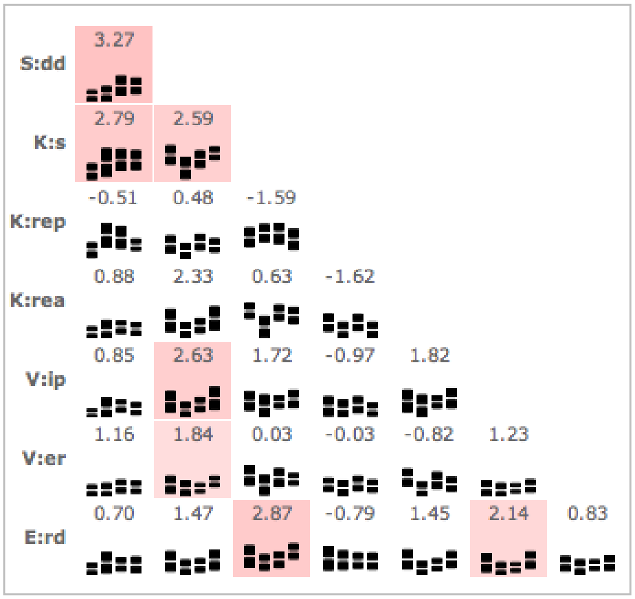
Design Challenge: Approaches to Visualizing Epistemic Frameworks

Chris Hinrichs
Nate Vack
Danielle Albers
CS838-2 Visualization

The Uncertain Matrix

This visualization is based on the presumption that in looking at epistemic frame network data, investigators will have a testable hypothesis about how the data will change across time, between subject populations, or in different assignment types. Additionally, this technique presumes that correlations between nodes in an epistemic frame are not specially meaningful.



S:dd	3.27						
K:s	2.79	2.59					
K:rep	-0.51	0.48	-1.59				
K:rea	0.88	2.33	0.63	-1.62			
V:ip	0.85	2.63	1.72	-0.97	1.82		
V:er	1.16	1.84	0.03	-0.03	-0.82	1.23	
E:rd	0.70	1.47	2.87	-0.79	1.45	2.14	0.83

In this visualization, the results of a statistical comparison between epistemic frames is shown in the same space as the original frame. The value of a statistical computation (in this case, the beta coefficient for a linear fit over time) is the primary numerical value. Below, the mean and standard deviation are shown for group of measurements. Finally, those cells in which the model fit passes a significance threshold (in this example, $p<0.05$) are colored, with saturation re-encoding the primary numerical value.

Implementation

To produce this visualization, the raw data that produced the small, dense networks were obtained, and a python program was written to parse the network data and produce the statistical values for the matrix. In this example, numpy, scipy, and a third-party Ordinary Least-Squares module were involved. The visualization code itself was written in HTML and Javascript; the mean and standard deviation bars are inline SVG elements produced with the Raphaël library.

Feedback and Iterations

In the first versions of this visualization, the matrix was always drawn as symmetric; much of the feedback about it suggested that that symmetry created false illusions of patterns, while not displaying more data. In this version, only the bottom half of the matrix is drawn; while it does indeed remove the redundant data, it creates a slight labeling problem – unless labeling both axes, one label will be missing.

More challenging was the warning from the Davids that other investigators in their field find matrix-type visualizations intimidating. Other than scrapping this method entirely, it was hard to think of ways to make that problem disappear.

Strengths and Weaknesses

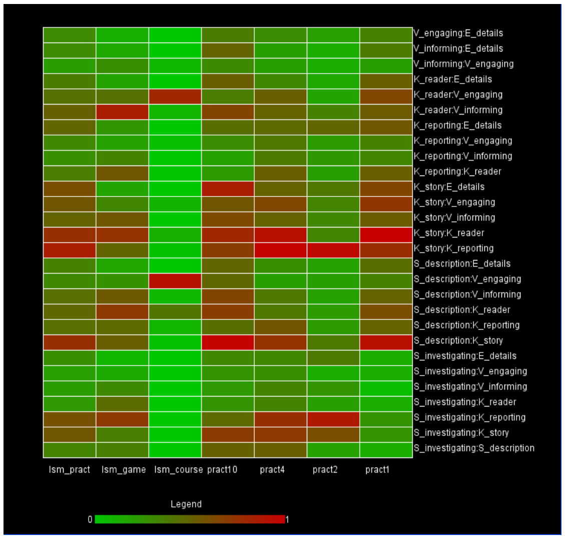
One of the main strengths of this approach is that it shows a relatively large number of data points in a small space – each cell contains a beta coefficient, four means, four standard deviations, and a binary indication of significance. More importantly, it encourages looking at the problem not as one of comparing relatively small numbers of dense networks of data, but rather, large numbers of related networks, many of which are very sparse. In many cases, subjects' networks had only a few nonzero values; seeing the correspondingly large errors bars tends to remind the viewer that the means are not estimated with a high degree of confidence. Additionally, it makes seeing the result of fitting a new model very fast, should more data or a better model be obtained.

There are several challenges with this approach, though. In using a matrix layout, horizontal and vertical position don't encode any particularly meaningful data – position is purely organizational. Additionally, these designs both fail to use the upper half of the matrix to any real advantage – either the data presented there are redundant, or that area of the display is unused. Designs in which the upper half of the matrix contain, perhaps, other matrix data are possible, but may be very hard to interpret.

In addition, color is probably not used to greatest effect. Saturation is used to re-encode a displayed value, but (as we recently learned) is not a particularly strong visual cue. Choosing colors in a more perceptually meaningful way may make this design stronger.

The Heat Map Vector Viewer

Description



The heat map vector viewer was created using the Processing language. In its essence, it stretches the upper triangle of a generic matrix into a vector with one entry for each entry in the matrix triangle. One such vector is created for each input data matrix. The values in the matrix are then normalized with respect to the entire data set and mapped to a red-green color ramp. Mousing over a particular cell provides the user with the raw data value associated with the cell.

Up to two rows or columns can be selected at any given time. Selection modes can be changed via 'r' (row) or ‘c’ (column) key presses. Up to two rows/columns can be selected at any given time and clicking on a selected row/column unselects the target. Two targets can be swapped by simply selecting them and pressing the ‘s’ key.

The underlying idea behind this visualization is to allow the user to examine any combination of properties quickly without the constraints of redundancy. By mapping numeric value to a relative color scale, patterns can be more rapidly perceived and are much easier to detect. Additionally, by utilizing the upper triangular matrix, redundant examples are removed and as a result more data can be fit on a single screen and redundant patterns are removed from the data to simplify analysis. The interaction allows the user to highlight fields of interest and compare for commonalities across either properties or matrices more easily by manipulating proximity.

The executable can be downloaded from http://pages.cs.wisc.edu/~dalbers/prototype/Bar_View.zip.

Purpose

The heat map vector viewer addresses the issue of being able to find patterns and approximate relationships across a fairly large number of matrices and properties. The color encoding allows for the relationships and patterns to really stand out, while the interaction effects allow for the user to customize the data groupings in order to more easily view desired comparisons.

Feedback and Iteration

The primary point of feedback for the heat map viewer is that it was very simple to interpret. However, the original blue-red color ramp was brought up as potentially problematic, which motivated the switch to a red-green ramp.

Color binning was suggested and could possibly be nice for a more summary-based interpretation of the view. However, the relational nature of using a color ramp made differences between different values in the different cells more natural to interpret. For a similar reason, the suggestion to bin data values was not implemented, although it would certainly prove useful if the matrices grew substantially larger in size or number.

Addressing Design Principles

The heat map vector viewer and circle segment viewer discussed next were heavily impacted by the discussion of color that took place in class. The selection of the red-green ramp was made after considering some of the literature in class and using that information to better understand the perceptual difficulties of the non-opposing color ramp. Similar reasoning was used to select the colors of the highlight for the selection mechanisms.

Further more, the implemented and proposed interaction effects help to manipulate some of the encoding concepts presented in class. By giving absolute information in the form of a mouseover, the user can get detail on demand. Thus, the visualization is not cluttered by extraneous data but instead displays the data when prompted by the user. Furthermore, the rearrangement of the data manipulates the power of positional encoding, allowing the user to chose the positional orderings that best facilitate their exploration. In this manner, the user has control over a very powerful dimension of the display in order to tailor the viewpoint to fit the queries being made.

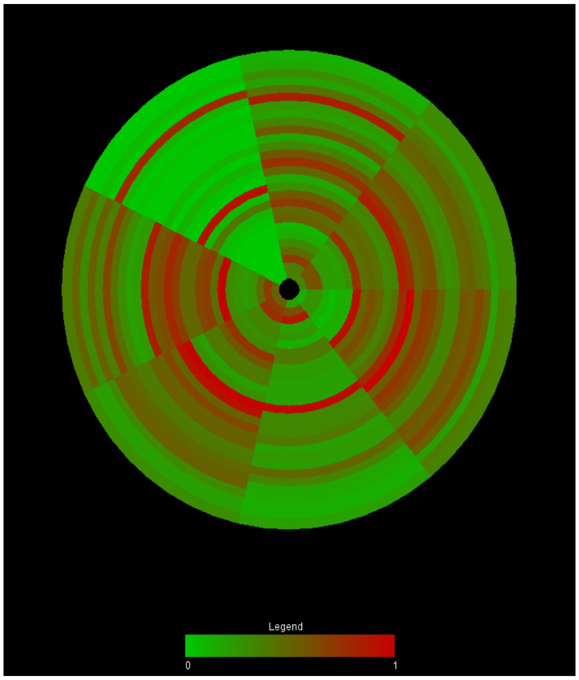
Strengths and Weaknesses

The heat map viewer is very good at quickly conveying relationships between both matrices and properties as well as within matrices and properties. The interaction effects allow the user to customize the viewpoint in order to maximize these comparisons with respect to their own queries.

However, because the visualization utilizes a color ramp, it may be difficult to distinguish beyond general value relationships. This issue is worsened as the number of matrices or properties is increased to large numbers (approximately 40 or larger). Additionally, the program requires that all the matrices be of identical dimensions, which could be a potentially limiting factor. Finally, individuals who are red-green colorblind may have a difficult time with this tool for the obvious reasons.

The Circle Segments Viewer

Description



The circle segment viewer was also created in Processing and manipulates the data from the upper triangle of the data matrices. Each property of the matrix is mapped to a particular radius and each matrix to a particular wedge of the wheel. The values are then normalized with respect to the entire data set and mapped to a red-green color map.

This idea is based on the idea of allowing the user to quickly draw comparisons between matrices and variation amongst properties. By mapping the relative values to color, the user can very quickly get a sense of the differences amongst values and can more easily detect patterns between different data sets.

Had there been more time to compute the reversals of the encoded geometries, ideally, interactions would have been integrated to greatly improve the usability of the visualization. A mouseover could be used to reveal which data matrix a particular wedge represents and the property pairing that any given segment represents. Additionally, the user could have been allowed to swap property segments and matrix wedges to manipulate the proximities of the different data points to more easily make certain comparisons.

The executable can be downloaded from http://pages.cs.wisc.edu/~dalbers/prototype/Wedge_View.zip.

Purpose

The circle segment viewer highlights differences between matrices for a reasonably large number of matrices. The color ramp very quickly allows the user to view similarities between different wedges. While the color ramp paired with the ring technique used to visualize the different properties does facilitate comparisons between properties, the general matrix differences are much more the focus of this visualization.

Feedback and Iteration

People generally seemed to like the idea of the wedge-based encoding of the matrices for the simplicity with which differences can be spotted using it. However, the initial encoding of the data all the way to the center of the circle drew questions about possible false exaggeration of the importance of the center property. To address this, the first data property was drawn out into a ring and the center was colored to match the background property.

Additionally, the difficulty of interpretation was suggested as a possible issue. With more time, this could have been addressed by the inclusion of mouseovers giving summary data of the value represented by the point. Finally, based on the input from the heat map view, the color scheme was adjusted. Comparative issues with respect to proximity were also brought up, which, as mentioned in the summary of the visualization, could be adjusted by allowing the user to manipulate the positioning of different elements in the view.

Addressing Design Principles

See the heat map vector viewer section on addressing design principles for a discussion of how course design principles were used for this viewer.

Strengths and Weaknesses

The primary strength of the circle segment viewer is that it is very easy to spot general trends between matrices. Additionally, should the changes mentioned in the design description be implemented, it would allow extremely flexible comparison between matrices and different properties.

However, the circular nature of the view can cause some perceptual distortion in the size of the bands. Additionally, it can be difficult to interpret changes between different properties pinned in the middle of the visualization may be difficult to interpret without moving them to a more extreme ring. The visualization also falls victim to the weakness of the red-green ramp described in the heat map vector.

Group Dynamics

As far as group cooperation went, it was relatively ad-hoc, with a few brainstorming and comparison meetings. Most of our ideas were generated and tested by the individual team members; a large part of the utility of group meetings was to see the fundamentally different approaches we took to the problem. We used a process of brainstorming and collaborative feedback from both the class feedback and each other in order to help to refine our ideas to develop the final products.