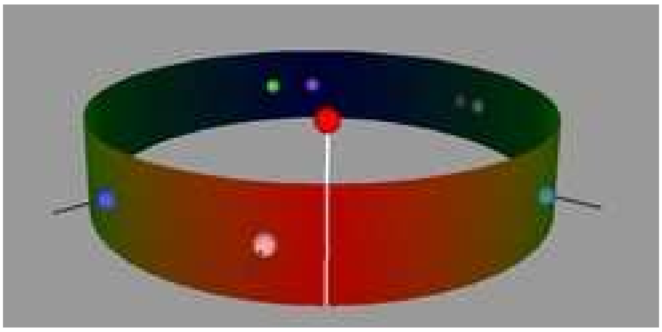
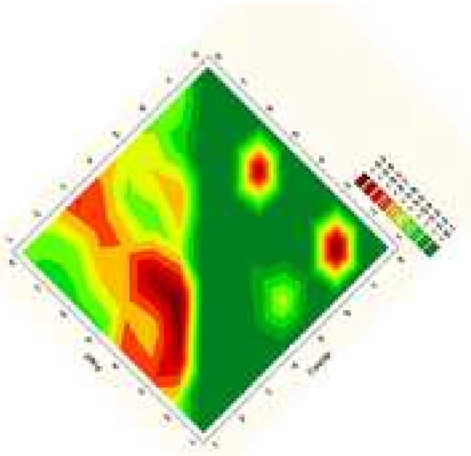
CS839:
Epistemic Net Visualization Challenge

Chaman Singh Verma*, Kim Nak Ho † and Emma Turetsky ‡

March 17, 2010





Abstract

Epistemic networks consist of sets of correlated concepts. The data is represented in matrices where entry $M(i, j)$ reveals the strength of association between concept i and concept j. Designing visualizations of these networks to make comparisons between different datasets poses many challenges, especially when the networks become large and relational ties are prevalent. In this project, we present two ways to represent these network aiming at large networks.

1 Introduction

Visualization of social networks is challenging because of the visual pollution generated by large number of colors, overlapping connections and space crunch. Especially in networks where the nodes are densely interconnected with diverse node strength, the prevalent visualization methods of social network diagrams are less than effective to read the data and derive meaningful patterns. We propose two visualization methodologies which can significantly remove the visual pollution and compare two or more networks efficiently.

2 Data and Task Abstraction

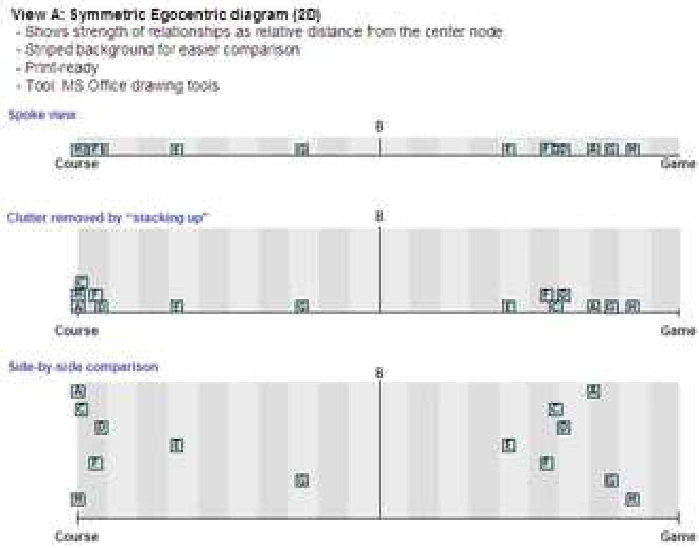
To develop solutions, our team chose collaborative brainstorming over aggregating individual member assignments. Ideas were discussed in each stage extensively through face-to-face meetings. The three principles we set were —em originality, clutter removal and *pattern recognition* effectiveness.
To generate ideas, we loosely implemented the four steps of Munzner's nested model(Munzner, 2009).

1. *Problem identification*, was to understand what merits the visualization should give to the domain science research problem. We identified the task as twofold: one is to browse the network ties of a specific data node, the other is to look at the shape of the general network to derive a pattern. In both tasks, differences

*csverma@cs.wisc.edu
†nkim3@wisc.edu
‡turetsky@cs.wisc.edu

between various datasets should be able to be compared easily. We concluded that combining the two tasks into a single visualization is hard to achieve and inefficient because the data gets easily cluttered. As a result, we decided to tackle the two tasks separately into an egocentric view and a general view.

2. *Abstraction*, started with critiques on existing models which were shown as part of the initial challenge explanation. The common problem was that egocentric views worked while general views did not, largely due to data cluttering. Moreover, we decided not to merely upgrade existing approaches but to focus on out-of-the-box thinking, given the purpose to the challenge. In the initial discussion, ideas to utilize social network diagrams and adding visual elements such as bar graphs to the data matrices were discarded. As a result, a *cylindrical view* was introduced which focused intensively on data clutter removal by limiting the dataload on one viewpoint and utilizing parallel coordinates.

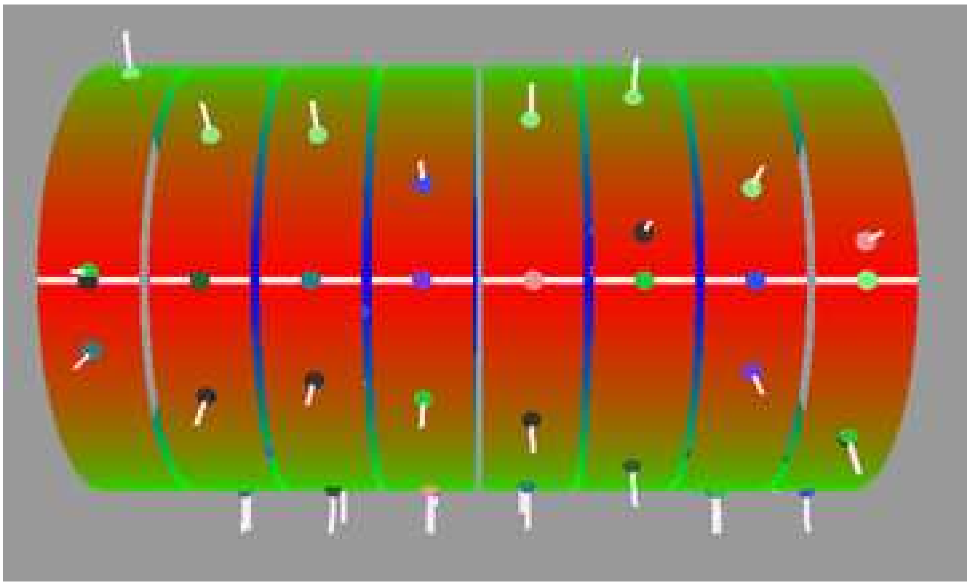


To deal with problems not addressed by the egocentric cylinder view, we expanded our ideas in two ways. One was to build a 2D representation suitable for print, and the other was to build a general view. For a non-interactive 2D version, various dataset comparisons were attempted including a symmetrical distance method. For the general view, an initial idea was to build a 3D surface graph on top of a data matrix. However, it was not effective to represent patterns and not easily comparable with other datasets. What eventually stayed was the top-view of the graph, which was essentially a wafer plot. To further remove data clutter, we decided to drop one of the two symmetrical triangular data cell areas inherent in the network data. Also we integrated the symmetrical comparison from our 2D approach by putting a different dataset in its place. It was based on the idea that if an human eye can detect symmetry easily, so it can detect non-symmetry where symmetry is expected.

3. *Encoding*, focused on position and color suitable for representing metric and categorical data. In the cylindrical view, we used superimposition to compare datasets using and used juxtaposition for the symmetrical view.

4. *Implementation*, we used simple drawing programs to test 2D layouts, programmed a cylindrical view prototype in OpenGL, and used existing statistical software to draw wafer plots which were reedited to symmetrical comparison mockups in graphic editors. In testing these methods, we used actual datasets instead of synthetically generated data to visualize real research results.

3 Cylindrical View



In this view, each concept is represented as a spherical ball (colored arbitrary but distinct for each concept). The center line which lies on the cylindrical surface and drawn with the white line represents the reference line. Association values monotonically decrease away from the reference line and becomes zero at diametrically opposite side. All the association values are normalized in the range $(0.0, 1.0)$ where one represents very high association and zero represents no association and with this range the surface of the cylinder is also colormapped. User can rotate the cylinder at any given angle and with the associated color it is easy to find an association value approximately. With this view, if an user is interested in knowing high correlated concepts, then he/she rotates the cylinder so that the reference line is in the front (Red Color) and if he/she wants to understand low correlation data, the cylinder is rotated so that blue color is in front.

3.1 Design Principles

It is a long-standing theory that the one of the purposes of visualization is to separate signal from noise in an effective and relevant manner (see Tufte, 1997). To discard a large number of nodes from the first view, a well-known heuristic that people are interested in knowing either high or low correlation data but not both at the same time can be used. Removing unwanted regions is the prime design goal of the cylindrical view.

3.2 Advantages of the Cylindrical View

The cylindrical view has advantages than many other views.

- **Sparsity** The data have been distributed along the circumference and if there are two or more concepts having the same association values, than they can be placed along the height of the cylinder and therefore no overlapping will occur.
- **Clustering** Data are naturally clustered on the surface and therefore, it is easy to find low/high correlated data immediately.
- **No Visual Pollution** Unlike graph representations, there is no visual pollution, no connection line and no cluttering, which results in a clean visualization.

3.3 Disadvantages of the Cylindrical View

However, there are some weaknesses in the cylindrical design.

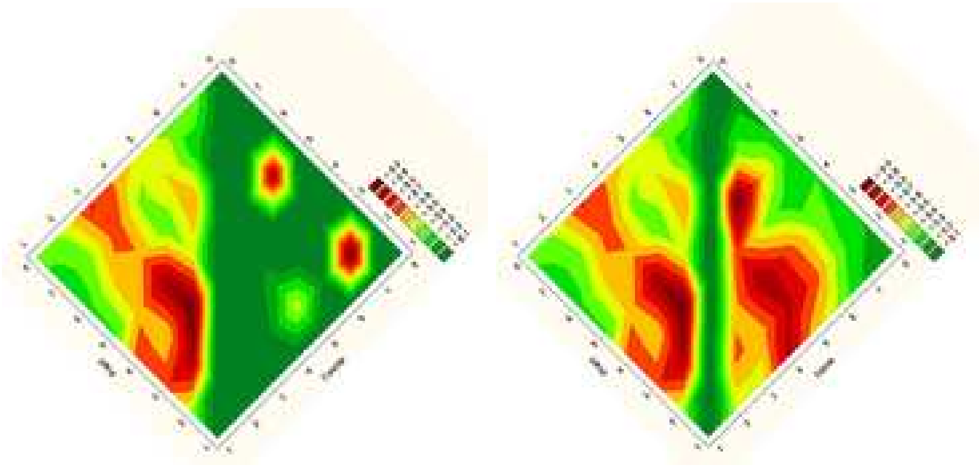
- **Symmetric Placement** Seen from the reference line, the concepts can be replaced either left or right side (because they have the same value). Therefore, an user have to rotate cylinder in front and back to see similar concepts. This could be cumbersome and inconvenient after some time.

- **Three Dimensional** This representation is not printer-friendly because of the 3D view and therefore could be used only on the display devices.

3.4 Implementation Details

This module was implemented with *libQGLViewer* which is one of the higher level library on top of OpenGL. The entire user defined code consists of less than 100 lines of code. The input to the program is any *csv* format datafile.

4 Heat-Map Visual Representation



In this approach, heat map view shows a general view of the data matrices. Since the matrices are always symmetrical, one triangle of one matrix of data was removed and replaced with another. Each triangle half was normalized from 0-1 and for each point, a vertex of data points on the graph, is mapped to a color based on that number (from a gradient of green to red). The graph is then rotated by 45 degrees so that the two triangles would be mirror images if the data were the same.

4.1 Design Principles

This view mainly focused on the idea of symmetry leveraging on highly developed cognitive perception skills in humans It sets up a view that easily allows for these comparisons. It was meant to be a simple picture that could be interpreted quickly and simply, to derive patterns in relatively low-level feature perception (see Ware, 2008).

4.2 Advantages of the Symmetrical View

The strength of the symmetrical view lies in the simplicity, it is very simple and easy to understand. It uses the redundancy of the data to allow for one image to compare with rather than two. Although our current prototype lacks interactivity, but symmetrical view has the potential to incorporate multiple view which our future versions may include. We may add new capabilities to move columns and rows and mirror it on the other side to allow for easier visualization of cluster. We may also experiment with various ways of displaying data triangles with same orientation, as it may open new possibilities for visual comparison.

4.3 Disadvantages of the Symmetrical View

Unfortunately, the method implies a connection between the data sets as the map is interpolated between the known points, leading to false positive similarities.For example, in the image four, there is a similar curve of red in both triangles. Their difference in the direction does not indicates that they share same characteristics, but the shape similarity can still mislead attention. Since the symmetry view may not allow comparing more than two data sets in one visualization, we discussed the possibility of implementing more symmetrical wedges to address this problem. However, we concluded it may likely result in data cluttering.

4.4 Implementation Details

The heatmap was implemented using *STATISTICA*, which is a data analysis and statistical software tool. It's interface is very much like excel, so the data was modified in the excel spreadsheet form and then transferred into *STATISTICA*, where the wafer plot was used. The picture was rotated afterwards.

5 Reference

Munzner, T. (2009), *A Nested Model for Visualization Design and Validation*, IEEE
Tufte, E. (1997), *Visual Explanations: Images and Quantities, Evidence and Narrative*, Graphics Press
Ware, C. (2008), *Visual Thinking for Design*, Morgan Kaufmann