

Macro DS2CONST implements Java Applet technologies to provide the human interaction features, such as pop-up data tips, drill-down links, and interactive menus.

The simulated hypothetical clinical laboratory test data, clinical AE and simulated functional magnetic imaging (fMRI) data are used in the paper for illustration purposes. Some of the specific ideas, techniques, and layouts have been suggested before, and some are novel. The main contributions of this paper are to introduce new applications as well as designs of constellation diagrams and to illustrate the correlation matrix iteration processes.

The SAS v9 products used in this paper are SAS Base, SAS/STAT, SAS Enterprise Miner (EM) 5.2 and SAS/GRAPH on the PC Windows platform.

I. TREE TYPE REPRESENTATION

A tree structure is a method of representing the hierarchical structure in a graphical form. It is named a “tree structure” because the graph looks like a tree. In graph theory, a tree is a collection of connected nodes. Every finite tree structure has a member that has no superior. This node is called the “root” node. The lines connecting nodes are called “branches” or “links”. Nodes without children are called “leaves” or “end-nodes”. A node is a “parent” of another node if it is one step higher in the hierarchy and closer to the root node.

Tree structures are used in many applications, such as hierarchical organizational structures, binary search tree, decision tree, partition tree, etc. There are many ways of visually representing tree structures. The most commonly used method is a classical node-link diagram, that connects nodes together with line segments. Figure 2 shows a tree structure from a Treeview applet with root node in the center of the presentation. One of the features from macro DS2CONST is to allow the user to assign the node's location in x-y coordinates. This feature makes the DS2CONST a powerful tool for tree presentations.

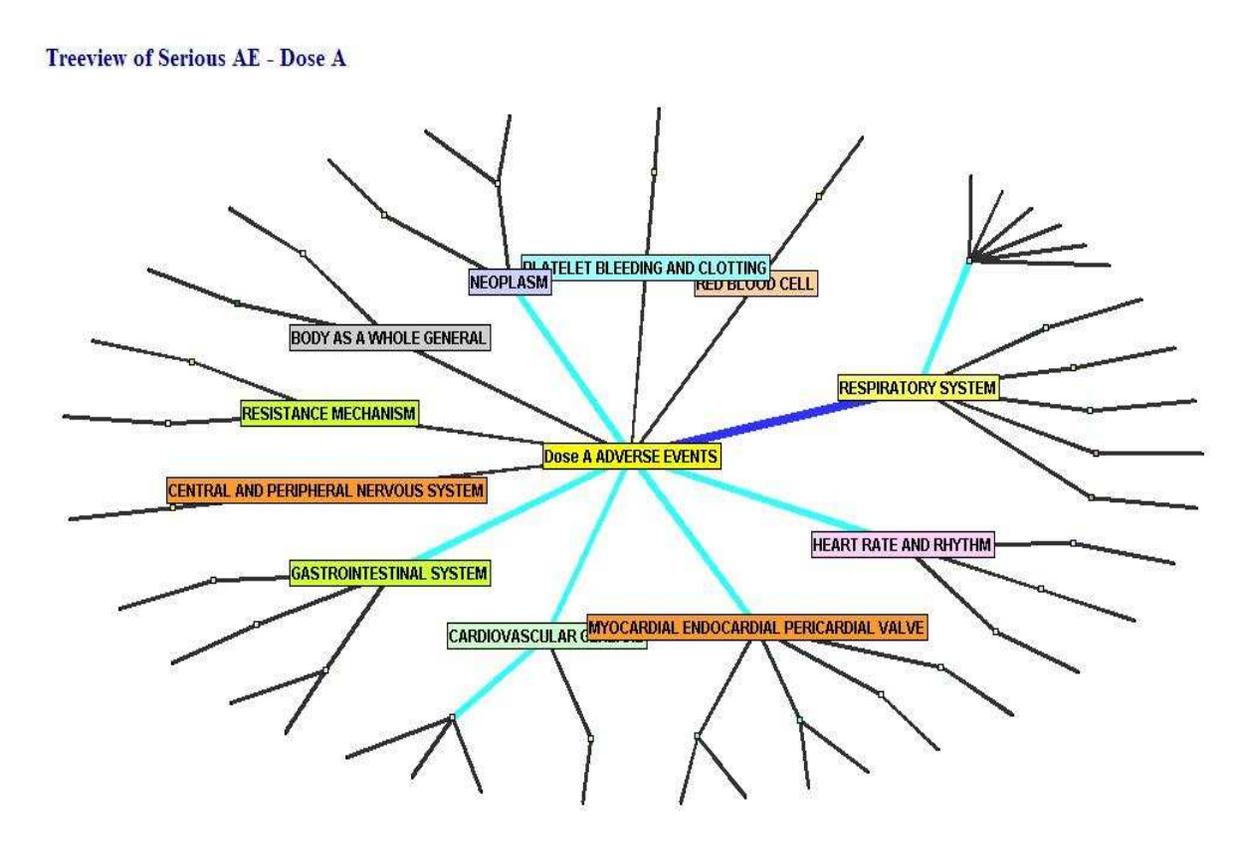


Figure 2. Sample Output from SAS Treeview Macro

Figure 3 shows a decision tree example from an application of a clinical AE predictive model. This model uses recursive partitioning to develop a decision tree for predicting “who is having a specific AE?”.

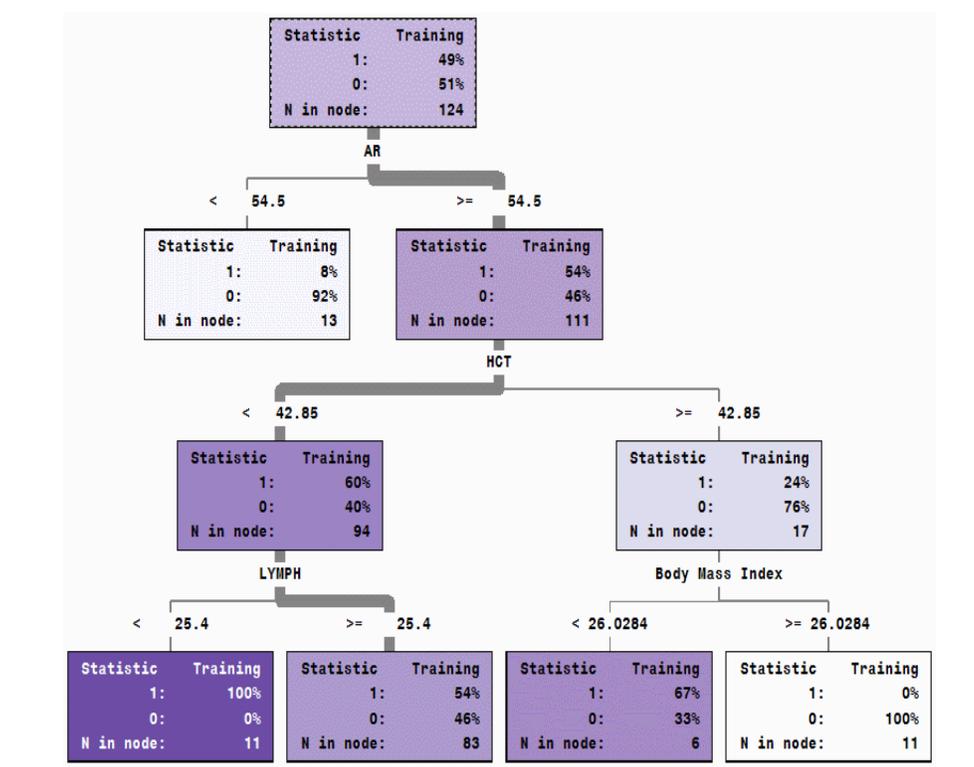


Figure 3. Sample Output of Decision Tree

Before calling the SAS DS2CONST macro, you need to prepare the data to feed into the SAS macro call. The data set to be used consists of two datasets – one dataset for node information, and the other for link information. The sample SAS code illustrates how to convert the decision tree output in Figure 3 to a SAS program for Constellation diagram presentation.

```
data linka;
  input from to value @6 ltip $20.;
datalines;
1 2 5 >= 54.5
1 3 2 < 54.5
2 4 5 < 42.85
2 5 3 >= 42.85
4 6 5 >= 25.4
4 7 4 < 25.4
5 9 2 < 26.028
5 8 2 >= 26.028
;
run;
data nodea;
  input nodeid @3 nlabel $12. @15 ntip $21. @37 xloc @41 yloc
  @45 ncolor $8.;
datalines;
1 AR 1: 49%, 0: 51%, N=124 440 50 liv
2 HCT 1: 54%, 0: 46%, N=111 560 150 biv
3 AR<54.5 1: 8%, 0: 92%, N= 13 310 150 white
4 LYMPH 1: 60%, 0: 40%, N= 94 360 250 biv
5 BMI 1: 24%, 0: 76%, N= 17 700 250 gray
6 LYMPH>=25.4 1: 54%, 0: 46%, N= 83 460 350 liv
7 LYMPH<25.4 1:100%, 0: 0%, N= 11 260 350 #0000cc
8 BMI>=26.03 1: 0%, 0:100%, N= 11 800 350 white
```

```

9 BMI<26.03 1: 67%, 0: 33%, N= 6 600 350 biv
;
data nodea;
set nodea;
length nshape $8. tname $10. fnwt $4.;
nshape='square  ';
nsize=60;
tname='sansserif';
fnwt='bold';
*proc print data=nodea;run;

%let f1=c:\anno\anewtre1.sas &sysdate &systemtime;

title1
"
footnotel "&f1";
ods listing;
proc cimport c=sashelp.javagr
infile='c:\temp\javagr.stc';
run;quit;
%ds2const (ndata=nodea,
ldata=linka,
datatype=assoc,
codebase=http://www2.sas.com/codebase/graph/v910/,
htmlfile=c:\inno\anewtre1.htm,
colormap=y,
nid=nodeid,
layout=user,
nlabel=nlabel,
nx=xLoc,
nfntname=tname,
nfntstyl=fnwt,
ny=yLoc,
nsize=nsize,
ntip=ntip,
nshape=nshape,
lfrom=from,
lto=to,
lvalue=value,
lwidth=value,
minlnkwt=1.6,
ltip=ltip,
linktype=line,
fntsize=17,
height=540,
width=960,
cback=white,
ncolor=ncolor,
cselect=red,
tface=arial,
fface=arial,
tcolor=black,
fcolor=black,
tsize=4,
fsize=3,
bg=white,
bgtype=color
);
run;
ods _all_ close;
run;

```

This program will produce a decision tree shown in Figure 4 with interactive features of data tips and subsetting.

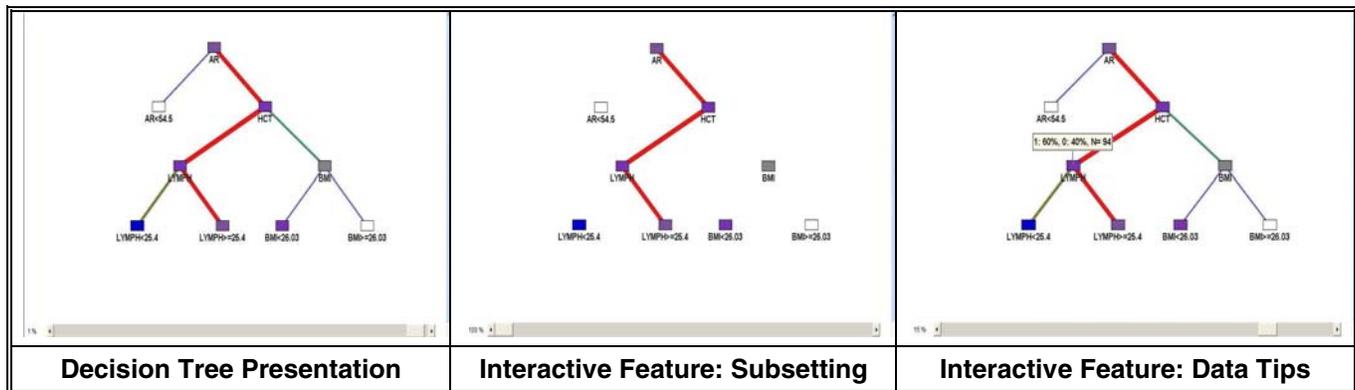


Figure 4. Decision Tree Presentation from DS2CONST

II. INTERACTIVE CORRELATION MATRIX GRAPH

The interactive graphs allow for user (human) interactions. The user interacts directly with the graphs, mainly through the user interfaces. The user interfaces are controlled by a mouse, buttons, or menus with buttons and keyboard. The interactive graphs provide the following desirable features. The user can:

- 1) change graphical properties, including modification of titles and axes through the interactive menus,
- 2) use the data tips feature to include extra information on the graph,
- 3) link to subsequent graphs, maps or tables,
- 4) change the scale by zooming techniques. Zoomable interfaces come in two types:
 - a) Geometric zooming where the scale is linear with the multiplier
 - b) Semantic zooming where objects may change shape at different zoom levels

Semantic zooming is a distortion technique that displays the object in a fisheye view. By moving around and changing the scale, the user can customize the data display.

- 5) produce slide show of static images

The SAS System implements several new technologies, such as Java applets and ActiveX Control and others, to produce the interactive graphs. SAS Graphics macro DS2CONST is selected for this application. The DS2CONST macro uses the Constellation Applet to produce a node-link constellation diagram. Each node can be linked to one or more other nodes. Unlike the Treeview Applet, the Constellation Applet does not require a hierarchical relationship between the nodes. Although it can be used to display hierarchical relationships, the Constellation Applet does not automatically place the root node at the center of the display.

The Constellation Applet supports the following interactive features. It allows:

- 1) zooming/panning or fish-eye distortion on the section of the diagram that is in the center of the display,
- 2) an embedded scroll bar to subset the links and nodes,
- 3) the user to click and drag the diagram as well as to change the section of the diagram,
- 4) a pop-up menu that has several functions, such as highlighting specific links and searching for specific nodes.
- 5) pop-up data tips for links and nodes.
- 6) hotspot links to Web sites or files.

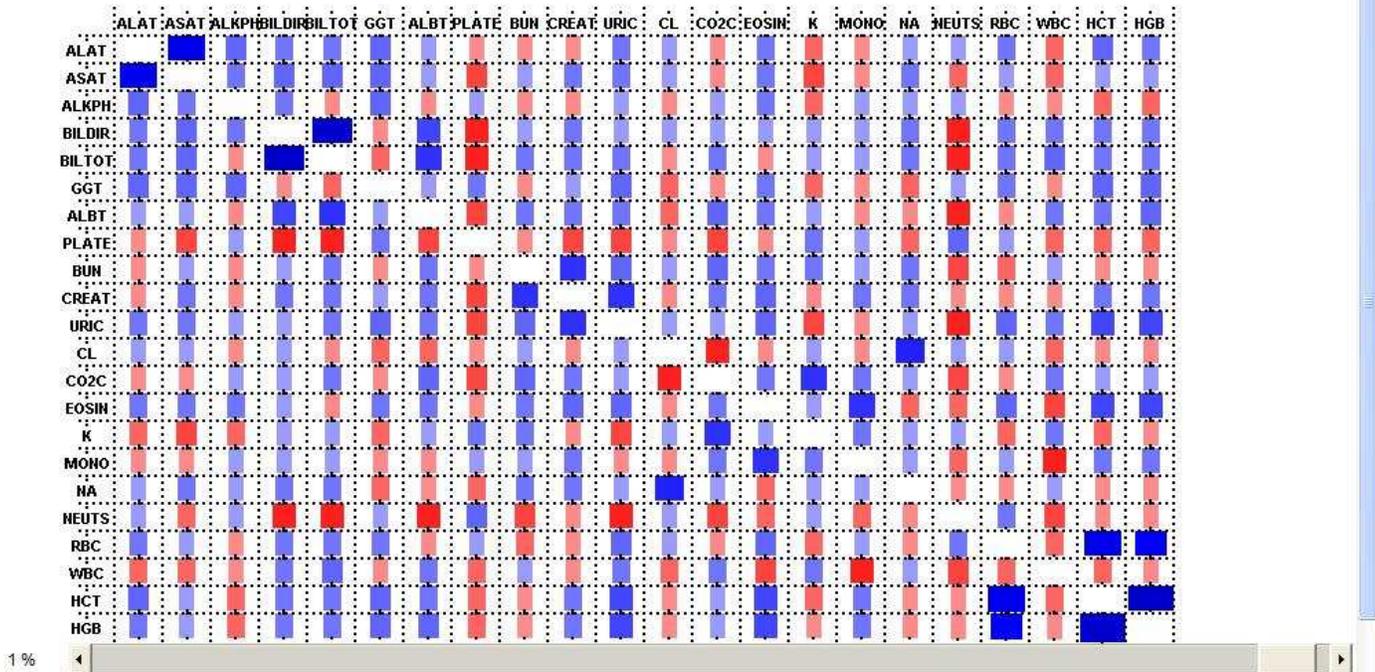
The design of the correlation matrix display focuses on the following features:

- 1) It converts the table presentation into a graphical display with blank diagonal cells.
- 2) The size of the bar symbol in each cell indicates the magnitude of the association.
- 3) Color is used to encode the sign of the correlation - blue for positive and red for negative. Scaled color is also used to show the magnitude of the correlation. The darker shade of color indicates a higher correlation value.

The interactive features for this application provide:

- 1) It provides variable labels as pop-up data tips for variable nodes,
- 2) It also provides correlation coefficients as pop-up data tips for bars (links), and
- 3) The embedded scroll bar allows the subset of nodes and links. Figure 5 shows the constellation display of the correlation matrix.

Displaying Correlations between Biomarkers



Note: Red Block: Neg. Corr., Blue Block: Pos. Corr. Ten types of block represent correlation values from +/- 0.999 to 0 by 0.1

Figure 5. Constellation Display of a Pearson Correlation Matrix

The pop-up data tip is an interactive feature which is activated when the viewer moves the mouse over a graphical element causing a drop-down text box display. You can decide on the content to be shown inside the text box. In this application, the correlation table cells are converted to link bar text boxes, and variable labels are converted to node text boxes.

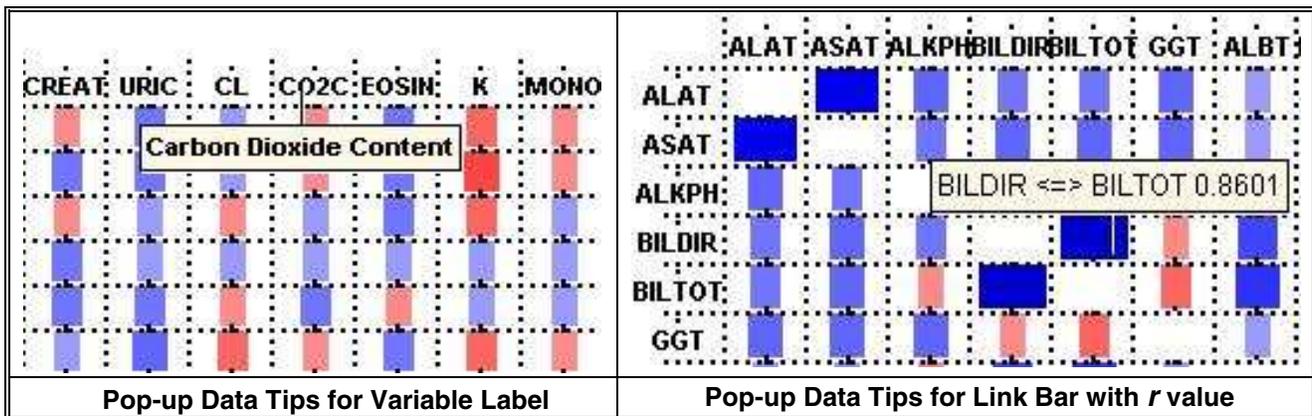


Figure 6. Illustration of Interactive Pop-up Data Tips

The other useful interactive feature is the embedded scroll bar for subsetting nodes and links. In this application, the absolute correlation values are grouped by increments of 0.1. When the viewer slides the scroll bar from right to left, the smallest absolute value groups will disappear from the display. This data reduction feature allows the viewer to focus on the associations that are significant. Figure 7 shows the process of data reduction by sliding the scroll bar from right to left.

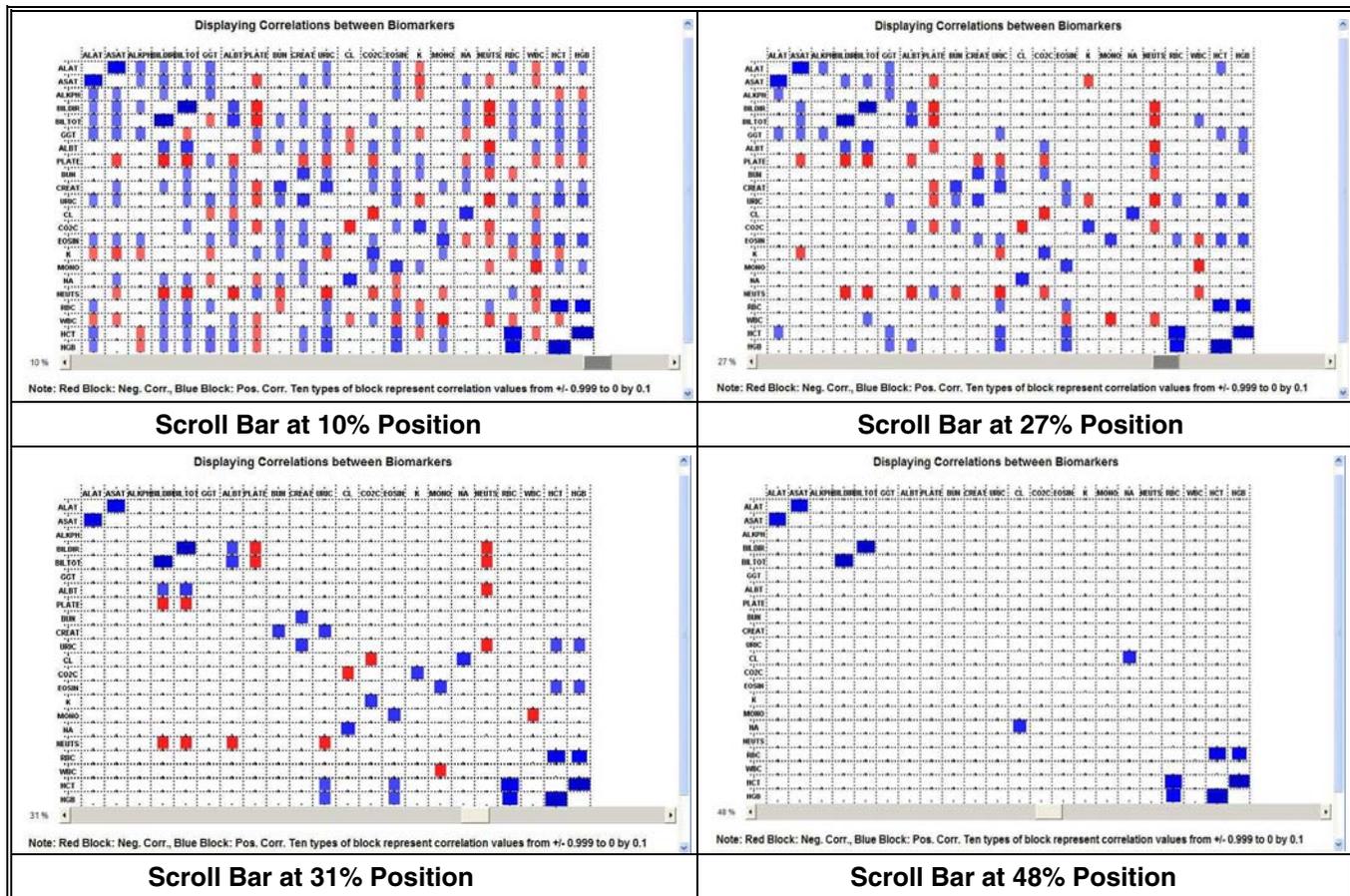


Figure 7. Using Scroll Bar for Subsetting Nodes and Links

III. VISUALIZATION OF THE CORRELATION MATRIX ITERATION PROCESS

The unordered data usually shows a random pattern. Good data ordering can rearrange the display, and relational patterns can be revealed. In addition, it will be easier to detect patterns of relations, similarities, trends, and data irregularities. The task of correlation ordering is to arrange the variables with similarities in a way that simplifies the pattern of relationships among variables. Friendly and Kwan (2002) proposed a general principle called “effect-ordered data display” which states that, with any data display (table or graph), unordered factors or variables should be ordered according to what we wish to show or see.

The other ordering principle is the placement of the variables onto the eigen-space for an optimal unidimensional or multiple dimensional order. This matrix iteration technique is based on the decomposition of a matrix to eigenvectors and eigenvalues for the correlation matrix \mathbf{R} as the entry matrix. This iteration process (or seriation) was first introduced by McQuitty (1968) and is stated as follows:

Given a proximity $p \times p$ matrix \mathbf{D} , a sequence of simple Pearson correlation matrices is iteratively generated from \mathbf{D} ,

$$\mathbf{R} = (\mathbf{R}^{(1)}, \mathbf{R}^{(2)}, \dots) \text{ where } \mathbf{R}^{(n)} = \Phi(\mathbf{R}^{(n-1)}), n > 1 \text{ and } \mathbf{R}^{(1)} = \Phi(\mathbf{D}).$$

The sequence of iterated correlation matrices will converge to a limiting matrix $\mathbf{R}^{(\infty)}$ in which all elements are 1 or -1. This limiting matrix $\mathbf{R}^{(\infty)}$ is in a convergence stage and partitions the p objects into two disjoint groups.

In mathematics, convergence describes limit behavior, particularly of an infinite sequence or series toward some limit.

This theorem has two components. 1) The Pearson correlation matrix is used as the entry for the iteration process, and 2) The iteration process will reach a convergence stage with a limiting matrix of $\mathbf{R}^{(\infty)}$. Exploratory visualization is a good approach to understand such a correlation matrix iteration process. Figure 4 is a representation of a Pearson correlation matrix. It is used as the entry for the iteration process. Figure 7 shows $\mathbf{R}^{(1)}$ to $\mathbf{R}^{(5)}$ from left to right in the first row, and $\mathbf{R}^{(6)}$ to $\mathbf{R}^{(10)}$ from left to right in the second row. It takes 10 iterations to reach the convergence stage. The limiting matrix $\mathbf{R}^{(10)}$ contains solid colors of blue and red with values of +1 and -1 in the matrix cells. $\mathbf{R}^{(10)}$ is also a representation of a correlation matrix map. The seriation cuts the correlation map into several blocks. Blocks for proximity matrices could be categorized as within-group blocks on the map diagonal and between-group blocks out of the diagonal.

The same iteration process can be observed in an eigen-space. In a visualization context, a convergent series provides the visual plots for observing the iteration process. When a series reaches a convergence stage, the plot of a current matrix is identical with the next sequence plot. The plot of a convergent matrix is also a limiting matrix plot.

Figure 8 shows the Pearson correlation matrix iteration process with the eigenvectors plotted in an eigen-space. The sequence reaches convergent stage at the 9th iteration.

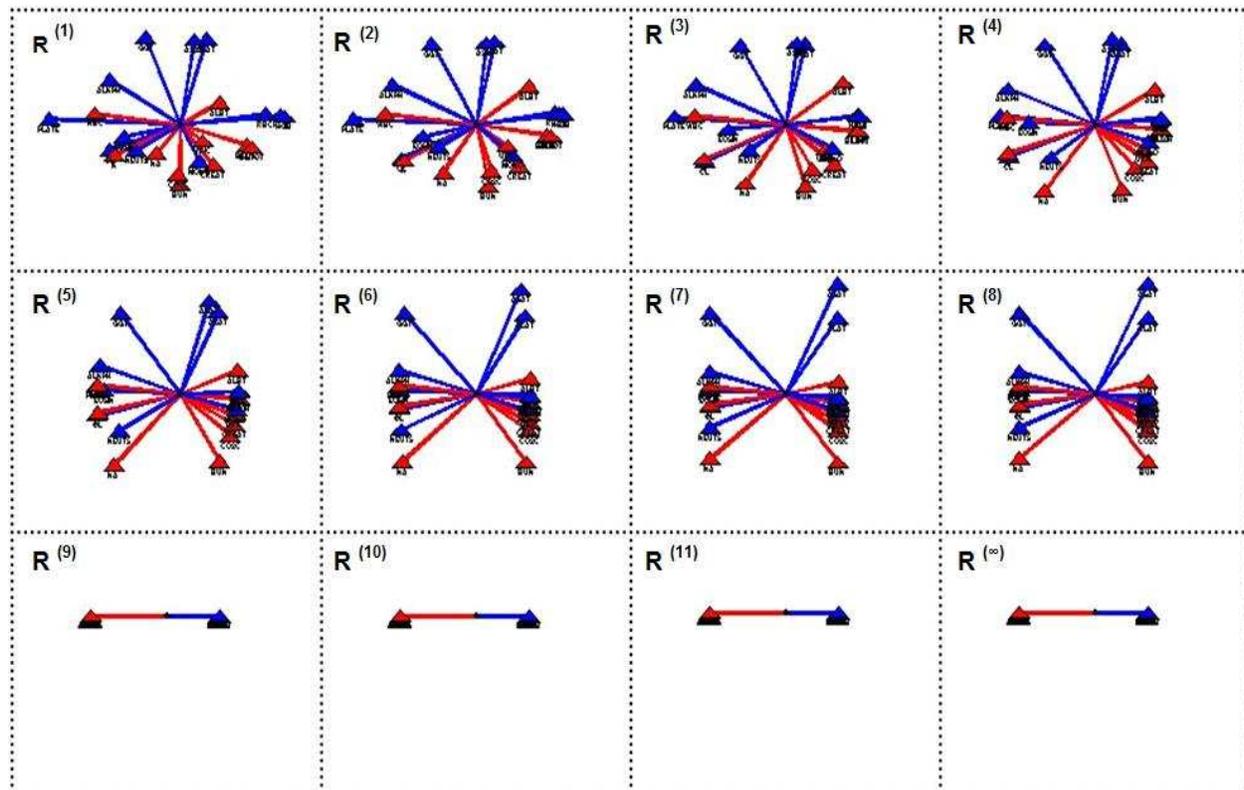


Figure 8. Pearson Correlation Matrix Converging Sequence in Eigen-Space

We have used dozens of different Pearson correlation matrices as entries for this matrix iteration process. In our experiments, all iterations can reach the convergence stage in 9 to 12 iterations. What happens to the other correlation matrices as entries for the iteration process? The matrices used as entries for these experiments are: Spearman rank-order correlations, and Kendall's Tau-bs. Figures 9 and 10 are produced by calling the

SAS/GRAPH DS2CONST macro. Figure 9 uses the Spearman correlation matrix as the entry for the iteration process.

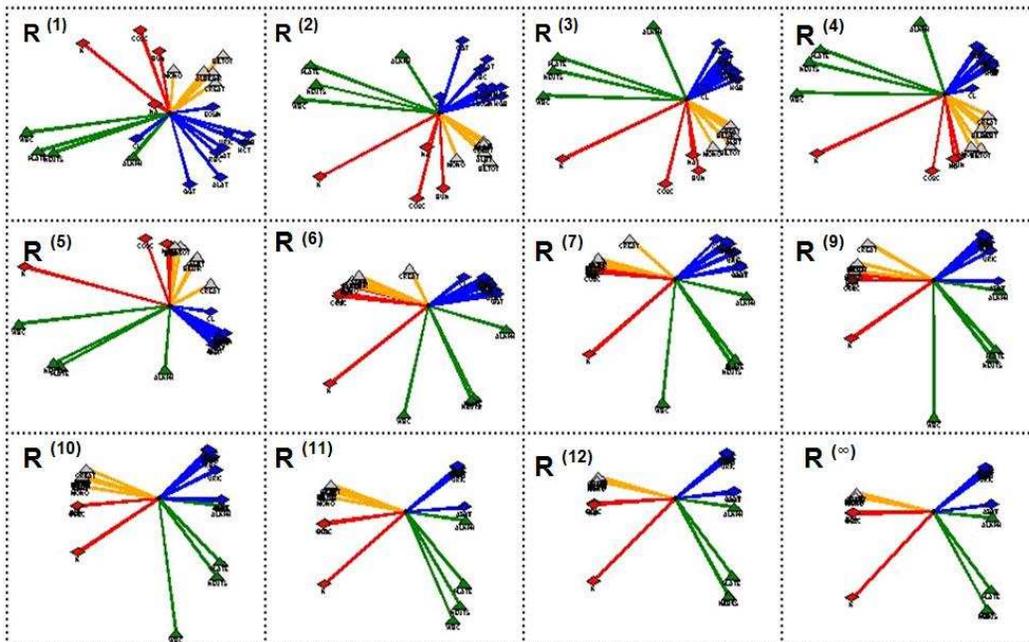


Figure 9. Spearman Correlation Matrix Converging Sequence in a Two-Dimensional Eigen-Space

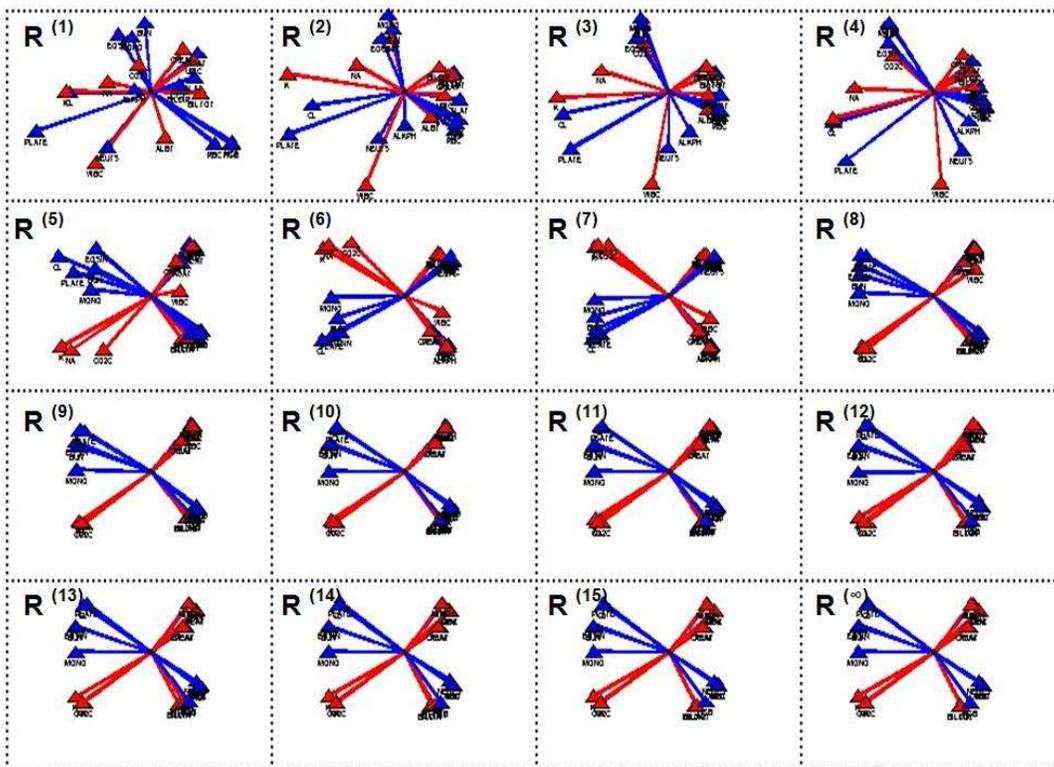


Figure 10. Kendall's Tau-b Matrix Converging Sequence in a Two-Dimensional Eigen-Space

IV DISPLAY OF MULTIPLE MATRICES WITH INTERACTIVE FEATURES

A clinical study usually involves multiple treatments and multiple time points. A desirable graph will display multiple correlation matrices on one page for easier comparison. This section discusses the display of multiple

matrices with interactive features. The display panel is divided into four sub-panels. Each sub-panel contains one matrix display. In the constellation diagram, the clinical laboratory test parameters are the nodes, and the correlation coefficients are the links. The colors of the link lines are the directions of the correlation - blue for positive and red for negative. The thickness of the line indicates the magnitude of the association. The interactive features in this application are data tips and an embedded scroll bar for subsetting nodes and links.

Figure 11 shows the correlation matrices of two treatments and two visits in a four-panel display for comparison. Figure 13 illustrates the interactive features of data tips for nodes and links.

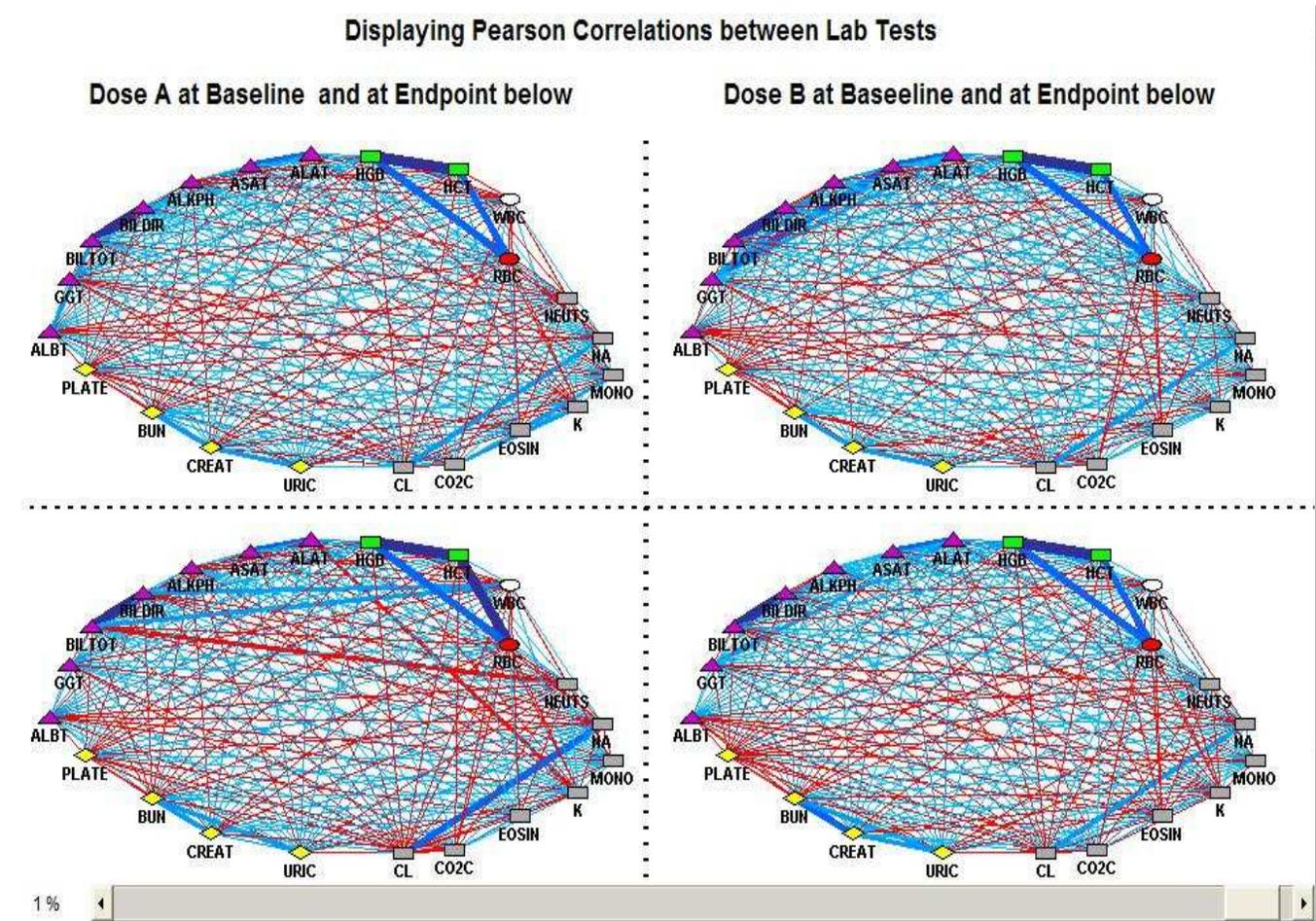


Figure 11. Multiple Correlation Matrices Display

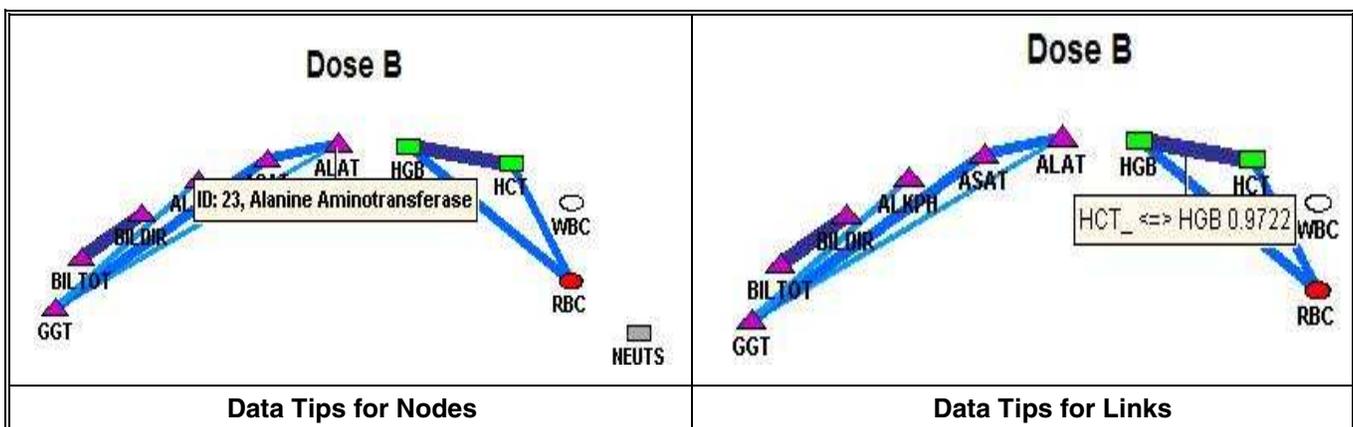


Figure 12. Interactive Features of Data Tips for Nodes and Links

Figure 13 shows the interactive features with the viewer's control of the scroll bar interaction. This interactive feature allows the viewer to remove the weak associations from the display. The viewer is able to focus on the stronger associations among the variables.

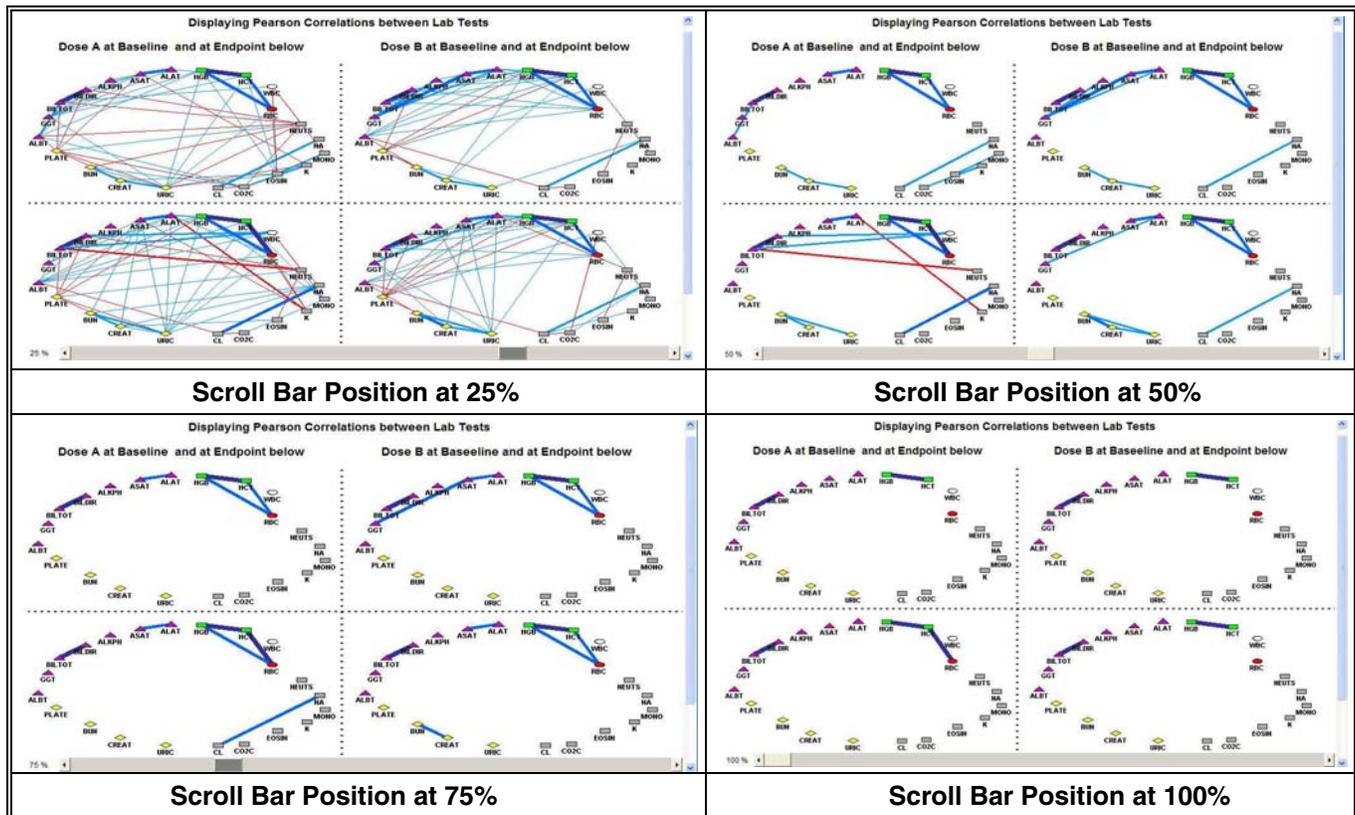


Figure 13. Using Scroll Bar for Subsetting Nodes and Links

V DISPLAY OF MRI CORRELATION MATRICES

Functional magnetic imaging (fMRI) is used to visualize brain function, by displaying changes in the chemical composition of brain regions. It can also display changes in the flow of fluids that occur over a timeframe of seconds or minutes. In fMRI studies, procedures are used to measure the hemodynamic effects of the metabolic changes in an active part of the brain during a task performed by a subject. Measurements are taken at regular intervals during the procedure. Then, the blood vessels show that the task changes are associated with specific regions. Various summary measurements are calculated across the time series to determine the hemodynamic response function (HRF) such as; positive area under the curve, negative area under the curve, maximum observed value, minimum observed value, etc. For these summary measures, correlations between the different brain regions can then be examined in order to determine inter-regional connectivity as reflected in the linear relationships among the brain activations.

The 24 specific brain regions that were activated in an auditory attentional paradigm are selected for analysis. The correlation matrix is a 24 x 24 matrix containing 276 observations to be examined. Each summary measure was considered for each subject subgroup. The SAS constellation diagram is used to summarize the correlation matrix visually. The display starts by mapping the 24 brain regions through the constellation diagram coordinates. Each node represents a region. Figure 14 illustrates the mapping of brain regions.

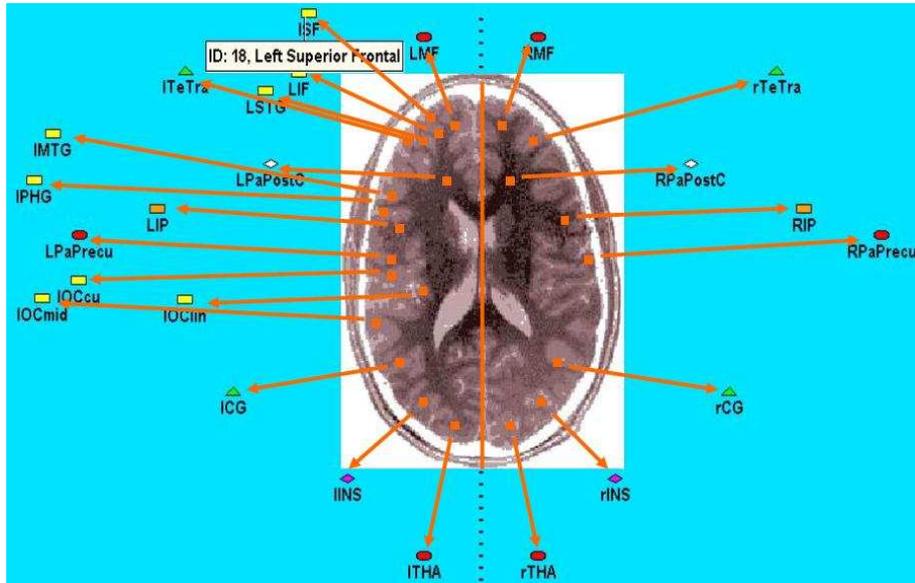


Figure 14. Mapping of Brain Regions to Chart Coordinates

The data from a fMRI correlation matrix are used to connect each of the pairwise combinations of brain regions. Different line widths are used to indicate the degree of correlation. Positive correlations are connected using blue solid lines, while negative correlations are connected with red solid lines.

From the output graph, we can easily see which regions are most strongly correlated in subjects with schizophrenia. The output also provides interactive graphics features for human interactions.

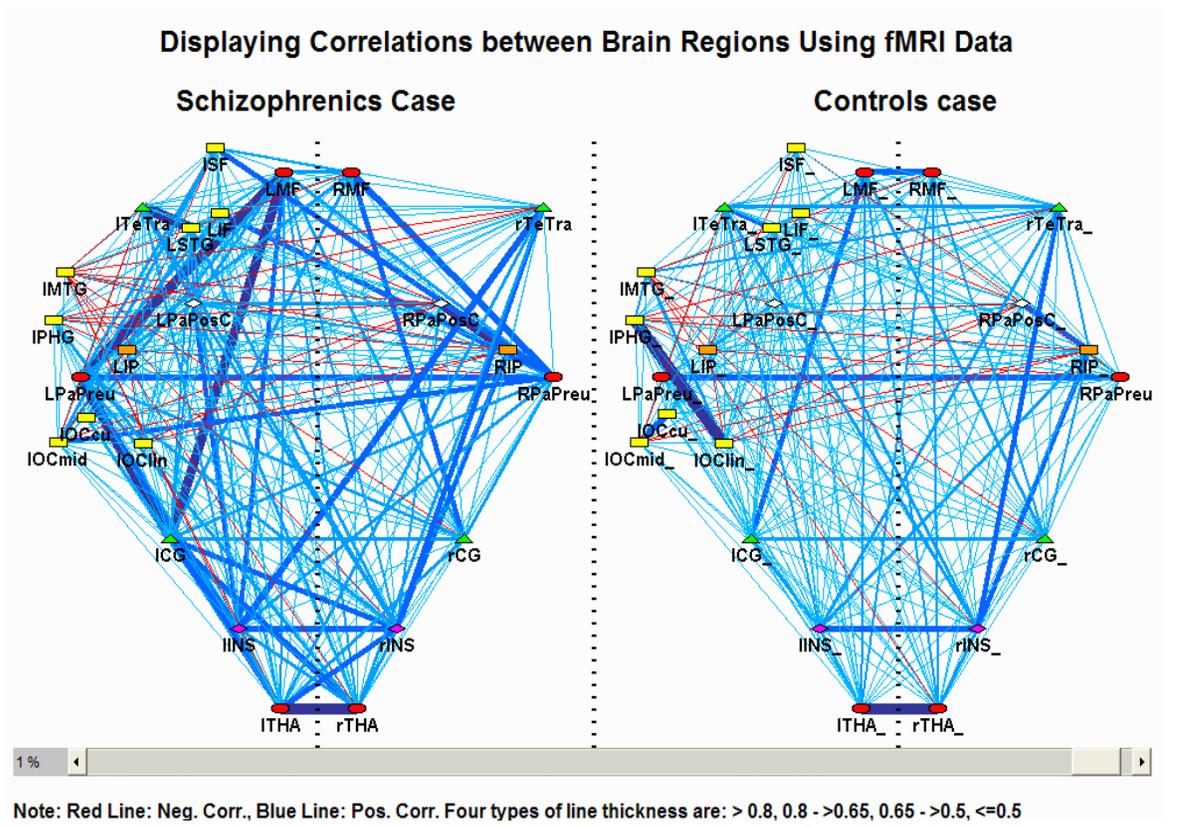


Figure 15. Displaying Correlations between Brain Regions Using fMRI Data

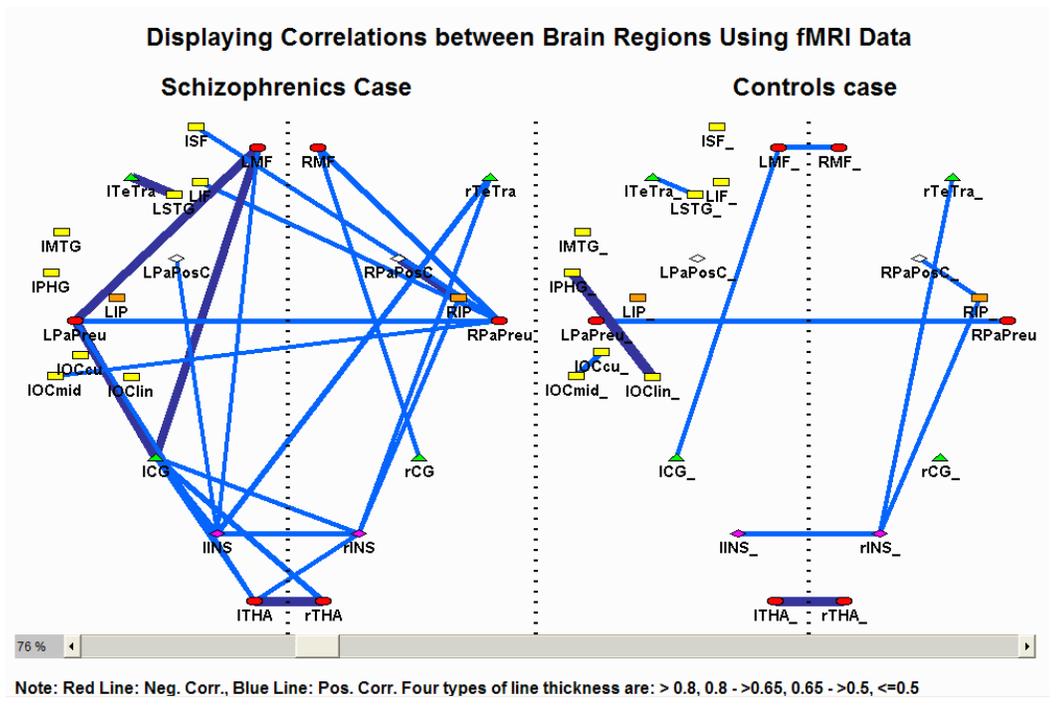


Figure 16. Interactive Features of Correlations between Brain Regions Using fMRI Data

VI. NEURAL NETWORK DIAGRAM

An artificial neural network is a computer application that attempts to mimic the neurophysiology of the human brain by finding patterns of data in a representative sample. Neural network is a class of flexible nonlinear regression models, discriminant models, and data reduction models, which are interconnected in a nonlinear dynamic system. This type of application can help an analyst make a prediction about clinical event for safety or clinical responder for efficacy. Figure 17 shows a theoretic neural network diagram.

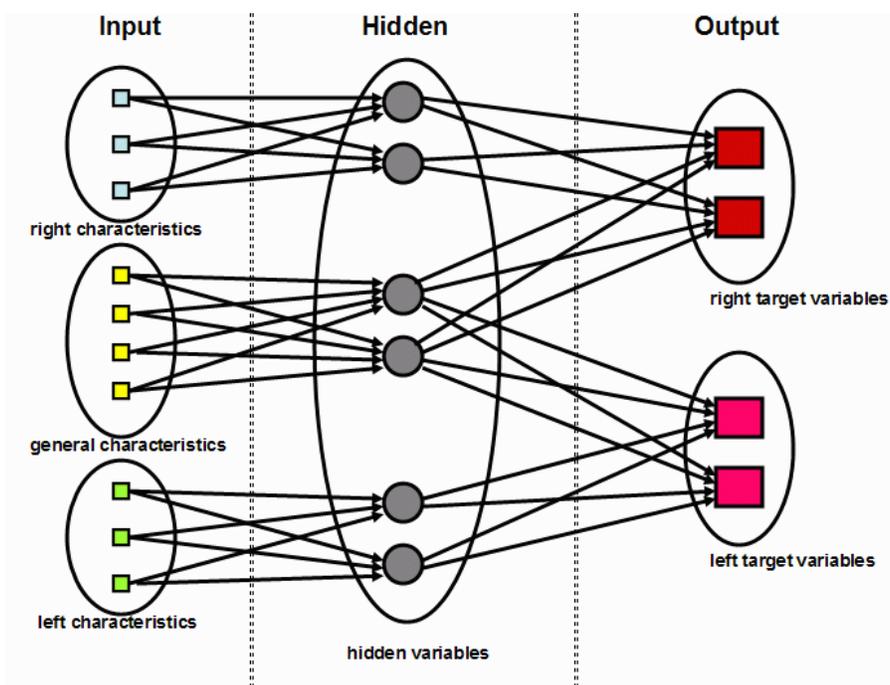


Figure 17. Theoretical Neural Network Diagram

Data Mining (DM) involves the use of software, statistical and graphical techniques to identify valid, non-trivial, previously unknown, interesting relationships/global patterns that exist in a large database. SAS EM provides a process flow diagram approach for DM tasks. The process flow diagram is a sequence of logical steps to build a DM project. Each step is defined by a visual flowchart depicting input, analyses and output. The figure below shows a sample process flow diagram.

Neural Network is one of the DM techniques used for supervised prediction problems. The building blocks of an artificial neural network are the input layer, hidden layers, and the output layer.

The input layer is composed of units that correspond to each input variable. The hidden layers are composed of hidden units. Each hidden unit outputs a non-linear function of a linear combination of its inputs – the activation function. The linear combination is the net output. The non-linear transformation is the activation function. The output layer has units corresponding to the target. With multiple target variables, there are multiple output units.

The network diagram is a representation of an underlying statistical model. The unknown parameters (weights and biases) correspond to the connection between the units. The data used for network diagram are from SAS EM neural network modeling. The weight and link output during the modeling process is saved as an SAS dataset and read into a SAS/GRAPH constellation diagram program to generate the network diagram. The diagram is illustrated below.

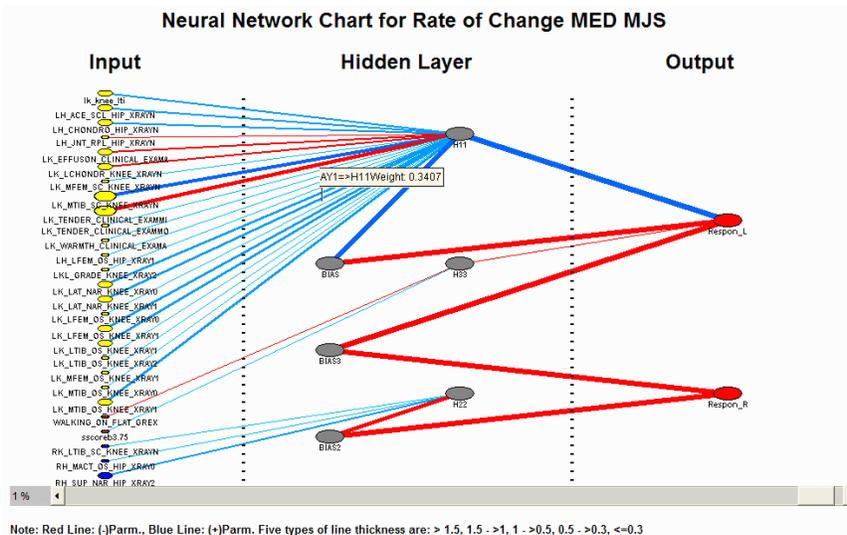


Figure 18. Sample Neural Network Diagram

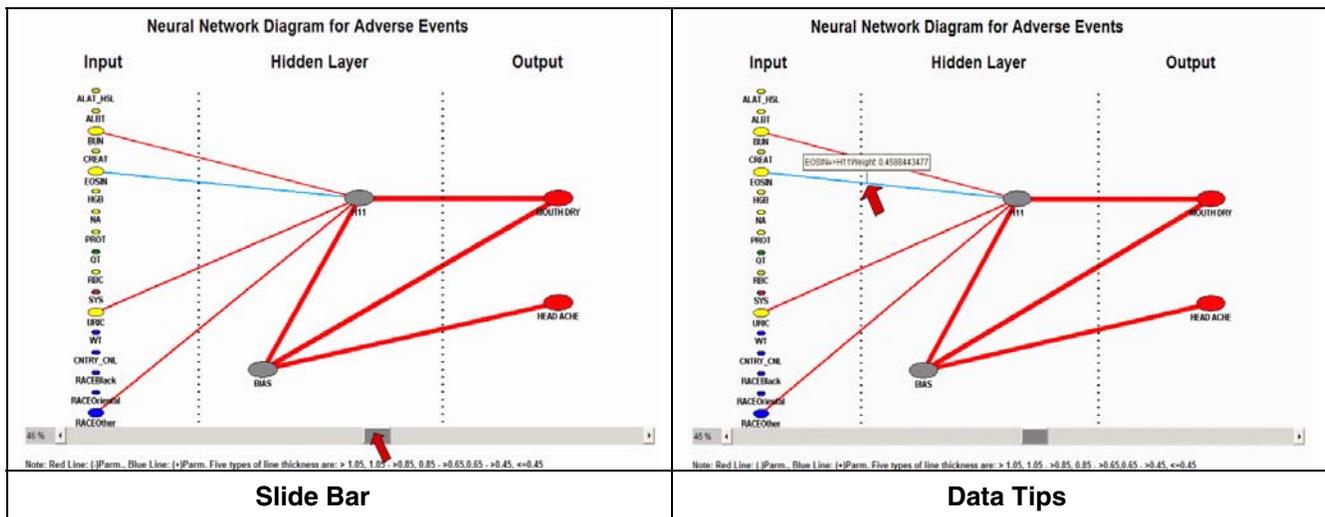


Figure 19. Interactive Features in Neural Network Diagram

Variables in the SAS dataset determine the size and color of nodes, as well as the width and color of the lines between nodes. The SAS Constellation diagram provides two interactive features: 1) slider bar which allows a user to choose how many of the links on the diagram are displayed, and 2) data tips – which is activated when the viewer moves the mouse over a graphical element causing a text box to display extra information.

CONCLUSION

The SAS constellation diagram is a very powerful tool for data visualization tasks. The DS2CONST macro allows users to specify the node's location on the display coordinates. The users can use this feature to design different application displays to fit their needs.

The interactive features are very desirable and enable the user to:

- 1) change graph properties without rerunning the job,
- 2) add additional graphical element information to the output, and
- 3) control the subsetting nodes and links so the viewer can focus on stronger association nodes.

The contributions of this paper are:

- 1) displaying of the constellation in tree structure form,
- 2) creating of interactive correlation matrix graphs,
- 3) illustrating of the correlation matrix iteration process,
- 4) displaying of multiple correlation matrices with interactive features on one page, and
- 5) displaying of a neural network diagram.

REFERENCES

- [1] Bilker, W.B., C.M. Brensinger, R.C. Gur, M. Herlin.(2005): *A SAS Macro for Visually Displaying Correlations between Brain Regions Using fMRI Data* , Proceedings of the 30th SAS Users Group International Conference, Paper 151-30
- [2] Bullmore, E.T., S. Rabe-Hesketh, R.G. Morris, S.C. Williams, L. Gregory, J.A. Grey, M.J. Brammer.(1996): *Functional Magnetic Response Image Analysis of a Large-Scale Neurocognitive Network* , NEUROIMAGE 4, pp. 16-33, #0026
- [3] Chen, C.H. (1996) *The properties and applications of the convergence of correlation matrices*. Proceedings of the Statistical Computing Section, pp. 49-54, Alexandria, VA, American Statistical Association.
- [4] Chen, C.H. (1999) *Extensions of generalized association plots (GAP)*, Proceedings of the Statistical Computing Section, pp. 111-116, Alexandria, VA, American Statistical Association.
- [5] Chen, C.H. (2002) *Generalized Association Plots: Information Visualization Via Iteratively Generated Correlation Matrices*, Statistica Sinica 12, pp-7-29, Academia Sinica, Taipei, Taiwan.
- [6] Friendly, M., (1991): *SAS System for Statistical Graphics*, SAS Institute, Cary NC,
- [7] Friendly, M., (2001): *Visualizing Categorical Data* , SAS Institute, Cary NC.
- [8] Friendly, M., (2002): *Corrgrams: Exploratory displays for correlation matrices*, The American Statistician, pp 1-18, V1.5, .
- [9] Friendly, M and E. Kwan, W.B., (2002): *Effect ordering for data displays*, Computational Statistics and Data Analysis, 37
- [10] Legat, D., (2005): *Matrix Visualization*, WDS'05 Proceedings of Contributed Papers, Part 1, pp 126-131

- [11] McQuitty, L. L., (1968): *Multiple clusters, types, and dimensions from iterative intercolumnar correlational analysis*, *Multivariate Behavioral Research* 3, pp 465-477.
- [12] SAS Institute Inc.,(2003): *SAS OnlineDoc 9.* , Cary NC. <http://support.sas.com/91doc/docMainpage.jsp>.
- [13] SAS Institute Inc.,(2004): *Base SAS® 9.1 Procedure Guide* , Cary NC. SAS Institute Inc.
- [14] SAS Institute Inc.,(2004): *Getting Started with SAS® Enterprise Miner™ 5.2* , Cary NC. SAS Institute Inc.
- [15] SAS Institute Inc.,(2004): *SAS/GRAPH® 9.1 Reference, Volumes 1 and 2* , Cary NC. SAS Institute Inc.
- [16] SAS Institute Inc.,(2004): *SAS/STAT® 9.1 User's Guide* , Cary NC. SAS Institute Inc.
- [17] Yeh, S.T.(2005): *Using Interactive Treeview Diagrams for Clinical Adverse Experiences Records* , Proceedings of the North East SAS Users Group 2005 Conference, Paper pos8
- [18] Yeh, S.T.(2005): *Customizing ODS Statistical Graphics* , Proceedings of the 30th SAS Users Group International Conference, Paper 183-30
- [19] Yeh, S.T.(2006): *Interactive Graphs from the SAS System* , Proceedings of the 31th SAS Users Group International Conference, Paper 181-31

TRADEMARKS

SAS is a registered trademark of SAS Institute Inc., Cary, NC, USA in the USA, and other countries.
® indicates USA registration.

Author
Shi-Tao Yeh, Ph. D.
(610) 787-3856 (W)
E-mail: Shi-Tao.Yeh@gsk.com