APPROVAL SHEET

Title of Thesis: Quantitative Data Visualization in Compartmented Force-Directed Graphs using Calibrated Columns

Name of Candidate: Aparna Viswanathan Kaliappan
Master of Science, 2017

Thesis and Abstract Approved: ________________________________
Dr. Jian Chen
Assistant Professor
Department of Computer Science and
Electrical Engineering

Date Approved: _________________

ABSTRACT

Title of Document: QUANTITATIVE DATA VISUALIZATION IN COMPARTMENTED FORCE-DIRECTED GRAPHS USING CALIBRATED COLUMNS

Aparna Viswanathan Kaliappan
M.S. in Computer Science, 2017

Directed By: Dr. Jian Chen
Assistant Professor
Department of Computer Science and Electrical Engineering

Networks are commonly used to represent data and relationships. However, when mapping nodes to quantitative data, it is often difficult to accurately identify both the values of individual nodes and the overall relationships among nodes. Simultaneous group detection and precise quantitative data reading are necessary for scientists interpreting critical data. Here, we hypothesize that the calibrated columns method for encoding large-range quantitative values will provide a more accurate reading of data values than the common approach of variable-area circles. We also hypothesize that the addition of subtle halos around nodes will support accurate grouping of spatially distributed nodes in a network. We have conducted a pilot study with seven critical tasks in order to understand quantitative data reading and group inferencing in networks having up to three-levels of complexity. Our results show that (1) network size has a significant effect on confidence levels in grouping tasks; (2) the grouping

encodings do not have a significant effect on confidence levels, but do significantly affect accuracy in the overall task; (3) the quantitative data encodings do not have a significant effect on confidence levels, but do significantly affect accuracy when determining exact node values; and (4) the colored halos and calibrated columns encodings are particularly useful in tasks involving both precise and global perception of the network. This work has contributed to understanding effective construction of quantitative networks and their groupings and our results suggest design guidelines broadly applicable to inform visualization design in domains such as biology and the social sciences.

QUANTITATIVE DATA VISUALIZATION IN COMPARTMENTED FORCE-DIRECTED GRAPHS USING CALIBRATED COLUMNS

By

Aparna Viswanathan Kaliappan

Thesis submitted to the Faculty of the Graduate School of the University of Maryland, Baltimore County, in partial fulfillment of the requirements for the degree of
Master of Science
2017

Dedication

This thesis is dedicated to my loving parents who have made several sacrifices in their own lives in order to make my life more comfortable. I would like to thank them for their constant support and unconditional love. I will forever be grateful to them for providing me with the opportunity to acquire a rigorous education, for teaching me my mother tongue, and for instilling cultural values in me. This thesis is also dedicated to my brother, my teachers and professors, and my well-wishers, who have all helped me to become the person I am today.

Acknowledgements

I would like to thank Dr. Jian Chen for her constant support and encouragement while I worked on my thesis. She was very prompt in answering emails and consistently made herself available, even when her schedule was very busy, so that I could ask questions and work harder on my research. She never hesitated to let me know if I was making a mistake in my research and also never hesitated to praise me if I had done something well.

I would also like to thank Dr. Penny Rheingans and Dr. Charles Nicholas for agreeing to be on my Master's thesis committee.

Furthermore, I would like to thank Wesley Chiou, Guohao Zhang, and Henan Zhao from the Interactive Visual Computing Lab (IVCL) at UMBC. They were always willing to help in a timely manner when I had questions about conducting the empirical study and analyzing the results.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Networks are commonly used to represent data and relationships in many fields, such as biology. However, when mapping nodes to quantitative data, it is often difficult to identify both overall relationships among nodes, as well as the values of individual nodes. This problem can become more complex when a node's value is difficult to distinguish from another node's value, due to the similarity in the shapes and sizes of the nodes.

Networks commonly represent quantitative data by varying a circular node's size based on its calculated area, which is proportional to its value. A circular node only allows the perception of size based on one variable, which is the radius of the node. Therefore, it is often difficult to convey precise quantities using this shape. On the other hand, a rectangular node can allow the perception of size based on two variables, which are the height and width of the node. We hypothesize that the method of calibrated columns, as explained by Jacques Bertin in his book "Semiology of Graphics", will be a visualization technique that can better resolve the aforementioned difficulties and provide a more accurate perception of data values (2010). Calibrated columns have been commonly used to visualize quantitative data on maps, but not on networks. So, this is a novel approach. This encoding is further discussed in Section 2.1.1 of this document.

Furthermore, another difficulty in networks is determining the group membership for a node or a set of nodes. This problem becomes especially difficult in networks containing several hundreds of nodes. Typically, nodes in the same group are clustered around the same parent node and colored with the same color. This coloring can be performed in two ways. The first method is to color the inside of a node completely. The second method is to add a slightly opaque colored halo around the node. We hypothesize that the method of adding halos around nodes could improve the accuracy of grouping tasks in a network, due to the fact that these nodes will occupy a larger spatial area on the monitor, than the colored nodes.

1.1 Thesis Contribution

For my thesis, I applied Jacques Bertin's technique of calibrated columns to visualizing nodes in a force-directed graph (Bertin, 2010). I compared this novel approach with the traditional variable-area circles method to test performance on quantitative tasks in a network. Furthermore, I compared the approach of adding colored halos around nodes with the approach of coloring the nodes to test performance on grouping tasks in a network. I conducted a pilot study with six participants to evaluate my hypotheses regarding these network visualization encodings.

Chapter 2: Background and Related Work

Much research has been done about node-link diagrams and the accuracy of completing different tasks using different visualization techniques applied to these diagrams. Furthermore, books have been published regarding specific terminology for describing components of a graphic, which are commonly used by graphics and visualization researchers today. Finally, research about the perception of size and color has also been done, to ensure the visible distinction between two different sizes or two different colors used in a visualization. This background and work is discussed in this chapter.

2.1 Graphics

In his book titled "Semiology of Graphics", Jacques Bertin details concepts and examples to illustrate methods that can be used to construct graphics in such a way that the data can be displayed efficiently and read quickly, in order to answer questions about the data accurately (2010).

2.1.1 Point Visualization

Typically, nodes in a network are represented using circles. The circles can be of fixed-size when the purpose of the visualization is to establish simple relationships between nodes. However, when nodes are mapped to quantitative data, it is common to vary the circular node's size based on its calculated area, which is proportional to its value. We know that the area of a circle can be written as

$$A = \pi r^2$$

where r is the radius of the circle. Since a node's area increases in proportion to its quantity, this means that if a node has a larger quantitative value associated with it, then it will have a larger area. Similarly, if a node has a smaller quantitative value associated with it, then it will have a smaller area. An example is shown in Figure 2.1.

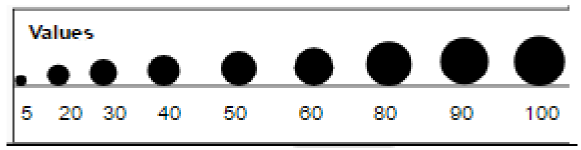

Figure 2.1: Legend for the Variable-Area Circles Encoding

Bertin discusses these methods in relation to geographic maps where fixed-size circular points can be used to illustrate the concentration of a quantitative attribute in a geographic location, and variable-size circular points can be used to illustrate the value of a quantitative attribute in a geographic location.

Bertin further discusses another method of point representation in maps called "calibrated columns". In this representation, data points are shaped as rectangles, and the heights and widths of the rectangles are varied so that the area of the rectangle is proportional to its value (Bertin, 2010). The area of a rectangle can be written as

$$A = H * W$$

where H is the height of the rectangle and W is the width of the rectangle. In this method, at a constant height, the areas of these bars are proportional to their widths. A series of quantities can be created which correspond to the bars in this same proportion. In Figure 2.2, the node heights are constant for the quantities 20, 40, and 80. The width varies according to the data quantity interval. Suppose that the data range is fixed at (0, 100], meaning that values are between 0 exclusive and 100 inclusive. Further suppose that rectangles mapped to values less than 20 have a width of x. Then, rectangles mapped to values between 20 inclusive and 40 exclusive will have a width of 2x, rectangles mapped to values between 40 inclusive and 80 exclusive will have a width of 4x, and rectangles mapped to values greater than 80 will have a width of 8x. Given these fixed widths, the heights can easily be computed using the area formula.

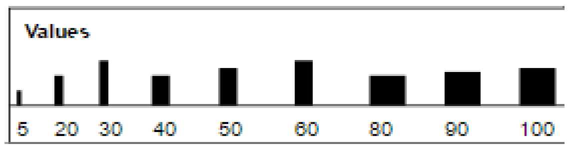

Figure 2.2: Legend for the Calibrated Columns Encoding

The variable-size circular points have changes in area that are only dependent upon a single variable, r, which is the radius of the circle. However, the calibrated columns have changes in area that are dependent on two variables, H and W, which are the height and width of the rectangle, respectively.

2.1.2 Retinal Variables

Bertin additionally discusses six retinal variables that can be used to provide additional encoding of information, other than the X and Y spatial dimensions in the graphic plane. These retinal variables are color, orientation, shape, size, texture, and value (Bertin, 2010).

2.1.3 Levels of Reading

Bertin further discusses three levels of reading at which questions can be asked about a graphic. The first level is called the “elementary level of reading”, where questions can be asked about a single data element or a single category in the graphic. The second level is called the “intermediate level of reading”, where questions can be asked about groups of data elements or groups of categories in the graphic. The final level is called the “overall level of reading”, where questions can be asked about the graphic as a whole (Bertin, 2010).

2.2 Force-Directed Graphs

Force-directed graphs are a type of visualization used to visualize relationships between nodes. In these graphs, the spatial locations of nodes are determined based on forces that are applied to nodes, so that nodes with several links will attract, while nodes with fewer or no links will repel. There are several different algorithms that can be applied to calculate these forces and draw such diagrams (Kobourov, 2013). However, existing tools such as the D3.js library can be used to easily draw the diagrams and control link and charge attributes.

2.2.1 Components and Usage

Although edges between nodes are typically undirected in a force-directed graph, they can easily be made directed if necessary to the application. For instance, to visualize gene or cellular pathways, directed links may be useful to determine specific transitions between nodes (Genc & Dogrusoz, 2003). Another application of force-directed graphs in biology is to create RNA secondary structure diagrams (Kerpedjiev, Hammer, & Hofacker, 2015). It can also be used to visualize social media data such as the relationships between tweets sent on Twitter (Morstatter, Kumar, Liu, & Maciejewski, 2013). Furthermore, it can be used to visualize relationships between different genres in literature (Simeone, 2012). In biology, collections of nodes having similar properties are often referred to as compartments, but these collections can also be referred to as clusters or groups (Schreiber, Dwyer, Marriott, & Wybrow, 2009).

2.2.2 Node Visualization

Nodes are typically shaped as circles, as was used in the previously mentioned applications of force-directed graphs. Nodes have also been shaped as ovals or ellipses in some studies (Schreiber et al., 2009). However, no existing studies mapping nodes to quantitative data have used rectangular nodes, as per our knowledge and research.

Nodes can be colored in different ways. One method is to color the inside of a node a specific color, to associate it with a specific group. For instance, a node's color could represent the group and importance of a certain protein in a protein interaction network (Schreiber et al., 2009). Another method is to use a double-encoding where the node's stroke color represents the node's group, and the intensity of the node's fill color represents a quantitative value. In this case, a darker fill color would represent a larger quantitative value, while a lighter fill color would represent a smaller quantitative value (Morstatter et al., 2013). Furthermore, a node's group can also be represented by displaying a colored halo around each node to enhance the visibility of clusters in the network (Simeone, 2012).

2.2.3 Graph Density and Size

The density of a graph affects the number of links that are displayed in the graph. Given a fixed number of nodes, a graph with a higher density will have a greater number of links than a graph with a lower density. Ghoniem et al. defined the link or edge density of a graph to be

$$d = \sqrt{\frac{l}{n^2}}$$

where d is the density, l is the number of links in the graph, and n is the number of nodes in the graph (2004).

A graph can have any number of nodes. In reality, graphs can have several hundreds of nodes. However, for the purposes of empirical studies, the number of nodes is usually fixed to a specific value, based on the desired graph size or complexity. In Ghoniem et al.'s study comparing graph readability, small-sized graphs had 20 nodes, medium-sized graphs had 50 nodes, and large-sized graphs had 100 nodes (2004). In Saket et al.'s study using node, node-link, and node-link-group diagrams, the researchers performed an empirical study to select the number of nodes that should be in the different graph sizes, so that the average task completion time would be between 5 and 30 seconds (2014). Their small-sized graphs had 50 nodes, medium-sized graphs had 100 nodes, and large-sized graphs had 200 nodes (Saket et al., 2014).

2.3 Node Perception

There are several factors that can influence the perception of nodes in a network. One factor is shape, such as whether a node is circular or rectangular. Another factor is size, such as whether a node is bigger or smaller than another node. A third factor is color, such as whether the color of one node is the same or different from the color of another node. There are basic thresholds at which humans can perceive even small changes in each of these factors. Just Noticeable Difference (JND) is the threshold at which humans can correctly differentiate between two stimuli at least fifty percent of the time. If two stimuli are presented below this threshold, then humans will generally be unable to discriminate between them. However, if the two stimuli are presented above this threshold, then humans will generally be able to consistently discriminate between them.

2.3.1 Size Discrimination

Size discrimination between two stimuli is generally modelled using Weber-Fechner's law, which is an equation that approximates the amount of change that must be present between two stimuli, in order for the difference to be perceived. Chung et al. performed a study about perceptually ordering different stimuli that were encoded using Jacques Bertin's retinal variables (2016). One of the encodings that they studied was size. They mention that the perception of two circles can be represented using Weber-Fechner's law written as

$$p = k \ln \frac{r_i}{r_j}$$

where r_i is the radius of circle i, r_j is the radius of circle j, k is an empirically derived constant, and p is the Weber fraction for discriminability between the two stimuli. In their study, they estimated k to be 0.23 and p to be 0.048. They further note that the JND for circle discrimination is at a p of 0.025 (Chung et al., 2016).

Furthermore, research has been done with regards to shape discrimination in terms of the aspect ratio of rectangles. For instance, Nachmias performed a study to estimate the Weber fractions for size and shape discrimination, under different conditions, based on a task which required participants to select the taller stimulus given a pair of stimuli (2011).

2.3.2 Color Selection

There are several methods to select distinguishable colors for a study. One method is to manually compute the distance between two colors in a color space, as was done by Chung et al. (2016). Another approach is to select colors from an existing color scheme selection tool such as ColorBrewer 2.0, which was created to allow researchers to select color schemes with distinguishable colors, for map region coloring (Brewer & Harrower, n.d.). These color schemes were developed for public use, based on research conducted by Harrower and Brewer (2003).

Chapter 3: Approach

3.1 Network Generation

My network visualizations use three of the six retinal variables that Bertin presented in his book (Bertin, 2010). I use size to encode the quantitative value of the node, color to encode the group to which a node belongs, and shape to encode either a circular node or a rectangular node.

3.1.1 Data Generation

Data was randomly generated for this study because we hope to be able to generalize results to visualizing quantitative compartmented data from any field. We only evaluate the effectiveness of the encodings themselves, and the generated data values are only used for the purposes of evaluating the accuracy of data readings using different quantitative data encodings. Therefore, the encoding methods can be applied to any data set where each node belongs to a single group based on a single categorical variable and has a quantitative value mapped to it. For the study, the data was generated using Python. Data values were generated using a Gaussian distribution. A Gaussian distribution was selected since our hypothesis reflects that we would like to evaluate the accuracy of quantitative value reading in a network. Therefore, this differentiation will become more apparent when values are similar to each other. By the properties of Gaussian distributions, since approximately 68% of the randomly generated values will be within one standard deviation of the mean, we can be confident that more values will likely be similar to each other than very different from each other, which will allow us to evaluate the quantitative encoding methods. Therefore, these encodings can be applied to datasets from any field, with the consideration that the sizes of the nodes are above the JND, when it is necessary to differentiate between the values of nodes.

For the study, we generated data using a Gaussian distribution with a mean of 50 and a standard deviation of 15. Due to the properties of Gaussian distributions, this means that approximately 99.7% of the randomly generated values will be between three standard deviations from the mean. So, 99.7% of the generated values will be in the range [5, 95]. Therefore, the target data range was selected to be (0, 100] because that would allow for a specific set of data values where differences in radius or height and width can be perceived. Since both the variable-area circles encoding and the calibrated columns encoding have areas which are proportional to the data values, this data range can easily be extended to support a smaller or larger proposed data range. For instance, if the desired data range is (0, 1000], then each value can be divided by 10 before computing the radius or height and width, so that the areas correspond to the sizes used in this study.

The grouping tasks and the overall task in this study presented both medium-sized networks containing 50 nodes and large-sized networks containing 200 nodes. These choices are consistent with the number of nodes generally selected for networks in the literature (Ghoniem et al., 2004; Saket et al., 2014). While the number of groups in the network can vary in reality, we have chosen to select a fixed number of groups for this study, so as not to introduce another variable. So, medium-sized networks contain four groups and large-sized networks contain eight groups. The density for medium-sized networks is 0.1 and the density for large-sized networks is 0.07. These densities were selected after generating networks with different densities, and selecting the densities which seemed to best minimize the amount of node occlusion.

Links were also assigned randomly between nodes, based on the selected densities. The previously mentioned formula for link density can be rewritten in terms of the number of links, in order to calculate the total number of links in the network.

$$d = \sqrt{\frac{l}{n^2}} \quad \rightarrow \quad d^2 = \frac{l}{n^2} \quad \rightarrow \quad l = n^2 d^2$$

When links were randomly assigned between pairs of nodes in the network, using a random number generator, very few links tended to be formed between nodes within the same compartment. Therefore, a fixed number of the links were allocated to be placed between nodes in the same compartment, and the remaining links were placed between nodes from different compartments. $\frac{l}{2}$ was set to be the total number of links between nodes in the same compartment in medium networks, and $\frac{l}{1.5}$ for large networks. The exact number was divided close to evenly among each compartment in the network. Further links were added between child nodes and their respective parent nodes, in order to cluster the children around their respective parents. These added links had a stroke width of 0, and therefore remained invisible in the visualizations.

Furthermore, parameters for the force-directed graphs were selected in D3.js, so as to support the clustering while minimizing the amount of occlusion between nodes. A many-body force with a strength of -120 was applied to the nodes, to allow for a greater repulsion between nodes within the same group (Bostock, Devinsuit, Watson, Heer, & Velikiy, n.d.). This repulsion helped to minimize the amount of occlusion between the nodes. Also, a link force was applied to the links with a distance of 100, to further enhance the overall perception of the network.

3.1.2 Network Visualizations

The network visualizations for the study were created using D3.js, a JavaScript library commonly used for creating interactive visualizations of data. For this study, I focused particularly on force-directed graphs, adapting from the sample code provided for creating such visualizations using Version 4 of the D3.js library (Bostock, 2017).

A network can have any number of groups, where each node in the same group is colored with the same color. Each group has one parent node, and each parent node is labelled with a number, representing the index of that group. All child nodes belonging to a group are clustered around the parent node for the group.

Additionally, a child node is connected to its parent node via a link that is drawn from the child to the parent. These links are represented by thin gray lines in the network. A child node can be connected to other child nodes in the same group or in different groups. This connection is via a link that is drawn between two child nodes. These links are represented by thick gray lines in the network.

3.1.3 Grouping Techniques

One grouping method that I used in the study is “colored halos”. In this technique, colored halos surround each node in a network. All halos are of the same size, and nodes with the same halo color are in the same group. The nodes themselves are colored black.

The other grouping method is “colored nodes”. In this technique, the nodes themselves are colored, and nodes with the same color are in the same group.

3.1.4 Quantitative Data Visualization Techniques

One quantitative data visualization method that I used in the study is “variable-area circles”, where as described in Chapter 2, quantities are mapped to the area of a circle. An example of this encoding is shown in Figure 3.1. The other quantitative data visualization method is “calibrated columns”, where as also described in Chapter 2, quantities are mapped to the area of a bar or rectangle. An example of this encoding is shown in Figure 3.2.

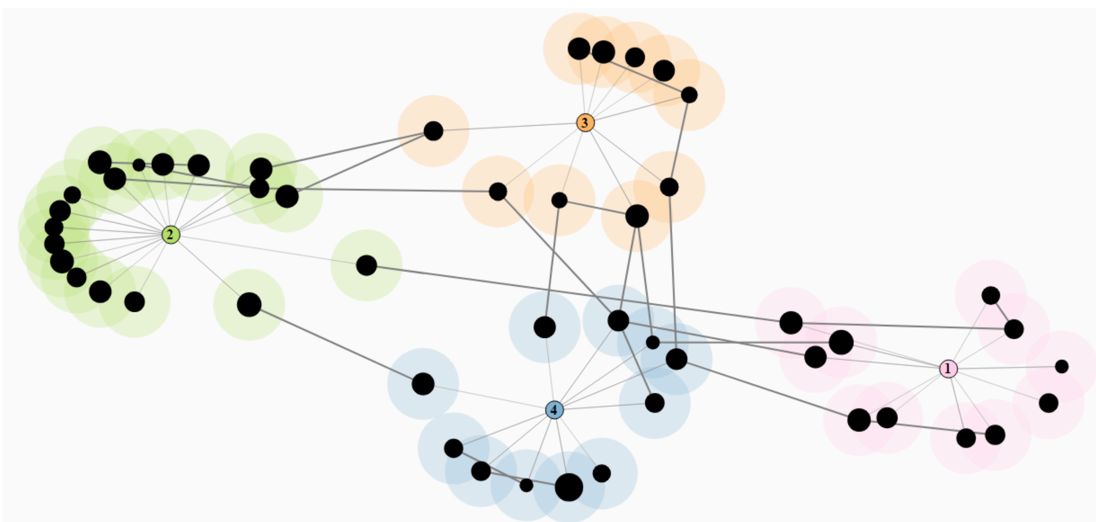

Figure 3.1: Variable-Area Circles Encoding of Quantitative Network

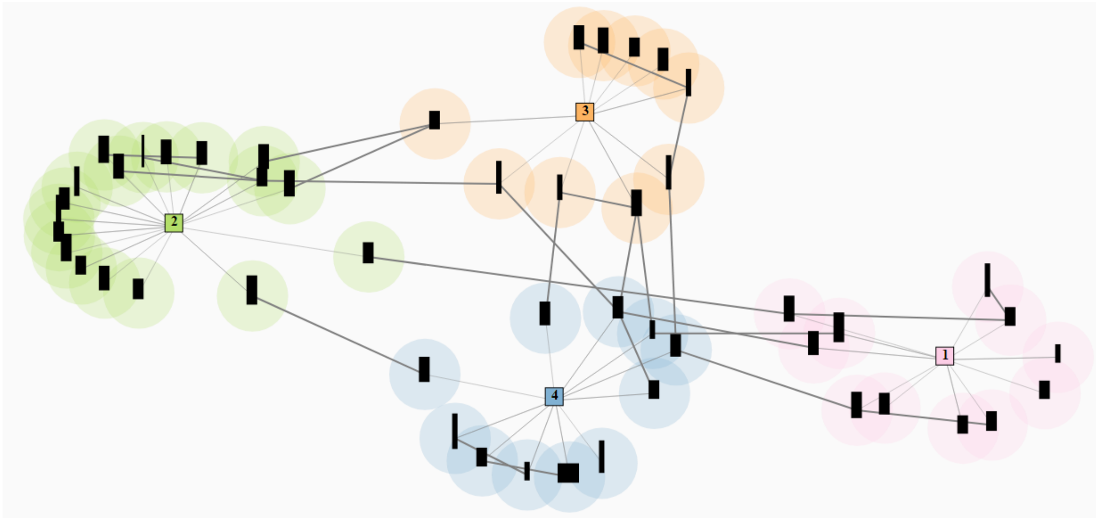

Figure 3.2: Calibrated Columns Encoding of Quantitative Network

3.1.5 Dimensions of the Encodings

In the grouping encodings and the calibrated columns encoding, the parent nodes are shaped as squares with a size of 20 pixels x 20 pixels. In the variable-area circles encoding, the parent nodes are shaped as circles with a radius of 10 pixels. These dimensions were chosen to ensure that the parent nodes have the same approximate size in all networks. Furthermore, these dimensions allow the parent nodes to be distinguishable in the network, but do not detract from the perception of the surrounding child nodes. In the colored halos encoding and the quantitative data visualization encodings, the halos have a radius of 40 pixels and an opacity of 0.25. This radius was chosen to allow for the halo to be visible, even when a node was mapped to the maximum data value of 100. This opacity value allowed the halos to be distinguishable between groups, while also showing the overlap between halos. In the grouping encodings, the nodes themselves have a size of 15 pixels x 15 pixels. These dimensions were selected to provide distinguishability between nodes in the colored nodes encoding, while providing an additional perception of the halos in the colored halos encoding.

In all of the generated networks, only the child nodes have quantitative values mapped to them. Parent nodes do not have quantitative values, since they are only used to group child nodes and provide the index numbers for the groups.

In the variable-area circles encoding, the nodes themselves have a radius of $3\sqrt{(V_i / \pi)}$, where V_i is the value of circle i. The radius is scaled by a factor of 3 to improve visibility of the encoding on the monitor. With this scale, a circle encoded with a value of 1 would have a radius of approximately 1.693 pixels, and a circle encoded with a value of 100 would have a radius of approximately 16.926 pixels.

Suppose that V_i now represents the value of rectangle i. The following data intervals represent the ranges of values for V_i. In the calibrated columns encoding, the nodes themselves have a width of 3 pixels if (0, 20), 6 pixels if [20, 40), 12 pixels if [40, 80), and 24 pixels if [80, 100]. The heights were calculated by first dividing by the scaling factor to compute the actual node width, and then computing normally using the area formula. The value was then scaled by a factor of 2 for visibility. With this scale, a rectangle encoded with a value of 1 would have a width of 3 pixels and a height of 2 pixels, and a rectangle encoded with a value of 100 would have a width of 24 pixels and a height of 25 pixels.

3.2 Hypotheses

Our hypotheses were:

1) For grouping, the colored halos encoding will support a more accurate grouping of nodes in the network than the colored nodes encoding.

2) For quantitative value estimation and comparison, the calibrated columns encoding will facilitate a more accurate reading of the quantitative values mapped to the nodes than the variable-area circles encoding.

3.3 Independent and Dependent Variables

For the pilot study, we used a within-subject design with three independent variables: network size (medium and large), grouping encoding (colored nodes and colored halos), and quantitative value encoding (variable-area circles and calibrated columns). We also had seven different tasks that participants completed. All combinations of the independent variables were not used in every task. These details are explained in Section 3.4.2 of this document. The dependent variables in this study include the accuracy of the responses, which is measured in terms of correctness for binary answers or relative error for continuous answers, and the confidence level ratings for the responses. More details are provided in Section 5.2 of this document.

3.4 Task Selection

Several tasks were initially proposed to evaluate the visualizations. The full list of our proposed tasks is presented in Table 3.1. Tasks G2, G3, G4, and G5 were adapted from the clustering study performed by Radu et al. (2014). Tasks were divided into three overarching categories: Grouping Tasks, Quantitative Tasks, and Overall Tasks. Grouping tasks were designed to evaluate the relationship between the colored nodes and colored halos encodings, while quantitative tasks were designed to evaluate the relationship between the variable-area circles and calibrated columns encodings. Overall tasks were designed to evaluate combinations of the grouping and quantitative encodings.

Grouping Tasks
G1. Please estimate the number of nodes in the highlighted compartment.
G2. Please identify the relative proportional ratio (1X, 1.5X, 2X, 3X) between the numbers of nodes in each of the highlighted compartments.
G3. Please enter the total number of distinct groups in the network.
G4. Please determine whether the two highlighted nodes belong to the same compartment.
G5. Please enter the number of nodes connected to the highlighted node.
Quantitative Tasks
Q1. Please identify the relative proportional ratio (1X, 1.5X, 2X, 3X) between the values of the highlighted nodes belonging to the same compartment.
Q2. Please identify the relative proportional ratio (1X, 1.5X, 2X, 3X) between the values of the highlighted nodes belonging to different compartments.
Q3. Please enter the exact value of the highlighted node.
Q4. Please identify whether the value of the highlighted node is greater than or less than a specific value.
Q5. Please identify the relative proportional ratio (1X, 1.5X, 2X, 3X) between the averages of the node values in the two highlighted compartments.
Q6. Please enter the number of nodes in the network having a specific value.
Q7. Please enter the number of nodes in the network that have a value greater/less than that of the highlighted node.
Combined Tasks
C1. Please identify the compartment having the largest variance in node values.
C2. Please identify the compartment having the greatest number of nodes with values greater/less than a specific value.
C3. Please estimate the percentage of nodes in the same compartment having values greater/less than the highlighted node.
C4. Please identify which node in the highlighted compartment has a specific value.
C5. Please identify which node in the highlighted compartment has the largest/smallest value.

Table 3.1: Proposed Empirical Study Tasks

After compiling the initial list of tasks, we filtered the list down to seven tasks (G4, G1, G2, Q4, Q3, Q1-Q2, C1), by eliminating tasks that had redundant components or were irrelevant for this study.

3.4.1 Eliminated Tasks

Task G3 was eliminated since my force-directed networks were designed such that child nodes belonging to the same compartment were clustered around the parent node for that compartment. Therefore, finding the number of distinct groups would simply be asking the participant to count the number of parent nodes in the network. Therefore, this network layout would not test the participant's visualization of the colors and groupings in the network in the intended manner. Task G5 was eliminated because edges in the network are irrelevant to testing the effectiveness of the encodings themselves.

In the quantitative tasks, Tasks Q1 and Q2 were combined into a single task for the study, since the goals of both tasks are the same. Task Q6 was eliminated because in a continuous range of data values, it is likely for two nodes to have similar values, but not the same values. Also, the goal of value identification is being tested in tasks Q3 and Q1-Q2. Therefore, this task would be redundant. Task Q7 was eliminated because the goal of this task is similar to that of Q4.

Restrictions needed to be placed on the number of overall tasks selected for the study because these tasks would necessitate a greater task duration, due to the increased task difficulty. Therefore, we were only able to select one overall task, in order to keep the empirical study's length to 1.5 hours per participant. Task C2 was eliminated because it was an extension on task Q4. Task C3 was eliminated because the calculation of a specific percentage may require a longer task duration, since it would require the judgement of node values, the counting of specific nodes, the estimation of the total number of nodes in the compartment, and the final calculation of the percentage. Tasks C4 and C5 were eliminated because they tested reading of visualizations at an intermediate level, while task C1 tested reading at the overall level.

3.4.2 Trials for Each Selected Task

We selected three grouping tasks, three quantitative tasks, and one combined task for the empirical study. Tasks 4 and 5 test an elementary level of reading, Tasks 1, 2, 3, and 6 test an intermediate level of reading, and Task 7 tests an overall level of reading. The final list of tasks used in the empirical study is shown in Table 3.2. The tasks are labelled Task 1 through Task 7, and will be referred to in this way, throughout the remainder of this document. Additionally, for the purposes of the empirical study, we chose to change the terminology of “compartment” to “group”, since both words have the same meaning in this context, but the latter would be a more familiar term to the participants.

Furthermore, we have chosen to add network size as an additional variable to the grouping tasks (Tasks 1-3) and the overall task (Task 7), but not to the quantitative tasks (Tasks 4-6). This is because a larger network size indicates that several more nodes will be present on the screen. Since we wanted to test the relationship between the accuracy of completing tasks using variable-area circles and calibrated columns in the pilot study, we did not want to have the size of the network detract from the observation of any relationships. However, Task 7 does include network size in combination with the other encodings. Also, all visualizations in Tasks 4-6 used the colored halos encoding to specify grouping, since the focus of these tasks was on estimating quantitative values, and not on determining group relationships.

Tasks for Empirical Study
Task 1: Please identify whether the two highlighted nodes are in the same group.
Task 2: Please estimate the number of nodes in the highlighted group.
Task 3: Please identify the relative proportional ratio (1X, 1.5X, 2X, 3X) between the numbers of nodes in each of the highlighted groups.
Task 4: Please identify whether the value of the highlighted node is greater than or less than 50.
Task 5: Please estimate the exact value of the highlighted node.
Task 6: Please identify the relative proportional ratio (1X, 1.5X, 2X, 3X) between the values of the highlighted nodes.
Task 7: Please identify the group having the largest variance in node values.

Table 3.2: Finalized Empirical Study Tasks

For all tasks, nodes, compartments, and graphs were selected with the following criteria in mind. Each network was randomly assigned to a trial number, and the order of the trials within a task was also randomized using a random permutation of the trials.

Task 1

Task 1 required participants to identify whether two highlighted nodes belong in the same group. The purpose of this task is to evaluate intuitive judgment of grouping. Therefore, we assigned this task's duration to be two seconds per trial, as was selected by Radu et al. (2014). A total of 32 trials was allocated for this task: eight per size and encoding combination (medium-colored nodes, medium-colored halos, large-colored nodes, and large-colored halos). Within each of those eight trials, four trials were for the two highlighted nodes belonging to the same compartment and four trials were for the two highlighted nodes belonging to different compartments. Within each of those four trials, two trials used nodes that appeared spatially close to each other in the network, while two trials used nodes that were more distant in the network. In total, the approximate time for this task was (2 sec / task) * (32 tasks) = 64 seconds = 1 minute and 4 seconds, excluding the time to select answers and confidence level ratings. During each trial, two red circles appear on the screen for about one second to denote the spatial locations of the two selected nodes for the trial. Afterwards, the rest of the network appears and the timer begins. In this way, participants do not need to search for the highlighted nodes once the timer begins.

Task 2

Task 2 required participants to estimate the number of nodes in the highlighted group. Participants were asked to "estimate" the number because depending on the size of the network in a trial, the participants may not have time to count all of the nodes within the time limit. The purpose of this task is to evaluate grouping with some deliberation.

We estimated this task's duration to be ten seconds per trial, because this task required a more deliberative response, but was not as difficult as an overall task. A total of sixteen trials was allocated for this task: four per size and encoding combination (medium-colored nodes, medium-colored halos, large-colored nodes, and large-colored halos). Within each of those four trials, two trials had networks with a relatively small number of nodes, compared to the overall network size, while two trials had networks with a much larger number of nodes (approximately one half of the number of nodes in medium-sized networks and approximately one sixth of the number of nodes in large-sized networks). In total, the approximate time for this task was (10 sec / task) * (16 tasks) = 160 seconds = 2 minutes and 40 seconds, excluding the time to enter answers and select confidence level ratings. During each trial, a red box appears on the screen for about one second to denote the spatial location of the parent node of the selected compartment for the trial, before the rest of the network appears and the timer begins.

Task 3

Task 3 required participants to identify the relative proportional ratio between the numbers of nodes in the given highlighted groups. The purpose of this task is to evaluate the grouping of nodes within a larger subset of the network.

We assigned this task's duration to be twenty seconds per trial, as was selected by Radu et al. (2014). A total of sixteen trials was allocated for this task: four per size and encoding combination (medium-colored nodes, medium-colored halos, large-colored nodes, and large-colored halos). Within each of those four trials, two trials had networks with large differences in the compartment sizes (2X or 3X size difference) while two trials had networks with much smaller differences in the compartment sizes (1X or 1.5X size difference). In total, the approximate time for this task was (20 sec / task) * (16 tasks) = 320 seconds = 5 minutes and 20 seconds, excluding the time to select answers and confidence level ratings. During each trial, two red boxes appear on the screen for about one second to denote the spatial locations of the parent nodes of both selected compartments for the trial, before the rest of the network appears and the timer begins.

Task 4

Task 4 required participants to identify whether the value of a highlighted node is greater than or less than 50. The purpose of this task is to evaluate intuitive judgment of a node's value. Therefore, we assigned this task's duration to be two seconds per trial, as estimated from Task 1. Similar to the reasoning for task 1, this short time is so that we can identify the participants' intuitive judgement as to whether the given node's value is greater than or less than 50. The value of 50 was selected because this was the mean of our data distribution, and would allow participants to better perform comparisons with a fixed node, during the short duration of each trial. A total of twenty trials was allocated for this task: ten for each encoding (variable-area circles and calibrated columns). Within each of those ten trials, four trials had nodes that were approximately one degree away from 50 (two below and two above), four trials had nodes that were approximately one to two degrees away from 50 (two below and two above), and two trials had nodes that were more than two degrees away from 50 (both below). In total, the approximate time for this task was (2 sec / task) * (20 tasks) = 40 seconds, excluding the time to select answers and confidence level ratings. During each trial, a yellow halo appears on the screen for about one second to denote the spatial location of the selected node for the trial, before the rest of the network appears and the timer begins.

The approximate range of values was selected by rearranging Weber-Fechner's law and rewriting it in terms of the quantitative values. Below is the derivation for circles. Since the quantitative values are proportional to the areas of the circles, suppose that A_i is the area of circle i, A_j is the area of circle j, V_i is the quantitative value to which circle i is mapped, V_j is the quantitative value to which circle j is mapped, and c is the proportionality constant. Then, $A_i = cV_i$ and $A_j = cV_j$. Rearranging the area formula we get

$$A = \pi r^2 \quad \rightarrow \quad r^2 = \frac{A}{\pi} \quad \rightarrow \quad r = \pm\sqrt{\frac{A}{\pi}} = \sqrt{\frac{A}{\pi}}$$

Rearranging Weber-Fechner's law and substituting these equations we get

$$p = k \ln\frac{r_i}{r_j} \quad \rightarrow \quad \ln\frac{r_i}{r_j} = \frac{p}{k} \quad \rightarrow \quad \frac{r_i}{r_j} = e^{\frac{p}{k}} \quad \rightarrow \quad r_i = r_j e^{\frac{p}{k}} \quad \rightarrow \quad \sqrt{\frac{A_i}{\pi}} = \sqrt{\frac{A_j}{\pi}} e^{\frac{p}{k}} \quad \rightarrow$$

$$\frac{A_i}{\pi} = \frac{A_j}{\pi}\left(e^{\frac{p}{k}}\right)^2 \quad \rightarrow \quad \frac{cV_i}{\pi} = \frac{cV_j}{\pi}\left(e^{\frac{p}{k}}\right)^2 \quad \rightarrow \quad V_i = V_j\left(e^{\frac{p}{k}}\right)^2$$

By approximating p to equal a maximum of 0.048 and a minimum of 0.025, and approximating k to be 0.23, the values that correspond to the approximate degrees of variance can be estimated (Chung et al., 2016).

A similar derivation can be achieved for rectangles. The widths of the rectangles are perceivable since they are doubled between the value intervals. Therefore, the two stimuli that need to be considered using Weber-Fechner's law is the heights of two rectangles. Since the quantitative values are proportional to the areas of the rectangles, suppose that A_i is the area of rectangle i, A_j is the area of rectangle j, V_i is the quantitative value to which rectangle i is mapped, V_j is the quantitative value to which rectangle j is mapped, and c is the proportionality constant. Then, $A_i = cV_i$ and $A_j = cV_j$.

Rearranging the area formula we get

$$A = H * W \quad \rightarrow \quad H = \frac{A}{W}$$

Rearranging Weber-Fechner's law and substituting these equations we get

$$p = k \ln \frac{H_i}{H_j} \quad \rightarrow \quad \ln \frac{H_i}{H_j} = \frac{p}{k} \quad \rightarrow \quad \frac{H_i}{H_j} = e^{\frac{p}{k}} \quad \rightarrow \quad H_i = H_j e^{\frac{p}{k}} \quad \rightarrow$$
$$\frac{A_i}{W_i} = \frac{A_j}{W_j} e^{\frac{p}{k}} \quad \rightarrow \quad \frac{cV_i}{W_i} = \frac{cV_j}{W_j} e^{\frac{p}{k}} \quad \rightarrow \quad \frac{V_i}{W_i} = \frac{V_j}{W_j} e^{\frac{p}{k}}$$

Therefore, there are two possible situations for rectangles.

1) If $W_i = W_j$, then

$$V_i = V_j e^{\frac{p}{k}}$$

2) If $W_i \neq W_j$, then

$$V_i = \frac{W_i * V_j}{W_j} e^{\frac{p}{k}}$$

By approximating p to equal 0.075 and k to equal 0.23, the values that correspond to the approximate degrees of variance can be estimated (Chung et al., 2016; Nachmias, 2011).

Task 5

Task 5 required participants to estimate the exact value of a highlighted node. Participants were asked to "estimate" the values because the actual values of the nodes are floating point numbers, and we cannot expect participants to enter answers that are correct to all digits after the decimal point. The purpose of this task is to evaluate quantitative value estimation with some deliberation.

We assigned this task's duration to be ten seconds per trial, as estimated from Task 2. A total of sixteen trials was allocated for this task: eight for each encoding (variable-area circles and calibrated columns). Within each of those eight trials, there were two trials each for node values up to two degrees away, between two and four degrees away, between four and six degrees away, and greater than six degrees away from the value five. The value of five was selected because this was one of the lowest perceivable data values in our selected data range. In total, the approximate time for this task was (10 sec / task) * (16 tasks) = 160 seconds = 2 minutes and 40 seconds, excluding the time to enter answers and select confidence level ratings. During each trial, a yellow halo appears on the screen for about one second to denote the spatial location of the selected node for the trial, before the rest of the network appears and the timer begins.

Task 6

Task 6 required participants to identify the relative proportional ratio between the values of two highlighted nodes. The purpose of this task is to evaluate quantitative value estimation and comparison between nodes in the network.

We assigned this task's duration to be twenty seconds per trial, as estimated from Task 3. A total of sixteen trials was allocated for this task: eight for each encoding (variable-area circles and calibrated columns). Within each of those eight trials, four trials were for nodes belonging to the same compartment and four trials were for nodes belonging to different compartments. Furthermore, within these divisions, two trials had nodes with large value differences (2X or 3X difference) while two trials had nodes with much smaller value differences (1X or 1.5X difference). In total, the approximate time for this task was (20 sec / task) * (16 tasks) = 320 seconds = 5 minutes and 20 seconds, excluding the time to select answers and confidence level ratings. During each trial, two yellow halos appear on the screen for about one second to denote the spatial locations of the selected nodes for the trial, before the rest of the network appears and the timer begins.

Task 7

Task 7 requires participants to identify the group having the largest variance in node values. In this task, “variance” is used to refer to the amount of variability in the node values. In particular, it is used synonymously with the term “range”, and is not used to refer to the amount of deviation from the mean node value. Therefore, this task required participants to identify the group whose largest value minus its smallest value is the maximum among that of all of the groups in the network. The purpose of this task is to evaluate both node grouping and quantitative value estimation.

We assigned this task’s duration to be twenty seconds per trial, since it required both implicit grouping of nodes, as well as the general estimation of node values. A total of 32 trials was allocated for this task: four for each size and combination of encodings (medium-colored nodes-circles, medium-colored nodes-calibrated columns, medium-colored halos-circles, medium-colored halos-calibrated columns, large-colored nodes-circles, etc.). Within each of those four trials, two trials used networks where the range was obviously greater in one or two compartments, while two trials used networks where the differences in ranges were less obvious among all of the compartments in the network. In total, the approximate time for this task was (20 sec / task) * (32 tasks) = 640 seconds = 10 minutes and 40 seconds, excluding the time to select answers and confidence level ratings.

3.5 Color Selection

The stroke color of the circles used to denote the selected nodes in Task 1 was chosen to be red, since using red circles is analogous to using red arrows to point to the nodes (Zhao et al., 2016). Similarly, red boxes were used to denote the selected compartments in Tasks 2 and 3 for similar identification purposes. The halo color for the selected nodes in Tasks 4-6 was selected to be yellow because the black color of the node was easily perceivable on top of the yellow halo (Gabriel-Petit, 2007). Therefore, the color change would not detract from the perception of the node's size, and therefore its value.

The color representing each compartment was selected using ColorBrewer 2.0, a tool that is commonly used in research studies to select distinguishable colors. I selected the "qualitative" data tab and selected the 12-class Set3 color scheme, which is shown in Figure 3.3 (Brewer & Harrower, n.d.; D3.schemeSet3, n.d.). I then eliminated the two yellow colors, since I used yellow to highlight the halos of the selected nodes in Tasks 4-6. I also eliminated the gray color, since that is used as the color of the links in the network. I finally eliminated the purple color, since I only needed eight colors to represent each of the eight compartments in the network. The color associated with each compartment was kept constant among all of the tasks.

Figure 3.3: Color Scheme for Compartment Colors (D3.schemeSet3, n.d.)

Chapter 4: Empirical Study

A pilot study was conducted with six participants to evaluate my hypotheses based on the participants' responses to the tasks mentioned in Section 3.4.2 of this document. Only a small number of participants was selected for this pilot study, so that preliminary results could be collected and errors in the design could be adjusted, before a formal study can be conducted with a larger sample of participants. Participants were first asked to read and sign an informed consent form. Then, they were asked to complete a demographics form, so that I could collect some basic information about their familiarity with the concepts involved in the study. Next, they were asked to do a short Ishihara Color Blindness test. Afterwards, they completed a training session by reading a training document about the visualizations and tasks, and completing practice trials for each task. Then, they completed the actual pilot study. Finally, they completed a post-questionnaire about their experience using the different encodings, and then participated in a short interview, where I asked questions about their experience in more detail. The maximum duration of the empirical study was 1.5 hours per participant. Each participant was paid at $12/hour with a maximum of $18 for their participation.

4.1 Study Setup

The following sections discuss the components of the empirical study in greater detail.

4.1.1 GUI Setup

The GUI for the empirical study was created using HTML, JavaScript, CSS, PHP, AJAX, and JQuery. One GUI was created for the training portion of the study and a separate GUI was created for the actual pilot study.

The layout of the GUI for my empirical study is based on a GUI that was created by other researchers working on a different study using eye tracking and the PathRings web application. The task is displayed in the bottom left half of the GUI, the remaining time for the task is displayed near the middle of the bottom right half of the GUI, and the "Next" button is displayed near the bottom right corner of the GUI (*GraphStudy3*, n.d.). This layout allows participants to see the countdown of the remaining time in their peripheral vision and to glance at the task description, while primarily focusing on the network visualization.

At the start of every new task, the GUI shows the task number and its description in the middle of the screen. The trials for the task begin once the participant clicks the "Next" button. Each trial features a static force-directed graph, where the node locations have been fixed beforehand. The participant can pan the visualization in any direction within the area in the GUI using left-mouse dragging. However, the participant is not provided with the capability to zoom in and out of a network. Although zooming is a common feature in visualizations, I do not provide this feature in the study, since zooming in may allow participants to better perceive small changes in the areas of two circles or rectangles, which may not be easily perceivable otherwise. This could then distort the results of the quantitative tasks, which rely on the nodes being perceived at a consistent scale among all of the tasks.

Additionally, zooming out may allow participants to better perceive groupings in the network due to the overall perception of the network based on the spatial locations of the nodes. This could then distort the results of the grouping tasks, which also rely on the nodes being perceived at a consistent scale.

The participant is not prompted to input or select an answer until the next page in the GUI. This allows the participant to use the entire task duration to analyze the visualization and determine an answer for a trial. However, if the participant is ready to input or select his/her answer before the given time duration has elapsed, then he/she can click the “Next” button to move to the next page. If the participant chooses to use the whole time duration, then he/she is automatically redirected to the next page, once the time duration has elapsed. The participant is additionally prompted to rate his/her confidence level for each response on a scale from 1 to 7, where 1 is the lowest confidence level, and 7 is the highest confidence level.

Participants used a BenQ GTG XL 2720Z 27” monitor with a resolution of 1920 × 1080 to view the visualizations and perform the tasks in this study.

4.1.2 Pre-Questionnaire

Before the training, participants were asked to fill out a short questionnaire asking about some demographic information and their experience with the overarching topics presented in the study. In particular, each participant was asked to input his/her age, gender, and optionally his/her area of study or research. Furthermore, he/she was asked to rate his/her familiarity with computers, network visualizations, bar graphs/charts, and node-link/vertex-edge graphs, on a scale from 1 to 7, where 1 referred to “beginner” and 7 referred to “expert”.

4.1.3 Training

During the training session, each participant was asked to read a training document which described the visualization methods and the tasks. An audio version of me reading the training document was prepared beforehand. Participants were provided with both a hardcopy and a softcopy of the training document, so that they could follow along with the audio. At specific points in the training document, participants were encouraged to pause the audio so that they could ask any questions about what they had read, or about the tasks they had completed.

For each task, participants were only presented with medium-sized networks, and were provided with four trials for practice. Tasks 1-3 had two trials each for the colored nodes encoding and the colored halos encoding. Tasks 4-6 had two trials each for the variable-area circles encoding and calibrated columns encoding. Task 7 had one trial each for every combination of both sets of encodings. After the participant submitted his/her answer, he/she was presented with the correct answer to the question. The participant was also allowed to view the visualization again, to make sure that he/she understood the task and the answer, before moving on to the next question.

4.1.4 Pilot Study

Before each participant began the pilot study, I left the room, so that he/she could complete the tasks without any disturbances or feelings of being judged or watched. For each trial, once the “Next” button was clicked or once the time for the task had elapsed, the participant was required to input or select an answer, and was not provided the option to view the visualization again. Sample network visualizations for each task from the study are presented in Figures 4.1 – 4.7.

Figure 4.1 shows an example of a trial from Task 1. The figure shows a large network using the colored nodes encoding. The two selected nodes for the trial are highlighted using two red circles that surround the nodes. In this trial, both nodes belong to the same group because they are colored with the same color.

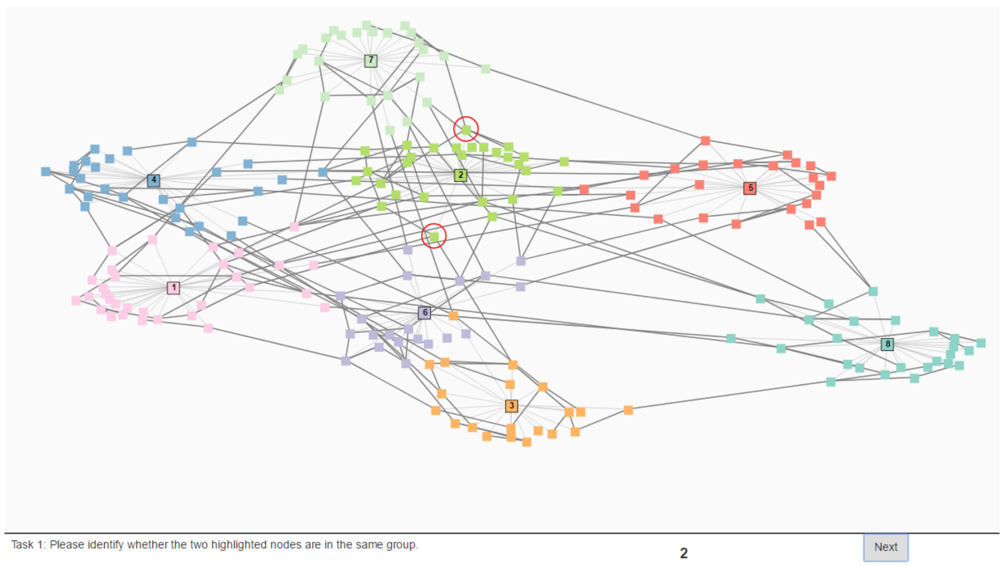

Figure 4.1: Task 1: Please identify whether the two highlighted nodes are in the same group. (This example uses a Large 200-node network with the Colored Nodes Encoding. Answer: Yes.)

Figure 4.2 shows an example of a trial from Task 2. The figure shows a large network using the colored halos encoding. Group 3 has been selected for this trial, since the parent node for that group has a red box surrounding it. In this trial, 31 nodes belong to this group because the halos for these nodes are colored with the same color as the selected parent node.

Figure 4.3 shows an example of a trial from Task 3. The figure shows a medium network using the colored halos encoding. Groups 1 and 2 have been selected for this trial, since the parent nodes for these groups have red boxes surrounding them. In this trial, nineteen nodes belong to group 1 and seven nodes belong to group 2. Therefore, the closest relative proportional ratio is 3X.

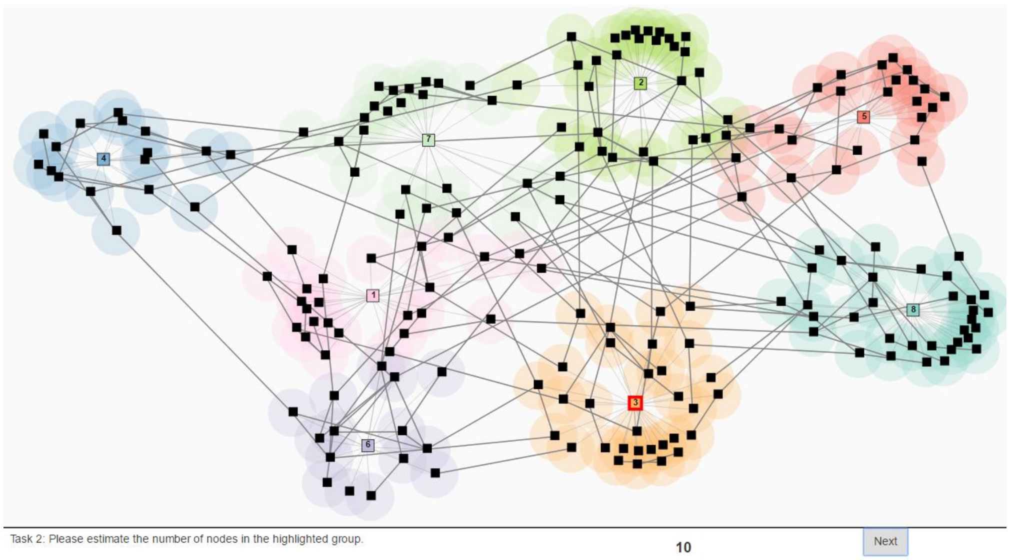

Figure 4.2: Task 2: Please estimate the number of nodes in the highlighted group. (This example uses a Large 200-node network with the Colored Halos Encoding. Answer: 31.)

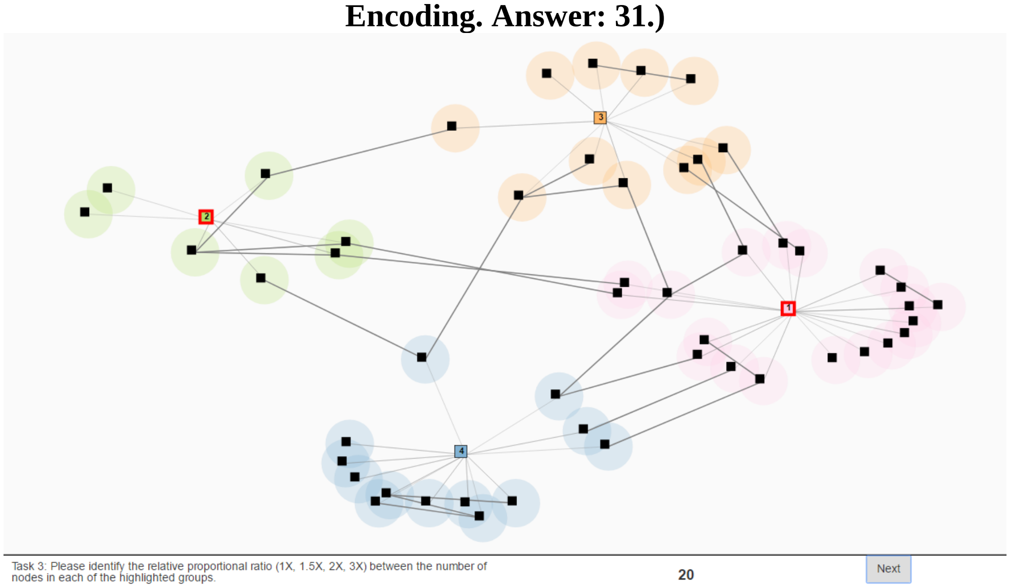

Figure 4.3: Task 3: Please identify the relative proportional ratio (1X, 1.5X, 2X, 3X) between the numbers of nodes in each of the highlighted groups. (This example uses a Medium 50-node network with the Colored Halos Encoding. Answer: 3X.)

Figure 4.4 shows an example of a trial from Task 4. The figure shows a network using the calibrated columns encoding. The selected node for this trial is highlighted using a yellow halo. In this trial, the highlighted node's value is greater than 50. This judgement can be made by noticing that the selected node's width is equivalent to that of values that lie in the interval [40, 80), and that the selected node's height is greater than that of the rectangle symbolizing the value of 50. In actuality, this node's exact value is approximately 70.

Figure 4.5 shows an example of a trial from Task 5. The figure shows a network using the variable-area circles encoding. The selected node for this trial is highlighted using a yellow halo. In this trial, the highlighted node's value is approximately 76.

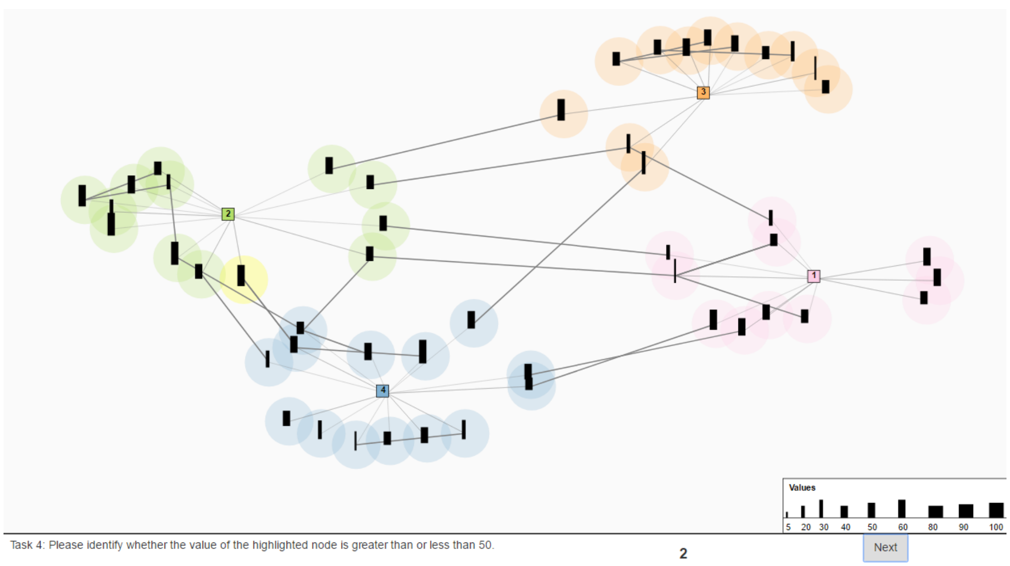

Figure 4.4: Task 4: Please identify whether the value of the highlighted node is greater than or less than 50. (This example uses the Calibrated Columns Encoding. Answer: Greater.)

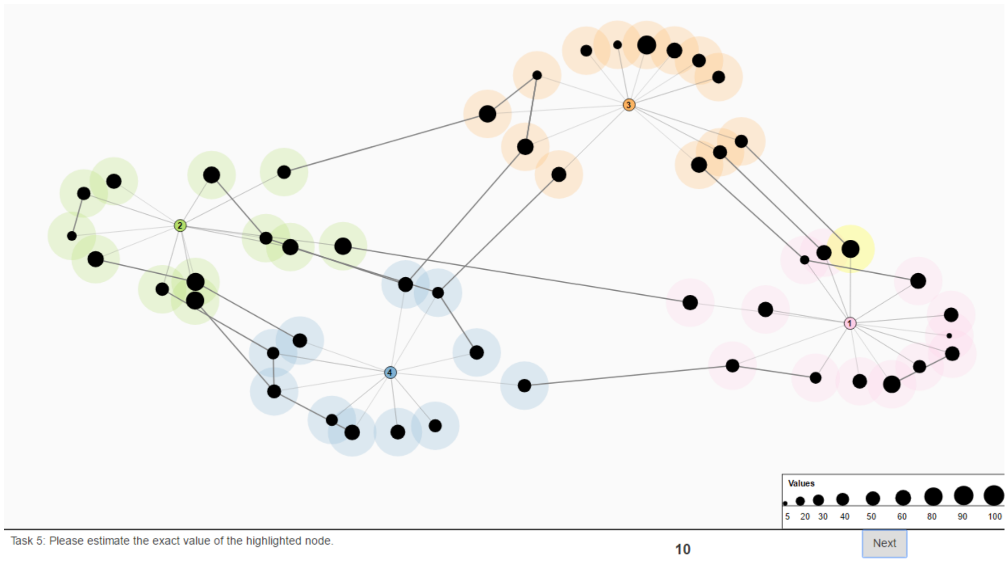

Figure 4.5: Task 5: Please estimate the exact value of the highlighted node. (This example uses the Variable-Area Circles Encoding. Answer: 76.16.)

Figure 4.6 shows an example of a trial from Task 6. The figure shows a network using the calibrated columns encoding. The two selected nodes for this trial are highlighted using yellow halos. In this trial, the relative proportional ratio between the values of the highlighted nodes is approximately 3X. This judgement can be made by noticing that the node from group 4 has a width that is equivalent to that of values that lie in the interval [20, 40), and that the node from group 1 has a width that is equivalent to that of values that lie in the interval [40, 80). Furthermore, the node from group 4 has a height that is approximately equivalent to that of the rectangle symbolizing the value of 20, and the node from group 1 has a height that is approximately equivalent to that of the rectangle symbolizing the value of 60. Therefore, the approximate ratio is 3X. In actuality, the node from group 4 has a value that is approximately 21, while the node from group 1 has a value that is approximately 58.

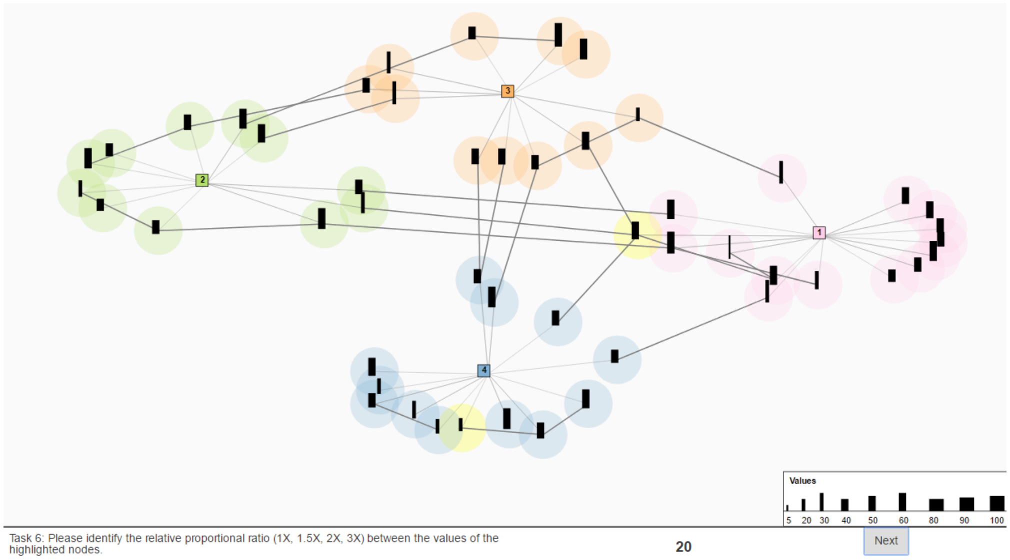

Figure 4.6: Task 6: Please identify the relative proportional ratio (1X, 1.5X, 2X, 3X) between the values of the highlighted nodes. (This example uses the Calibrated Columns Encoding. Answer: 3X.)

Figure 4.7 shows an example of a trial from Task 7. The figure shows a large network using the colored nodes encoding and the variable-area circles encoding. In this trial, the group having the largest range in node values is group 7.

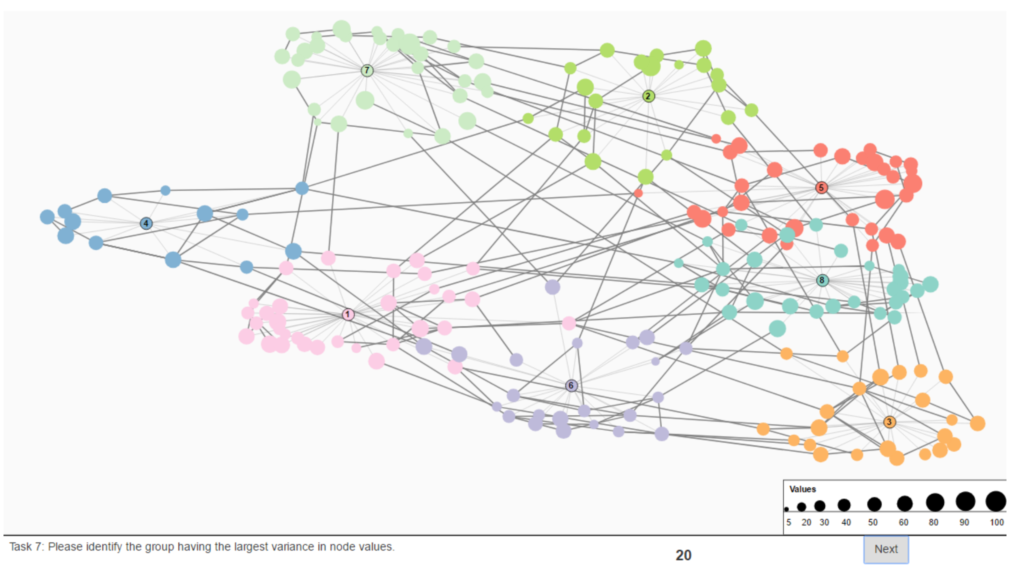

Figure 4.7: Task 7: Please identify the group having the largest variance in node values. (This example uses a Large 200-node network with the Colored Nodes and Variable-Area Circles Encodings. Answer: 7.)

4.1.5 Post-Questionnaire

After the pilot study, participants were asked to fill out a short questionnaire asking them to rate their experience using the different visualizations in the study, under different criteria. First, the questionnaire presented the real names for the encoding methods used in the study, as well as images of sample network visualizations using these encodings, in order to help jog their memory. For each encoding method (colored halos, colored nodes, variable-area circles, and calibrated columns), participants were asked to select a rating on a scale from 1 to 7, where 1 referred to "very low" and 7 referred to "very high", in response to questions asking about mental demand, physical demand, temporal demand, performance, and frustration. These questions are presented in Appendix A.1.

4.1.6 Short Interview

The last step of the empirical study was a short verbal interview with each participant where I asked him/her several questions about his/her experience using the different visualizations. These questions are presented in Appendix A.2. The participant was allowed to answer each question with however much detail he/she wanted to provide.

4.2 Data Collection

During the study, the participant's response and confidence level rating are written to a text file after every trial, when he/she clicks the "Submit" button. I do not consider the participant's response time because it can be misleading. For instance, if a participant pressed the "Next" button to input his/her answer, before the task duration had elapsed, then it would be difficult to determine whether the participant had clicked the button since he/she knew the answer or since he/she was just moving quickly through the tasks. Similarly, if a participant used the whole task duration, then it would be difficult to determine whether the participant had known the answer and was using the remaining time to review the visualization again or whether the participant had used the whole time and still did not know the answer.

The questionnaires were completed using Google Forms, so results were saved to Excel spreadsheets from there.

Chapter 5: Results and Discussion

After the pilot study concluded, I performed the data processing and results analysis. I used Java to write code to combine the separate text files containing participants' answers, correct answers, and trial information into tab separated text files for each task. Then, I used both SAS and Microsoft Excel to perform statistical analysis about the data and generate the graphs for the results. These findings as well as discussions about the results for each task are presented in this chapter.

5.1 Pre-Questionnaire Results

A total of six participants, one female and five male, took part in the pilot study. All participants were not colorblind. Their ages ranged from 19 to 35 (mean = 25.2, std. dev. = 5.78). Their areas of research or study were varied: Computer Science, Mechanical Engineering, Data Visualization, Electrical Engineering, and Biology. Their familiarity with computers ranged from a rating of 4 to a rating of 7 (mean = 5.5, std. dev. = 1.225). Also, their familiarity with network visualizations ranged from a rating of 1 to a rating of 5 (mean = 2.833, std. dev. = 1.472). Furthermore, their familiarity with bar graphs/charts ranged from a rating of 2 to a rating of 7 (mean = 5.333, std. dev. = 1.751). Finally, their familiarity with node-link/vertex-edge graphs ranged from a rating of 1 to a rating of 5 (mean = 2.667, std. dev. = 1.633). Since the ratings are on a scale from 1 to 7, where 1 refers to "beginner" and 7 refers to "expert", these results indicate that the participants were more familiar with computers and bar graphs/charts than they were with network visualizations and node-link/vertex-edge graphs.

5.2 Pilot Study Results

In the pilot study, I only measured the accuracy of a participant's response to each trial of a task, as well as the participant's corresponding confidence level for that trial. Tasks 1, 3, 4, 6, and 7 only had one correct answer for each trial, since the participants were able to select only one option from a list of radio buttons. So, the answers were binary, meaning that a correct answer could be labelled as 1, while an incorrect answer could be labelled as 0. For these tasks, accuracy was measured in terms of correctness, which was whether the answer was correct (1) or incorrect (0).

Tasks 2 and 5 involved the user inputting a numeric value. Correct answers for Task 2 were discrete while correct answers for Task 5 were continuous. Task 2 required participants to estimate the number of nodes in a group, so it is likely that participants entered some answers that were close to the correct answer, although not exact. Similarly, since Task 5 involved continuous values, it is unreasonable to expect participants to enter answers that are correct to all digits after the decimal point. Therefore, these types of inputs cannot be penalized by assigning a binary 0 or 1 label. For these tasks, accuracy was measured in terms of relative error, which can be written as

$$\text{RelativeError} = \frac{|\text{CorrectAnswer} - \text{ParticipantAnswer}|}{\text{CorrectAnswer}}$$

in order to normalize the deviations of participants' answers from the correct answers (Zhao, Bryant, Griffin, Terrill, & Chen, 2016). When the participant's answer is the same as the correct answer, the relative error value will be zero, meaning that there was no error involved in the answer. Relative error values farther from zero indicate a greater deviation from the correct answer.

I used the Logistic Procedure in SAS to analyze the results for tasks 1, 3, 4, 6, and 7, since the answers were binary values. The associated odds ratios are presented for most findings. The Wald χ^2 and *p* values are reported in Table 5.1. For tasks 2 and 5, I used the General Linear Model (GLM) in SAS to analyze the results, since the relative errors were continuous values. The associated Tukey test is performed for post hoc analysis for significant findings. The *F* and *p* values are reported in Table 5.1. The analysis of the confidence level ratings for each task was also done using GLM. The *F* and *p* values for confidence level ratings are reported in Table 5.2. The results that are significant at a 95% confidence level are bolded in the tables. Error bars in all graphs are shown to 1.96 standard errors about the mean to represent a 95% confidence interval.

The results and relevant graphs are presented in the following subsections, separated by each task. A discussion about these results and graphs is presented in Section 5.5.

5.2.1 Task 1

Task 1 required participants to identify whether two highlighted nodes were in the same group. Trials in this task used two variables which were the network size (medium or large) and the encoding method (colored nodes or colored halos).

In total, among all participants, 192 data points were collected for this task. However, only four out of these 192 trials were answered incorrectly. Therefore, neither size nor the encoding method was significant for this task in terms of accuracy. The joint combination of size and encoding method was also not significant for this task. All variables had a mean of 0.979 and a standard deviation of 0.144. A graph of the average accuracy for all independent variables is shown in Figure 5.3. The joint combinations of the variables had the following statistics: medium – colored nodes (mean = 1.000, std. dev. = 0), medium – colored halos (mean = 0.958, std. dev. = 0.202), large – colored nodes (mean = 0.958, std. dev. = 0.202), and large – colored halos (mean = 1.000, std. dev. = 0).

Task	Independent Variable	Significance
Task 1	Network Size	$\chi^2(1, 192) = 0.00$, $p = 1.000$
	Grouping Encoding	$\chi^2(1, 192) = 0.00$, $p = 1.000$
	Network Size and Grouping Encoding	$\chi^2(1, 192) = 0.0046$, $p = 0.946$
Task 2	Network Size	$F(1, 92) = 2.03$, $p = 0.158$
	Grouping Encoding	$F(1, 92) = 0.69$, $p = 0.41$
	Network Size and Grouping Encoding	$F(1, 92) = 0.59$, $p = 0.445$
Task 3	**Network Size**	$\chi^2(1, 96) = \mathbf{4.45}$, $p = \mathbf{0.035}$
	Grouping Encoding	$\chi^2(1, 96) = 0.76$, $p = 0.384$
	Network Size and Grouping Encoding	$\chi^2(1, 96) = 0.76$, $p = 0.384$
Task 4	Quantitative Data Encoding	$\chi^2(1, 120) = 0.68$, $p = 0.411$
Task 5	**Quantitative Data Encoding**	$F(1, 94) = \mathbf{5.79}$, $p = \mathbf{0.018}$
Task 6	Quantitative Data Encoding	$\chi^2(1, 96) = 1.08$, $p = 0.299$
Task 7	**Network Size**	$\chi^2(1, 192) = \mathbf{5.90}$, $p = \mathbf{0.015}$
	Grouping Encoding	$\chi^2(1, 192) = \mathbf{4.64}$, $p = \mathbf{0.031}$
	Quantitative Data Encoding	$\chi^2(1, 192) = 1.68$, $p = 0.194$
	Network Size and Grouping Encoding	$\chi^2(1, 192) = 1.12$, $p = 0.290$
	Network Size and Quantitative Data Encoding	$\chi^2(1, 192) = 0.56$, $p = 0.454$
	Grouping Encoding and Quantitative Data Encoding	$\chi^2(1, 192) = 0.22$, $p = 0.637$
	Network Size, Grouping Encoding, and Quantitative Data Encoding	$\chi^2(1, 192) = 0.23$, $p = 0.635$

Table 5.1: Summary Statistics for Accuracy by Tasks

However, with regards to the relationship between the network size and the encoding method on the confidence level ratings, the size variable was significant at the 95% confidence level, while the encoding variable was not significant. Also, the interaction between size and encoding was not significant. The post-hoc analysis using the Tukey test classified the means for the size variable as being part of two groupings, meaning that the means were significantly different. However, medium-sized networks had a mean of 7.000 and a standard deviation of 0, while large-sized networks had a mean of 6.917 and a standard deviation of 0.402. The means for the encodings were not significantly different, and were placed into the same Tukey group. The colored nodes encoding had a mean of 6.927 and a standard deviation of 0.391 for 96 data points, while the colored halos encoding had a mean of 6.99 with a standard deviation of 0.102 for 96 data points. A graph of the average confidence level ratings for all independent variables is shown in Figure 5.4.

Task	**Independent Variable**	**Significance**
Task 1	**Network Size**	$F(1, 188) = \mathbf{4.19}$, $p = \mathbf{0.042}$
	Grouping Encoding	$F(1, 188) = 2.36$, $p = 0.126$
	Network Size and Grouping Encoding	$F(1, 188) = 2.36$, $p = 0.126$
Task 2	**Network Size**	$F(1, 92) = \mathbf{15.75}$, $p = \mathbf{0.0001}$
	Grouping Encoding	$F(1, 92) = 0.48$, $p = 0.492$
	Network Size and Grouping Encoding	$F(1, 92) = 0.12$, $p = 0.731$
Task 3	**Network Size**	$F(1, 92) = \mathbf{8.68}$, $p = \mathbf{0.0041}$
	Grouping Encoding	$F(1, 92) = 0.18$, $p = 0.675$
	Network Size and Grouping Encoding	$F(1, 92) = 1.59$, $p = 0.21$
Task 4	Quantitative Data Encoding	$F(1, 118) = 0.29$, $p = 0.591$
Task 5	Quantitative Data Encoding	$F(1, 94) = 1.30$, $p = 0.257$
Task 6	Quantitative Data Encoding	$F(1, 94) = 0.48$, $p = 0.492$
Task 7	Network Size	$F(1, 184) = 0.71$, $p = 0.399$
	Grouping Encoding	$F(1, 184) = 1.76$, $p = 0.186$
	Quantitative Data Encoding	$F(1, 184) = 1.98$, $p = 0.161$
	Network Size and Grouping Encoding	$F(1, 184) = 2.22$, $p = 0.138$
	Network Size and Quantitative Data Encoding	$F(1, 184) = 1.36$, $p = 0.245$
	Grouping Encoding and Quantitative Data Encoding	$F(1, 184) = 1.56$, $p = 0.214$
	Network Size, Grouping Encoding, and Quantitative Data Encoding	$F(1, 184) = 0.04$, $p = 0.841$

Table 5.2: Summary Statistics for Confidence Level Ratings by Tasks

The joint combinations of the variables had the following statistics: medium – colored nodes (mean = 7.000, std. dev. = 0), medium – colored halos (mean = 7.000, std. dev. = 0), large – colored nodes (mean = 6.854, std. dev. = 0.545), and large – colored halos (mean = 6.979, std. dev. = 0.144). Furthermore, all trials with medium networks had been given a confidence level rating of 7, while for large networks the ratings ranged from 4 to 7 for the colored nodes encoding, and 6 to 7 for the colored halos encoding.

5.2.2 Task 2

Task 2 required participants to estimate the number of nodes in the highlighted group. Trials in this task used two variables which were the network size (medium or large) and the encoding method (colored nodes or colored halos).

With regards to accuracy, neither the size nor the encoding had a significant effect. The combination of size and encoding was also not significant. Furthermore, the Tukey test for both independent variables classified the sizes into the same Tukey group and the encodings into the same Tukey group. Therefore, the means were not significantly different. Medium-sized networks had a mean of 0.03 and a standard deviation of 0.084, while large-sized networks had a mean of 0.056 and a standard deviation of 0.095 for 48 data points each. The colored nodes encoding had a mean of 0.036 and a standard deviation of 0.089, while the colored halos encoding had a mean of 0.051 with a standard deviation of 0.091. A graph of the average relative error for all independent variables is shown in Figure 5.1. The joint combinations of the variables had the following statistics: medium – colored nodes (mean = 0.03, std. dev. = 0.103), medium – colored halos (mean = 0.031, std. dev. = 0.06), large – colored nodes (mean = 0.042, std. dev. = 0.074), and large – colored halos (mean = 0.071, std. dev. = 0.111).

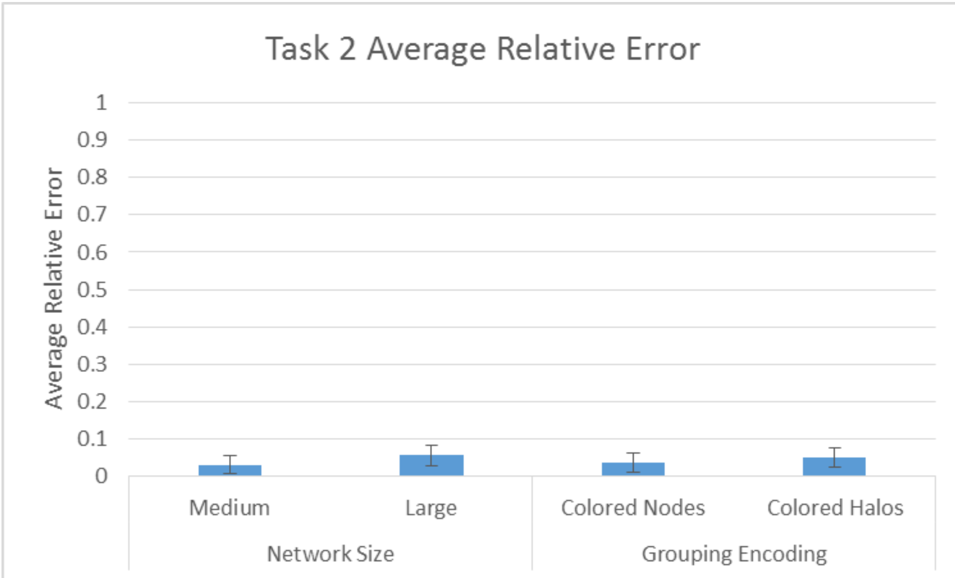

Figure 5.1: Average Relative Error in Task 2

The size variable was significant in relation to the participants' confidence levels. However, the encoding was not significant, and neither was the interaction between size and encoding. The interaction plot from SAS showing the interaction between the network size and the confidence levels is shown in Figure 5.2.

Furthermore, a post-hoc analysis using the Tukey test showed that the medium and large sized networks belong to separate Tukey groups, meaning that the difference in the means was significantly different. Trials with medium networks had a mean of 6.417 and a standard deviation of 0.986 for 48 data points, while large networks had a mean of 5.458 and a standard deviation of 1.336 for 48 data points. The means for the encodings were not significantly different, and were placed into the same Tukey group. The colored nodes encoding had a mean of 5.854 and a standard deviation of 1.414 for 48 data points, while the colored halos encoding had a mean of 6.021 with a standard deviation of 1.101 for 48 data points. A graph of the average confidence level ratings for all independent variables is shown in Figure 5.4.

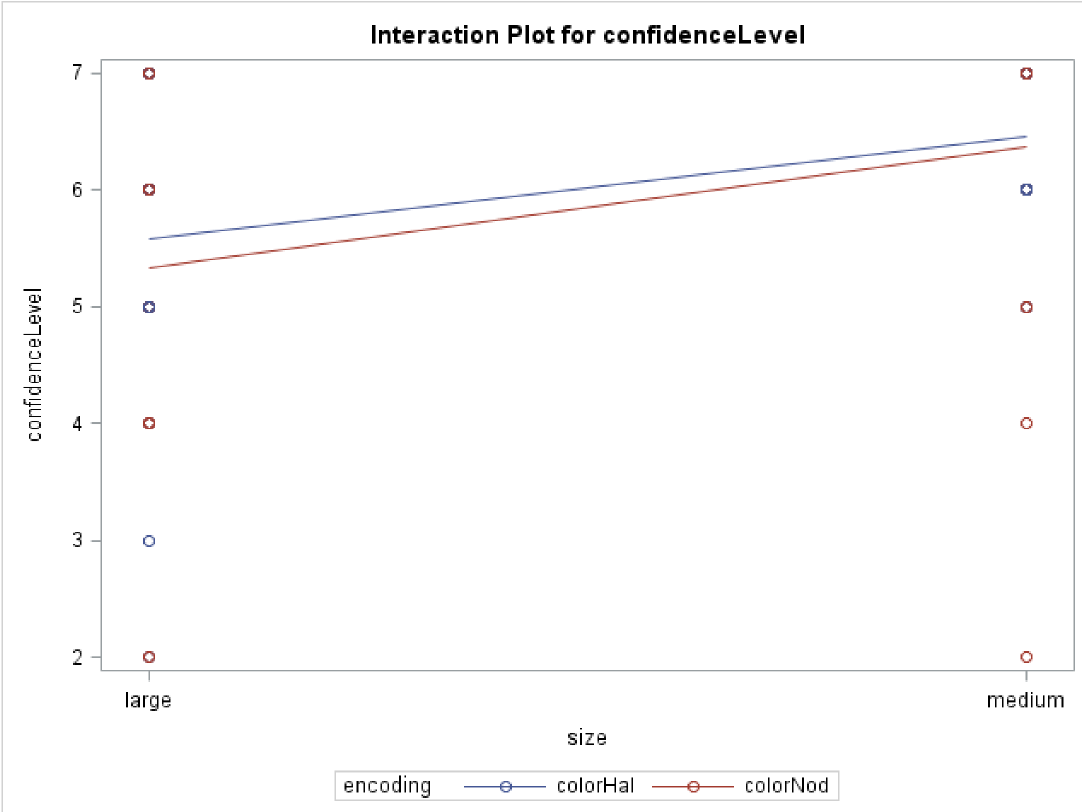

Figure 5.2: Interaction between Grouping Encoding and Network Size in Task 2

The joint combinations of the variables had the following statistics: medium – colored nodes (mean = 6.375, std. dev. = 1.245), medium – colored halos (mean = 6.458, std. dev. = 0.658), large – colored nodes (mean = 5.333, std. dev. = 1.404), and large – colored halos (mean = 5.583, std. dev. = 1.283). Furthermore, all comparisons of size and encoding combinations had a minimum confidence level rating of 2 and a maximum confidence level rating of 7, except for the medium-sized network with the colored halos encoding, which had a minimum confidence level rating of 5.

5.2.3 Task 3

Task 3 required participants to identify the relative proportional ratio between the numbers of nodes in each of two highlighted groups. Trials in this task used two variables which were the network size (medium or large) and the encoding method (colored nodes or colored halos).

With regards to accuracy, the network size was significant at the 95% confidence level. Medium networks had a mean of 0.917 and a standard deviation of 0.279 for 48 data points, while large networks had a mean of 0.75 and a standard deviation of 0.438. However, the encoding was not significant, nor was the joint combination of size and encoding. Colored nodes had a mean of 0.854 with a standard deviation of 0.357, while colored halos had a mean of 0.813 and a standard deviation of 0.394. A graph of the average accuracy for all independent variables is shown in Figure 5.3. The joint combinations of the variables had the following statistics: medium – colored nodes (mean = 0.958, std. dev. = 0.204), medium – colored halos (mean = 0.875, std. dev. = 0.338), large – colored nodes (mean = 0.750, std. dev. = 0.442), and large – colored halos (mean = 0.750, std. dev. = 0.442).

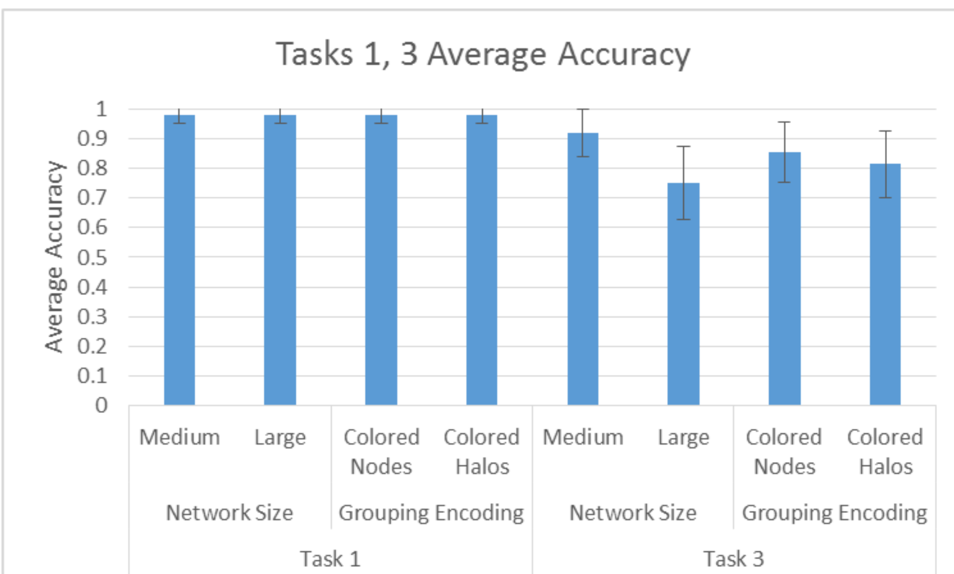

Figure 5.3: Average Accuracy in Tasks 1 and 3

When the network size is large, the odds ratio is 1, indicating that either encoding is equally likely to produce a correct answer. However, when the network size is medium, the odds ratio is 0.304, indicating that a correct answer is more likely with the colored nodes encoding. When using either a colored nodes encoding (odds ratio = 0.130) or a colored halos encoding (odds ratio = 0.429), since both ratios are less than 1, this means that a correct answer is more likely with medium-sized networks.

The results for confidence level ratings show that size was again significant, while encoding and the combination of size and encoding were not significant. The Tukey test showed that the means for medium and large networks were significantly different, and were placed into separate Tukey groups. Trials with medium networks had a mean of 6.208 and a standard deviation of 1.091 for 48 data points, while large networks had a mean of 5.479 and a standard deviation of 1.321 for 48 data points. The means for the encodings were not significantly different, and were placed into the same Tukey group. The colored nodes encoding had a mean of 5.896 and a standard deviation of 1.207, while the colored halos encoding had a mean of 5.792 with a standard deviation of 1.32. A graph of the average confidence level ratings for all independent variables is shown in Figure 5.4. The joint combinations of the variables had the following statistics: medium – colored nodes (mean = 6.417, std. dev. = 0.504), medium – colored halos (mean = 6.000, std. dev. = 1.445), large – colored nodes (mean = 5.375, std. dev. = 1.469), and large – colored halos (mean = 5.583, std. dev. = 1.176).

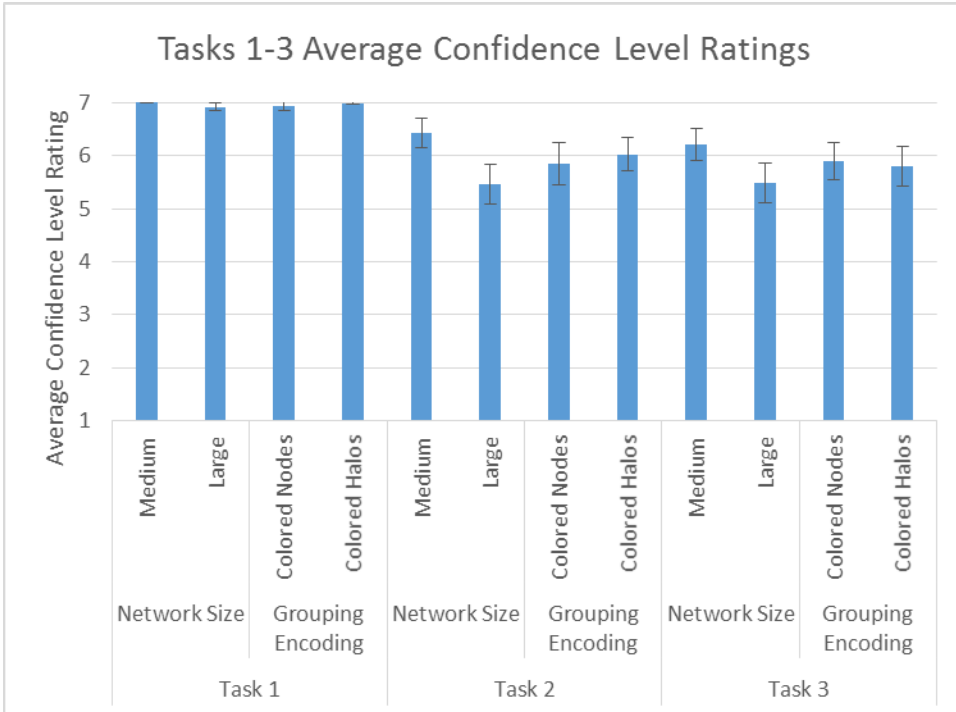

Figure 5.4: Average Confidence Level Ratings in Tasks 1-3

5.2.4 Task 4

Task 4 required participants to identify whether the value of the highlighted node was greater than or less than 50. Trials in this task used one variable which was the encoding method (variable-area circles or calibrated columns). All networks were medium-sized and used the colored halos encoding.

In total, among all participants, 120 data points were collected for this task. However, only six out of these 120 trials were answered incorrectly. So, the encoding method was not significant in terms of accuracy. Variable-area circles had a mean of 0.967 and a standard deviation of 0.181, while calibrated columns had a mean of 0.933 and a standard deviation of 0.252. The odds ratio for encoding was 0.483, indicating that a correct answer is more likely with the variable-area circles encoding. A graph of the average accuracy is shown in Figure 5.6.

With regards to the distribution of randomly selected node values, the minimum and maximum node values for variable-area circles was 18.071 and 85.004 respectively, while for calibrated columns it was 14.881 and 74.367 respectively. Furthermore, variable-area circles had a mean node value of 46.727 and a standard deviation of 24.361, while calibrated columns had a mean node value of 44.524 and a standard deviation of 22.933.

With regards to the confidence level ratings, encoding was not significant. Furthermore, the Tukey test placed both encodings in the same Tukey group, indicating that their means were not significantly different. Variable-area circles had a mean of 6.417 and a standard deviation of 1.03, while calibrated columns had a mean of 6.317 and a standard deviation of 0.9999. Both methods had the same minimum rating of 3 and maximum rating of 7. A graph of the average confidence level ratings is shown in Figure 5.7.

5.2.5 Task 5

Task 5 required participants to estimate the exact value of a highlighted node. Trials in this task used one variable which was the encoding method (variable-area circles or calibrated columns). All networks were medium-sized and used the colored halos encoding.

With regards to accuracy, encoding was significant at the 95% confidence level. Tukey's test determined that the means for both encodings were significantly different, and placed them into two separate Tukey groups. Variable-area circles had a mean of 0.14 and a standard deviation of 0.127, while calibrated columns had a mean of 0.247 and a standard deviation of 0.281. A graph of the average relative error is shown in Figure 5.5.

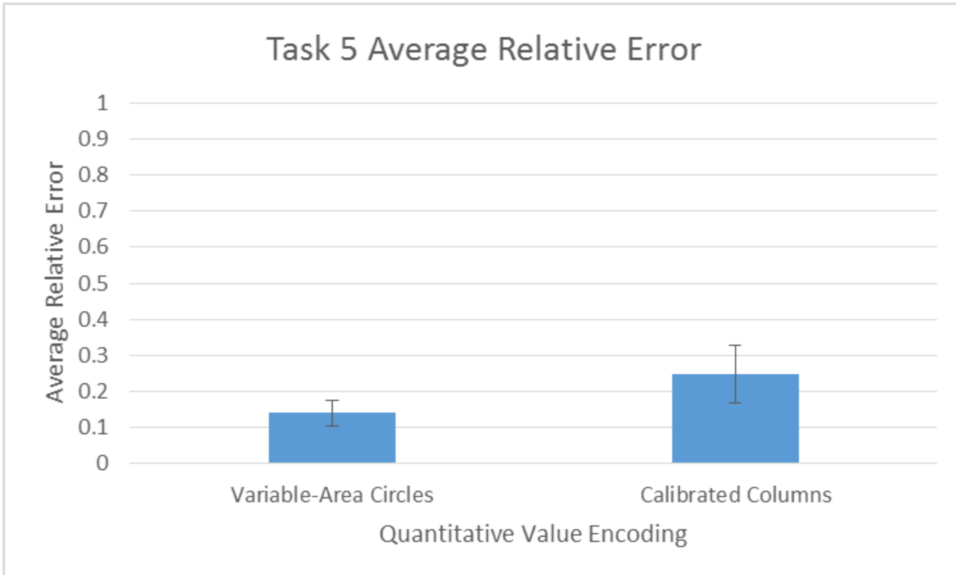

Figure 5.5: Average Relative Error in Task 5

With regards to the distribution of randomly selected node values, the minimum and maximum node values for variable-area circles was 9.54 and 76.16 respectively, while for calibrated columns it was 5.201 and 85.528 respectively. Furthermore, variable-area circles had a mean node value of 34.651 and a standard deviation of 24.069, while calibrated columns had a mean node value of 33.977 and a standard deviation of 26.914.

However, encoding was not significant with regards to confidence level rating. Tukey's test placed the means for the two encodings in the same group, since they were not significantly different. Variable-area circles had a mean of 4.75 and a standard deviation of 1.564, while calibrated columns had a mean of 5.104 and a standard deviation of 1.477. Both had a minimum confidence level rating of 2 and a maximum rating of 7. A graph of the average confidence level ratings is shown in Figure 5.7.

5.2.6 Task 6

Task 6 required participants to identify the relative proportional ratio between the values of two highlighted nodes. Trials in this task used one variable which was the encoding method (variable-area circles or calibrated columns). All networks were medium-sized and used the colored halos encoding.

Encoding was not significant in this task when evaluating accuracy. Variable-area circles had a mean of 0.854 and a standard deviation of 0.357, while calibrated columns had a mean of 0.771 and a standard deviation of 0.425. The odds ratio estimate was 0.574, meaning that the circles encoding would more likely produce accurate results for this task. A graph of the average accuracy is shown in Figure 5.6.

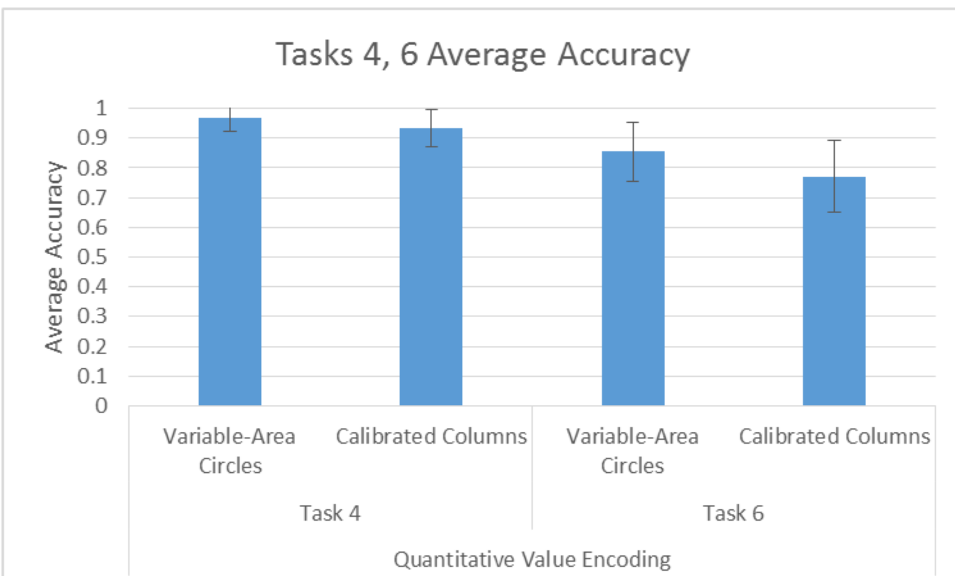

Figure 5.6: Average Accuracy in Tasks 4 and 6

With regards to the distribution of randomly selected pairs of node values, variable-area circles had a mean relative proportion of 1.688 with a standard deviation of 0.665, while calibrated columns had a mean relative proportion of 1.875 and a standard deviation of 0.747.

Encoding was also not significant in terms of confidence level ratings. The Tukey test classified both means to be in the same Tukey group, meaning that the means were not significantly different. Variable-area circles had a mean of 6.208 and a standard deviation of 0.922 for 48 data points, while calibrated columns had a mean of 6.063 and a standard deviation of 1.137. A graph of the average confidence level ratings is shown in Figure 5.7.

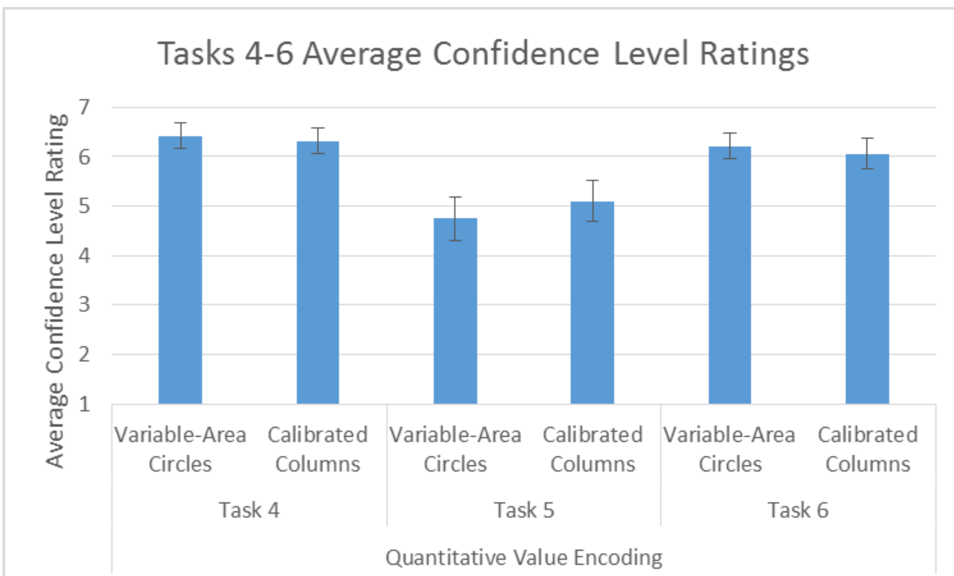

Figure 5.7: Average Confidence Level Ratings in Tasks 4-6

5.2.7 Task 7

Task 7 required participants to identify the group having the largest variance. Trials in this task used three variables which were the network size (medium or large), the encoding method for grouping (colored nodes or colored halos), and the encoding method for quantitative data (variable-area circles or calibrated columns).

With regards to accuracy, both size and the encoding method for grouping were significant at the 95% confidence level. The encoding method for quantitative data was not significant. Furthermore, the joint combinations of the sizes and the encodings were not significant as well. Medium networks had a mean of 0.521 and a standard deviation of 0.502, while large networks had a mean of 0.354 and a standard deviation of 0.481. Colored nodes had a mean of 0.365 and a standard deviation of 0.484, while colored halos had a mean of 0.51 and a standard deviation of 0.503. Also, variable-area circles had a mean of 0.396 and a standard deviation of 0.492, while calibrated columns had a mean of 0.479 and a standard deviation of 0.502. A graph of these average accuracies is shown in Figure 5.8.

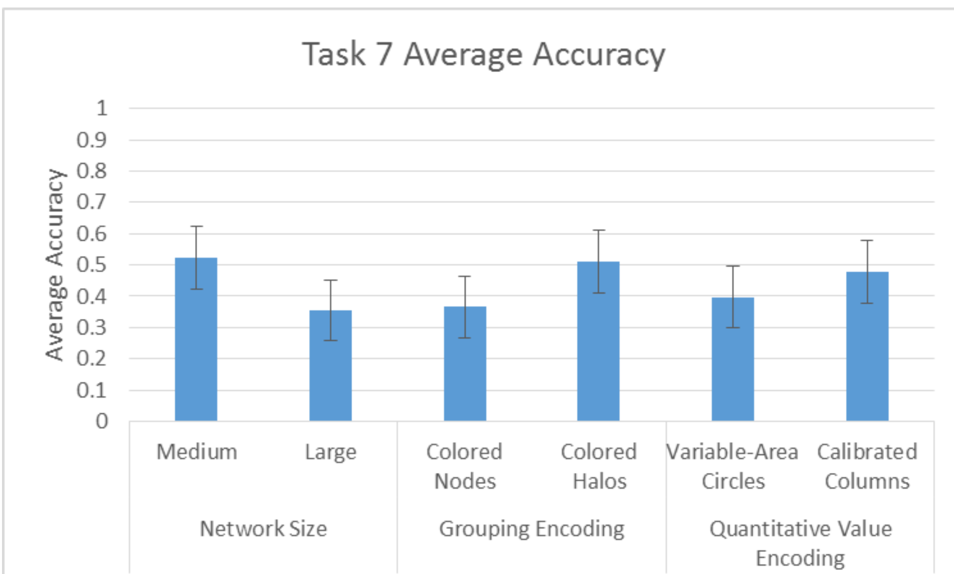

Figure 5.8: Average Accuracy in Task 7

The joint combinations of the encodings provide further statistics with regards to accuracy. The size and grouping encoding combinations were medium – colored nodes (mean = 0.479, std. dev. = 0.505), medium – colored halos (mean = 0.563, 0.501), large – colored nodes (mean = 0.250, std. dev. = 0.438), and large – colored halos (mean = 0.458, std. dev. = 0.504). The size and quantitative data encoding combinations were medium – circles (mean = 0.500, std. dev. = 0.505), medium – calibrated columns (mean = 0.542, 0.504), large – circles (mean = 0.292, std. dev. = 0.459), and large – calibrated columns (mean = 0.417, std. dev. = 0.498). The grouping and quantitative data encoding combinations were colored nodes – circles (mean = 0.313, std. dev. = 0.468), colored nodes – calibrated columns (mean = 0.417, 0.498), colored halos – circles (mean = 0.479, std. dev. = 0.505), and colored halos – calibrated columns (mean = 0.542, std. dev. = 0.504).

Network size had an odds ratio of 0.494, indicating that it was more likely for a medium-sized network to be associated with a correct answer. The grouping encoding method had an odds ratio of 1.859, indicating that the colored halos encoding would more likely produce a correct answer. Finally, the quantitative data encoding method had an odds ratio of 1.43, indicating that the calibrated columns encoding would more likely produce a correct answer. The SAS plot of odds ratios for this task is shown in Figure 5.9.

Analyzing the distribution of randomly selected compartment value ranges showed that medium networks had a mean range of 65.191 and a standard deviation of 8.898, while large networks had a mean range of 78.51 and a standard deviation of 7.366. The range values for the combinations of encodings was also analyzed: colored nodes – circles (mean = 71.719, std. dev. = 14.769), colored nodes – calibrated columns (mean = 72.378, std. dev. = 9.108), colored halos – circles (mean = 72.33, std. dev. = 8.781), and colored halos – calibrated columns (mean = 70.975, std. dev. = 8.448). Furthermore, analysis of the difference between the ranges of the correct compartment and the compartment that was selected by the participant showed that the range difference had a mean of 18.496 and a standard deviation of 10.004, with a minimum difference of 3.027 and a maximum difference of 42.447.

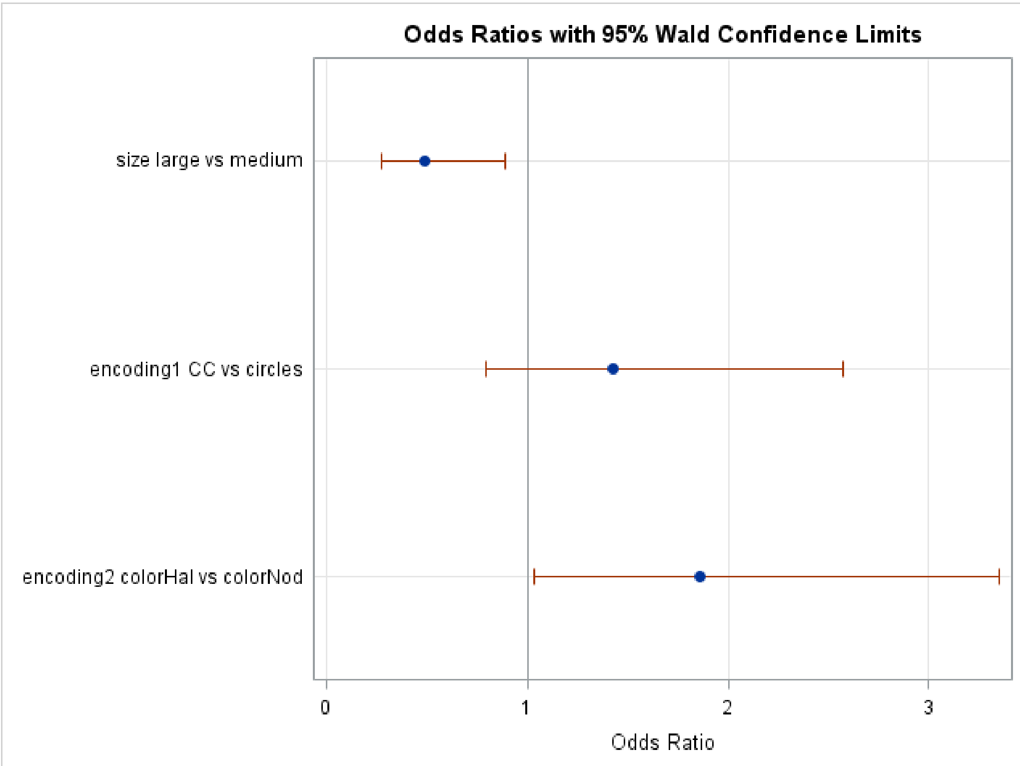

Figure 5.9: Odds Ratios in Task 7

The network size, the grouping encoding, and the quantitative data encoding were not significant for confidence level ratings. Furthermore, the joint combinations of the sizes and the encodings were not significant as well. In all Tukey tests, the variables were placed into the same Tukey group, indicating that the means within each variable were not significantly different. Medium networks had a mean of 4.969 and a standard deviation of 1.744 for 96 data points, while large networks had a mean of 4.75 and a standard deviation of 1.869. Colored nodes had a mean of 4.688 and a standard deviation of 1.831, while colored halos had a mean of 5.031 and a standard deviation of 1.774. Furthermore, variable-area circles had a mean of 5.042 and a standard deviation of 1.812, while calibrated columns had a mean of 4.677 and a standard deviation of 1.792. A graph of these average confidence level ratings is shown in Figure 5.10.

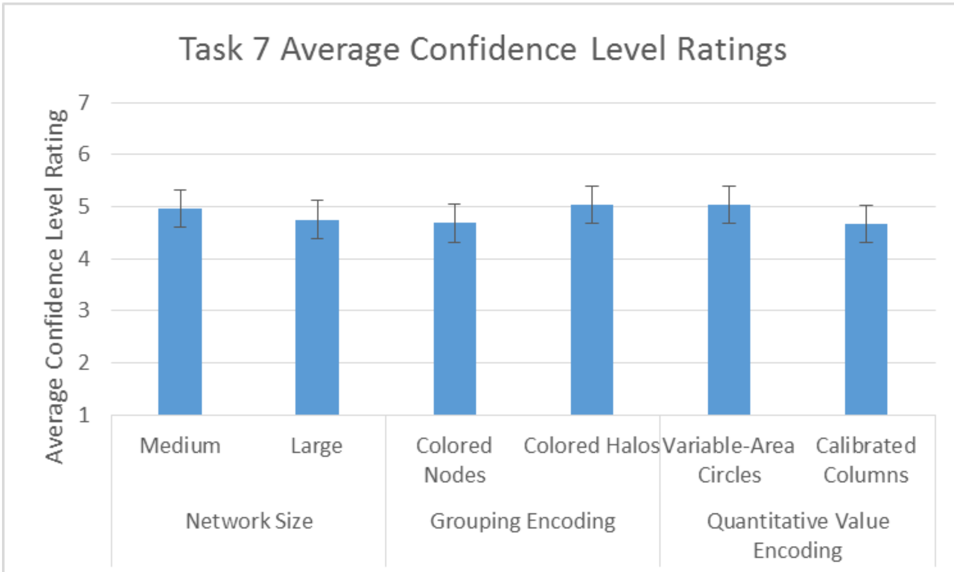

Figure 5.10: Average Confidence Level Ratings in Task 7

<u>5.3 Post-Questionnaire Results</u>

Figure 5.11 shows the average rating among the six participants for their experience using the different encodings in the study. A rating of 1 indicates "very low" while a rating of 7 indicates "very high". Error bars representing 95% confidence intervals are shown on the graph, but it is important to note that the sample size was only six participants. So, the standard deviations were comparatively high for the given rating scale. A discussion about these results is presented in Section 5.5.4.

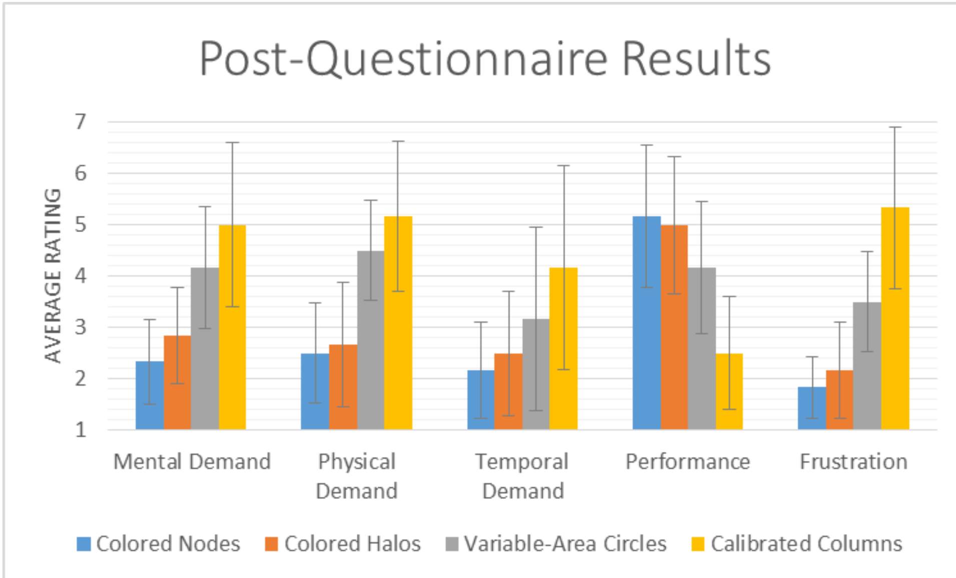

Figure 5.11: Post-Questionnaire Results

5.4 Short Interview Results

Five out of the six participants mentioned that they ran out of time while completing Task 7. Some participants found the task especially difficult for large networks, usually when the calibrated columns encoding was used. In general, participants found that large networks were more difficult to analyze within the given task duration due to the greater number of nodes present on the screen at once, when compared to medium networks. Additionally, one participant mentioned that large networks generally reduced the participant's confidence level ratings. Another participant mentioned that large networks were difficult for tasks involving the counting of nodes (Tasks 2 and 3).

Five out of the six participants found the colored nodes encoding easier to perform grouping tasks with than the colored halos for a few reasons. In the colored halos encoding, one participant noted that the black node was noticed first, before the surrounding halo was perceived. Also, the colored nodes encoding had less "noise" to look at. Furthermore, colors were more easily perceivable and differentiable with the colored nodes. In particular, when nodes from different compartments overlapped, the overlapping of the black nodes made it difficult to determine the associated halos for some nodes. However, one participant noted that the presence of the halos did not inhibit his/her ability to find nodes and groupings. In addition, participants generally believed that they were more accurate in completing grouping tasks when the colored nodes encoding was used.

Participants had mixed responses regarding the variable-area circles and calibrated columns encodings. For different scenarios, they generally preferred one or the other. Some participants found that after performing a few trials using the circles encoding, they were able to answer tasks without frequently referring to the legend. However, other participants mentioned that values were a bit difficult to differentiate for circles when the values were large, such as above 50. With regards to calibrated columns, some participants found that they were able to easily identify the general value category for the calibrated columns, based on a node's width, but needed to refer to the legend to differentiate between different node heights within the same value category. Most participants generally believed that they completed tasks more accurately using the calibrated columns method, especially in tasks involving quick comparison of node values (Task 4) and determination of exact node values (Task 5). However, most participants mentioned that although more accurate, analyzing networks using the calibrated columns encoding took more time than analyzing networks using the variable-area circles encoding.

One participant preferred the variable-area circles encoding since the participant compared values by panning nodes to the legend, and comparing the nodes' heights to estimate values. For circles, the participant only needed to focus on one variable, while for calibrated columns, the participant need to focus on two variables. In large networks where some nodes were overlapping, participants found the variable-area circles more effective, since the radii were more easily perceivable. Calibrated columns were more difficult in this case, since the bases of the rectangles were difficult to perceive. Two participants thought that Task 7 was a little easier to complete when the variable-area circles encoding was used, since it was easier to find the smallest circle in the network, as opposed to the smallest rectangle.

Three out of the six participants mentioned that the light blue and light green colors used in the study were sometimes difficult to distinguish if nodes were overlapping. This was especially pointed out with regards to overlapping halos using these colors. Some minor issues faced by participants were that one participant accidently pressed the "Next" button twice during a few trials, thereby making the participant move to the answer input screen without viewing the associated visualization. Another participant sometimes missed the flashed yellow halo in tasks 4-6, and therefore needed to use some of the trial duration to find the specified nodes. This detracted from the participant's time to perform the given task.

One participant provided the suggestion that participants should be allowed to tilt the monitor if necessary, since based on a participant's height, the perception of the colors used in the study can slightly change. Another participant suggested that participants should be allowed to adjust the screen brightness, since participants may not be accustomed to using a monitor with a high resolution.

5.5 Discussion

Results from the pilot study, the post-questionnaire, and the short interview were presented in Sections 5.2-5.4. In this section, I provide a discussion and analysis of the results with suggestions for potential changes to the study.

5.5.1 Grouping Tasks

Results from the grouping tasks (Tasks 1-3) show that the encoding method of colored nodes or colored halos has no significant effect on the accuracy of the responses. Furthermore, for Tasks 1 and 2, the network size had no significant effect on the accuracy. In Task 3, only the network size had a significant effect on accuracy. It is clearly noticeable that medium networks had a larger average accuracy and a smaller standard deviation than large networks for this task. These results could possibly suggest that medium networks allowed participants to spend a comparatively longer time to perform the tasks, since there were much fewer nodes and compartments to view. In particular, medium networks had only a quarter of the number of nodes and a half of the number of compartments than was in the large networks.

Results from these tasks also show that network size had a significant effect on confidence level ratings for all three grouping tasks. It is interesting to note that in Task 1, all trials using medium networks had been given a confidence level rating of 7. For all three tasks, medium-sized networks had a higher average confidence level rating and a lower standard deviation than for large-sized networks. These findings support the general consensus from the short interview that participants had greater confidence in their answers when presented with tasks using a medium network.

To account for these observations, a possible change in future studies is to allocate less time for completing a task using medium networks, and allocate a proportionately longer time for completing a task using large networks. Although this change may seem reasonable, it will only serve to normalize the timings based on the difficulty of the task, thereby making all trials, regardless of the network size, to have about the same level of difficulty. Results from Tasks 1 and 2 show that network size was not significant in terms of accuracy, but it was significant in terms of confidence level ratings. Furthermore, results from Tasks 3 and 7 show that network size was significant in terms of accuracy, but was only significant in terms of confidence level ratings for Task 3. Therefore, these observations suggest that as the task gets more difficult and requires a higher level of reading of the network, the network size becomes significant in terms of accuracy, but does not necessarily affect confidence level ratings. Since the performance differences become more evident under increased task difficulty, both medium networks and large networks should continue to be allocated the same task durations.

For all three tasks, the encoding method had no significant effect on the accuracy nor the confidence level ratings. While the interaction plot for Task 2 shows that the colored halo encoding had a higher average confidence level rating than the colored node encoding in both network sizes, the difference is not significant. These findings leads us to believe that there is no significant difference between using the colored nodes and the colored halos, in the three grouping tasks that were conducted. So, this means that one encoding method does not significantly detract from or enhance the perception of grouping in the networks any more than the other encoding method does. Therefore, in future studies, either of the two encodings could be used to further evaluate just the impact of the quantitative value encodings in simple tasks involving either an elementary or intermediate level of reading.

Out of a total of 192 trials in Task 1, only the answers to four trials were incorrect, thereby suggesting that participants were able to intuitively judge grouping using both encoding methods. Trials were selected such that nodes appeared in both same and different compartments, while also including trials where nodes appeared both near and far from each other spatially on the screen. Since the purpose of this task was to intuitively judge whether two nodes belonged to the same group, we propose that perhaps the organization of the nodes in the network prevented either encoding method from having a significant effect. This is because in the generated networks, nodes belonging to the same compartment were located around the parent node for the compartment. Therefore, regardless of the encoding method, the nodes would still appear near their parent nodes, which could immediately suggest a correct answer. So, in future studies, we propose that the same grouping tasks could be repeated, but with the nodes appearing in more random locations on the screen. Such a comparison between nodes being clustered around the parent node, and nodes appearing in random locations on the screen, could perhaps suggest a more significant difference between the encodings.

5.5.2 Quantitative Tasks

Results from Tasks 4 and 6 show that the quantitative data encodings of variable-area circles and calibrated columns had no significant effect on the accuracy. It is important to note that out of 120 data points collected in Task 4, only six out of the 120 trials were answered incorrectly. This may be due to our choice of p and k in the Weber-Fechner's law. For this study, we had estimated the values from relevant studies conducted by other researchers (Chung et al., 2016; Nachmias, 2011). We suspect that the task was too easy for the participants, due to the JND being set to too high a threshold. In future studies, it would be useful to conduct a small study to empirically derive values for p and k that are relevant to the tasks in this study. In this way, degrees of variance from 50 would reflect more minute variations, which would allow us to possibly see more significant differences between the encodings. This is because the added perception of both width and height using calibrated columns would more likely allow for the determination of accurate values, compared to just the perception of radius using variable-area circles. This supposition is justified by the results of the short interview, where participants suggested that they answered tasks more accurately using the calibrated columns encoding.

For Task 5, the encoding was significant in terms of accuracy at the 95% confidence level. Although, the significance was clear, it was contrary to our hypothesis that participants would perform better using the calibrated columns. Figure 5.5 shows that the calibrated columns had a higher relative error and a higher standard deviation than the variable-area circles. Further analysis of the randomly selected node values show that the range of generated values was smaller for variable-area circles than it was for calibrated columns. Although the mean node values and the standard deviations were approximately close in value, this difference could mean that some trials were easier for variable-area circles than it was for calibrated columns. The short interview with each participant showed that most participants had difficulty differentiating between circles with greater values in the data range. Therefore, providing some trials that used circles with smaller values than that of the calibrated columns could have underestimated the relative error values associated with the variable-area circles encoding. Similar discrepancies in the randomly assigned node values were also observed in Tasks 4 and 6. For instance, in Task 4, variable-area circles had a larger overall data range, thereby suggesting that some trials specified nodes with values much farther from 50, than were specified in trials using calibrated columns. Since values closer to 0 and 100 are more easily differentiable than values closer to 50, some trials could have been easier with the variable-area circles.

In future studies, in order to prevent this issue, it would be useful to generate a specific set of random node values and assign exactly half to each node encoding. All participants will complete tasks using the same data values, however one half of the participants will complete the tasks using one set of encoding and value assignments, while the other half of the participants will complete the tasks using the opposite set of encoding and value assignments. In this way, both encoding methods can be equally tested on the same data values, and it would prevent similar situations where one encoding gets assigned an easier or harder trial than another encoding. Such a study would be more useful given a large sample of participants. At least 12 participants may be necessary for an enhanced pilot study, so that six participants would be able to work with each version of the value assignments. A larger sample of participants would not be necessary in this case, since the purpose of this enhanced pilot study would be to retrieve preliminary results using the revised data assignment method and to determine any further errors in the study's design.

Results from all three quantitative tasks show that the quantitative data encodings have no significant effect on the confidence level ratings. This is consistent with the results of the short interview, where participants preferred one encoding or the other depending on several factors. There was no strong consensus regarding both encoding methods.

5.5.3 Overall Task

Results from Task 7 show that the network size was significant in terms of accuracy. This observation is consistent with the general findings from the grouping tasks, and the short interview with the participants. Furthermore, the grouping encoding method was significant, but not the quantitative data encoding method. Figure 5.8 shows that colored halos had a higher average accuracy by about fifteen percent than the colored nodes. This result supports our hypothesis that colored halos provide an improved perception of grouping. Additionally, Figure 5.8 shows that while both variable-area circles and calibrated columns had an average accuracy less than fifty percent, calibrated columns had an average accuracy that was about eight percent higher than that of the variable-area circles. This result supports our hypothesis that calibrated columns allow for an improved perception of node values than the variable-area circles.

Further analysis of the joint combinations of the encodings provided additional confirmation for the observed results. For instance, analysis of the size and grouping encodings showed that regardless of the network size, the colored halos encoding had a greater average accuracy than the colored nodes encoding. Furthermore, analysis of the size and quantitative data encodings showed that regardless of the network size, the calibrated columns encoding had a greater average accuracy that the variable-area circles encoding. Additionally, analysis of the grouping and quantitative data encodings showed that regardless of the grouping encoding, the calibrated columns encoding had a greater average accuracy than the variable-area circles encoding. Similarly, regardless of the quantitative data encoding, the colored halos encoding had a greater average accuracy than the colored nodes encoding.

Although these results support our hypotheses, the statistics for the randomly selected data ranges show that although the average mean ranges are similar among the pairs of encodings, there are some differences in the standard deviations. Such variations can be prevented by performing a paired study using the same data values, as mentioned in the discussion for the quantitative tasks, in order to further validate these preliminary results. Additionally, differences between the ranges of the correct groups and the ranges of the groups selected by participants show that the average differences were often very small, as shown by the large standard deviation, as well as the small minimum difference value. This problem can be rectified in future studies by selecting groups that have ranges that are more distinguishable from the ranges of the other groups in the network. In this way, minute differences between the ranges of groups in a network can be avoided. Additionally, it is possible to evaluate the results for this task in terms of relative error, in order to normalize responses in terms of the deviation from the range of the correct answer.

The network size, grouping encodings, and quantitative data encodings did not have a significant effect on the confidence level ratings. This observation may have been due to the difficulty of the task, in comparison to the difficulties of the other six tasks. While the other tasks involved an elementary or intermediate level of reading, this task involved an overall level of leading that required both precise data reading as well as global perception of the network. Therefore, the task's difficulty could have masked participants' confidence level ratings in terms of the network sizes or the encodings themselves.

5.5.4 Post-Questionnaire

Results from the post-questionnaire were largely varied in range, due to the small sample of participants. Since there were only six participants, the average ratings could easily be skewed given an outlier. More concrete relationships can be determined once a study with a greater number of participants is performed. Overall, the results are mostly consistent with the participants' opinions from the short interview. One future area of analysis is the question regarding performance. Although participants generally believed that they were more accurate when using the calibrated columns method, it is surprising that their average rating was much lower than that for the variable-area circles method. A possible reason for this observation is that the participants had completed the post-questionnaire immediately after completing Task 7. This task had presented all combinations of the network sizes as well as the grouping and quantitative data encodings. As shown in the results from the short interview, the participants had several concerns about the calibrated columns in relation to Task 7. For instance, when the networks were large, participants found the task especially difficult when the calibrated columns encoding was used, since it was difficult to perceive the bases of the rectangles when nodes were overlapping. Furthermore, some participants had found it easier to find the smallest circle, as opposed to the smallest rectangle. Therefore, it is possible that the "recency effect" had influenced the outcomes in the post-questionnaire, since the participants might have remembered their performance using the encodings in Task 7 better than they had remembered their performance in the middle set of tasks (Tasks 4-6) which also involved the quantitative encodings ("recency effect", 1998). In future studies, this question could be reworded, or another question could be added, asking participants to rate how accurate they believed their responses were when using each of the visualization techniques.

5.6 Summary of Changes to the Empirical Study Design

The pilot study revealed several preliminary results about network size, the grouping encodings, the quantitative data encodings, and their relationships with participants' accuracy and confidence level ratings in the seven different tasks. These findings are summarized in section 5.7. The purpose of a pilot study is to additionally identify areas of improvement in the study's design. These changes can be implemented in the design, in order to gather precise results from a larger sample of participants in future studies. Here, I summarize some changes to the study's design, as became evident after analyzing the results from the pilot study.

First, the results from this study need to be interpreted carefully, especially for tasks involving the quantitative data encodings. The most significant issue in this pilot study is that some tasks could have been easier or harder for trials using one encoding than for trials using the other encoding, due to the random assignment of data values, and the random selection of nodes for the trials. Analysis of statistics about the values of the selected node(s) in each trial, as reported in Section 5.2 and discussed in Section 5.5, provide evidence for this problem. For instance, both Tasks 4 and 5 could have been easier for trials using variable-area circles than for trials using calibrated columns, as previously discussed. To resolve this issue in future studies, the same set of randomly generated data values could be assigned to trials using the different encodings. A paired data assignment scheme can be used to ensure that trials have a consistent level of difficulty among all of the encodings. This change will allow for results to be equally comparable within independent variables, which will provide concrete and justifiable conclusions.

Additionally, a study can be conducted to empirically select p and k values for use with Weber-Fechner's law. These constants were estimated in this pilot study, based on studies conducted by other researchers, as discussed in Section 5.5. Empirically selecting the constants would allow us to reflect more minute variations between the quantitative values, such that the differences between the node sizes will more closely approximate the JND for the nodes in the quantitative tasks. This change will be useful in determining the differences in data reading accuracy for precise values, when using the quantitative data encodings.

5.7 Summary of Preliminary Results

The significant preliminary results from the pilot study are given below. Due to some of the caveats discovered in this pilot study, as mentioned in Section 5.6, these preliminary results must be interpreted carefully. Further studies are necessary after implementing the proposed changes, in order to validate these results.

1. Network size has a significant effect on confidence level ratings in tasks involving only grouping. In particular, participants were more confident when trials involved medium networks.
2. The grouping encodings do not have a significant effect on accuracy nor confidence level ratings in tasks involving only grouping.
3. The quantitative data encodings have a significant effect on accuracy when the task involves the estimation of a node's exact value.
4. The quantitative data encodings do not have a significant effect on confidence level ratings in any tasks using these encodings.
5. Network size and the grouping encodings have a significant effect on accuracy in the overall task.
6. Participants' answers were more accurate when using the colored halos and calibrated columns encodings in the overall task.
7. Network size and the grouping and quantitative data encodings did not have a significant effect on confidence level ratings in the overall task.

Chapter 6: Conclusion and Future Work

The perception of groupings in a network as well as values of individual nodes in the network is important in data visualization. In reality, a data set can have several hundreds or thousands of nodes, and it is important to be able to perceive both specific details about nodes, as well as overall relationships. For my thesis, I performed a pilot study using six participants to evaluate grouping encoding methods (colored nodes and colored halos) and quantitative data encoding methods (variable-area circles and calibrated columns). We hypothesized that colored halos would provide a better perception of groupings in the network, due to the larger surface area that a node will occupy on a screen. We additionally hypothesized that calibrated columns would provide a more accurate perception of a node's value, due to the dual perception of both changes in the height and width of a node.

Several preliminary results were gathered from the study, but these results must be interpreted carefully due to some caveats discovered in the pilot study's design. Changes such as implementing a paired data assignment scheme and empirically selecting constants are necessary to further validate these preliminary results. Results from the study showed that the network size had a significant effect on accuracy for one grouping task and a significant effect on confidence level ratings for all tasks involving only grouping. Furthermore, results showed that there was no significant effect of the grouping encoding methods on the accuracy and the confidence level ratings in tasks involving only grouping. For quantitative tasks, the encodings did not have an effect on the confidence level ratings. However, the encodings were significant in terms of accuracy for one task, but contrary to our hypothesis, thereby prompting some discussion for proposed changes to our design in future studies. Finally, the overall task showed results that supported both of our hypotheses, possibly indicating that these encodings are useful for more difficult tasks involving both precise and global perception of the network. The preliminary results from this study can be used to improve and expand the scope of the study in the future, in order to improve the visualization of complex data in all fields.

6.1 Future Work

As discussed in Sections 5.5 and 5.6, future studies can use a paired data assignment scheme to ensure that equivalent data values are being assigned to all of the encoding methods. For the grouping encodings, a study could be conducted to evaluate the impact of grouping child nodes around the parent nodes versus allowing child nodes to appear more randomly on the screen. Additionally, an empirical study can be conducted to select more precise constants to be used with Weber-Fechner's law. Furthermore, a small addition could be made to the post-questionnaire to capture participants' ratings of their accuracy when using each of the visualization techniques.

Also, a revised version of the empirical study presented in this document can be performed using a larger sample of participants. Also, more trials could be used per task. These adjustments would help to better identify patterns and anomalies in the data. Additionally, a small empirical study could be conducted to adjust the task durations based on the time it takes for a small group of participants to complete each task within some threshold level of accuracy. In this way, the task durations could better fit the tasks in the study, with consideration for the different network sizes and visualization encodings.

Furthermore, an eye tracker could be incorporated into the study, in order to track the movement of each participant's eyes as he/she completes each task. This would be useful to understand which surrounding nodes the participant may be visually comparing in the network, in order to determine the values of specific nodes while completing quantitative tasks. Additionally, the eye tracker may help to evaluate whether the participant was perhaps visually grouping nearby nodes from a different category, based on the participant's gaze time in a specific location of the network, while completing grouping tasks. A final addition would be to store and analyze all of the interactions that the participant performs with the GUI while completing tasks. This would include actions such as panning, button clicking, and other mouse movements.

Appendix A: Post-Questionnaire and Short Interview Questions

A.1 Post-Questionnaire Questions

The following are the questions that were presented to the participants in the post-questionnaire.

1. (Mental demand) How mentally demanding were these visualization techniques?
2. (Physical demand) How much physical demand (e.g. hand movement or eye strain) was required by these visualization techniques?
3. (Temporal demand) How hurried or rushed were you when completing tasks using these visualization techniques?
4. (Performance) How successful were you in accomplishing the given tasks using these visualization techniques?
5. (Frustration) How discouraged, irritated, stressed and/or annoyed were you when using these visualization techniques?

A.2 Short Interview Questions

The following are the questions that were asked to the participants during the short interview.

1. Which visualization technique did you find the hardest to use? Why? For which tasks?
2. Which visualization technique did you find the easiest to use? Why? For which tasks?
3. Which visualization was more effective for grouping tasks? Why?
4. Which visualization was more effective for quantitative tasks? Why?

5. Using which technique do you think you answered tasks more accurately (i.e. correctly) for grouping tasks? Why?

6. Using which technique do you think you answered tasks more accurately (i.e. correctly) for quantitative tasks? Why?

7. Did the presence of medium graphs as well as large graphs affect the way you answered the tasks? If so, in what way?

8. Did you find yourself running out of time for some tasks? If so, which ones? Why?

9. Did you find yourself having time left over for some tasks? If so, which ones? Why?

10. Do you have any suggestions for improvement to the study (font size, color perception, potential bugs)?

11. Do you have any other comments about the study?

Bibliography

Bertin, J. (2010). *Semiology of Graphics* (W. J. Berg, Trans.). Redlands, CA: Esri Press.

Brewer, C., & Harrower, M. (n.d.). ColorBrewer 2.0. Retrieved April 7, 2017, from http://colorbrewer2.org/

Bostock, M. (2017). Force-Directed Graph. Retrieved April 7, 2017, from https://bl.ocks.org/mbostock/4062045

Bostock, M., Devinsuit, Watson, N., Heer, J., & Velikiy, A. (n.d.). d3-force. Retrieved April 26, 2017, from GitHub website: https://github.com/d3/d3-force/blob/master/README.md

Chung, D. H. S., Archambault, D., Borgo, R., Edwards, D. J., Laramee, R. S., & Chen, M. (2016). How Ordered Is It? On the Perceptual Orderability of Visual Channels. *Eurographics Conference on Visualization*, *35*(3). https://doi.org/10.1111/cgf.12889

d3.schemeSet3 [Illustration]. (n.d.). Retrieved from https://github.com/d3/d3-scale-chromatic

Gabriel-Petit, P. (2007, January 20). Applying Color Theory to Digital Displays. Retrieved April 16, 2017, from UXmatters website: http://www.uxmatters.com/mt/archives/2007/01/applying-color-theory-to-digital-displays.php

Genc, B., & Dogrusoz, U. (2003). A Constrained, Force-Directed Layout Algorithm for Biological Pathways. *International Symposium on Graph Drawing*. https://doi.org/10.1007/978-3-540-24595-7_29

Ghoniem, M., Fekete, J.-D., & Castagliola, P. (2004). A Comparison of the Readability of Graphs Using Node-Link and Matrix-Based Representations. *IEEE Symposium on Information Visualization*. https://doi.org/10.1109/INFOVIS.2004.1

GraphStudy3. (n.d.). Retrieved from https://github.com/wchiou1/GraphStudy3

Harrower, M., & Brewer, C. A. (2003). ColorBrewer.org: An Online Tool for Selecting Colour Schemes for Maps. *The Cartographic Journal*, *40*(1). https://doi.org/10.1179/000870403235002042

Jianu, R., Rusu, A., Hu, Y., & Taggart, D. (2014). How to Display Group Information on Node-Link Diagrams: An Evaluation. *IEEE Transactions on Visualization and Computer Graphics*, *20*(11). https://doi.org/10.1109/TVCG.2014.2315995

Kerpedjiev, P., Hammer, S., & Hofacker, I. L. (2015). Forna (force-directed RNA): Simple and effective online RNA secondary structure diagrams. *Bioinformatics, 31*(20). https://doi.org/10.1093/bioinformatics/btv372

Kobourov, S. G. (2013). Force-Directed Drawing Algorithms. In R. Tamassia (Ed.), *Handbook of Graph Drawing and Visualization* (pp. 383-408). CRC Press.

Morstatter, F., Kumar, S., Liu, H., & Maciejewski, R. (2013). Understanding Twitter Data with TweetXplorer. *ACM SIGKDD Conference on Knowledge Discovery and Data Mining.* https://doi.org/10.1145/2487575.2487703

Nachmias, J. (2011). Shape and size discrimination compared. Vision Research, 51(4), 400-407. https://doi.org/0.1016/j.visres.2010.12.007

"recency effect." (1998). *A Dictionary of Sociology*. Retrieved April 26, 2017 from Encyclopedia.com: http://www.encyclopedia.com/social-sciences/dictionaries-thesauruses-pictures-and-press-releases/recency-effect

Saket, B., Simonetto, P., Kobourov, S., & Borner, K. (2014). Node, Node-Link, and Node-Link-Group Diagrams: An Evaluation. *IEEE Transactions on Visualization and Computer Graphics*, *20*(12). https://doi.org/10.1109/TVCG.2014.2346422

Schreiber, F., Dwyer, T., Marriott, K., & Wybrow, M. (2009). A generic algorithm for layout of biological networks. *BMC Bioinformatics*. https://doi.org/10.1186/1471-2105-10-375

Simeone, M. (2012, December 2). Visualizing Topic Models with Force-Directed Graphs [Blog post]. Retrieved from The Stone and the Shell website: https://tedunderwood.com/2012/12/02/visualizing-topic-models-with-force-directed-graphs/

Zhao, H., Bryant, G. W., Griffin, W., Terrill, J. E., & Chen, J. (2016). Validation of SplitVectors Encoding for Quantitative Visualization of Large-Magnitude-Range Vector Fields. *IEEE Transactions on Visualization and Computer Graphics*. https://doi.org/10.1109/TVCG.2016.2539949