HELPING ANALYSTS REASON ABOUT SOCIAL NETWORK STRUCTURE: EVIDENCE FROM TRAINING AND INTERACTIVE VISUALIZATION

BY

CAROLYN R. RATCLIFFE

THESIS

Submitted in partial fulfillment of the requirements for the degree of Master of Science in Human Factors in the Graduate College of the University of Illinois at Urbana-Champaign, 2011

Urbana, Illinois

Adviser:

Professor Alex Kirlik
ABSTRACT

Communications through phone, email and social networking create increasingly vast webs of data that can be analyzed for research in areas such as marketing, forecasting, and criminal network analyses. Trained analysts can manipulate algorithmically processed data to find conclusions for further investigation, and visualization of this data can enable analysts to draw conclusions more quickly and accurately than from text alone. To examine various methods for manipulating visualized, networked data an experiment was conducted in which participants used an interactive program to determine if hierarchical sub-graph hypotheses conformed to network data. To determine what types of training best prepare analysts to interact with such networked data sets, participants were divided into four groups, each completing one of four training exercises prior to the interactive visualization portion of the experiment. Training sessions required participants to examine or actively create visual, verbal, or visual and verbal hierarchical structures. Groups that were trained with multiple representations of hierarchical structures correctly completed more hypothesis testing tasks and were able to do so more quickly. All groups showed an overall reduction in time needed to correctly complete the hypothesis testing tasks as they proceeded through trials, resulting in learning curves for each group. Various solution methods were identified that were used by high-performing subjects, but differences in method suggest that analysts create individualized mental-models, and there may be no single “best” solution method.
ACKNOWLEDGEMENTS

Thank you to my adviser, Professor Alex Kirlik, for his guidance in navigating graduate school and learning in a research environment, as well as his assistance in creating this thesis. From his direction and encouragement during my application to the program, through the final days of completing my studies, his expertise and patience has been invaluable. I give my sincere appreciation and thanks.

I would also like to thank my fellow lab group members. Thank you to Sven Bertel, whose help with visual displays and suggested readings helped me to form a strong background to understand the visualization of data. Thank you to Wei Dong, whose thoughtful insights in lab discussions helped to illuminate issues that may have never otherwise arisen. Thank you to Jennifer Tsai, who provided assistance with everything from writing IRBs, to sending lists of journal articles; I am grateful for her willingness to share her experience. A thank you is also owed to Scarlett Herring for her readiness to lend an ear for questions, problems, and brainstorms. Our coffee shop sessions enabled me to complete this thesis on schedule; her motivation in all things has been an immeasurable gift.

I would like to thank my parents not only for their support of my graduate studies as I completed this work, but for their dedication to my education from the beginning. Without their guidance, encouragement and sometimes strict principles I would not have accomplished as much as I have today. Their experiences in higher education have motivated me and guided me more than they know.

Finally I would like to thank my fiancé Mark for his unwavering love and patience as I continued my education. I am beyond grateful for the countless drives over many miles and across state lines. I am unbelievably lucky to know and love someone who is so willing to help me through my panics and problems, and is there to share in the celebration of my successes. I look forward to a future of returning the support, patience and love you have provided to me these past years.
# TABLE OF CONTENTS

**CHAPTER 1: INTRODUCTION** .................................................................1  
1.1 Social Network Analysis ..............................................................2  
1.2 Application: Criminal Network Detection and Analysis .....................3  
1.3 Data Visualization ......................................................................4  
1.4 Interactive Visualization .................................................................6  
1.5 Human-Computer Interaction .........................................................7  

**CHAPTER 2: TRAINING AND TRANSFER** ..............................................9  
2.1 Transfer ......................................................................................9  
2.2 Multiple Representations ............................................................10  
2.3 Learner-Generated Drawing ..........................................................12  

**CHAPTER 3: PILOT STUDY** ...............................................................14  

**CHAPTER 4: EXPERIMENTAL HYPOTHESES** ....................................15  

**CHAPTER 5: METHODS** ..................................................................17  
5.1 Preliminary Training: Sessions 1 Through 3 ....................................18  
5.1.1 Training Group A: Verbal-Verbal Comparison ..............................18  
5.1.2 Training Group B: Visual-Visual Comparison ..............................19  
5.1.3 Training Group C: Verbal-Visual Passive Comparison ..................20  
5.1.4 Training Group D: Verbal-Visual Active Comparison ....................21  
5.2 Experiment: Sessions 4 and 5 ........................................................22  
5.2.1 VUE Program Training ............................................................23  
5.2.2 Trials ....................................................................................24  

**CHAPTER 6: RESULTS** ....................................................................27  
6.1 Criteria I: All Nodes and Positions Correct ....................................28  
6.1.1 One-Way ANOVA for Time, by Group and Difficulty ..................30  
6.1.2 Generalized Linear Model for Accuracy, by Group and Difficulty ....31  
6.2 Criteria II: All Nodes Correct, Position Can be Incorrect ..................33  
6.2.1 One-Way ANOVA for Time, by Group and Difficulty ....................36  
6.2.2 Generalized Linear Model for Accuracy, by Group and Difficulty ....37  

**CHAPTER 7: DISCUSSION** ...............................................................39  
7.1 Latency Analysis ........................................................................40  
7.2 Accuracy Analysis .......................................................................41  
7.3 Difficulty Analysis .......................................................................43  
7.4 Method Analysis ..........................................................................44  
7.5 Criteria Analysis ..........................................................................48  
7.6 Future Direction ..........................................................................48  
7.7 Conclusions .................................................................................49  

**REFERENCES** .................................................................................50  

**APPENDIX A** ..................................................................................54  

**APPENDIX B** ..................................................................................56
CHAPTER 1: INTRODUCTION

As technology continues to proliferate throughout society the amount of data being created grows rapidly. One such type of data is that which connects people, creating traceable links and webs, forming virtual networks and communities. These communications through phone, email and social networking create increasingly vast webs of data that can be analyzed for research in areas such as marketing, forecasting, and criminal network analyses (Natarajan, 2006; Krebs, 2002; Lu, Polgar, Luo, & Cao, 2010).

Often these analyses center on the identification of key people or entities that create a sub-group within larger networks. Through various data mining techniques the larger networks can be filtered by algorithms into more manageable networks, and these smaller networks can then be manipulated and interpreted through the use of tools such as interactive visualizations to further allow analysts to make inferences about their structure (Han, Lakshmanan, & Ng, 1999). In many cases, a hypothesis about this structure has been provided, and the analyst's task is to determine whether or not the hypothesis conforms to the network structure.

The aim of this thesis is to investigate how analysts provided with these types of hypothesis testing tasks should be trained before they use interactive visualizations, how various training methods can affect solution time and accuracy, and how training may impact the number or variety of solution methods attempted. This experiment falls at the crossroads of many fields of study, so a brief overview of each topic is provided to aid in understanding the methods and techniques utilized in this thesis. The remainder of Chapter 1 is a review of social network analysis, criminal network investigation applications, data visualization, interactive visualization, and human-computer interaction. Chapter 2 provides a literature review and background on training and transfer, particularly regarding representations of data. Chapter 3 details the pilot study, while Chapter 4 details the hypotheses for the thesis experiment. Chapter 5 outlines the method of the experiment that was conducted for this thesis. Chapter 6 and 7 provide results, conclusions and major findings as well as future directions for this area of research.
1.1 Social Network Analysis

Social network analysis is a way of investigating patterns of relationships and interactions between groups of people to analyze underlying structures. In social network analysis (SNA) networks are typically displayed as graphs called sociograms, with nodes or points representing network members connected by relationships represented as links. Nodes can be of various shapes, sizes and colors to indicate varying levels of characteristics such as age, connectivity, title, rank, etc. Links, also called edges, can be of various colors, thicknesses and styles indicating variables such as the strength or recency of the connection. Links can also be directional or non-directional.

When looking at graphs of social networks, there are several terms that specifically describe relationships within the network (Freeman, 1978). If two nodes are directly connected, they are adjacent; a set of two nodes and a link between them is called a dyad; if two nodes can be connected by a path through other nodes they are reachable from each other; when every node can be reached from any other node the graph is connected; the distance of a path between two nodes is a sum of the lengths of the links on that path; a geodesic of two nodes is the shortest paths between those two nodes.

Most studies of social network analysis use measures of betweenness centrality, and closeness centrality, clustering data into groups with key or “gateway” players. Betweenness centrality measures the degree to which an actor is situated between two other actors; if an actor often occurs on the shortest path between two other actors then that actor has a high level of betweenness. These nodes with high levels of betweenness may be “gatekeepers” for a network, serving as communication bridges.

Closeness centrality is the measure of how far an actor is from all other actors in the network. The closer an actor is to other actors, the more quickly they can interact with all others. This makes them a source of information in networks and means that they do not have to rely on information from others (Wasserman & Faust, 1994). These members serve as communication sources.

There is another way to look at centrality: point centrality vs. graph centrality (Freeman, 1978). Point centrality occurs when a node is the center of a wheel or star and is considered to be structurally more central than other nodes in that star or wheel, adjacent to the highest number of nodes. Graph
centrality refers to the compactness of the graph, where higher graph centrality corresponds to shorter distances between pairs of points. While other definitions exist for these terms (see Freeman, 1978, for an in-depth review), the ones provided here are commonly accepted for social network analysis.

All of these measures help to define networks which allows for clustering of nodes based on certain measures to identify sub-graphs. However, the question exists: can these sub-networks be found without clustering? Even without grouping based on quantity or type of connection, can analysts identify key subjects based on provided hypotheses about sub-graph structures? In this experiment all relationships are non-directional with no measure of the amount or type of connectivity, only retaining dichotomous relations, i.e. the presence or absence of a connection. The goal is to see if this type of hypothesis testing task is still solvable.

1.2 Application: Criminal Network Detection and Analysis

This experiment was initially driven by interest in the 2009 IEEE VAST challenge, a visual analytics competition to explore visualization tools and techniques to analyze visualized data (IEEE, 2009). Challenge 2 from 2009 details a scenario where an embassy employee is leaking information to a criminal network using a fictional micro-blogging tool, Flitter. The challenge provides a table of user-to-user, two-way connections and two sub-networks with various fuzzy constraints as potential social structures of the criminal network. The goal is to identify which social structure more closely matches the data, and the names of those in the criminal network.

Though the VAST challenge that originally created this experiment was fictional in nature, the use of network analysis to examine criminal behavior has long been a part of law-enforcement and intelligence organizations. Warr (2002, as cited in McGloin & Kirk, 2010) states that crimes are social behavior, and that the strongest single predictor of criminal behavior is the number of the individual’s delinquent friends. Beginning with manual link analysis to examine connections between people and events (Harper & Harris, 1975), investigators have continued to utilize advancing technology, updating to systems such as COPLINK that automatically scan for links between people, vehicles, locations and
organizations (Hauck, Atabakhsh, Ongvasith, Gupta, & Chen, 2002). Recently, computer programs such as COPLINK have been used as the background program for other projects that aim to visualize this data for more accurate, timely and intuitive analyses (Chen, Zeng, Atabakhsh, Wyzga, & Schoeder, 2003; Xu & Chen, 2003; Chen, Chung, Xu, & Wang, 2004).

These advances in crime database analysis have been applied to a wide array of current and past cases including using wiretap data to investigate heroin distribution networks (Natarajan, 2006), analyzing relationship strength to investigate the terrorist networks involved in the hijackings of September 11th, 2001 (Krebs, 2002), and observing activity to investigate a criminal hacker community (Lu, Polgar, Luo, & Cao, 2010), to name just a few.

1.3 Data Visualization

Data visualization is the display of information for communication and analysis. Abstract data and large tables can be presented through visualizations that allow for quick and accurate understanding and drawing of conclusions. Raw data can be placed into data tables which can then be transformed into visualized data structures using programs. Often, data visualization illustrates relationships between quantitative data, such as in typical line or bar graphs, though other more complicated visualizations can display networks, hierarchies, trees, etc.

While visualizations can assist in data interpretation, they can also hinder it if the visualizations are not done properly. There are certain ways to display data that our eyes can easily see and our brains can comprehend, which have been discovered through the study of human perception. In order for a graph to effectively convey information it must: clearly indicate relationships between values, represent quantities accurately, allow for easy comparison between quantities, allow for easy observation of the rank of values, and make clear how the information can and should be used (Few, 2010).

To be able to accomplish these goals an understanding of human perception is required. Kosslyn’s *Graph Design for the Eye and Mind* (2006) details eight psychological principles for constructing effective visualizations of data, divided into three sets for (1) connecting with the reader of
the graph, (2) directing the reader’s attention through the display, and (3) promoting understanding and memory.

Included in the first set, connecting with the reader, is the principle of relevance, that graphs and displays are most effective when only relevant information is included, and all relevant information is included. The second in this set is the principle of appropriate knowledge, that the creator of a graphic must understand the audience’s level of knowledge and familiarity with material.

The second set, directing and holding attention, includes three principles. The first is the principle of salience, that the most important information should be the most salient and draw the most attention. The second is the principle of discriminability, that two properties must be different enough for the reader to be able to tell they are different, with the example provided that \( m \) and \( mn \) are much more difficult to distinguish than \( m \) and \( o \). This principle also applies to color, hue, size, shape, etc. The final principle in this set is the principle of perceptual organization, that people automatically group items into units. This grouping effect was clearly defined by the Gestalt school of psychology as principles that suggest that the whole is greater than the sum of its parts. These principles specify that we group items based on characteristics such as proximity, similarity, enclosure, continuity, and connection (Figure 1).

![Figure 1: Examples of Gestalt principles of human perception that should be considered when creating effective visualizations. These principles reflect how humans group items into sets or objects.](image)

The graph, (2) directing the reader’s attention through the display, and (3) promoting understanding and memory.

Included in the first set, connecting with the reader, is the principle of relevance, that graphs and displays are most effective when only relevant information is included, and all relevant information is included. The second in this set is the principle of appropriate knowledge, that the creator of a graphic must understand the audience’s level of knowledge and familiarity with material.

The second set, directing and holding attention, includes three principles. The first is the principle of salience, that the most important information should be the most salient and draw the most attention. The second is the principle of discriminability, that two properties must be different enough for the reader to be able to tell they are different, with the example provided that \( m \) and \( mn \) are much more difficult to distinguish than \( m \) and \( o \). This principle also applies to color, hue, size, shape, etc. The final principle in this set is the principle of perceptual organization, that people automatically group items into units. This grouping effect was clearly defined by the Gestalt school of psychology as principles that suggest that the whole is greater than the sum of its parts. These principles specify that we group items based on characteristics such as proximity, similarity, enclosure, continuity, and connection (Figure 1).
Kosslyn’s final set of principles aims to promote understanding and memory. The first, the *principle of compatibility*, states that the form of the display should match the meaning, such as the idea of “more is more”, referring to the idea that larger bars, shapes and points should always designate higher quantities. The second is the *principle of informative changes*, that if something changes or is different, then the reader will assume that this is significant and is intended to convey information. Lastly is the *principle of capacity limitations*, stating that people have a limited information capacity for retention and processing, so only a certain amount of data should be displayed at one time.

1.4 Interactive Visualization

Visual analytics research and applications have increasingly focused on utilizing these principles of good design and expanding them to interactive displays and visualizations. These interactions can occur in one-, two-, or three-dimensions with work occurring to add a fourth dimension such as sound (Smith, Bergeron, & Grinstein, 1990; Hermann & Hunt, 2005). Studies have shown that interactive visualizations allow for greater understanding during navigation and analysis of data than visualizations alone (Noppens & Liebig, 2008). Interactive visualization tools have also been studied, such as for Bayesian reasoning, a traditionally difficult and abstract problem-set, also with successful results (Tsai, Miller, & Kirlik, 2011).

Buja, Cook, & Swayne (1996) suggest a taxonomy for interactive data visualization to classify approaches and aspects of visualization techniques. The foundation of their taxonomy is divided into rendering, what to show in the visualization, and manipulation, what to do with the visualizations. Manipulation is then divided into three tasks for data exploration: *Finding Gestalt*, identifying local or global groups based on Gestalt principles; *posing queries*, determining what the discovered Gestalt features imply; and *making comparisons*, the creation of plots that are organized in a way to draw meaningful conclusions.

These tasks will represent themselves in this experiment as subjects attempt to find which provided hypothesis conforms to the network structure, essentially looking for groupings within the
network based on color and organization. They will then pose a query to see if a particular actor meets the criteria necessary, and they will then organize any actors that meet the criteria into a solution sub-graph to enter into the hypothesis template.

1.5 Human-Computer Interaction

Human-computer interaction is an experimental methodology that involves the interaction of a human with a computer, allowing collaboration that capitalizes on the strengths of each. Han, Lakshmanan, and Ng (1999) promote constraint-based data mining in which the user guides the computer through a search by providing constraints. These constraints are divided into five categories: knowledge constraints, data constraints, dimension/level constraints, rule constraints, and interestingness constraints. By allowing the manipulation of these constraints the user is able to precisely specify what they would like to analyze.

An example of an application where human-computer collaboration is effective is the identification of sub-sets of data within a network. In very large sets of data, algorithms are used to find solutions that match desired inputs, and most often they are very good at this solution query. There are times, however, when values are unknown or fall in an uncertain range, or when exact solutions are not present in the data at all. In these cases the algorithm can filter results to a few best solutions, but without exact inputs the algorithm cannot give an exact output.

In our example of a criminal ring we may be looking for someone that has about 100 cell phone contacts, based on previous knowledge of other criminal rings, or an existing hypothesis regarding sub-graph structure. But what does ‘about’ mean? Is there a margin of 5 contacts, for a range of 95 to 105? Or does this mean there can be a range of 20 contacts, for a range of 80 to 120 contacts? Prior to running analyses is it difficult to know where to set such boundaries. What if there are no actors in the data that have anything close to 100 contacts, but we know that someone in the phone tree is the main target? Here algorithms have problems with uncertain or ‘fuzzy’ values, and in some cases returning zero results if the right numbers and ranges are not properly entered. If there are even only a handful of variables (and often
there are many more), it can take many iterations and a great amount of time to pinpoint where numbers are incorrect.

It is in these types of ambiguous situations where a human analyst can enter the picture and use information that the computer may be unaware of. This knowledge could be, for example, new information learned about the ring, resulting in a simple change in the point of view of the analyst but could mean a rewrite of the algorithm. The analyst might instead be able to use prior knowledge that cannot be entered into the algorithm to weight information.

Additionally, humans have been shown to be very good with ambiguity and filling in missing information, and can complete tasks that computers cannot such as online word verification security measures (von Ahn, Blum, & Langford, 2004) or matching numerical values on checks to the amount written out in longhand (Thompson, 2007).

Beyond cases of ambiguity, humans have also proven their abilities to outperform computers in visualization tasks. Imagine you are sitting in front of a screen that contains 10,000 grey circles, and one red circle. You would be able to locate the red circle nearly immediately; in fact, it would be difficult to not notice the red circle. If a computer were assigned the task to find which circle was different it would need to cycle through circles until it found the red one, resulting in an average search through 5,000 circles (for every case where it found the circle first there would be a case where it found the circle last). It is this tolerance for ambiguity and high visual capacity that enables analysts to look at large amounts of data and draw conclusions quickly in situations where a computer cannot.
CHAPTER 2: TRAINING AND TRANSFER

It is of interest in this experiment how different training techniques affect solution time and accuracy in later trials. In order for the previous training to make a difference, however, what is learned in the training task must be carried over to the experiment task. This is referred to as transfer of training (Wickens & Hollands, 2000).

2.1 Transfer

Questions often arise during training research about how representations affect transfer between tasks, or how learning or knowledge in one task can affect performance and understanding in a subsequent task. It has been determined through years of research that transfer between tasks is unlikely to occur naturally between tasks, but Sternberg and Frensch (1993) provide four mechanisms that can affect the degree of transfer. These are: encoding specificity, the retrieval of information depends on how it was encoded; organization, the retrieval of information depends on how it is organized in memory; discrimination, the retrieval of information depends on whether information is considered or described as relevant; and set, the application of information or methods depends whether the previous method is seen as fitting for the task.

In our experiment we addressed the third mechanism, discrimination, by specifically referring to all representations, whether verbal or visual, as hierarchies. This naming was used both in the training tasks and in the experiment, with the included statement “Today you will be completing a task just like the one you did before, but on a higher level”. Additionally, examples of a paragraph of text and a visual hierarchy were provided, representative of the training task, along with an equivalent representation in a new form such as would be used for the experiment task.

The fourth mechanism, set, has been famously proven by Gick and Holyoak’s version (1980, 1983) of the Duncker (1945) radiation problem. In their version of the problem the participant is provided with a problem about attacking a fortress with many roads radiating from it, all laden with landmines. While smaller groups of men could navigate the roads, larger groups would detonate the mines. The
participant is asked how sufficient force could be used against the fortress, with the solution requiring several smaller groups of men approaching from several sides.

The subject is then given the Duncker radiation problem about treatment methods for a tumor. Any X-ray powerful enough to destroy the tumor would also damage the healthy tissue between the emission point and the tumor, and the patient would die. The participant is asked what type of procedure should be used to destroy the tumor but not harm healthy tissue around it, with the solution requiring several weaker rays to converge at the tumor, originating from several points.

It was found that working on the fortress problem helped participants solve the radiation problem only when the participants were primed to use the former problem to help them solve the latter. If, however, the subjects are told that the problems are a “memory experiment” there is almost no transfer between the two tasks. This shows the importance of mental set on transfer between tasks, and illustrates another reason why such care was taken in our experiment to emphasize the relation between the training task and the experiment task to participants.

Similarly, Brown and Palincsar (1989) specify that transfer of knowledge can occur when learners are shown problems that resemble each other, and when learners had their attention directed towards the goal structure of similar tasks. Greeno, Smith, and Moore (1993) believe, according to their “situated cognition theory”, that transfer between tasks requires a transformation process, occurring when the learner perceives the subsequent task to retain similar constraints and affordances as the first task.

Regarding the effect of representation on transfer, it is our belief that all subjects will have transfer due to our efforts of specifying the relation of the two tasks, but there is an expectation that those participants who interacted with visual representations will have higher transfer, and therefore better performance, on the visual experiment task.

2.2 Multiple Representations

Representations, symbolized versions of the external world, can occur in the mind as mental representations, or materially expressed, such as on paper, as expressed representations. In our
experiment we provided the participants in group C with multiple expressed representations to compare at a structural level. These were, specifically, a paragraph of text describing a hierarchical structure and a graphical representation of a hierarchy similar to a business’ organizational chart.

Multiple representations of data have been shown to positively affect text comprehension conceptual recall (Mayer, 1989b; Mayer, 1993; Mayer & Gallini, 1990), as well as deeper comprehension (Butcher, 2006; Mayer & Gallini, 1990).

Multiple representations can theoretically contain the same information, but different representations will innately offer different advantages (Ainsworth, 2006). These differences can combine in a complementary way, with each representation varying on which processes are supported, their ability to support computational offloading (reductions in cognitive effort), and to graphically constrain inferences (Scaife and Rogers, 1996; de Jong et al., 1998). Stenning and Oberlander (1995) maintain that diagrams are more effective for use in determinate tasks because they lack the ambiguity that text alone can permit and will aid “processibility”.

Studies show the importance of representations in learning mathematics and science (Ainsworth, 2006; Mayer, 1989a), and Mayer, Bove, Bryman, Mars, & Tapangco (1996) showed that students provided with textbook chapter summaries with simple illustrations recalled information better and performed better on transfer tasks than those student who received just the summary, the whole chapter, or both the chapter and the summary. Oberlander, Cox, Monaghan, Stenning, and Tobin (1996) found that those individuals that were classified as diagrammatic reasoners, such as engineers or network analysts, were able to translate information between multiple representations more successfully than those who were not classified as such.

Mayer’s Generative Theory of Textbook Design is a model for learning from text with illustrations, or multiple representations of similar information (Mayer, Steinhoff, Bower, & Mars, 1995). Mayer’s model theorizes that key elements from adjacent text and images are gathered and organized into verbal and nonverbal mental representations, which then are integrated into a single mental model (Mayer, 1993; Mayer, Bove, Bryman, Mars, & Tapangco, 1996). Work by Schnitz & Bannert (2003)
supports this theory on mental model creation, with additional suggestions that only task-appropriate graphics will aid mental model construction and that task-inappropriate graphics with interfere with model construction.

Multiple representations encourage diverse strategy methods, benefitting students by allowing them to switch between representations, compensating for strategy weaknesses (Tabachneck, Koedinger, & Nathan, 1994). It is predicted that subjects in group C, those comparing paragraphs of text with graphical representations, will attempt a higher number strategy methods per session than those who looked at only text or graphical representations.

2.3 Learner-Generated Drawing

In the fourth training condition subjects were asked to create visual representations of hierarchies from verbal representations in the form of paragraphs. Learner-generated drawing is a learning technique where participants create representative drawings from text in order to illustrate content. Learner-generated drawings hold the requirement that participants believe the final drawing is a correct representation, in effect declaring the passage and the drawing as having the same structure once they have finished their drawing. Additionally, there is an expectation that creating external drawings can only occur once a mental model is constructed internally.

Several studies have been conducted to examine the effect of self-generated concepts or representations on knowledge transfer (Terwel, Van Oers, Van Dijk, and Van den Eeden, 2009; Ainsworth, 2006; Rosenshine, Meister, & Chapman, 1996; diSessa, 2004; Van Dijk, Van Oers, & Terwel 2003). Terwel, et. al (2009) found that designing a representation helps the designer to reflect on core elements and their mutual relations and allows them to use the representations for their original purpose or transformed for new situations.

Just as Mayer’s Generative Theory on Textbook Design applies to provided multiple representations, it applies also to drawing an additional representation from text. Again, key elements are gathered from the text and organized into a mental representation (Mayer, Steinhoff, Bower, & Mars,
1995; Van Meter & Garner, 2005) as the participant activates old, and creates new, associative links between elements, again creating a mental model. This model enables the participant to apply knowledge to a greater range of applications (Kintsch, 1994). As the participant reads through a passage they continually gather more information, updating their mental model (Van Meter, et al., 2006). This recursive process serves as self-monitoring of solutions and increases their ability to detect errors (Van Meter, 2001).

As previous studies have found (Van Meter et al., 2006; Hall, Bailey, & Tillman, 1997), we anticipate that participants who generate drawings from text will construct a better mental model and perform better on subsequent higher-level tasks than those who only inspect text or diagrams.
CHAPTER 3: PILOT STUDY

A pilot study was conducted with five subjects as proof of concept, verifying that it was possible to test hypotheses about social network structure by matching provided sub-graph hypotheses to randomized larger webs within a reasonable amount of time. In the pilot study there were no training conditions. The subjects were introduced to the problem from a social networking perspective and informed that the webs represented a messaging system for societies at a university. They were instructed to match one of two provided sub-graphs with webs of 25, 50, and 100 nodes. Here the sub-graphs remained constant for all trials, with either A or B being present in each web.

Results showed that finding solutions in the randomized webs was not only possible, but that subjects were able to do so rather quickly. Solution times for a correctly completed trial of the 25 node set ranged from 75 seconds to 378 seconds with an average time for a correct solution of 237 seconds. Solution times for the 50 node set ranged from 368 seconds to 900 seconds with an average time for a correct solution of 636 seconds. Solution times for the 100 node set ranged from 372 seconds to 890 seconds with an average time for a correct solution of 581 seconds.

The pilot study helped to shape the full experiment in several ways. It was noted in the pilot study that each subject had a different method for finding the correct hypothesis in the web. This led to the question of “How do solution methods affect solution time and accuracy?” Additionally, the question was posed “How might an interactive tool help subjects gain an understanding of the task” which then became “How might various training methods, including passive and active training tasks, affect solution methods, times and accuracy?” Finally, questions were posed about how the difficulty of the sub-graph solutions might affect solution methods, time, and accuracy.
CHAPTER 4: EXPERIMENTAL HYPOTHESES

Hypothesis 1: Subjects presented with only visual representations in training will be the lowest performers with the slowest times and the fewest trials solved correctly. This is due to the expectation that, in comparing two hierarchical visual representations, the subject will only need to count the number of levels and the numbers of key player at each level, thereby not creating a mental representation or mental model of social network hierarchies.

Hypothesis 2: Subjects presented with only text based representations in training will be the next lowest performing group on measures of time taken for trials and the number of trials completed correctly. This is due to the expectation that the subject will make mental representations and mental models of the presented paragraphs describing hierarchies, but that without external visual representations the subject will not transfer knowledge as well, if at all, to a visual task.

Hypothesis 3: Subjects presented with a verbal representation and a visual representation in training will perform better than those subjects who are presented with a single type of representation on measures of trial time and the number of trials completed correctly. This is due to the expectation that the subject will create stronger mental models for hierarchy comparison, mentally transforming one type of representation into the other.

Hypothesis 4: Subjects asked to create visual representations of verbal descriptions of hierarchies in training will be the best performing group with the fastest times and most trials completed correctly. This is due to the expectation that they will need to form mental models to transform one representation into the other, just as is expected of Group C, and additionally they will reinforce this mental model through actively externalizing the visual representation.

Hypothesis 5: The difficulty of trials will affect time and accuracy. The trials categorized as most difficult will take longest to complete and will be completed correctly the least often. Trials categorized as having “medium” difficulty will take less time than those categorized as “hard” and will be completed correctly more often. The least difficult trials will take the least time to complete and will be completed correctly the most often.
Hypothesis 6: As subjects proceed through trials they will develop solution methods and will improve on time measures, even as difficulty of sub-graph level may change. This trend will become apparent as a learning curve where time generally decreases as the subjects proceed through trials.
CHAPTER 5: METHODS

Thirty-two students from the University of Illinois participated in this experiment. Each received $10 per hour for their participation. Each participant completed five one-hour sessions, for a total of approximately five hours and compensation of approximately $50. An additional prize of $50 was awarded to the top performer of each of four groups based on time and accuracy performance.

Participants were recruited using flyers placed in areas to target a specific demographic of engineers, and through a posting on the Psychology Department’s website. Due to the structural and spatial nature of the experiment participants were recruited from the college of engineering. Due to the complexity and fine granularity of verbal structure comparisons participants were required to be native English speakers.

The thirty-two participants were divided into four groups of eight participants, with each group participating in one of four preliminary training conditions.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Session</td>
<td>Introduction, Training 1 hour</td>
<td>1 hour</td>
<td>1 hour</td>
<td>1 hour</td>
</tr>
<tr>
<td></td>
<td>Training 1 hour</td>
<td>1 hour</td>
<td>1 hour</td>
<td>1 hour</td>
</tr>
<tr>
<td></td>
<td>Training 1 hour</td>
<td>1 hour</td>
<td>1 hour</td>
<td>1 hour</td>
</tr>
<tr>
<td></td>
<td>Experiment 1 hour</td>
<td>1 hour</td>
<td>1 hour</td>
<td>1 hour</td>
</tr>
<tr>
<td></td>
<td>Experiment 1 hour</td>
<td>1 hour</td>
<td>1 hour</td>
<td>1 hour</td>
</tr>
</tbody>
</table>

Table 1: Participation in each of five sessions, by group.
The preliminary training differed for each group. The following sections will provide details on the different sessions and groups. The final two sessions were the same for all participants. These last two sessions of the experiment were conducted on a Bravia® LCD 720p 60Hz HDTV with a 32” screen connected to a U450p 1.30GHz Lenovo Laptop with 4 GB of RAM.

5.1 Preliminary Training: Sessions 1 Through 3

After reading and signing a consent document, all thirty-two participants completed one of four preliminary training exercises during the first three sessions. The preliminary experiment exercises had participants examine either pairs of verbal hierarchical structures in the form of paragraphs, pairs of visual hierarchical structures in the form of charts composed of nodes and links, or pairs of one verbal structure and one visual structure. A fourth group created visual representations from verbal hierarchical structures. To maintain uniform complexity each pair of structures was identical with only the form of presentation varying, either verbally or visually. The group that created visual representations was provided with the first verbal structure from each pair.

5.1.1 Training Group A: Verbal-Verbal Comparison

The participants in this group were provided pairs of verbal descriptions of hierarchical structures in the form of paragraphs. Subjects were asked to compare the two structures based on the number of hierarchical levels, the number of players at each level, and the connections between players. They were told to ignore differences of gender, age, name, and title between the two structures unless it affected level. The two passages below show examples of two structures that were provided to participants.
1A. Happy Hamsters is a small business in Tolono, IL that makes hamster wheels. This business includes the CEO and three managers. Two of the managers have one engineer reporting to them. One of these engineers is then in charge of an intern. The other manager has two engineers reporting to them and no interns.

1B. Cutie Pies is a bakery in Tolono, IL. This business consists of the owner of the bakery and three head bakers who each direct one of three decorators. One of these decorators is currently mentoring a student from the local culinary school who is on her working rotation.

This example shows a pair of structures that are not identical. The first scenario, Happy Hamsters, has four individuals at the third level of the hierarchy (engineers) while the second scenario, Cutie Pies, has only three individuals at the third level of the hierarchy (decorators). Twenty pairs of structures were provided to each participant in each session for a total of sixty pairs throughout three sessions.

5.1.2 Training Group B: Visual-Visual Comparison

The participants in this group were provided two visual representations of hierarchical structures in the form of charts composed of nodes and links. Subjects were asked to compare them in the same way as the first group, on the number of hierarchical levels, the number of players at each level, and the connections between players. Again they were told to ignore differences of gender, age, name, and title between the two structures unless it affected level. Figures 2 and 3 show two visual representations of structures, 1A and 1B, that were provided to participants for comparison. These are the visual representations of the 1A and 1B verbal examples above.
Twenty pairs of structures were provided to each participant in each session for a total of sixty pairs throughout three sessions.

5.1.3 Training Group C: Verbal-Visual Passive Comparison

The participants in this group were provided a verbal hierarchical structure and a visual representation of a hierarchical structure and were asked to compare them in the same way as the previous two groups, on the number of hierarchical levels, the number of players at each level, and the connections
between players. They were told to ignore differences of gender, age, name, and title between the two structures unless it affected level. The passage below and Figure 4 show the first set of representations that were provided to participants, verbal structure 1A and visual representation 1B.

1A. Happy Hamsters is a small business in Tolono, IL that makes hamster wheels. This business includes the CEO and three managers. Two of the managers have one engineer reporting to them. One of these engineers is then in charge of an intern. The other manager has two engineers reporting to them and no interns.

Figure 4: Visual representation of hierarchy 2B provided to participants in the third group, to compare with verbal representation 1A.

Twenty pairs of structures were provided to each participant in each session for a total of sixty pairs throughout three sessions.

5.1.4 Training Group D: Verbal-Visual Active Comparison

The participants in this group were provided with verbal hierarchical structures and were asked to draw visual representations of hierarchical structures from them. The passage below shows verbal hierarchical structure 1A that was provided to the participants, and figure 5 shows an example of a drawing that was completed to match this description, considered 1B.
1A. Happy Hamsters is a small business in Tolono, IL that makes hamster wheels. This business includes the CEO and three managers. Two of the managers have one engineer reporting to them. One of these engineers is then in charge of an intern. The other manager has two engineers reporting to them and no interns.

Figure 5: A correct visual representation of a hierarchy, created by a participant in the fourth group from verbal representation 1A.

By completing a drawing and moving on to the next verbal representation, or by completing the session, the subject was considered to be expressing judgment that the visual representation they created was a match to the verbal description. Twenty verbal structures were provided to each participant in each session for a total of sixty pairs throughout three sessions.

5.2 Experiment: Sessions 4 and 5

For the remaining two sessions a computer program was used that allows the user to create visual representations of networks and allows manipulation in several ways. The program is called Visual Understanding Environment (VUE) and was created at Tufts University. The program can be downloaded at http://vue.tufts.edu/. This program graphically displays data sets reflecting a web of people or objects and the connections between them, with people or objects represented as nodes of various shapes, sizes and colors, and with the connections represented as lines connecting two nodes. For all trials the nodes were circular in shape.
5.2.1 VUE Program Training

To begin this session each participant was given an introduction to the program, including practice using the features of the program that might help with the experiment task such as zooming, dragging and arranging nodes, highlighting nodes that are linked to one another, and changing the size of nodes. Program introduction was identical for all subjects across all four groups. The full program introduction can be found in appendix A. The introduction additionally explains the goal of the task, specifically to find one of two provided hypothesis sub-structures in a larger web of circular nodes and links. For example, a subject might be provided with hypothesis sub-structures A and B representing hierarchies.

![Figure 6](image.png)

Figure 6: Example hypothesis sub-structures provided to subjects which are compared with the network to see which hypothesis conforms to the network structure.

They would then attempt to find one of these exact structures in a larger web. Figure 7 shows an example of a web where structure A is present, shown by the darker, thick, orange links.
Figure 7: An image from the program training illustrating how one of two provided hypothesis structures, A or B, might be found or embedded in a larger web of nodes. Structure A is highlighted in the web with darker, thick, orange lines.

5.2.2 Trials

After the participants completed the program training they were asked to then complete up to ten trials for the first session, and then up to ten trials for the second session, as many as they could complete within the hour. For each session the subject was provided with a ten-page packet; each page of the packet shows two hierarchical hypothesis structures, A and B. Figure 8 shows the provided structures for Trial 1.

Figure 8: The provided hierarchical structures for Trial 1, one and only one of which was present in the larger web.
All twenty sets can be found in Appendix B. Using the features from the program training and practice, participants attempted to determine which of two provided hypothesis structures conformed to the structure within each larger web of 50 nodes, such as the one shown in figure 9.

Figure 9: Trial 1, before any manipulation, consisting of 50 numbered nodes. Each trial contains one and only one of the two provided hierarchical hypothesis structures, A or B.

After the participant believed they had found which hypothesis conformed to the network structure they wrote the corresponding numbers in their packets. Time and accuracy were recorded for each trial. Screen capture software was used for later examination of strategies and solution methods. Notes were also taken regarding solution methods used by the participants.

At the end of the fifth session participants were provided with a debriefing summarizing the research and what was hoped to be gained from the study.

Sub-graph level difficulty was determined for each trial in order to examine how different groups might respond differently to trials of different difficulty. The difficulty of each trial was based the three factors: the number of nodes in the correct hypotheses, the number of links in the correct hypothesis, and
the degree of similarity between hypothesis pairs. The degree of similarity was measured on the whether the hypotheses had similar number of nodes, similar number of links, and if there was a lack of a distinguishable feature that would be easy to search for.

Using this system the trials were divided into easy, medium, and hard difficulty categories. There were seven hard trials, seven medium trials, and six easy trials.
CHAPTER 6: RESULTS

Two different criteria were used for accuracy evaluations. The first criteria required the participant to correctly identify all nodes and correctly fit them into the provided sub-graphs. The second criteria required the participant to correctly identify all nodes but allowed them to place them incorrectly in the sub-graph. In some sub-graphs a switch of two nodes would result in a different solution as links may be different for different nodes of the same level, as illustrated in figure 10.

Figure 10: Sub-graph I shows an example of a completely correct result while sub-graph II shows an example where all nodes are correctly identified but incorrectly inserted into the sub-graph. The first evaluation criteria only considers sub-graph I correct, while the second evaluation criteria considers both sub-graphs I and II correct.

This type of error, switching two nodes of the same level, is being investigated due to its relevance to applied research. In social network analysis it is most often the case that the person needs to be identified first, and their role in the network can be determined thereafter. Other errors, those where the subject did not find the correct nodes, were not investigated due to the fact that the correct person must first be identified in an investigation. This point is elaborated on in the discussion.
6.1 Criteria I: All Nodes and Positions Correct

All four training conditions show a general learning curve when time taken for a correctly completed trial is plotted against trials. There are two curves for each group, one for trials completed on day one and one for trials completed on day two. On day one each subject had the opportunity to complete trials 1 through 10. On day two each subject had the opportunity to complete trials 11 through 20. A trial without an average value indicates that no subjects from that group made it to that trial, or if they did, they did not complete the hypothesis testing task correctly. Figures 11 through 14 show the average times for correct answers for each trial, by group, with standard error. Figure 15 shows average times for each group and the overall average for all subjects. Plots for each subject, by group, can be found in Appendix C.

![Average Time per Trial, Group A, Criteria I](image)

Figure 11: Trial time learning curves for group A over two days of trials for criteria I. This group received only verbal representations in training.
Figure 12: Trial time learning curves for group B over two days of trials for criteria I. This group received only visual representations in training.

Figure 13: Trial time learning curves for group C over two days of trials for criteria I. This group was provided with verbal and visual representations in training.
Figure 14: Trial time learning curves for group D over two days of trials for criteria I. This group was provided with verbal representations and created visual representations from them.

Figure 15: Trial time learning curves for all groups as well as overall average times for all subjects, over two days of trials, for criteria I.

6.1.1 One-Way ANOVA for Time, by Group and Difficulty

A one-way between subjects ANOVA was conducted to compare the effect of training on time to correctly complete a hypothesis testing task in four training conditions. There was a significant effect of
training on time to complete a correct trial at the p<.05 level for the four conditions \( F(3, 255) = 5.59, p = .001 \). A one-way ANOVA was also conducted to compare the effect of difficulty of task on time to correctly complete the hypothesis testing task in easy, medium, and hard conditions. There was no significant effect of difficulty of task on time to complete a correct trial at the p<.05 level for the three conditions \( F(2, 255) = 1.91, p = .15 \).

A Tukey HSD post-hoc pair-wise comparison test was completed to investigate the significance found for training effects on time to complete the hypothesis testing test correctly. This test was conducted to compare each training condition, A, B, C and D, to every other condition. Group A received only verbal representations as training, group B received only visual representations as training, group C received both verbal and visual representations as training, and group D received verbal representations and were required to create visual representations.

The test found that the mean time for a correct task in group A \( (M = 544.51, SD = 344.35) \) was not significantly different than for group B \( (M = 582.83, SD = 370.48) \). It was also found that there was a significant difference between groups A \( (M = 544.51, SD = 344.35) \) and C \( (M = 397.63, SD = 283.96) \). There was no significant difference between groups A \( (M = 544.51, SD = 344.35) \) and D \( (M = 403.98, SD = 293.51) \), but results are close to significance.

The test also found a significant difference in the mean time to correctly complete the hypothesis testing task between groups B \( (M = 582.83, SD = 370.48) \) and C \( (M = 397.63, SD = 283.96) \), as well as between groups B \( (M = 582.83, SD = 370.48) \) and D \( (M = 403.98, SD = 293.51) \). It was determined there was no significant difference between groups A \( (M = 544.51, SD = 344.35) \) and B \( (M = 582.83, SD = 370.48) \), or between groups C \( (M = 397.63, SD = 283.96) \) and D \( (M = 403.98, SD = 293.51) \).

6.1.2 Generalized Linear Model for Accuracy, by Group and Difficulty

A Chi-square logistic regression was completed to compare the effect of training on the number of trials completed correctly in four training conditions. The test determined that there was a significant effect for group on the number of trials correctly completed \( (\chi^2(3, 636) = 36.63, p < .001) \). A test was then
completed to compare the effect of difficulty on the number of trials completed correctly for easy, medium and hard conditions. The test found that there was a significant effect for difficulty on the number of trials correctly completed ($\chi^2(2, 634) = 7.17, p = .03$).

A Tukey HSD post-hoc pair-wise comparison test was conducted to examine the significance found for the effect of training on the number of trials completed correctly. This test was conducted to compare each training condition, A, B, C and D, to every other condition. Group A received only verbal representations as training, group B received only visual representations as training, group C received both verbal and visual representations as training, and group D received verbal representations and were required to create visual representations.

This test showed a significant difference in the number of correct trials completed between groups A ($M = 6.00, SD = 3.59$) and C ($M = 11.63, SD = 4.41$) and between groups A ($M = 6.00, SD = 3.59$) and D ($M = 9.00, SD = 3.85$). The test also found a significant difference between groups B ($M = 6.00, SD = 2.07$) and C ($M = 11.63, SD = 4.41$) and between groups B ($M = 6.00, SD = 2.07$) and D ($M = 9.00, SD = 3.85$).

The test found no significant difference in number of trials completed between group A ($M = 6.00, SD = 3.59$) and group B ($M = 6.00, SD = 2.07$) and no significant difference between group C ($M = 11.63, SD = 4.41$) and group D ($M = 9, SD = 3.85$).

A Tukey HSD post-hoc pair-wise comparison test was also completed to examine the significance found for the effect of difficulty on the number of trials completed correctly. This test was conducted to compare trials categorized as easy with those categorized as medium and hard, and to compare trials categorized as medium with those categorized as hard.

This test showed a significant difference in the number of trials correctly completed that were categorized as easy difficulty ($M = 16.75, SD = 5.19$) and those that were categorized as hard difficulty ($M = 26.25, SD = 6.85$). There was no significant difference between the number of trials correctly completed that were categorized as easy difficulty ($M = 16.75, SD = 5.19$) and those categorized as medium difficulty ($M = 22.50, SD = 9.75$). There was also no significant difference between the number
of trials correctly completed that were categorized as medium difficulty ($M = 22.50$, $SD = 9.75$) and those categorized as hard difficulty ($M = 26.25$, $SD = 6.85$).

6.2 Criteria II: All Nodes Correct, Position Can be Incorrect

As expected, including the trials where subject have flipped the correct location of the correct nodes does not greatly affect the results. All four training conditions again show a general learning curve when time taken for a correctly completed trial is plotted against trials. Figures 16 through 19 show the average times for correct answers for each trial, by group, with standard error. Figure 20 shows average times for each group and the overall average for all subjects. Plots for each subject, by group, can be found in Appendix C.

![Average Time per Trial, Group A, Criteria II](image)

Figure 16: Trial time learning curves for group A over two days of trials for criteria II. This group received only verbal representations in training.
Figure 17: Trial time learning curves for group B over two days of trials for criteria II. This group received only visual representations in training.

Figure 18: Trial time learning curves for group C over two days of trials for criteria II. This group was provided with verbal and visual representations in training.
Figure 19: Trial time learning curves for group D over two days of trials for criteria II. This group was provided with verbal representations and created visual representations from them.

Figure 20: Trial time learning curves for all groups as well as overall average times for all subjects, over two days of trials, for criteria II.
6.2.1 One-Way ANOVA for Time, by Group and Difficulty

A one-way between subjects ANOVA was conducted to compare the effect of training on time to correctly complete a hypothesis testing task in four training conditions, including cases where all nodes were correctly identified but where position in the sub-graph could be incorrect. There was a significant effect of training on time to complete a correct trial at the p < .05 level for the four conditions \( F(3, 274) = 7.18, p < .001 \). A one-way ANOVA was also conducted to compare the effect of difficulty of task on time to correctly complete the hypothesis testing task in easy, medium, and hard conditions, again including cases with incorrect node position. There was no significant effect of difficulty of task on time to complete a correct trial at the p < .05 level for the three conditions \( F(2, 274) = 1.79, p = .17 \).

A Tukey HSD post-hoc pair-wise comparison test was completed to investigate the significance found for training effects on time to complete the hypothesis testing test correctly. This test was conducted to compare each training condition, A, B, C and D, to every other condition. Group A received only verbal representations as training, group B received only visual representations as training, group C received both verbal and visual representations as training, and group D received verbal representations and were required to create visual representations.

The test found that the mean time for a correct task in group A \((M = 562.82, SD = 343.46)\) was not significantly different than for group B \((M = 592.30, SD = 360.41)\). It was also found that there was a significant difference between groups A \((M = 562.82, SD = 343.46)\) and C \((M = 396.67, SD = 282.56)\). There was a significant difference between groups A \((M = 562.82, SD = 343.46)\) and D \((M = 405.68, SD = 288.08)\), which are different results than those from criteria I.

The test also found a significant difference in the mean time to correctly complete the hypothesis testing task between groups B \((M = 592.30, SD = 360.41)\) and C \((M = 396.67, SD = 282.56)\), as well as between groups B \((M = 592.30, SD = 360.41)\) and D \((M = 405.68, SD = 288.08)\). It was determined there was no significant difference between groups A \((M = 562.82, SD = 343.46)\) and B \((M = 592.30, SD = 360.41)\), or between groups C \((M = 396.67, SD = 282.56)\) and D \((M = 405.68, SD = 288.08)\).
6.2.2 Generalized Linear Model for Accuracy, by Group and Difficulty

A Chi-square logistic regression was completed to compare the effect of training on the number of trials completed correctly under criteria II in four training conditions. The test determined that there was a significant effect for group on the number of trials correctly completed ($\chi^2(3, 636) = 32.07, p < .001$). A test was then completed to compare the effect of difficulty on the number of trials completed correctly for easy, medium and hard conditions. The test found that there was a significant effect for difficulty on the number of trials correctly completed ($\chi^2(2, 634) = 11.82, p < .001$).

A Tukey HSD post-hoc pair-wise comparison test was conducted to examine the significance found for the effect of training on the number of trials completed correctly. This test was conducted to compare each training condition, A, B, C and D, to every other condition. Group A received only verbal representations as training, group B received only visual representations as training, group C received both verbal and visual representations as training, and group D received verbal representations and were required to create visual representations.

This test showed a significant difference in the number of correct trials completed between groups A ($M = 6.63, SD = 3.16$) and C ($M = 11.75, SD = 4.40$) and between groups A ($M = 6.63, SD = 3.16$) and D ($M = 10.00, SD = 4.00$). The test also found a significant difference between groups B ($M = 6.63, SD = 1.92$) and C ($M = 11.75, SD = 4.40$) and between groups B ($M = 6.63, SD = 1.92$) and D ($M = 10.00, SD = 4.00$).

The test found no significant difference in number of trials completed between group A ($M = 6.63, SD = 3.16$) and group B ($M = 6.63, SD = 1.92$) and no significant difference between group C ($M = 11.75, SD = 4.40$) and group D ($M = 10.00, SD = 4.00$).

A Tukey HSD post-hoc pair-wise comparison test was also completed to examine the significance found for the effect of difficulty on the number of trials completed correctly. This test was conducted to compare trials categorized as easy with those categorized as medium and hard, and to compare trials categorized as medium with those categorized as hard.
This test showed a significant difference in the number of trials correctly completed that were categorized as easy difficulty ($M = 16.50, SD = 5.51$) and those that were categorized as hard difficulty ($M = 28.25, SD = 5.74$). There was no significant difference between the number of trials correctly completed that were categorized as easy difficulty ($M = 16.50, SD = 5.51$) and those categorized as medium difficulty ($M = 25.25, SD = 9.36$). There was also no significant difference between the number of trials correctly completed that were categorized as medium difficulty ($M = 25.25, SD = 9.36$) and those categorized as hard difficulty ($M = 28.25, SD = 5.74$).
CHAPTER 7: DISCUSSION

The purpose of this research was to investigate interactive visualization of data related to social network analysis. One goal was to determine how different forms of training affect analysts’ ability to perform a task, testing hypotheses for conformity to a provided social network data set. Additionally, this research aimed to determine how the difficulty of a hypothesis testing task would affect solution time and accuracy. Finally, this research sought particular solution methods that an analyst might use for solving such a task.

These research questions were addressed through the conduction of an experiment that focused on training with varied representations of hierarchical structures, and then the transfer of that training to an interactive task involving visualized social network data. Subjects were provided with two hypotheses of what a sub-structure in a network might look like, and they attempted to see which hypothesis conformed to the network, with time and accuracy of these hypothesis testing trials recorded.

The data was analyzed under two set of criteria. Criteria I only considered a trial correctly completed if the subject found all of the correct nodes and correctly entered them into the sub-graph hypothesis template. Criteria II still required the discovery of all correct nodes but allowed for them to be placed in the incorrect spaces in the sub-graph hypothesis template.

Due to the fact that this is a study in applied research, to be used for applications such as the analysis of social network structure, the results from criteria II will be the main focus of discussion. As stated in the results, in social network analysis it is most often the case that the person needs to be correctly identified first, and their role in the network can then be determined thereafter.

In all cases positions were switched only within the same level, i.e. two green nodes were flipped. Through initial filtering and categorizing these would most likely be the errors that would occur in actual analyses, as one would be looking for two characters of a certain description and may simply flip which has a different role within that level. In an investigation these differences could be uncovered if the correct people are investigated, but someone with a correct description but not in the sub-network would
be unimportant and insignificant. It is for these reasons that criteria II will be the focus of the discussion. Differences between the two sets of results, however, will be discussed.

7.1 Latency Analysis

The results for the linear regression of the time taken for each correctly completed trial as the response variable show no significance between groups A and B, or C and D, but does show significance between groups A and C, and B and C, as well as for A and D, and B and D. This effectively divides the groups into two categories: those with single representations of hierarchies and those with multiple representations of hierarchies.

No significant difference in the trial time for groups A and B means that we cannot say that training with only verbal representations or only visual representations are better than one another. This means that we cannot confirm hypothesis 1 or 2 individually, that rank these two training methods. Similarly, no significant difference between the trial time for groups C and D means that we cannot say that training with multiple provided representations is different than training with the creation of multiple representations. This means that we cannot confirm hypotheses 3 or 4 individually, either.

If instead, however, hypotheses 1 and 2 are grouped together, and hypotheses 3 and 4 are grouped together, we can examine the idea that those receiving single representations of hierarchical structures in training will take more time to correctly complete a hypothesis testing task than those receiving multiple representations of hierarchical structures in training. This grouping still retains the inferences as the initial hypotheses.

We can then look at our Tukey post-hoc tests to investigate our significant effect of training on trial time. To confirm this new theory on representations, group A must perform worse than both groups C and D. The Tukey post-hoc tests indeed confirm a significant difference between each of these pairs, A and C, and A and D. Looking at the mean time to complete a trial for each group we see that group A ($M = 562.82$, $SD = 343.46$) takes longer than both groups C ($M = 396.67$, $SD = 282.58$) and D ($M = 405.68$, $SD = 288.08$) to correctly complete the hypothesis testing task.
Additionally, group B must perform worse than both groups C and D. Again the Tukey post-hoc tests confirm a significant difference between each of these pairs, B and C, and C and D. Looking at the mean time to complete a trial for each group we see that group B ($M = 592.30, SD = 360.41$) also takes longer than both groups C ($M = 396.67, SD = 282.58$) and D ($M = 405.68, SD = 288.08$) to correctly complete the hypothesis testing task.

From these analyses we can confirm that subjects receiving single-representations of hierarchical structures will take more time to correctly complete a hypothesis testing task than those who receive or create multiple representations of hierarchical structures. This result agrees with past findings on the benefit of training with multiple representations for success on subsequent tasks that involve deeper comprehension (Butcher, 2006; Mayer & Gallini, 1990). This is believed to stem from a greater ability to create and utilize a mental model if both representations are externally available during training (Mayer, 1993; Mayer et al., 1996; Schnozt et al., 2003).

This finding also supports the findings of Oberlander, et al. (1996) that diagrammatic reasoners, such as the engineers in this study, are able to translate information successfully and effectively between multiple representations.

### 7.2 Accuracy Analysis

The results for the logistic regression with accuracy as the binary response variable (correct vs. incorrect) show no significance between groups A and B, or C and D, but does show significance between groups A and C, and B and C, as well as for A and D, and B and D. This again effectively divides the groups into two categories: those with single representations of hierarchies and those with multiple representations of hierarchies.

No significant difference in accuracy for groups A and B means that we cannot say that training with only verbal representations or only visual representations are better than one another for accuracy, just as this conclusion could not be drawn for latency analyses. This means that we again cannot confirm hypothesis 1 or 2 individually. Likewise, no significant difference between the trial time for groups C and
D means that we cannot say that training with multiple provided representations is different than training with the creation of multiple representations. For a second time, this means that we cannot confirm hypotheses 3 or 4 individually.

If, again, hypotheses 1 and 2 are grouped together, and hypotheses 3 and 4 are grouped together, we can investigate the theory that those receiving single representations of hierarchical structures in training will correctly complete a hypothesis testing task less often than those receiving multiple representations of hierarchical structures in training. This new hypothesis still retains the inferences as the initial hypotheses.

We can then look at our Tukey post-hoc tests to investigate our significant effect of training on accuracy. For this new theory regarding representations to be true, group A must not complete as many trials as both groups C and D. The Tukey post-hoc tests confirm a significant difference between each of these pairs, A and C, and A and D. Looking at the mean number of trials correctly completed for each group we see that group A (\( M = 6.63, SD = 3.16 \)) completes fewer trials correctly than both groups C (\( M = 11.75, SD = 4.40 \)) and D (\( M = 10.00, SD = 4.00 \)).

Additionally, group B must complete fewer trials correctly than both groups C and D. The Tukey post-hoc tests confirm a significant difference between each of these pairs, B and C, and C and D. Looking at the mean number of trials correctly completed for each group we see that group B (\( M = 6.63, SD = 1.92 \)) also completes fewer trials correctly than both groups C (\( M = 11.75, SD = 4.40 \)) and D (\( M = 10.00, SD = 4.00 \)) to correctly complete the hypothesis testing task.

From these analyses we can confirm the theory that subjects receiving single-representations of hierarchical structures will correctly complete a hypothesis testing task less often than those who receive or create multiple representations of hierarchical structures. Once again, this supports the existing literature that multiple representations help users to perform better with a deeper comprehension than single representations, likely due to a lack of mental model creation of those users who are only presented with external representations of one kind.
7.3 Difficulty Analysis

The linear regression analysis showed no significant effect of difficulty of hypothesis on time to correctly complete a trial. This fails to confirm the portion of hypothesis 5 that predicted a significant effect of difficulty on the time to correctly complete a hypothesis testing task. However, in part this supports the idea that even as difficulty levels change, the time taken on each trial will follow a learning curve, with time for each task decreasing with progression through trials. This aligns with hypothesis 6.

In contrast, the logistic regression showed a significant effect of difficulty on accuracy. The Tukey post-hoc tests show a significant difference between easy and hard trials, but not between easy and medium trials, or between medium and hard trials. While at first this intuitively makes sense, closer examination shows easy trials ($M = 16.50$, $SD = 5.51$) were completed correctly less often than hard trials ($M = 28.25$, $SD = 5.74$).

There could be several reasons for this. First, the distribution of difficulty was not even throughout the trials. In an attempt to match which trials were completed across groups as often as possible, for the greatest overlap in trials attempted, the trials were not randomized within sessions. Secondly, the manner in which trials were deemed easy, medium, and difficult were based on assumptions of solution methods (that distinguishing features made it easier to correctly complete trials) and thought processes.

In order to more closely examine the effect of difficulty, an experiment is suggested where participants receive randomized trials with difficulty more evenly distributed within sessions. Fewer trials (or looser time constraints) should allow participants to complete all trials so that a high number of trial overlap is still achieved. Additionally, the subjects should then rate the hypothesis pairs after each trial as easy, medium or difficult in an attempt to better understand how subjects view the difficulty of various trials.
7.4 Method Analysis

Hypothesis 6, that subjects will develop solution methods and perform better as they progress through trials, was confirmed. Plots showing solution time averages by group (found in the Results section) confirm the presence of learning curves for each group, as well as for all subjects overall. The break between the first day and the second day creates two learning curves per subject.

Additionally, as stated in the previous section, the failure of difficulty to effect solution time illustrates the independence of time from difficulty. The reduction in solution time, then, can be attributed to the formation of solution methods and application of them to subsequent trials.

Solution methods were analyzed for the top performers from the second day of trials to examine which methods could be used to best train analysts. It was theorized that by the second day solution methods would be formed and put into use to complete trials, and those with the most correct nodes would have the most successful methods. The screen capture videos of the top six performing subjects were analyzed.

One of these subjects, from group C, began the trials by using the second degree feature where any nodes linked to the node of interest were highlighted. They used this feature to identify potential candidate nodes that might fit one of the two hypotheses, starting typically with the red nodes. Once a node was determined to be a possible fit, it was pulled out of the web, off to the side, to distinguish it from the others in the web. The nodes that were linked to that node were then pulled out, and these nodes (typically yellow) were then examined to see if they met the criteria. This continued until the initial node was rejected for either hypothesis, or the solution was found.

It was noted that this subject was very quick with using the keyboard shortcuts to transition from single highlighting of nodes to second degree highlighting, and back again, to deftly manipulate the web and view connections. Additionally, this subject utilized a work-around for the second-degree highlighting feature by simply bring a potentially connected node close to the other node of interest and “wiggling” the first node. If the subject was able to see a moving link among the other stationary links, then they knew the links were connected without having to use the second-degree highlighting function.
Figure 21: Screen shot from trial 12 of a high-performing subject from group C, illustrating their solution method of pulling potential nodes outside of the web for consideration.

Another high-performing subject, this one from group D, also used the second-degree highlighting to initially find potentially important nodes, pulling them out of the web for further investigation. This subject, however, rather than starting from the top or the bottom of the hierarchy, identified a distinguishing feature between the sets and looked for nodes matching that criterion first. For example, in trial 14 solution A has three green nodes, each connected to the same two purple nodes, which would likely not occur commonly in the web, so this subject searched for green nodes first and observed their connections, using scratch paper to quickly filter which nodes should be pulled to the side for further progression through the trial.
Figure 22: Trial 14, with a distinguishable feature in 14A of three green nodes each connected to the same two purple nodes. This is an easy feature to use for node acceptance or rejection.

Figure 23: Screen shot from trial 14 of a high-performing subject from group D, illustrating their solution method of pulling potential nodes out of the web for further consideration, with the additional strategy of searching for distinguishable features from the hypothesis pairs.
One high-performing subject from group C began the trials by sorting all nodes by color. They then started with an identifying feature from one of the hypotheses to determine potential solution nodes. Once these nodes were identified they were pulled to the other side of the screen for further investigation. A process was then used to work vertically through the hierarchy, confirming or rejecting nodes and hypotheses until a solution was found.

![Diagram](image)

Figure 24: Screen shot from trial 11 of a high-performing subject from group C, illustrating their solution method of initially sorting by color and then pulling potential nodes out of the web for consideration.

Three high-performing subjects, one each from groups A, C and D, did not manipulate the web in any way, they only used the second-degree highlighting feature to find nodes that were linked, and then worked on scratch paper to draw hierarchies and confirm or reject nodes or hypotheses.

From these method analyses, there is no clear best solution method. Some successful subjects were very interactive with the webs, with a high level of manipulation, and some barely interacted with the webs, with no manipulation at all. This result is not surprising, as different individuals will create different mental models and will come up with different solution methods.

This disconnect between success and a particular solution method reinforces the effect of training, with a significant difference between groups, despite different solution methods. These solution method
analyses suggest that analyst training should utilize the findings in the rest of this thesis, and be conducted with multiple representations, but that analysts should be left to find their own solution methods, as these may vary between analysts as they form their mental models.

7.5 Criteria Analysis

There was only one difference of significance between criteria I and criteria II in the analysis results. In criteria II there was found to be a significant difference between the solution times of group A and D, while in criteria I there was not. In criteria I, however, the difference is approaching significance with \( p = .08 \). Overall the means, standard deviations, and significant effects are very similar between the two sets of criteria.

7.6 Future Direction

There is much opportunity for advancement of this research. As discussed previously, the difficulty level of the pairs of hypotheses should be better dispersed throughout trials, and subjects should provide judgment of difficulty after completing each trial in order to accurately determine difficulty level.

Additionally, training techniques can be further investigated and refined, such as how time limits or provided examples might influence performance. The interactive visualization tool could additionally be scaled upward to determine the size of networks that can be correctly completed with various training methods. In the study of human-computer interaction, the question becomes, “How big of a network is too big, before a human cannot be an effective analyst?”

Another area of this study that should be investigated in more detail is the formation of mental models in the solution of hypothesis testing tasks related to social network analysis, and how training might influence these mental models. While this study saw differences in solution methods that implied differences in mental models, no measures were taken to accurately record subject reasoning. This is a large opportunity for further investigation.
Finally, while this experiment allowed the subject to interact with the links and nodes in many ways, there are continuously emerging technologies that allow for additional forms and types of interaction. One logical path would be three-dimensional representations, allowing a user to manipulate a network in an immersive environment.

7.7 Conclusions

The major findings of this experiment are two-fold. First, subjects who received training with multiple representations of hierarchical structures performed significantly better on a task where two subgraph hypotheses were provided and the subject attempted to match one of them to a visualization of a social network than subjects who received training with single representations of hierarchies. Those subjects that were trained using multiple representations were able to correctly complete the hypothesis testing task more often and more quickly than those trained with single representations.

Secondly, as subjects progressed through trials they formed solution methods and were able to correctly complete hypothesis testing tasks more quickly as they progressed. Based on varied solution methods, it is suspected that analysts form varied mental models to make judgments about the hypothesis testing task, which allows them to improve through practice as they continue to update a mental model. This suggests that analyst’s training should be centered on the types of representations used in training and not on what specific solution method might be best.

These two findings should be useful in guiding the training of analysts who interact with social network data, particularly in tasks where hypothesis structure is suggested and the network is being investigated to find sub-graphs that might match known structures.
REFERENCES


APPENDIX A

Sessions 4 & 5: VUE Program Introduction

Today you will be completing a task just like the one you did before, but on a higher level. Today you will be comparing pairs of hierarchical structures just as before, only the structures in this task are slightly different in nature. The structures today are all made of red, yellow, green and purple circles, where the colors define the levels of the hierarchy. Take the example of Fran and her family.

Before, the hierarchical structure might have looked like this:

Fran
  Liz
    Child
  Audrey
    Child
  Jack
  Child

In today’s task Fran and her family might be represented like this:

So two hierarchies we compare might look like this:

A
  B

Before, we were looking at just the hierarchies themselves, such as A or B.

In this task we will be looking at a hierarchy with other people and objects included:

Notice how hierarchy A is present in this web, shown using orange lines.

You will begin each trial looking at one of these groups of people or objects, displayed randomly on the screen as a web of circles. For each trial there is a page in your packet with two visual representations of hierarchies, A and B. These do not have particular scenarios associated with them, like the previous task, but if it helps you to think of a particular hierarchy, feel free to do so.

In each “web” we know that one, and only one, of the two hierarchies is present, A or B. We would like to figure out which one is present. To do this you will manipulate the web in a variety of ways.
The program you are using has several helpful shortcuts and features. Try practicing them now:

1) Click and Drag a circle
   a. Use CTRL + Z to UNDO a move

2) See which circles are linked to a circle. Select any circle – notice that it is highlighted blue
   a. Hold CTRL (●) and tap / (●)
      i. See how all of the circles connected to this circle are now highlighted blue too
      ii. Look at the top toolbar, see that the chain is now active
   b. Hold CTRL (●) and tap / (●)
      i. See how any circle connected to the previously blue circles are now highlighted
      ii. Look at the top toolbar, see that the third chain link is now active
   c. Hold CTRL (●) and tap . (●)
      i. See how only circles connected to the original circle are highlighted
      ii. Look at the top toolbar, see that we have gone back to the second chain link
      iii. Try to drag a circle while its connections are highlighted. Note what happens
   d. Hold CTRL (●) and tap . (●)
      i. See how only the original circle is now highlighted
      ii. Look at the top toolbar, see that the chain is no longer active

3) Zoom in and out, and zoom-to-fit
   a. CTRL - zooms out
   b. CTRL + zooms in
   c. CTRL ] zooms-to-fit the screen

4) Change the size of a circle. Click on a circle, and drag one of the white squares
   a. Try using CRTL Z... nothing happens! You have to change it back manually

5) Select the text: Select a circle, and click again on the number. The number is now in an edit box
   a. DO NOT EDIT THE NUMBERS
   b. If you accidentally change or delete a number, use CTRL Z to restore it

6) Deselect everything: when items are selected, either individually or as a group, a thin blue box appears around them (this is the same box that allows you to stretch a circle). To deselect everything, click outside the box

The numbers within the circles have no meaning, they are only for me to see if you are correct.

Remember, if there is no link between circles in one of the hierarchies then there should not be a link between those circles in the web. The absence of a link is as important as the presence of a link.

You will have until the end of the hour to complete up to ten (10) trials. I will let you know when to begin the first task. Once you have written your answer and finished the task, verbally confirm that you are done. I will record your time and will start you on the next task.

Use as much scratch paper as you need. If you need more paper, let me know and I will give you more.

Please ask any questions that you may have at this time.
APPENDIX C

Figure 25: Trial times for correctly completed trials for each subject in group A, under criteria I, as well as the average time for each correct trial for group A. This group received only verbal representations in training.

Figure 26: Trial times for correctly completed trials for each subject in group B, under criteria I, as well as the average time for each correct trial for group B. This group received only visual representations in training.
Figure 27: Trial times for correctly completed trials for each subject in group C, under criteria I, as well as the average time for each correct trial for group C. This group received verbal and visual representations in training.

Figure 28: Trial times for correctly completed trials for each subject in group D, under criteria I, as well as the average time for each correct trial for group D. This group received verbal representations and created visual representations from them in training.
Figure 29: Trial times for correctly completed trials for each subject in group A, under criteria II, as well as the average time for each correct trial for group A. This group received only verbal representations in training.

Figure 30: Trial times for correctly completed trials for each subject in group B, under criteria II, as well as the average time for each correct trial for group B. This group received only visual representations in training.
Figure 31: Trial times for correctly completed trials for each subject in group C, under criteria II, as well as the average time for each correct trial for group C. This group received verbal and visual representations in training.

Figure 32: Trial times for correctly completed trials for each subject in group D, under criteria II, as well as the average time for each correct trial for group D. This group received verbal representations and created visual representations from them in training.