provides color schemes designed mainly for choropleth maps. We selected the colors from the scheme which has enough saturation to recognize the difference of colors in a small size of points in the views. Also, we ensured that each view does not share the same colors to avoid misleading users. The shape of the nodes, e.g. triangular or round, is used to distinguish between averaged data and individual scans.

As the MDS results sometimes may contain cluttered regions, zooming and panning of the view by using a mouse are supported to display certain regions more clearly or to reduce overlapping of nodes. In the case when the user wishes to focus the analysis on

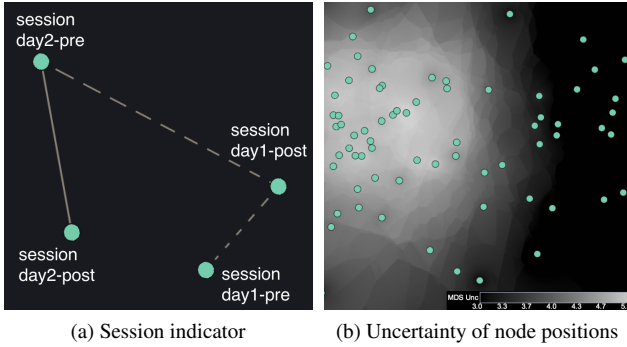(a) Session indicator   (b) Uncertainty of node positions

Figure 4: Visual cues for presenting additional information. (a) Session information indicated by the type of edges connecting two nodes. (b) The uncertainty of node positions introduced by MDS. The placement of nodes on the left have higher uncertainty values than the nodes on the right side.

certain nodes (scans) of interest, our system provides mouse click and lasso selection techniques for the user to select a subset of nodes (scans). After the selection is done, the system offers the option to re-calculate the MDS positions and refresh the MDS view using only the selected nodes. Or, the user can decide to show the corresponding correlation matrices as glyphs so as to obtain a quick examination to the correlation matrices of the subjects of interest. The size of the glyphs can be adjusted by the user as needed. For representing the correlation values in a correlation matrix, we use the divergent colormap proposed by Moreland [25], which is designed for general use in scientific visualization based on theories of color perception. The user also can find the detailed information of a subject and the meta-information by hovering a mouse on a node.

There are two additional visual cues employed in the MDS view. The user can decide whether to turn on/off either of them as desired. First, in order to identify the different nodes representing the same subject scanned in different sessions, the nodes of the same subject are connected with (dashed) lines that have different lengths of intervals to indicate the session sequence as shown in Figure 4(a). We also considered using an arrow head or a color gradient. However, both visual encodings would make the visualization more cluttered because they both require thicker and more saturated lines for one to recognize the difference between sessions. With the session indicator, the user can quickly grasp the information of how much each subject's brain activity changed across multiple sessions. Depending on the positions of the nodes, a session indicator line may overlap with other node(s) that does not belong to the session sequence. To identify which nodes actually represent the same subject, the connected nodes and session indicator lines will be highlighted when the mouse is hovered over a node.

The other visual cue is introduced to present the uncertainty of the MDS result, as applying any kind of dimensionality reduction method, including MDS, will inevitably produce uncertainty. Several ways have been proposed to visualize the uncertainty produced by dimensionality reduction methods [4, 23, 28]. The uncertainty value is calculated in a similar way as in Schrek et al. [28]. The uncertainty value of a node i can be computed as

$$unc_i = \sqrt{\sum_{j \in P}(d_{ij} - d'_{ij})^2} \qquad (1)$$

where P, d_{ij}, d'_{ij} denote the set of all nodes in the current view, the distance between $node_i$ and $node_j$ in the original dimension, and in the MDS-mapped 2D space, respectively. d_{ij} and d'_{ij} are first nor-

malized by dividing by the maximum distance in their corresponding dimension. As in [28], we interpolate the uncertainty value at each display pixel by calculating the weighted average value of neighbors from the pixel and then visualize the value as a white background, as shown in Figure 4(b). A brighter white color represents higher uncertainty. This representation emphasizes areas where the distances in the original dimension are not well preserved and raises user's awareness of potential issues of uncertainty in the dimensionality reduction results. Alternatively, the user can map the uncertainty value to reflect on the size of a node, as described in Section 4.1.1. To reduce the uncertainty in such areas, the user can update the visualized result by selecting a subset of nodes and then re-calculate the MDS positions for only the selected nodes.

Dimensionality reduction techniques, including MDS, have been widely studied and developed in recent decades, and are still an active research topic [30]. We chose MDS for the following reasons. First, the computation cost of MDS is low. Because we assume the analytical process needs to be flexible and interactive, the recalculation of dimensionality reduction results is often needed and thus needs to be relatively quick. The other reason is that MDS takes distance matrices as its input, which allows us to conveniently apply different types of distance metrics to measure the similarity between data from multiple scans.

4.2 Review Detailed Information (DG3)

Correlation matrices and node-link diagrams are commonly used in neuroscience for the purpose of helping neuroscientists inspect the detailed brain network information of certain selected regions and/or subjects.

#### 4.2.1 Correlation Matrix View (Figure 2C)

A correlation matrix presents an overview of how the brain regions correlate with each other. An effective matrix reordering and community detection algorithm can help reveal and identify modules of brain regions that have higher correlation within-modules than between-modules.

In our system, we provide complete-linkage clustering as a hierarchical clustering and the Louvain method [7]. Users are allowed to define their own customized matrix order and/or assign the module to which each node belongs to as needed. The number of modules can also be set based on a selected number when applying the hierarchical clustering method, or based on a detected number of modules with the Louvain method. The modules are indicated with colored perimeters of rectangles in the correlation matrix. Values in the matrix are colored with the divergent color proposed by Moreland [25] as described in Section 4.1.2. Since the dataset includes 137 ROIs, it is difficult to label all names of them in each row or column in the limited space. Moreover, following line by line visually to identify an ROI corresponding to the value is not a reasonable solution. The correlation matrix view supports mouse hovering to solve these problems. The hovered element is highlighted and shows the names of ROIs and the correlation value.

4.2.2 3D Graph View (Figure 2D)

A 3D node-link graph is provided as a complement to the correlation matrix since useful information regarding the anatomical position of the nodes is lost in the matrix visualization. Researchers commonly use 3D node-link graphs to convey results of RSFC analysis, because relationships between specific brain regions can be followed easily in the 3D space. We allow interactions such as zooming and rotation to help grasp the spatial positions more easily and correctly. In addition, mouse hovering interaction shows the name of ROI for further confirmation.

In a 3D graph view, two points are worth noting and should be taken care of. One is the visual clutter caused by edge crossing. Applying edge bundling techniques may help alleviate the clutter
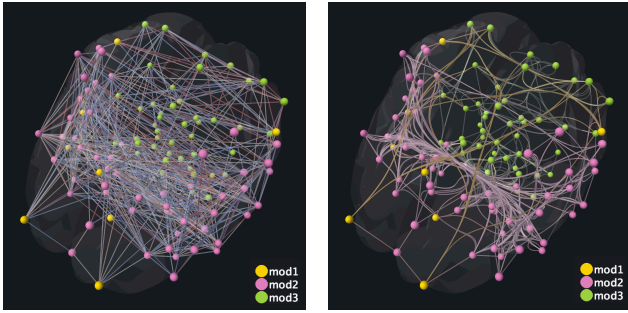
(a) Edges colored by correlation values and without edge bundling

(b) Edges colored by modules and with edge bundling

Figure 5: Comparison of two different edge coloring schemes and with/without edge bundling for the same brain network. In (b), the symmetry of the graph and the high connectivity between the nodes in module 2 (represented by pink) can be seen more clearly as compared to (a).

of the graph and improve the clarity of the graph structure. Our system provides the user with the option to apply mean-shift edge bundling [8]. The other notable point is the color representation of the nodes/edges. By default, the colors of the nodes present the module information described in Section 4.2.1 and correspond to the colored perimeters of rectangles in Figure 2(c) while the colors of edges present correlation values between nodes. Nonetheless, the color of the edges can also be switched to reflect the module information to emphasize the connections between or within the modules. Figure 5 shows an example to demonstrate the effect of using color and an edge bundling technique that can help emphasize certain patterns in the graph. As can be seen in Figure 5(b), the symmetricity of the graph and the high connectivity within the pink group nodes are displayed more clearly as opposed to in Figure 5(a).

4.2.3 Information View (Figure 2E)

The information view shows the lists of detailed information. Users can choose from one of the four options: 1) "Files" shows a file list of subjects displayed in MDS view, 2) "Corr Matrix Orders" shows the order of the regions displayed in the correlation matrix, 3) "Selected Brain Regions" shows the list of brain regions selected/displayed in the correlation matrix or 3D graph, 4) "Saved Images" shows the images saved from the other three views for further comparative analysis.

4.3 Linked Interactions between Views (DG5)

By observing the results displayed in each individual view, users are able to gain insight from different aspects of the data. However, sometimes neuroscientists would want to incorporate as much information as possible and examine the interplay of brain activity patterns and different attributes or aspects of data. To fulfill such a need, the system provides linked interactions between the views. That is, the interactions applied in one of the views will reflect and change the content in the other views accordingly.

4.3.1 Interactions from MDS View

In the MDS view, when a user selects one or a group of specific nodes, the system allows the user to see or compare the detailed information of the selected nodes in the three other views. If only one node is selected, the user can see the correlation matrix and 3D graph of the corresponding scan and the file name of the selected scan will be highlighted. If two nodes are selected, the user can

choose to see the averaged or the subtracted correlation matrix (i.e., subtracting the values in one matrix from the other) and 3D graph. For the cases where more than two nodes are selected, three options are provided for comparing the selected scans: average, standard deviation, and maximum-to-minimum difference. The average method can be used to identify which brain regions are more active. The standard deviation method is suitable for presenting the similarity of brain regions within the selected scans. The maximum-to-minimum difference method shows the range of the correlation values in each brain region. With the interactions described above, users are able to intuitively compare the differences of brain networks in a more flexible fashion.

4.3.2 Interactions from Correlation Matrix View and 3D Graph View

In some cases, a user may want to focus on specific brain regions or the brain regions that obtain a certain range of correlation values for further analysis. Two interaction methods are provided in the correlation matrix view. The user can use the sliders provided under the correlation matrix to filter out the correlation values that are not within a certain range. Or, the users can click-and-drag on the correlation matrix to select certain brain regions. The correlation matrices and the 3D graph view will be updated simultaneously after either operation is applied. The matrix cells and the graph edges are displayed only if their values are within the selected range. These operations can help reduce clutter in the 3D graph view. In addition, the user can also use the filtered or selected correlation values to re-calculate the MDS results. It allows the user to verify if certain brain regions are the reason that some subjects differ in overall RSFC. When the user hovers the mouse over a certain matrix cell, the corresponding nodes and link will be highlighted in the 3D graph view to provide a reference of their anatomical positions. On the other hand, if the user selects a node in the 3D graph, the row and column corresponding to the selected node are highlighted in the correlation matrix view with blue rectangles as shown in Figure 2(c).

4.3.3 Interactions with Information View

The information view displays detailed information about data selected in the other views. It can also be used as an inverse selection interface so that a user can directly find a specific data item (such as a region or a subject) and then check its behavior in the other views.

5 CASE STUDIES

We present case studies on the dataset that motivated the design and implementation of this system. Examples are demonstrated to show the diverse utility of the system. We paraphrase the remarks made by over a dozen collaborating neuroscientists after using the system.

5.1 Case Study 1: Comparison between Averages and Individuals

Many studies of RSFC involve averaging RSFC correlation matrices across all subjects in a group, then comparing the group-averaged RSFC to the averaged RSFC of another group. However, it is well established that there is a great deal of variability in RSFC across individuals. The comparison of group averages conceals this variability across individuals, and thus may as well conceal interesting aspects of a RSFC dataset. In this case study, we start by examining the entire dataset in the MDS view. In particular, we inspect the average RSFC for each group (stress, control) and each scan session (day1-pre, day1-post, day2-pre, day2-post), as well as the average RSFC across all participants (i.e., ignoring group membership) for each scan session. This resulted in the twelve average matrices displayed at the top of Figure 6(a). With a standard approach that examines the group differences in RSFC by comparing average RSFC of each scan session for the two groups, one might simply draw a conclusion that there is little difference in RSFC between the stress
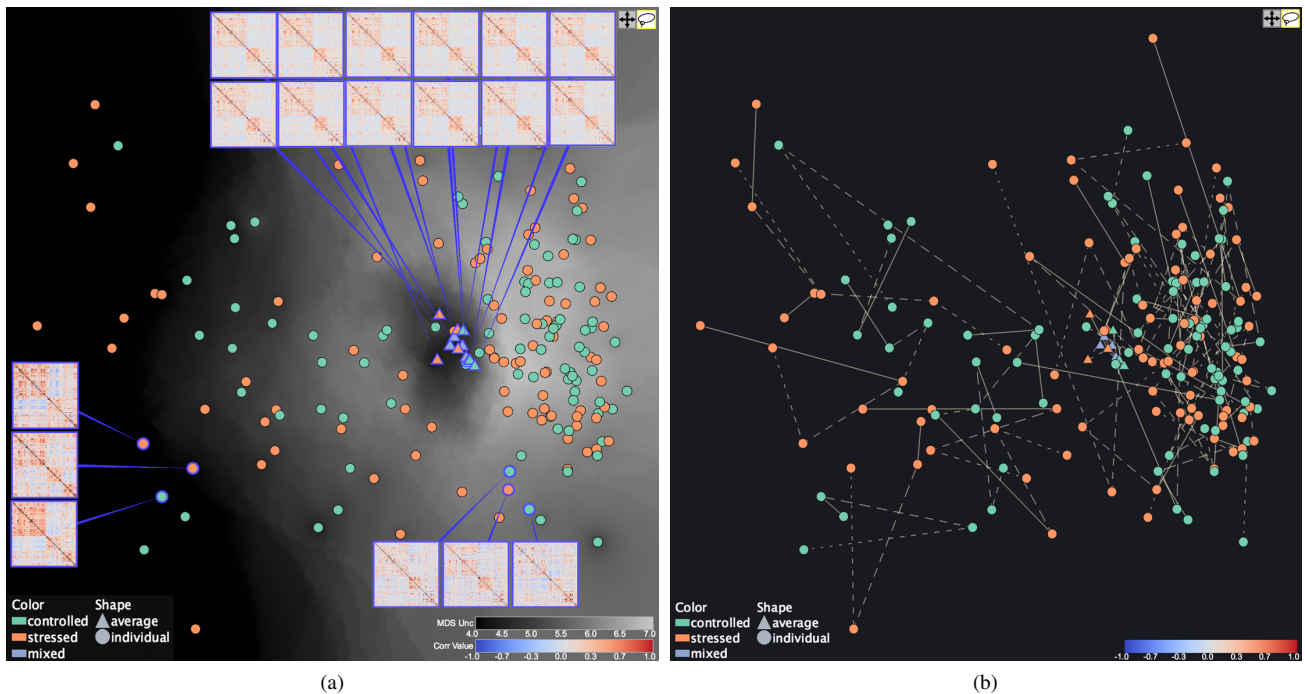
(a)

(b)

Figure 6: (a) An MDS view showing the overview of the differences between scans in the entire dataset, including the averaged scans. Averaged scans tend to be similar to each other (top), while individual scans vary (bottom and left). Several scans are selected and their corresponding correlation matrices are displayed for visual comparison. (b) The session indicator lines allow the user to visualized the self-stability or similarity of RSFC matrices of each subject in multiple sessions.

and control groups. This result is apparent by a quick glance of the twelve average RSFC matrices, as they all appear relatively similar to one another.

However, as described above, RSFC is known to vary across individuals. A totally different pattern may be found if the differences between every individual are taken into consideration. Using the functions provided in the MDS view, one can easily see that there is in fact great variability across individuals, no matter which group the subjects belong to. For example, the user can highlight some of the nodes and easily examine the correlation matrices of those highlighted nodes to explore the individual differences. As can be seen at the bottom left and right of Figure 6(a), correlation matrices of several different subjects (from both control and stress groups) are displayed. We can see that some control subjects' RSFC matrices are more similar to the RSFC matrices of stressed subjects than to some other control subjects. The patterns and insight found in Figure 6(a) suggest that it might be misleading to make comparisons on group averages while ignoring individual variances as the two approaches could lead to contradictory conclusions.

The rs-fMRI data are collected from each subject in multiple sessions (e.g., before and after cognitive tasks). The representation of the time indicator, as presented in Figure 4(a), can help identify the self-stability or similarity of RSFC matrices of each subject in multiple sessions. Figure 6(b) shows the MDS result when the session indicator feature is turned on. We can see that the matrices for any given individual tend to cluster together, showing the strong influence of individual variation on RSFC similarity, regardless of group membership (i.e., stress or control) or the timing of the scan session (i.e., day 1 or day 2). In addition to the influence of individual variation in RSFC, these patterns of (dis)similarity across days and scan sessions reflect the influence of time on an individual's RSFC.

The ability to easily examine individual RSFC matrices, in this case from multiple scan sessions with the same subject, is useful for better understanding the degree to which various factors influence RSFC.

5.2 Case Study 2: Differences between Pre- and Post-Task Scans

Researchers are often interested in examining the variation in RSFC from before to after some cognitive tasks. The changes observed in RSFC can be used to make inferences about the cognitive processes involved in the tasks. In this case study, we present a representative example that the neuroscientists used the system for examining the changes in RSFC from before to after a task. The goal of this example is to analyze the changes in RSFC of different subject groups (i.e., control and stress) between the pre- and post-task scans from day 1. We start by selecting "Difference Between Pre and Post" and the data from only "day1" in the settings panel, and then perform MDS calculation. The MDS view is shown in Figure 7(a) where orange nodes tend to be placed on the left side of the screen while green nodes are mostly placed on the right side of the screen. This pattern suggests that the stress and control tasks induced changes in RSFC in a different manner.

Because this method uses the correlation matrices that reflect the difference in pre-post task RSFC, the average effect that an experimental task induces on RSFC of multiple groups can easily be identified. As can be seen in Figure 7(a) the averaged pre-post correlation matrices for both control and stress groups are displayed and show substantial differences from each other. We then explore the distance between them in more detail by viewing the subtracted correlation matrix in both matrix and 3D graph views to find out which brain regions diverge the most between the two groups. After
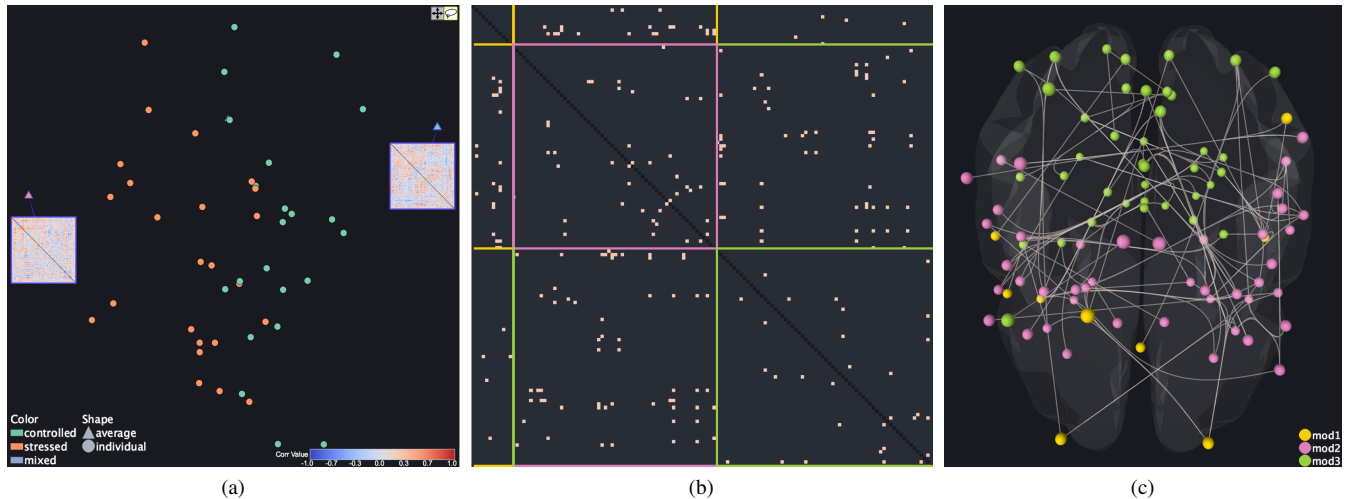
Figure 7: (a) An MDS view showing that the RSFC of the control and stress groups between the pre- and post-task scans in day 1 are changed in a different manner. After setting up a relatively high threshold value, the brain regions where connectivity differentiates the two groups are easily identified. The difference of the averaged pre-post correlation values for both groups is shown in the correlation matrix (b) and 3D graph views (c). Note that the three perimeters shown in (b) and the nodes shown in (c) are colored based on the modularity information. Using consistent color encoding in these two views provides easy reference for the user.
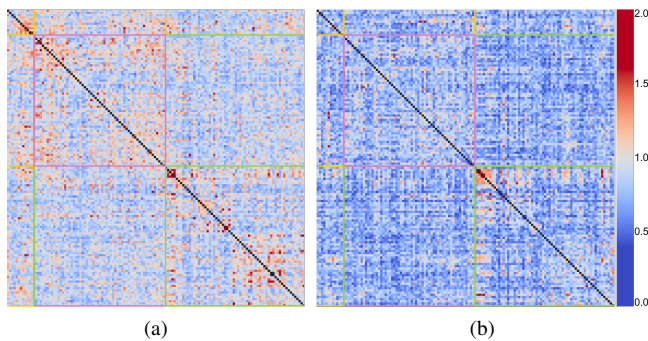

Figure 8: (a) The maximum-to-minimum correlation matrix of subjects in the stress group who showed a positive cortisol response. (b) The maximum-to-minimum correlation matrix of subjects in the stress group who did not exhibit a physiological stress response. (a) tends to have higher values than (b) in general. The values in both matrices are between 0 and 2 because the range of values in the original matrices are −1 and 1.

setting up a relatively high threshold, we are able to see the brain region pairs that have most different correlation values between the control and stress groups, as shown in Figures 7(b) and (c). Some further analysis can be possibly done by also examining how an individual subject's RSFC compares to the group averages.

5.3 Case Study 3: Influence of Stress-related Physiological Measure on RSFC (Dis)similarity

Most rs-fMRI experiments are based on theories of cognition and prior evidence of the effects of some experimental manipulation, and thus are conducted with a set of clearly defined hypotheses in mind. Nonetheless, it is often useful to examine the influence of additional variables that were not thought a priori to influence RSFC. In the

experiment from which the current dataset was drawn, we collected a physiological measure of stress, which is known as the cortisol response, after subjects completed either a stress-induction task or a control task. As stress can be induced by a variety of factors beyond the experimental manipulation (i.e., the stress-induction task), it may also be useful to examine RSFC based on the physiological measure of stress rather than based solely on the experimental manipulation. We would like to know whether RSFC differs for subjects in the stress group who exhibited a positive cortisol response compared to the subjects who completed the stress-induction task but did not exhibit a physiological stress response.

We configure the system as following to conduct such an analysis. First, we utilize the data filtering function in the settings panel to select only scans of subjects in the stress group. Then, we set the attribute filter as "Cort Delta" > 0 in the settings panel to include only the scans of subjects who show positive cortisol response. 18 subjects (68 scans) meet this condition. After that, we derive the maximum-to-minimum correlation matrix of the remaining scans, as shown in Figure 8(a). Similarly, we can obtain the maximum-to-minimum correlation matrix of those subjects who are in the stress group but did not exhibit a physiological stress response (i.e., "Cort Delta"≤ 0). 6 subjects (24 scans) are left with this condition. The resulting correlation matrix can be seen in Figure 8(b). By comparing the two correlation matrices, we can see that the subjects who exhibit positive cortisol response has a wider range of correlation values in most brain regions than the subjects who did not exhibit physiological stress response, suggesting that physiological stress responses indeed influenced RSFC. This efficient approach to exploring the influence of additional factors on RSFC can help neuroscientists to specify hypotheses regarding a wide variety of influences on RSFC. The researchers are then in a better position to devise some more advanced statistical methods for testing those hypotheses.

6 DISCUSSION

We discuss some potential issues of the MDS results, feedback and comments received from our team of collaborating neuroscientists, and the generalizability of our system.

6.1 Potential Issue with Dimensionality Reduction Methods

In order to present data of high dimensionality in a 2D screen, dimensionality reduction methods are commonly applied to transform the data in the high-dimensional space to a space of fewer dimensions. In this process, it is inevitable for any dimensionality reduction method to introduce distortion to the visualized data. As a result, there might be some risk for the viewer to misinterpret the visualization results. To address this issue, the system employs the uncertainty visualization to raise the user's awareness of uncertainty in the presented visualization. In addition, if the user is interested in seeing in more detail some part of the visualization that has high uncertainty, the system also provides the functionality for the user to select a subset of displayed nodes and then re-apply the MDS calculation. Such a capability not only reduces the uncertainty of the presented visualization, but also allows the user to focus the analysis on nodes that are of particular interest.

6.2 Feedback from Users

Our collaborating neuroscientists provided positive feedback for various aspects of the system. In general, they found the visualization representations intuitive for them to understand and interpret. In addition, they like the interactivity of the system, which allows them not only to flexibly examine a considerably large and complex dataset but also to quickly explore alternative hypotheses and see if a certain attribute could have an important influence on the data. More specifically, they appreciate two aspects of the system the most: 1) the intuitive and inter-connected features of the multiple views, and 2) the capability of the system to allow them to view and compare the data in multiple levels, i.e., from one or a few individuals to a subset of individuals (as defined by any number of variables of interest), and to pre-defined groups of individuals. The systems used previously by the neuroscientists only allow for comparison at the group-level. Our software provides them with the unique ability to thoroughly examine RSFC data with respect to researchers' ever-changing hypotheses regarding variables that may influence patterns of functional connectivity. Below, we paraphrase some of the researchers' feedback about the three main views.

The MDS view is a user-friendly tool for examining the similarity between RSFC datasets. Whether examining the similarity of scans within an individual, similarity of experimentally-defined groups of individuals, or similarity based on ranges of a continuous measure such as the cortisol response, the MDS view allows the user to easily select subsets of data and efficiently re-examine the data based on a vast number of variables of interest.

The correlation matrix view shows RSFC data across the entire brain and is familiar to most neuroscientists. They especially appreciate the ease with which the user can display data from an individual subject or from groups of scans or participants. The tool also allows the user to specify the order of the regions in the correlation matrix, which will facilitate the user's interpretations.

The 3D graph view is another view that neuroscientists are familiar with, while it provides a means for quickly visualizing the anatomical sources of the correlations presented in the correlation matrix view. They appreciate the zoom and rotate functions of the 3D graph view, as well as the edge bundling feature. A general feature they greatly appreciated was the inter-connections between views, for example when a region is selected from the information view (Figure 2(e)) and the corresponding node in the 3D graph view is highlighted (Figure 2(d)), as well as the corresponding row and column of the correlation matrix view (Figure 2(c)).

While our collaborating neuroscientists are pleased with the functionality of the system, they have provided several suggestions to improve the system. First, they wish to obtain more flexibility on the mapping of the visual elements in the MDS view. Currently, the size, shape and color of a node and different patterns of an edge are each used to present some pre-defined aspect/attribute of the data. The neuroscientists suggest that in some cases it would be helpful if some of the mappings can be switched according to their analytical needs. For example, when the analysis only involves individual scans but not averaged scans, the shape of the nodes can then be used to represent some other aspect of the data.

They also made some suggestions on the 3D graph visualization. One comment is to give users the option to view only one of the hemispheres (right or left) at a time, or to view both. This would help reduce clutter, because in most cases, the connectivity within each hemisphere is often very similar.

Finally, the neuroscientists think that with some added features the system could be very useful for them to prepare materials for teaching and communicating results with other researchers. While the system currently supports a "save image" function for each view, it is desired to have the ability to save those images with a much higher resolution (300 dpi or higher) as well as to add annotations on the saved images. They also requested the function to extract a video showing the 3D graph with different scales and angles according to the user's interaction.

6.3 Implications for Future Usage of the System

We have demonstrated several uses, based on one specific dataset, to show the strengths and utility of the system. The dataset used is representative of many rs-fMRI datasets, as it contains typically measured data including correlations between ROIs, ROI location information, data collected from subjects in different groups and data collected in multiple scan sessions for each subject, as well as demographic and physiological data for each subject. We believe that the components currently designed in this system can be easily generalized to apply to other rs-fMRI datasets. In brief, the visualization tool requires a simple input of correlation matrices, which is a form of data highly familiar to any RSFC researcher. Upon startup, the system allows the user to define unique aspects of their data structure (such as the number of unique scan sessions per individual, membership in experimental groups, and individuals' physiological variables). Thus, the system provides the foundation of a uniquely optimizable platform for visualization of rs-fMRI data from many different types of experiments. The user-friendly interface provides familiar views of the data, in correlation matrices and 3D graph views, as well as a novel visualization that allows the user to efficiently examine how RSFC data groups according to a variety of variables of interest. By facilitating quick visualization of RSFC similarity according to a variety of factors, the MDS view is particularly useful for the data-driven examination of interactions of factors influencing RSFC, thus providing a means for researchers to examine the influence of a multitude of factors not predicted a priori to influence their results. The three views are interactive and coordinated in their displays, and are further supported by the user-friendly settings panel and information view. We have already reviewed several different datasets and have started to work on extending our system for use with general RSFC datasets in future work.

7 CONCLUSION

In this paper, we present a visual analytics design that meets the growing needs of brain researchers to extract meaning from data of high dimensionality and high volume. The resulting system integrates data manipulation methods, visualization designs, and interaction techniques to support visualization and analysis tasks on brain functional connectivity data. The case studies conducted together with our collaborating neuroscientists show that the combined functionalities of the system greatly enhance their ability to gain new insights into their data. The capabilities the system offers can help accelerate brain research by allowing researchers to easily perceive and understand the (dis)similarities between individuals (and groups of individuals), brain regions (and networks of brain regions), and

scan sessions (whether before and after an experimental task, or on different days). As brain research continues to advance with growing ambition and data collection capabilities, interactive visual analytics systems like ours will play a more crucial role in the overall research process.

REFERENCES

[1] R. D. Airan, J. T. Vogelstein, J. J. Pillai, B. Caffo, J. J. Pekar, and H. I. Sair. Factors affecting characterization and localization of interindividual differences in functional connectivity using MRI. *Human Brain Mapping*, 37(5):1986–1997, 2016.

[2] E. A. Allen, E. Damaraju, S. M. Plis, E. B. Erhardt, T. Eichele, and V. D. Calhoun. Tracking whole-brain connectivity dynamics in the resting state. *Cerebral Cortex*, p. bhs352, 2012.

[3] B. Alper, B. Bach, N. Henry Riche, T. Isenberg, and J.-D. Fekete. Weighted graph comparison techniques for brain connectivity analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 483–492. ACM, 2013.

[4] M. Aupetit. Visualizing distortions and recovering topology in continuous projection techniques. *Neurocomputing*, 70(7):1304–1330, 2007.

[5] B. Bach, N. Henry-Riche, T. Dwyer, T. Madhyastha, J.-D. Fekete, and T. Grabowski. Small MultiPiles: Piling time to explore temporal patterns in dynamic networks. In *Computer Graphics Forum*, vol. 34, pp. 31–40. Wiley Online Library, 2015.

[6] B. Bach, C. Shi, N. Heulot, T. Madhyastha, T. Grabowski, and P. Dragicevic. Time curves: Folding time to visualize patterns of temporal evolution in data. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):559–568, 2016.

[7] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, (10):P10008, 2008.

[8] J. Bottger, A. Schafer, G. Lohmann, A. Villringer, and D. S. Margulies. Three-dimensional mean-shift edge bundling for the visualization of functional connectivity in the brain. *IEEE Transactions on Visualization and Computer Graphics*, 20(3):471–480, 2014.

[9] V. D. Calhoun, R. Miller, G. Pearlson, and T. Adalı. The chronnectome: time-varying connectivity networks as the next frontier in fMRI data discovery. *Neuron*, 84(2):262–274, 2014.

[10] G. Deco, V. K. Jirsa, and A. R. McIntosh. Resting brains never rest: computational insights into potential cognitive architectures. *Trends in Neurosciences*, 36(5):268–274, 2013.

[11] J. Dubois and R. Adolphs. Building a science of individual differences from fMRI. *Trends in Cognitive Sciences*, 20(6):425–443, 2016.

[12] E. S. Finn, X. Shen, D. Scheinost, M. D. Rosenberg, J. Huang, M. M. Chun, X. Papademetris, and R. T. Constable. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature Neuroscience*, 2015.

[13] L. Geerligs, M. Rubinov, R. N. Henson, et al. State and trait components of functional connectivity: Individual differences vary with mental state. *The Journal of Neuroscience*, 35(41):13949–13961, 2015.

[14] S. Gerhard, A. Daducci, A. Lemkaddem, R. Meuli, J.-P. Thiran, and P. Hagmann. The connectome viewer toolkit: an open source framework to manage, analyze, and visualize connectomes. *Frontiers in Neuroinformatics*, 5:3, 2011.

[15] M. Harrower and C. A. Brewer. ColorBrewer.org: an online tool for selecting colour schemes for maps. *The Cartographic Journal*, 40(1):27–37, 2013.

[16] M. Hassan, M. Shamas, M. Khalil, W. El Falou, and F. Wendling. EEGNET: An open source tool for analyzing and visualizing M/EEG connectome. *PloS One*, 10(9):e0138297, 2015.

[17] B. He, Y. Dai, L. Astolfi, F. Babiloni, H. Yuan, and L. Yang. eConnectome: A MATLAB toolbox for mapping and imaging of brain functional connectivity. *Journal of Neuroscience Methods*, 195(2):261–269, 2011.

[18] N. Henry, J.-D. Fekete, and M. J. McGuffin. NodeTrix: a hybrid visualization of social networks. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1302–1309, 2007.

[19] S. H. Hosseini, F. Hoeft, and S. R. Kesler. GAT: a graph-theoretical analysis toolbox for analyzing between-group differences in large-scale structural and functional brain networks. *PloS One*, 7(7):e40709, 2012.

[20] R. M. Hutchison, T. Womelsdorf, E. A. Allen, P. A. Bandettini, V. D. Calhoun, M. Corbetta, S. Della Penna, J. H. Duyn, G. H. Glover, J. Gonzalez-Castillo, et al. Dynamic functional connectivity: promise, issues, and interpretations. *NeuroImage*, 80:360–378, 2013.

[21] J. Kruschwitz, D. List, L. Waller, M. Rubinov, and H. Walter. GraphVar: A user-friendly toolbox for comprehensive graph analyses of functional brain connectivity. *Journal of Neuroscience Methods*, 245:107–115, 2015.

[22] R. A. LaPlante, L. Douw, W. Tang, and S. M. Stufflebeam. The connectome visualization utility: Software for visualization of human brain networks. *PloS One*, 9(12):e113838, 2014.

[23] S. Lespinats and M. Aupetit. CheckViz: Sanity check and topological clues for linear and non-linear mappings. In *Computer Graphics Forum*, vol. 30, pp. 113–125. Wiley Online Library, 2011.

[24] D. S. Margulies, J. Böttger, A. Watanabe, and K. J. Gorgolewski. Visualizing the human connectome. *NeuroImage*, 80:445–461, 2013.

[25] K. Moreland. Diverging color maps for scientific visualization. In *International Symposium on Visual Computing*, pp. 92–103. Springer, 2009.

[26] H. Pfister, V. Kaynig, C. P. Botha, S. Bruckner, V. J. Dercksen, H.-C. Hege, and J. B. Roerdink. Visualization in connectomics. In *Scientific Visualization*, pp. 221–245. Springer, 2014.

[27] M. Rubinov and O. Sporns. Complex network measures of brain connectivity: uses and interpretations. *NeuroImage*, 52(3):1059–1069, 2010.

[28] T. Schreck, T. Von Landesberger, and S. Bremm. Techniques for precision-based visual analysis of projected data. *Information Visualization*, 9(3):181–193, 2010.

[29] L. Shi, H. Tong, and X. Mu. BrainQuest: Perception-guided brain network comparison. In *IEEE International Conference on Data Mining*, pp. 379–388, 2015.

[30] C. O. S. Sorzano, J. Vargas, and A. P. Montano. A survey of dimensionality reduction techniques. *arXiv preprint arXiv:1403.2877*, 2014.

[31] W. S. Torgerson. Multidimensional scaling: I. theory and method. *Psychometrika*, 17(4):401–419, 1952.

[32] M. Xia, J. Wang, and Y. He. BrainNet Viewer: a network visualization tool for human brain connectomics. *PloS One*, 8(7):e68910, 2013.

[33] X. Yang, L. Shi, M. Daianu, H. Tong, Q. Liu, and P. Thompson. Blockwise human brain network visual comparison using NodeTrix representation. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):181–190, 2017.