THE IMPACT OF VISUAL AESTHETICS ON THE UTILITY, AFFORDANCE, AND READABILITY OF NETWORK GRAPHS

by

Patrick M. Dudas

Physics, Westminster College, 2006

Master of Information Science, University of Pittsburgh, 2009

Submitted to the Graduate Faculty of

the Department of Information Science in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2016

UNIVERSITY OF PITTSBURGH

DEPARTMENT OF INFORMATION SCIENCE

This dissertation was presented

by

Patrick M. Dudas

It was defended on

August 14th 2015

and approved by

Dr. Stephen Hirtle, Professor, School of Information Sciences

Dr. Michael Lewis, Professor, School of Information Sciences

Dr. Yu-Ru Lin, Assistant Professor, School of Information Sciences

Dr. Sean Everton, Associate Professor, Department of Defense Analysis, Naval Postgraduate School

Dissertation Director: Dr. Stephen Hirtle, Professor, School of Information Sciences

THE IMPACT OF VISUAL AESTHETICS ON THE UTILITY, AFFORDANCE, AND READABILITY OF NETWORK GRAPHS

Patrick M. Dudas, PhD

University of Pittsburgh, 2016

The readability of networks how different visual design elements affect the understanding of network data has been central in network visualization research. However, existing studies have mainly focused on readability induced by topological mapping (based on different layouts) and overlooked the effect of visual aesthetics. Here is proposed a novel experimental framework to study how different network aesthetic choices affect users' abilities of understanding the network structures. The visual aesthetics are grouped in two forms: 1) visual encoding (where the aesthetic mapping depends on the underlying network data) and 2) visual styling (where the aesthetics are applied independent of underlying data). Users are given a simple task identifying most connected nodes in a network in a hybrid experimental setting where the visual aesthetic choices are tested in a within-subject manner while the network topologies are tested in a between-subject manner based on a randomized blocking design. This novel experimental design ensures an efficient decoupling of the influence of network topology on readability tests. The utility of different visual aesthetics is measured comprehensively based on task performance (accuracy and time), eye-tracking data, and user feedback (perceived affordance). The results show differential readability effects among choices of visual aesthetics. Particularly, node based visual encoding significantly enhances network readability, specifically glyphs based on their ability to be utilized as a means to allow participants to create more robust strategies in their utilization. The study contributes to both the understanding of the role of visual aesthetics in network visualization design and the experimental design for testing the network readability.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

I would like to thank my advisor, Dr. Stephen Hirtle, and Masters advisor, Dr. Hassan Karimi, for helping me in my academic endeavors. Also to my committee members, Dr. Yu-Ru Lin, Dr. Michael Lewis, and Dr. Sean Everton, for their help and guidance. Dr. Petr Pancoska and Dr. Robert Branch for their patience with me, as the Clinical Pharmacology department was a wonderful means of getting into research that felt like it had a greater importance than just publications. Finally, Dr. Terri Lenox for continued support of me, well outside my time at Westminster College, thank you.

I will be forever grateful to my mother, Kathy, and sister, Katie, for their love and support throughout this very long journey. To my father, George, whom this thesis is dedicated to in memorial, for being an inspiration to me as both a man and a teacher. Thank you very much to my in-laws, Jackie and Michael, for always supporting Dannie and me over these last many years. Finally to my friends and family members, thank you for keeping me grounded and supporting me.

Finally, and most importantly, thank you to my wife, Dannie, and my beautiful daughter, Charlotte. Dannie, your love and kindness has meant the world to me and I am so excited to be starting our new adventure in State College with you and Charlotte. Even in my darkest day, your smile in the library and now when I get home, reminds me of how lucky I am. I find that I am a better person for meeting you and sharing our lives together. Even though I continue to trip over them at home, always remember,"I like your shoes."

1.0 INTRODUCTION

This study examines the readability, usability, and affordance of network graphs as a function of the topological structure and the application of visual aesthetics. Both the readability and visual encoding/styling relies on pre-processing in regards to the network topology and the aesthetic rule-set based on this topology. The usability (based on task performance) and perception phases both rest on the shoulder of the observer and their ability to work with and understand the network structure with the various visual aesthetic properties. Presented is work that looks at these two major relationships between how the network topological structure affects readability and usability and how the visual aesthetics affects encoding and styling to enhance the perception for the observer. Utilizing eye-tracking hardware and software, both of these relationships will be looked at quantitatively and qualitatively. A better, more complete understanding of the network graphs and graphical visual aesthetics that can aide in its affordance will be showcased and critiqued.

Traditional means of understanding network aesthetics have typically been focused upon the formation of the topological structure of the network itself, or syntactic aesthetics (Purchase, 2002). The work of this type has produced a plethora of different heuristics to showcase networks in a more visually appealing way, such as the formation of the network topological structure by means of minimization of edge bends, edge crossing, and node overlap (Bennett, Ryall, Spalteholz, & Gooch, 2007; Huang & Eades, 2005; Purchase, 1997; Purchase, Carrington, & Allder, 2002; Ware, Purchase, Colpoys, & McGill, 2002), the utilization of orthogonality, collinearity, and symmetry frameworks (Marriott, Purchase, Wybrow, & Goncu, 2012), and minimizing edge curve (Xu, Rooney, Passmore, Ham, & Nguyen, 2012). Breaking from this mold is the incorporation of visual aesthetics, such as the use of color, shape, and size to highlight the topological component of the graph, broadening network aesthetics into more of the domain of network visualization.

Visualizations themselves can be separated into two classes of perception: planar and retinal (Andrienko & Andrienko, 2006), where planar resides at levels of point, line, and area (topological) and retinal is more traditionally understood as the visual encoding and visual styling of data with aesthetic attributes for quantitative data (Bennett et al., 2007). This work provides an extension of this design to gain a more complete understanding of how graphs and visual aesthetics can enhance a participant's ability to complete a local, simplistic task, such as finding the most connected node in a network graph. When combined with eye-tracking, there is a more visceral understanding of both the aesthetics and usability of visual aesthetics in a network graph.

1.1 CONTRIBUTIONS

This thesis presents a novel examination of understood concepts in the domain of network visualizations. The application of visual aesthetics is not a new subject matter, but its utility and perceived affordance when applied to network graphs is an understudied field that this work hopes to bring into focus. The way this problem was approached was by first defining a task that can be accomplished independently of any additional visual aesthetics, as in finding the most connected node(s) in a network graph. The visual aesthetics themselves can be seen as simple filters that do not change or alter the task, but either enhance or hinder a user's ability to complete the task. These visual aesthetics are applied to three different levels of the graph (node, link, or groups) and are based on two different means of visual enhancements (either encoding data or stylizing the graph). This is applied to two sets of graphs (small graph - 40 nodes, and large graph - 80 nodes) in a novel research design to incorporate both the topological structure of the network and the visual aesthetics in a randomized blocking design. Gathering user generated data is believed to provide a better understanding of the readability and usability of network graphs when visual aesthetics are applied, based on their perspective of the network and its visual aesthetics, coupled with the utilization of eye-tracking data.

Utilizing this approach, the study systematically investigates the connection between the network formation based on its topological structure and the applications of various visual aesthetics. This provides a novel methodology for testing networks when applying an independent variable and identifying when it is necessary to remove the network's topological structure bias. Showcased is the use of this design, supporting that users prefer task dependent visual aesthetic encoding, even if the participants themselves may be novice to the structure of the experiment (novice to network graphs and their utility), and that task-dependent aesthetic encoding lends itself to higher accuracy values. Both results may be seen as obvious conclusions, but with the use of eye-tracking results can provide more context to how the participants' saccade and fixation points alter based on three levels (node, links, and groups).

Also, breaking from the norm in terms of visual encoding using size, shape, and color at the local level, it is demonstrated that shapes (glyphs) provide the highest accuracy level based on their ability to be used in exploration and discovery within the graph structure. Encoding at the link level can limit exploration but heighten accuracy levels based on focusing the eye-movements to areas of interest.

The focus of this work was not done for sheer testing alone, but as an extension of a previous project called Dynamic Twitter Network Analysis (DTNA), which was developed in conjunction with researchers at the Naval Postgraduate School (CORE Labs) in Monterey, CA. This platform allows users the ability to collect, analyze, and better understand social media messages via a real-time network visualization. The visualization attempts to change the paradigm in handling "big data" datasets by allowing more refined filtering in real-time to collect a more targeted search by showing tweets in a visually cognitive manner. This provides smaller snapshots of larger datasets to allocate in learning and understanding. This thesis work outlines how to better understand these smaller networks (40 and 80 nodes) with the application of visual aesthetics.

Provided in this work:

- A novel approach to combat the significance typically imposed by the topological structure of the network by use of a randomized blocking structure, where a corresponding network is created for each visual aesthetic. Participants were cycled through this matrix design to create a within-subject test based on the visual aesthetic utilized and independent of the topological structure of the network.

- The use of rotation to create redundancy in our design to help validate eye-tracking. This was done by rotating the networks by 90°, 180°, and 270° degrees (Pilot Study), and 90° (Thesis Experiment) to create similar results, but not replication.
  - Also, based on complexity of the network graph structure and the rotation, it showed that an effect similar to change blindness can occur.
- A direct comparison of task-related visual encodings, non-task-related visual encodings, and visual styling (Pilot Study), and a task-related visual encoding and visual styling (Thesis Experiment), and their effects on the measures of accuracy and time. Also, examined is how these visual aesthetics are perceived by the users based on feedback and qualitative analysis in regards to their affordances of each visual aesthetic.
- A highly-structured, quantitative approach to eye-tracking data that removes a majority of human error in their coding, which was employed based on three types of eye-tracking usability calculations: average saccade length, task abandonment, and task focusing ratio (inverse of task abandonment).
- A retrospective talk-aloud protocol was used in conjunction with affinity propagation cluster heatmaps, fixations, and saccades to understand strategies employed by the participants.
- A direct comparison of the subjective beauty and the task focusing time to better understand engagement when visual aesthetics are applied.

1.2 DISSERTATION ROAD MAP

In Chapter 1, the background is first introduced for both network aesthetics tied to the topological structure and visual aesthetics in terms of readability and usability. Also included in the background are numerous projects of how readability can be increased by changing how a network graph is illustrated, as well as eye-tracking studies in regards to usability. The motivation is presented to showcase why this study was put together. Chapter 2 provides the pilot study, its design, results, and implication on the main study. Chapter 3, the thesis study, provides detail on how the study was developed, implemented, and the hypothesis for the study. Finally, Chapter 4 provides the results and detailed discussion on what was learned throughout this work and its implication in other domains.

1.3 BACKGROUND

1.3.1 Network Topological Structure

For network aesthetics based on the topological structure, Purchase et al. (1997, 2000), concentrated on investigating the usability of automatic graph layout algorithms in relation to aesthetic properties. Based on the nature of the graph (i.e., syntactic graph or semantic graph) and the particular effect to be tested (individual aesthetic or individual algorithm), experimental results showed that: 1) for both syntactic and semantic graphs, reducing the number of crossing edges is the most important aesthetic; 2) orthogonality has a relatively high priority for semantic graphs despite it not being true for syntactic experiments; and 3) there is no best algorithm when syntactic graphs are employed.

Continuing this work, Ware et al. (2002) defined a measure for cognitive costs relative to the more commonly known network topology aesthetics by networks that violated these criteria and used an online survey to find commonality in their flaws. Using stepwise multiple regression, they determined that the shortest path length, continuity of the shortest path, the number of crosses on the shortest path, and the number of branches from intermediate nodes on the shortest path produced the highest cognitive dependencies in traversing the network. These works show the necessity of network aesthetics (based on topology).

Huang provided multiple experiments in the area of network aesthetics based on the topology of the network (Huang & Eades, 2005; Huang, Hong, & Eades, 2006; Huang, 2007). Huang et al. (2005) examined different types of layouts utilizing a radial and circular layout, having them answer questions based on a path related task relative to the position of the node or path of interest, and used eye-tracking to analyze the user's progression through the task. Huang et al. (2006) expanded the graph layouts used by including five sociogram drawing conventions (i.e., radial, hierarchical, circular, group, and free), where they employed user task performance and usability references. Their experimental results showed that although both conventions and edge crossings have no impact on subjects' perceptions of actor status, they are important for users to find groups and usually have significant roles on subjects' preferences. Huang (2007) tested in-depth the relationship between eye-movement and network aesthetics based on the topology. Their findings suggested that edge crossing leads to high degree of misdirection when tasks were focused on paths, yet when focused on groups, edge crossing could help illuminate these areas based on how well edge crossing fixates eye-movement.

Pohl, Schmitt, & Diehl (2009) compared three separate layouts (force-based diagram, hierarchical layout, and orthogonal layout), though the study asked similar questions in regards to local tasks (centrality) while evaluating eye movement. Using heat-maps from the eye-tracking data, they showed common patterns users would traverse within the network graph. The force-based graph showed superior results over the hierarchical layouts and orthogonal layouts. Again, no empirical methodology was utilized and all results were subjective to the researcher's judgment when it came to the eye-tracking data.

Saffrey and Purchase (2008) tested the connection between a user's mental model and the map in dynamic graphs. Acknowledging traditional key features such as orthogonality, clusters, and topology, they proposed two algorithms to adhere preserving node placement, which, as they proposed, would aid in building the user's mental map. Archambault et al. (2011) tested various dynamic graph drawings with different maps, which included animation, slide show, and small multiples of time-slices. They were interested in tasks such as evolution of nodes' degree, the introduction of new edges, global trends based on the number of edges, and topology readability. Differentiating from that, this thesis utilizes a number of visual aesthetics on a static graph, but using eye-tracking software to monitor users' eye movements while traversing network structures.

To some degree, visual aesthetics implementation is not a novel idea, based on the works of Jianu et al. (2009) and Rusu et al. (2011) where an emphasis was placed on the minimization of error caused by edge-crossing. Jianu et al. (2009) used a coloring solution, where areas of high overlap were varied using CIE L*a*b to create perceptually opposing colors. Rusu et al. (2011) again focused on the edge-overlap issue by focusing on the utilization of Gestalt principles, particularly closure. This was evaluated and found to have no significant improvement based on difficulty of finding neighboring nodes. Even though there was little significance found, these ideas of the utilization of Gestalt principles is applied to this work at all three levels of visual aesthetics (node, link, and groups).

Two more recent works by Holten et al. (2011) and Netzel et al. (2014) incorporate various link aesthetics in path related tasks. Holten et al. (2011) applied tapered edging, animation, and glyphs with a variation in compression. They had participants determine if two nodes (highlighted) were connected and found that participants will prefer tapered edges to the other visual aesthetics. Similarly to this, Netzel et al. (2014) applied various aesthetics to directed-GPS tracked records to raise readability levels of the network graph. These came in the form of standard arrows, equidistant arrows, equidistant comets, and tapered links, which were analyzed based on path following, longest link, and number of nodes in a given cluster. They concluded the best means for directed graphs of this type would be the incorporation of tapered links. The Pilot Study did not use tapered edges, as it was not using directed groups (Holten et al., 2011; Netzel et al., 2014), but did use the ideas of using visual styling to relate whether it could help or hinder a participant's ability to complete the task. The Thesis Experiment applies a variation of the tapered edge, but augmented to meet the needs of an undirected graph. Also, improvements were made on these studies by incorporating visual encoding, both related and non-related to the task.

Burch et al. (2011) compared traditional, orthogonal, and radial tree diagrams based on the task of finding the least amount of ancestors on randomly selected leaf nodes. As noted by the authors, this was done as the task would require a specific strategy and using eye-tracking would help in understanding these strategies. They also applied rotation to the non-radial tree designs and found that orientation of the root to the top of page preferred. This would not be applicable to this work since these networks used in the study do not follow a hierarchical or tree-design, so preference to orientation would be moot. Instead, the rotations are done to collect redundant data and discussed later on that, based on saccadic suppression, orientation does not matter based on spatial memory.

1.3.2 Network Aesthetics

A great deal of research in terms of network graphs has centered on networks and readability. In terms of understanding the scope of this project, a section has been dedicated to a sampling of how network graphs were tweaked, altered, or re-imagined in the pursuit of better understanding their underlining topology, and in some cases, hidden social networking features.

1.3.2.1 Edge Reduction With much of the work in network graph readability focusing on the topology itself and its inherited "confusability" based on edge crossing and orthogonality (Bennett et al., 2007; Huang & Eades, 2005; Purchase, 1997; Ware et al., 2002), when it comes to visualizing and understanding highly connected or highly dense networks, there is a certain human cognitive limit to the amount of data that can be illustrated without further advancements to gain insight into its design. Zinsmaier et al. (2012) provide a density-based visualization where nodes and edges are bonded together to reduce clutter and highlight both high degree nodes and highly traveled or weighted edges. This paper refers to Carr's (1987) work in scatter plots, where overlapping of data or areas of high density are represented by larger hexagons than other areas of lower overlap. Figure 1 shows two plots, from Carr's (1987), the left being the original data set and the right representing their approach in gray scale. Figure 2 showcases this same design, but uses color to discriminate between positive (blue) and negative (yellow) values.

Zinsmaier et al.'s (2012) application distinguishes the network topology as a density field by mapping vertices of the graph to a viewpoint of the display. This is done using a two-step process designated as Seed Point Method where: 1) nodes that fall in the same pixel are set to the same bin; and 2) weighted texture for every bin size greater than 0. This creates an Accumulation Field that feeds to edge aggregation method. For edge aggregation, a hill climbing method is computed to map two edge types where: 1) edges that start and end in the same bin (inner-cluster): and 2) edges that start in one bin, but end in another bin (inter-cluster). For inner-cluster, this aggregates to nullification and are not shown in the final graph. For inter-cluster, edges are based on local maximums. Figure 3 shows this by

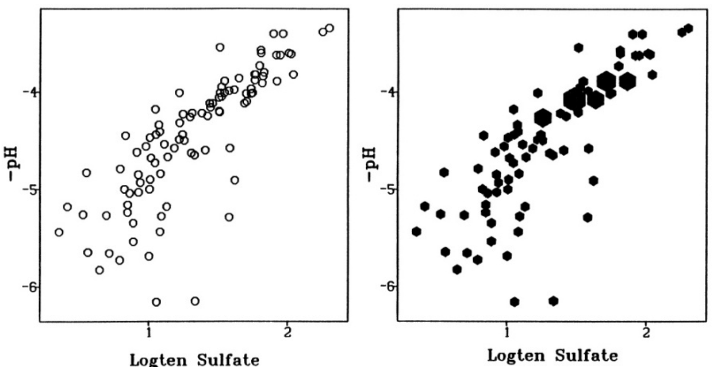


Figure 1: Updated scatter plot to reduce overlapped data

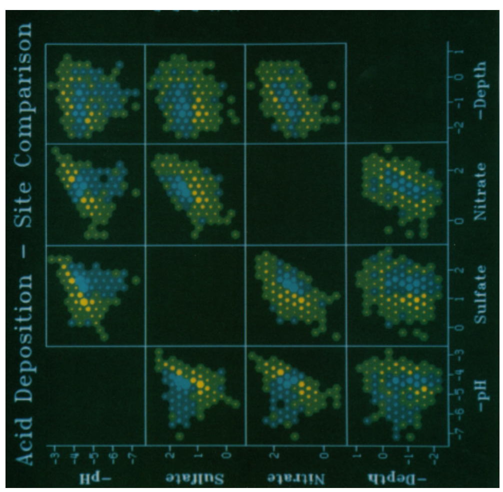


Figure 2: Using the same reduction in overlap and color to highlight positive and negative values

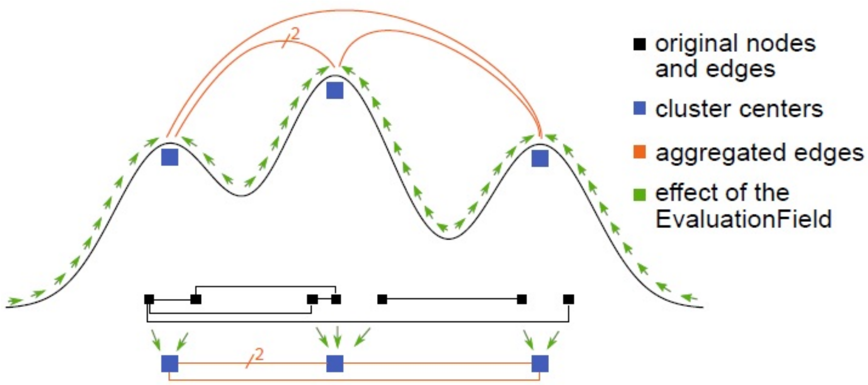


Figure 3: Local maximums and edge reduction

the local maximums of visual clustering; showcasing how inner-cluster nodes dissipate and inter-cluster nodes are highlighted.

Contrary to this are methods, such as work done by Ersoy et al. (2011), of edge bundling techniques. These techniques require specific data structures to be created for each bundle and can break down quickly with larger graphs due to these additional structures. Ersoy et al. (2011) also use edges that echo those of confluent drawing. An example can be found in Figure 4 from Dickerson et al.'s (2005) work and shows that edges can be drawn to be locally-monotone, in that they do not provide any sharp turns and never overlap. This is due to the fact that when it comes to constructing the graph, edge overlap can be the more disorienting aesthetic property (Purchase, 1997) to deal with, and when faced with this problem without altering the edge shape, can lead to a NP-hard problem (Garey & Johnson, 1983). Even though both Zinsmaier et al. (2012) and Ersoy et al. (2011) are both similar in spirit, the ability to compute graphs that affix to visual displays and can be done without pre-processing make Zinsmaier et al.'s (2012) work more expansive in the long run.

Selassie et al.'s (2011) network diagram used constraints provided in a force-directed graph. Edges contain control points that group nodes based on the direction of the edge to provide directionality to edges within the graph. Figure 5, from Selassie et al. (2011), shows an example of this transaction, where lines that flow in parallel, group at the potential minimum $m_{j=} q_j$ and when anti-parallel, at the potential minimum $m_{j=} q_j + lN_j$, creating

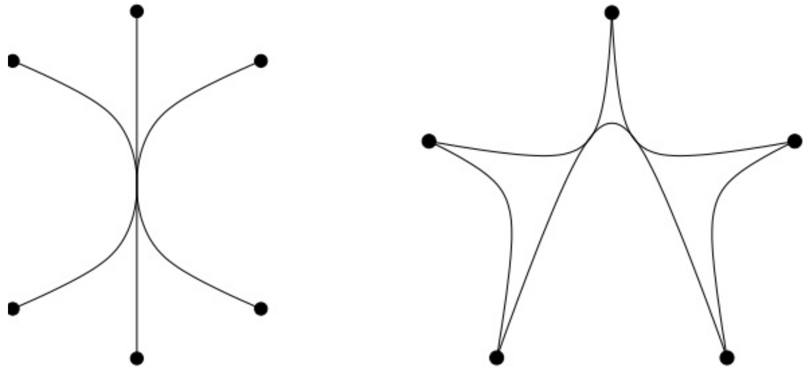


Figure 4: Two demos of confluent drawings

a "two lane highway" where both directions can be shown in the graph.

1.3.2.2 Usability/Interaction Instead of focusing efforts on augmenting the network graph itself via pre-processing, other approaches have focused on interactions to allow the user a better reader-driven experience with the graph structure (Segel & Heer, 2010). McGuffin & Jurisica (2009) presented a visualization rich in interactions to allow a user to manipulate a graph to be more "readable." Using brushing, lassoing, or selection defined by neighbor radius size (a slider is provided to increase/decrease size), users can select multiple nodes at a given time and change the display of the nodes based on: changing the glyphs of the nodes;

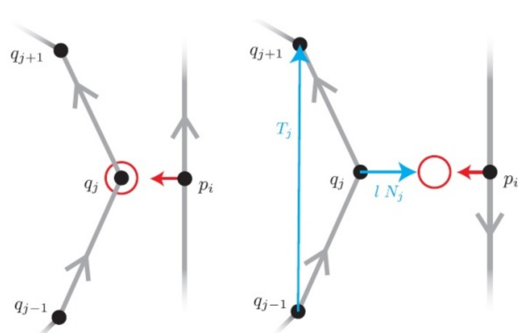


Figure 5: Grouping or Repulsion Based on Flow Direction

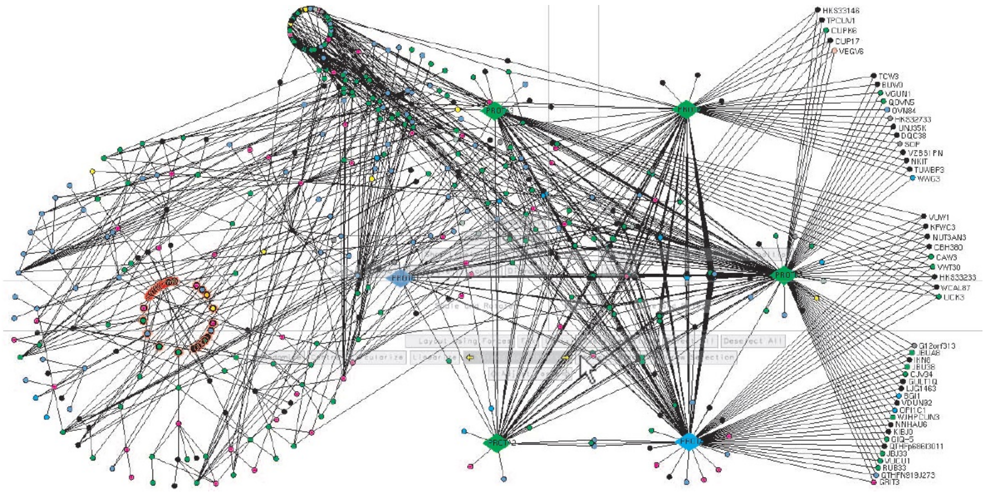


Figure 6: Layout made by designers (McGuffin & Jurisica, 2009)

arranging nodes in a straight line; separating nodes away from the graph; arranging nodes in a circular display; and collapsing nodes. These commands, plus others, were selected using a popup widget or hotbox presented, allowing for easier affordance to novice and more advanced action for expert users. Disappointingly though, none of the users in the usability testing were able to get to expert level within the two hour time window allotted to the participants. Figure 6, from McGuffin & Jurisica (2009), presents one of their layouts.

SocialNetSense (Gou, Zhang, Luo, & Anderson, 2012), which utilized TreeNetViz (Gou & Zhang, 2011), again provides hierarchical networking visualization support by allowing the user to arrange the domain specific information alongside network based data reflecting this information. This allows for bottom-up or top-down learning, understanding, and manipulation of the network graphs. The top-down provides the user with an overarching view of the topology to help with exploration and expansion of the network. The bottom-down requires the user to find additional "evidence" to append to the graph to solidify observations. This hierarchical network is showcased using a traditional node-link diagram and more specialized circular layout called TreeNetViz (Gou & Zhang, 2011). TreeNetViz showcases areas of connections, similar to Mizbee (Meyer, Munzner, & Pfister, 2009), but with the addition of the hierarchical structure. Figure 7 shows an example TreeNetViz where the inner most

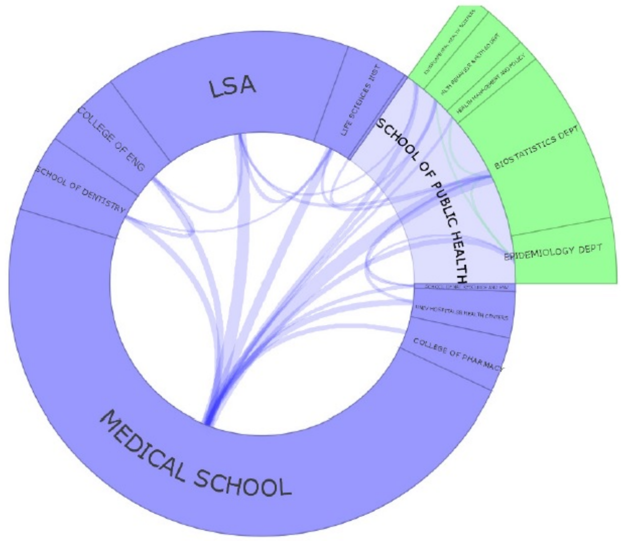


Figure 7: Circular display depicting inner-relations between School of Public Health and Departments

layer of circular design is the Medical School and the School of Health, then departments within the School of Health (Gou & Zhang, 2011).

1.3.2.3 Alternative Displays Breaking away from the more "traditional" force-based diagram were SocialNetSense (Gou et al., 2012) and Mizbee (Meyer et al., 2009). This can stem from radial or circular designs, spectral layouts, integrated statistical plots, etc. Similar to SocialNetSense and Mizbee is Hive Plots (Krzywinski, Birol, Jones, & Marra, 2011), which utilized a circular design, but allows for better overall structural understanding especially for larger, hierarchical networks. Figure 8 shows a network as it migrates to each step of the process, where Figure 8-b illustrates a more traditional force-based diagram. Then moving from Figure 8-c to Figure 8-d, steps are taken to reduce clutter by the removal of an edge and node. Figure 8-d shows the hive plot representation.

One of the more recent adaptations of network graphs is work done by Dwyer et al. (2013), which is slightly different then confluent drawing and edge reduction techniques. To remove redundancy from a network graph, identical neighbors, or Matching Neighbors,

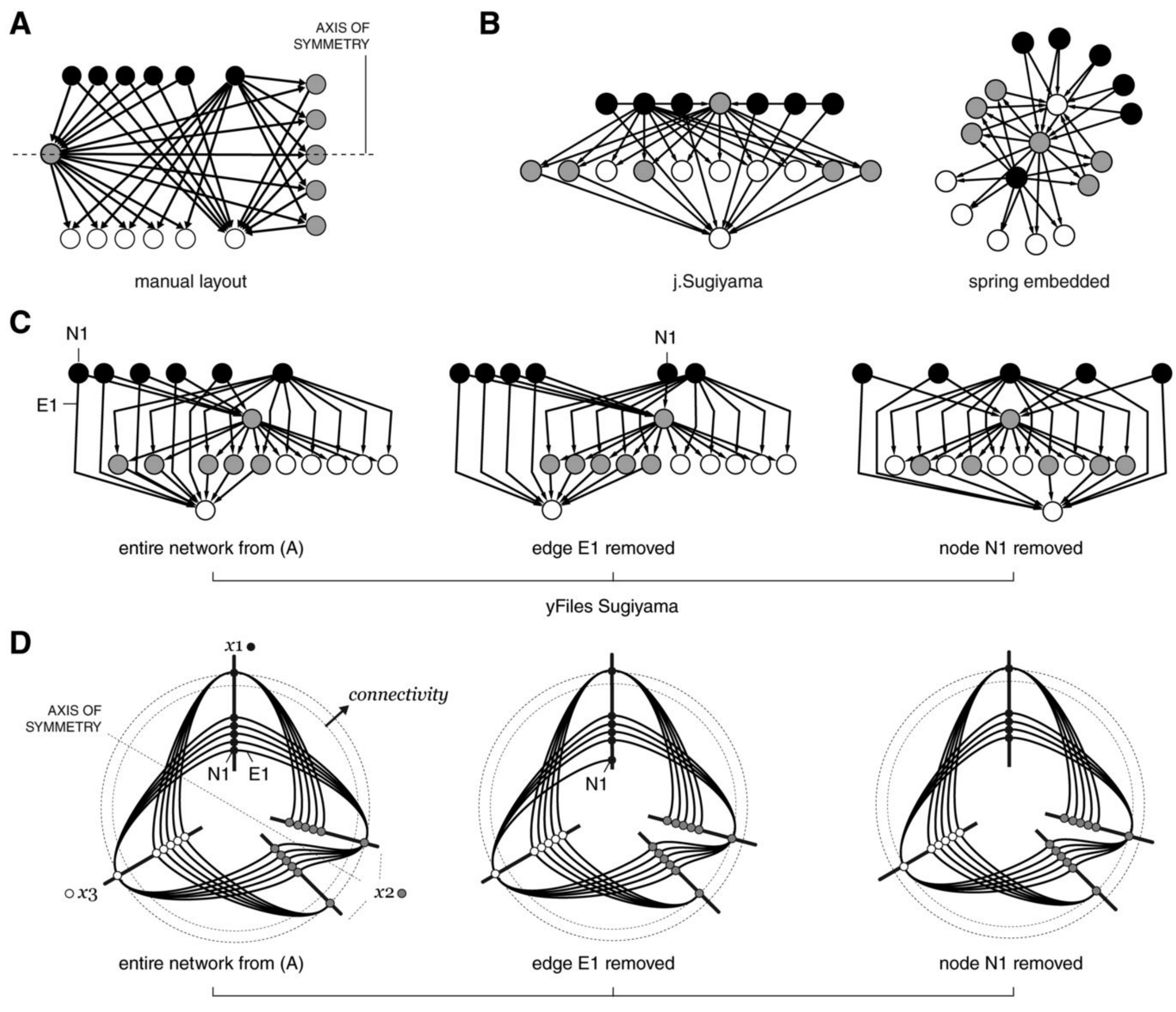


Figure 8: From Krzywinski et al. (2011) - edited, illustrates a more Traditional Force-Based Diagram to more specific Hive Plot Graphic

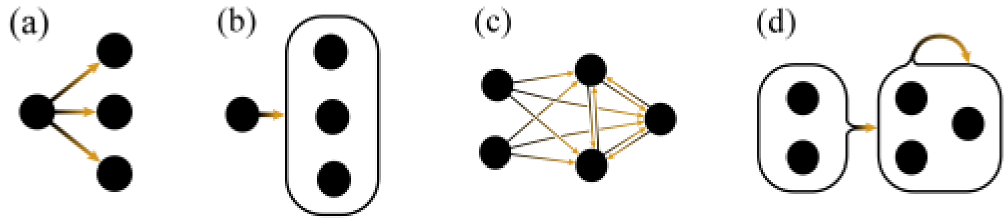


Figure 9: (a) and (c) represents the existing network structure and (b) and (d) their corresponding equivalents based on Matching Neighbor and Modular Decompression

are mapped together in a single "bubble" or "convex." Along with this constraint, two others were implemented called Modula Decomposition, where similar internal structures are mapped as single entities, and Power Graph Analysis, where clusters of nodes are considered "candidates" for decompression. Figure 9, from Dwyer et al. (2013), shows this Matching Neighbor and Modular Decompression.

Jianu et al. (2014) focused on the group level of data, incorporating four means of showcasing this data, including a colored node-link diagram, LineSets (color of node and links, and using curved lines), GMap (akin to maps of locations), and BubbleSets (similar to convex halls, but described more as enclosed contour). Figure 10 shows an example of each of these networks. They evaluated these designs using ten tasks, grouped over (group task, network task, combine group-network tasks, and memory tasks) using Amazon Mechanical Turk to collected data from over 788 participants (between-subject design). Considering this thesis looks at how convex hulls can be used to subdivide the problem set into more manageable cognitive parts, and finding validation that BubbleSets seemed to outperform in terms of group related tasks (including highest degree-within a group) based on its design, helped solidify the decision to use convex hulls in both experiments (1 and 2).

To this point, the visualization/graphs showcased represented to some degree the original network structure. This type of visualization can allow for better understanding of both temporal and geographical data that can be lost in the traditional network graph. In this regards, Brandes & Nick (2011) created a technique called Gestaltlines based on Gestalt

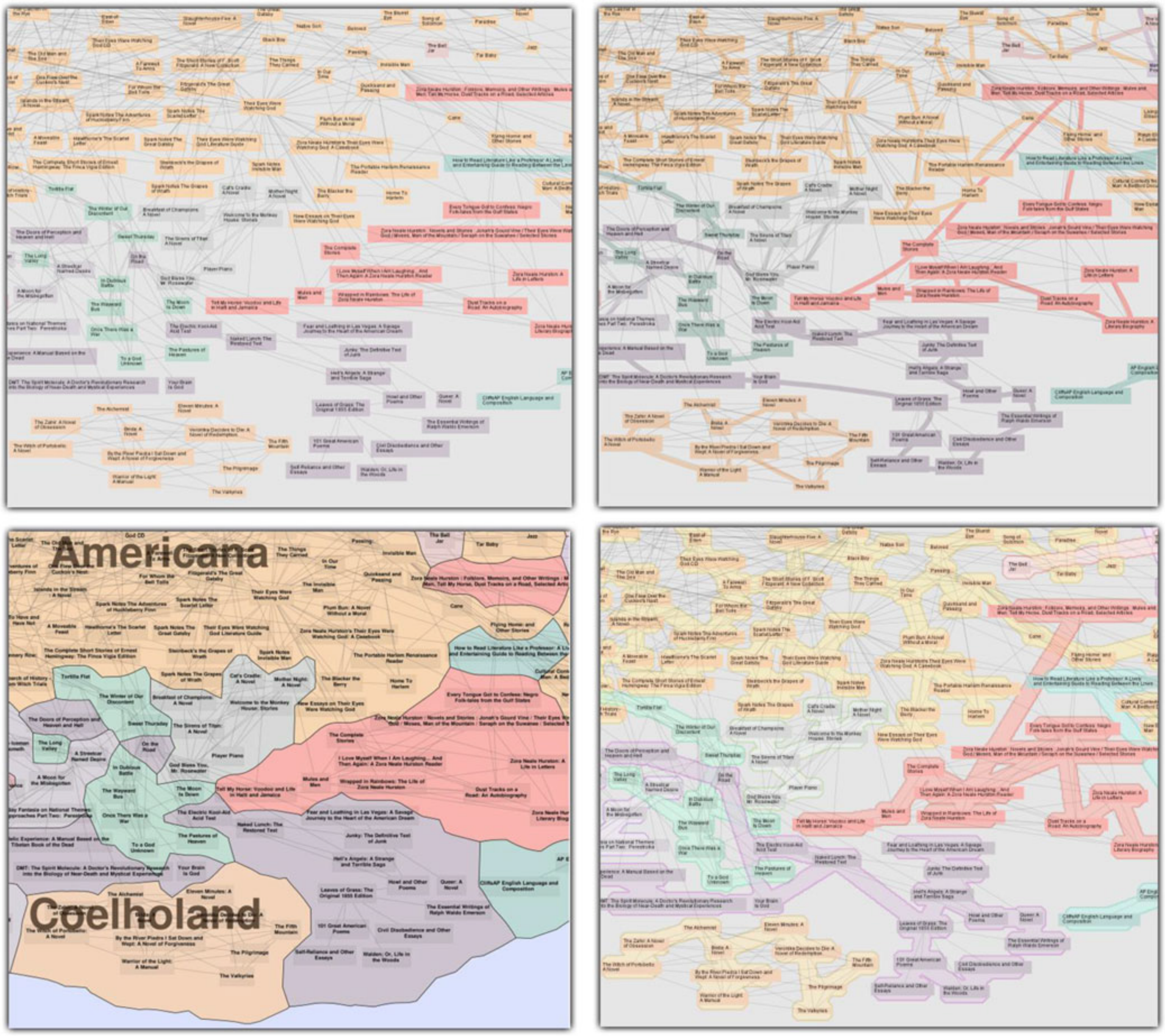


Figure 10: Colored node-link (top-left), LineSets (top-right), GMap (bottom-left), and BubbleSets (bottom-right)

Figure 11: Gestaltlines representing data read from bottom to top

principles and inspired by Edward Tufte's Sparklines (Tufte, 2006). These multivariate glyphs showcase both distinct "values" for ego-alter dyads, their differentials, and time in a single dataword. Figure 11, from Brandes & Nick (2011), illustrates values based on color, and should be read from bottom to top. Using these glyphs allow for a better understanding of social relations over-time in a static medium.

1.3.3 Eye-Tracking

Considering the subject matter of this work relates to visual perception, visual aesthetics, eye-tracking, visualizations, and perceived affordance, a brief elucidation of vision and perception are presented to solidify fundamental ideas moving forward. While we can utilize all of our senses to make cognitive decisions, our vision requires more bandwidth than all other senses combined at around 20 billion neurons (Ware, 2012), which is roughly 100 megabytes per second (Görg, Pohl, Qeli, & Xu, 2007). The human eye, our lens to the outside world, allows light via the pupil and is controlled (the amount of light) via the iris. When the light source is intense, the iris will contract, and when the light source is dim, it will then expand. Most of the light that hits the eye is immediately condensed and focused by the cornea (which also acts as a clear, protective filter). To provide more fine grain reduction of light, and hence focusing, the lens and ciliary muscles will work jointly to thicken or thin the lens to adjust for more visual details, called accommodation.

The eye will accommodate to provide as much detail to the back portion of the eye called the retina, which includes the fovea. Within the retina, rods and cones are used to provide color, detail, and motion sensitivity (the cones mostly for color and detail, and the rods for motion sensitivity). Most of the cones are centered on fovea and are dispersed around the rest of the retina and gradually dissipate). To make up for this, rods are almost the reverse of this construction and provide our peripheral vision, which is less detailed but allows us to detect motion. Considering the eye does not have a shutter like a typical camera, images are impressed 30 per second. This information is then transferred to the brain via the optical nerve (Giancoli, 2008; Wickens, Gordon, & Liu, 2004). The conversation from this point changes distinctly from the biologically defined to the more cognitively understood, not to say the cognitive is independent of the biological, but more in the sense of how light/vision is used to understand various cognitive processes.

David Marr first pioneered the idea that perception is defined by both a low-level and high-level vision (Marr, 1982), where low-level perception allows for people to abstract details of objects within sight and make judgments on what these objects are, where they are located, and to where they are moving. The high-level perception provides us the ability to detect features, structures, alignment, and much more high-level thinking (Douglas, Brian, & Arthur, 2005). When it comes to cognition, visual systems can affect the value of this type of information via bottom-up or top-down processes (Wickens et al., 2004). Bottom-up, like low-level, resides in the pre-attentive level of cognition and vision and can be handled in parallel versus its counterpart top-down, which needs to be handled sequential. These pre-attentive objects, or visual primitives, do not require previous knowledge (as top-down processing does) and can be distinguished even to the degree of the presence or absence of these low-level objects, which can apply parallel or sequential searching, respectively (Treisman, 1985). These cognitive processes can govern our understanding of structures (topologies in network graphs) and our visual cues (aesthetics in visualizations), and to this degree, that "art itself should be regarded as a specific kind of cognitive engineering (Turner, 2006)."

As there is a link to vision and cognition, usability testing with the incorporation of eye-tracking methodologies provides a better means of understanding how a user/participant engages with a visual display in a given task. Poole and Ball (2006) defined the use of

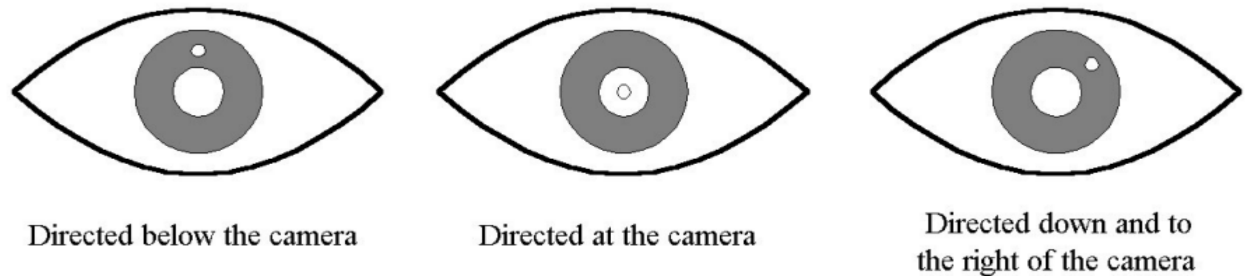


Figure 12: Three measurement examples applied to eye-tracking

eye-tracking as a, "technique whereby an individual's eye movements are measured so that the researcher knows both where a person is looking at any given time and the sequence in which their eyes are shifting from one location to another (Poole & Ball, 2006)." The technique specified relies on a camera and light omitting unit (which in this case, within the borders of the monitor). To minimize the amount of cumbersome direct light used, an infrared light is used to avoid distracting the participant. Once the infrared light contacts the eye, it causes two effects: 1) a slight dilation of the pupil, causing a "bright pupil" effect; and 2) a reflection on the cornea called the "corneal reflection." Both of these points are utilized to calculate the position of the eye, relative to its view on the screen. Figure 12, from Poole & Ball (2006), provides an illustration of the three measurement points on the eye to give examples of how this calculation is recorded to provide an accurate eye-position.

The use of eye-tracking provides a greater understanding of the participant's cognitive process, as the eye-movement exemplifies the participant's attention given as they fixate on the screen to complete the task. There are a number of eye-tracking usability terms worth defining at this point, which includes fixations, saccades, gaze, and scanpaths. Fixations are specific (focused) areas where the user maintains direct eye contact with the screen. In a cognitive sense, these can be seen as areas of interest based on task and on attention centered design (Poole, Ball, & Phillips, 2005) and cognitive processing, both caused by confusion and complexity in general search patterns (Just & Carpenter, 1976).

With the collection of eye-tracking data becoming more simplified with time, its application on a variety of information technologies have become more second nature. The connection between eye-tracking and usability can be seen in part to general decision making (Goldberg, Stimson, Lewenstein, Scott, & Wichansky, 2002) and adaption to different usability changes (Ehmke & Wilson, 2007). This work takes both of these processes into consideration by creating a repeated task but a consistent domain (network graph), so participants can develop strategies throughout the study. However, the study varies the visual aesthetics, which alters the participant's ability to complete the task in the same manner (as there may be a better strategy with a given visual aesthetic).

Huang (2007) showed a direct issue (slowness and perceived confusion) in the eye-tracking analysis in graphs with more edge-crossing. Netzel (2014) analyzed average fixation duration and average saccade length based on their various link aesthetics, where average fixation duration indicates depth of search and average saccade length based on long scans as exploratory and short scans as confusion. This thesis incorporates the average saccade length based on waiting to better understand the difference (if they are presented) between the levels of visual aesthetics applied (node aesthetics, link aesthetics, and group aesthetics).

Pohl, Schmitt, & Diehl (2009) compared three separate layouts (force-based diagram, hierarchical layout, and orthogonal layout). In the study they would ask similar questions in regards to local tasks (centrality) and global tasks (shortest path) while evaluating eye movement. Using heat-maps from the eye-tracking data, they showed common patterns users would traverse within the network graph. The force-based graph showed superior results over the hierarchical layouts and orthogonal layouts. No empirical methodology was utilized and all results were subjective to the researcher's judgment when it came to the eye-tracking data.

Granka, Joachims, & Gay (2004) utilized eye-tracking as a means to better understand user behavior when results are given. They found that for a given ranked list on Google search results, both the 1st and 2nd links had nearly the same fixation time, but the user would still substantially select the 1st result. Also, intuitively, user saccades more rapidly declined when the user needed to scroll to that selection and they would search the list from top to bottom. This indicates that when developing this experiment, page breaks should be minimized (in this experiment there are no page breaks). Also, the behavior mentioned by the experimenters could be seen more as a strategy based on their methodology and results, specifically when participants view search results using the ranked list (most relevant on top) as specified by the designers (Google).

With the most typical usability tests, participants are asked to provide details of their accounts while completing a given task, either by means of think-aloud protocol or more of post-hoc self-reporting. As mentioned and studied by Schiessl et al. (2003) there can be a disconnect between what a participant communicates to the experimenter and what their eye-movements actually reveal in their behavior. Using 120 participants, half female and half male, they examined how each participant engaged in a website where both text and pictures were utilized. The female participants reported that pictorial stimuli was more crucial in their judgmental process and the male participants reported the opposite (the textural information was more crucial). However, the eye-tracking data exhibited the opposite relationship, where females looked at the textural information more thoroughly and the males, the pictorial information. This indicates a direct disconnect to the participants' perceived behavior and the eye-tracking results, which is examined in this work.

Conati and Merten (2007) expanded on this paradigm of confliction or synergistic nature of eye-movements and cognition in the exploratory and active learning process. Their work highlighted a better understanding of the user's thought process with the utilization of eye-tracking verses a lower-level predictor such as time based on the accuracy of self-explanation and exploration. Also reported, and crucial to the design of this experiment, was a comparison of the exploratory task and the use of self-explanation, which significantly increased the time needed (roughly 13.1 seconds). As the design of the experiment limits the time a user can spend exploring the network to find the most connected node(s), this thesis opted to not use a talk aloud protocol as this could greatly increase the time necessary to complete the task.

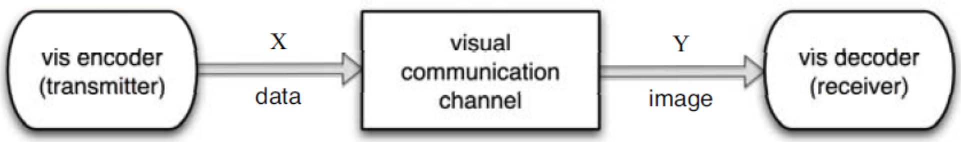


Figure 13: Analogy of data visualization, from encoder to decoder

1.4 DEFINITION OF TERMS

The term "aesthetics," or more specifically, "network aesthetics," is usually focused on the construction based on the topological structure of a network. The paper is diligent in using the phrase "visual aesthetics" to describe aesthetics applied to the network graph structure, akin to visual variables (Bertin, 1983). To flesh out this definition more, the utilization of a previous, simplified analogy to entropic relationship and visualization channel is presented and altered to be more specific to this study. Wang and Shen (2011) provide an analogous means of understanding the process of encoding and decoding visual information, similar to message transmission, where a message information is maximized when a transmitter has direct access to the receiver. This could hold true for visual communication when data can be mapped directly to a visual image (modeling), but not the case in terms of data visualizations as data needs to be encoded or mapped to an appropriate representation to maximize the image or graphic received by the decoder. In Figure 13, this is presented as a three step process, which provides an appropriate guideline to use in defining this project. Figure 14 shows the visual communication channel as more of an "interaction" level, where the equipment (eye-tracking and user-interface) can be utilized to better understand how well data (network graphs + visual aesthetics) is understood by the decoder (participant).

1.4.1 Visual Encoding/Styling

For the visual encoding/styling, included are two major components of the study: the network itself and the visual aesthetics. The network is discussed in greater detail later in the

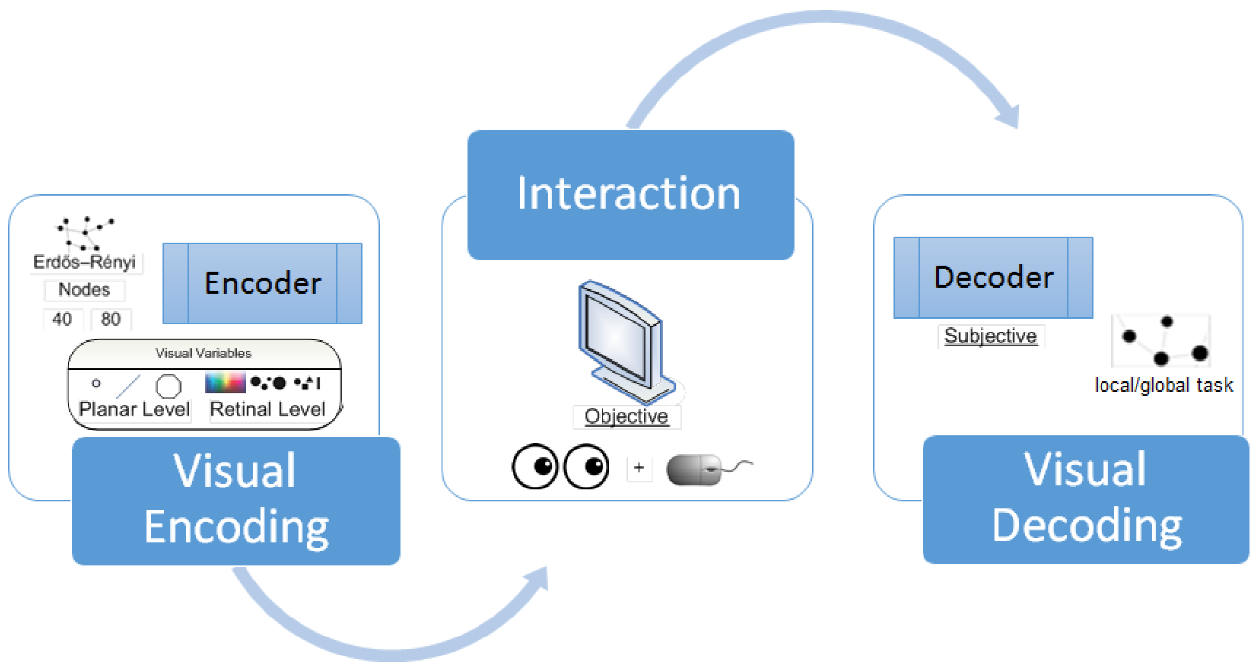


Figure 14: Three tiers (Encoding/Styling, Interaction, Decoding)

paper, but includes the randomized graphs (Erdös-Rènyi) of size 40 and 80 nodes for both the Pilot Study and the Thesis Experiment. This randomized graph was specifically used to emulate datasets seen in DTNA, but to the degree that reduced the likelihood of having a scale-free design, as a scale-free design would greatly focus the attention to a single node based on this power law distribution. Erdös-Rènyi is a probabilistic model that provides a network with equal probability of a centralization amount close to 0, between 0 and 1, or close to 1 (the 1st and 3rd examples are both extreme cases). Based on this, the visual aesthetics have a higher utility versus a scale-free network. The visual aesthetics are an extension of visual variables, which stem from two class perceptions: planar and retinal (Andrienko & Andrienko, 2006).

1.4.1.1 Planar Level Planar resides on the x, y coordinate system and at the levels of point, line, and area. This is considered to be a spatial-based component of the visualization, which makes the network diagram ideal for this situation with nodes, links, and groups. The reason is that if the network graph was strictly dependent on the planar dimension, that extra dependency would need to be included in the final analysis of the visual aesthetics. In another context, imagine a common scatter plot with both an x and y component, the observer can gain additional information based on where the markers reside on the plot itself. If the axes are kept static, but the data points are rotated 90 degrees, ultimately there would be the complete opposite visual trend as without the 90 degree rotation. Network graphs are not subject to this restriction based on this rotation. Rotating a network graph or augmenting the position of the nodes or links would not change the topological structure, as long as nodes and links were not hidden to the observer. The network graph is not 100% independent of this level, but all networks were operated on using the same force-directed layout algorithm so all nodes and links were set by the same restrictions.

1.4.1.2 Retinal Level The retinal variables are more traditionally understood as the visual encoding of data with aesthetic attributes for numerical values (Bennett et al., 2007). "One must determine which questions to ask, identify the appropriate data, and select effective visual encodings to map data values to graphical features such as position, size, shape, and color (Heer, Bostock, & Ogievetsky, 2010)." This definition is expanded to include visual styling as well, as long as the styling's creation is completed a priori to display. These styles are not determined as the visual encoding based on the topological structure of the network, but uniformly designed to enhance overall readability and applied to the node, link, and group levels.

1.4.2 Interaction Level

The interaction or display level of the experiment both showcases the network graph and the accompanied visual encodings/stylings. The participant will utilize the mouse to select nodes, which will change the outline of the node to showcase the selection. The keyboard will not be used because some people require looking at the keyboard to use it as an input, which will cause issues/errors with the eye-tracking software (Huang & Eades, 2005). Mouse movements and clicks will be recorded, along with eye-tracking software. Participants will also have the ability to supply subjective measures after each network to provide numerical data on the network's aesthetic quality.

1.4.3 Decoding Level

The decoding phase is where the participant is cognitively engaged with the networks and their corresponding visual encodings/stylings. This can be defined as the graphical perception level, as per original work of Cleveland and McGill (1984) on this subject matter in terms of graphics. This is where the term aesthetics becomes more applicable because it is a subjective measure. The main objective of this research is to better understand how users utilize the decoding phase to find the most connected node in the network. The overall theme throughout this paper is that the network itself can be used to solve the task independently of the visual aesthetic and the visual encodings/stylings can be seen as filters. What is gained by this work is figuring how these filters change the way the participants approach this task.

1.5 MOTIVATION

This project was designed specifically in response to a previous project called Dynamic Twitter Network Analysis (DTNA). DTNA looked at understanding social media in smaller, more cognitively accessible chunks (Ware, 2012) and was done by creating smaller network graphs based on specific location identities, keywords, and temporal data. This thesis extended this paradigm by analyzing how network graphs can be reimaged using visual aesthetics to help relay the information within the topological structure of the graph. Provided is an introduction into the realm of social network analysis as a means of highlighting DTNA's importance in this domain of network visualizations.

Much of the original work in social networks was based on sociograms (Moreno, 1960) within the confines of social conversation, where actors (people being represented) and their ties (the connection between these actors) can be illustrated by these sociograms (Newman, 2010). Extending these efforts, force-directed graphs were established to enhance the dynamic nature of these networks (Tutte, 1963) and to present a better representation of the occurrence or co-occurrence of these actors.

Social media analysis should not be constricted or limited to the actual, simple, 1-to-1 conversation or connections that come about through social websites and their services. A multitude of additional support and analysis can be included, like understanding individual roles in a network or how people socially interact in groups or communities formed on and offline. By redefining roles of 1-to-1 connections within the domain, social media analysis can enhance both online and offline connections

Examining the roles of actors in a social network, a plethora of the research has focused primarily on the development of groups, communities, cliques, and mutually-relevant, homogeneous clusters of individuals based on metadata about the actors (Ahn, Han, Kwak, Moon, & Jeong, 2007; Heer & Boyd, 2005; Matsuo et al., 2006). Social media is a diverse media set where real-life social constructs can be mirrored with the ability to easily find and suggest more people based primarily on homophily; meaning that based on a small sample size of friends, additional friends can be found by network connections following the simple mantra of, "birds of a feather, flock together" (McPherson, Smith-Lovin, & Cook, 2001). These social connections based on homophily can bring like-minded individuals together into groups or cliques, and allow the idea that people can and will belong to multiple groups or social networking societies.

These types of groups are considered virtual communities and foci in many research studies surrounding social media analysis. Blanchard (1998) examines groups that form from virtual communities by either presenting an already embedded user within a community with common resources, as in a local government or community groups, or by the creation of non-geospatial communities built on common topics or interests (Blanchard & Horan, 1998); whereas the former benefits from having both computer mediated communication and face-to-face communication (McPherson et al., 2001; Shah, Kwak, & Holbert, 2001; Wellman, Haase, Witte, & Hampton, 2001). Virtual communities, even dispersed ones, can thrive due to feelings of equality between participants, anonymity, a feeling of group membership and camaraderie, the removal of stereotypes based on physical traits, and other beneficial attributes. There are many negative attributes as well based on this anonymity and user created personas, but these types of deceptions have diminished over time due to the application of social connections in most virtual communities. Users feel more of a necessity to adhere to social norms found in most offline, face-to-face communities based on their public displays of connections and thus be more truthful about themselves to avoid exposure of their falsehood (Donath & Boyd, 2004).

A specific problem set of social media and subsequent virtual communities are terroristic or criminal networks called dark networks. Dark networks (Raab & Milward, 2003), or terrorist informatics (Cheong & Lee, 2011), provide their own unique problem set because immediate reaction is critical to the success of authorities looking for terroristic or criminal activity. Most traditional social network analysis come into play when working with dark networks with the caveat that brokers and bridges, and roles and positions within the domain of the network are enhanced when finding terroristic or criminal organizations (Everton, 2012). Also, within some network types, an important measure can include the identification of key players (Borgatti, 2006) and finding susceptible actors, which help influence the network to a certain agenda (Kempe, Kleinberg, & Tardos, 2003). Within the domain of both SNA and dark networks, a popular choice is the utilization of Twitter. It provides user-generated content, or micro-blogging, in a real-time environment as a social media outlet that is propagated from one individual to another in terms of text, links, pictures, or videos.

Twitter started in 2006 and has since gained 255 million users (Whitehouse, 2014). Its simple design has made it a household name in most countries. Limiting the message size to 140 characters, its first installment allowed users to text micro-blog posts, called tweets, for friends, family, and even the world to see. Twitter handles its association between actors slightly different than other communication or social media platforms, which typically requires mutual parties to agree on their social media "friendship" by creating bi-directional connections. In Twitter, people have the avenue to "follow" other actors without the necessity to reciprocate this connection, creating a much more scale-free network (Barabási & Albert, 1999), based on highly "followed" users and power-law distribution, which can range in the millions (Katy Perry has the most followers, currently at 52,546,289 (Wikipedia, 2014)).

When looking at patterns in large data sets, projects for this type of networked data is usually data-mined afterwards because either researchers are interested in a certain past event (Lotan et al., 2011; Yardi & Boyd, 2010) or topics of interest (Tumasjan, Sprenger, Sandner, & Welpe, 2010). There may also be a need for a very large dataset and backtracking provides the best means of collecting ample amounts of information versus real-time information (Bruns, 2011; Kwak, Lee, Park, & Moon, 2010). SocialAction (Perer & Shneiderman, 2008) extends this model by interweaving the statistical properties in the visualization itself, imploring a higher rate of success of exploratory SNA. They even extended this expert knowledge to the analysis of terrorist connections. Also, being a Shneiderman work, they followed the mantra of, "Overview first, zoom and filter, and then details on demand (Shneiderman, 1996)," to create the basic design of the force-based graph, but then extended this by depicting a scatter plot with SNA specific metrics for the graph itself (see Figure 15).

In specific instances when data is mined in real-time, the network structure is usually abandoned for either content-based natural language processing or content-based visualization. A real-time data approach to terrorism informatics was developed by Cheong & Lee (2011), who provided a detailed look at parsing and filtering the data from Twitter. They provided some markups on how they would visualize this information for users, including: 1) WEKA timeline analysis (Hall et al., 2009); 2) a Google Map with geo-tagged tweets; and 3) a self-organizing map for unsupervised clustering (Kohonen, 1984). The timeline analysis is a great mechanism for visualizing terrorism or dark networks, but lacks the ability to provide network analysis in real-time.

Whisper (Cao et al., 2012) utilized visual metaphors by creating a visualization that represented social media data diffusion in real-time and coupled this information to an analogous design of a sunflower. This allowed them to showcase both search topics and retweeted messages on a single visualization, which can overlay with geographical, temporal, and user-based data. Figure 16 showcases the main interface. The center position showed both retweeted and non-retweeted tweets, where retweeted tweets will migrate outside of the center circle and non-retweeted tweets will dissipate in the middle. Retweeted tweets will also connect to outside groups, which can be distinguished by either location or by common topic interests. They also illustrate temporal data by using arcs to connect the center circles

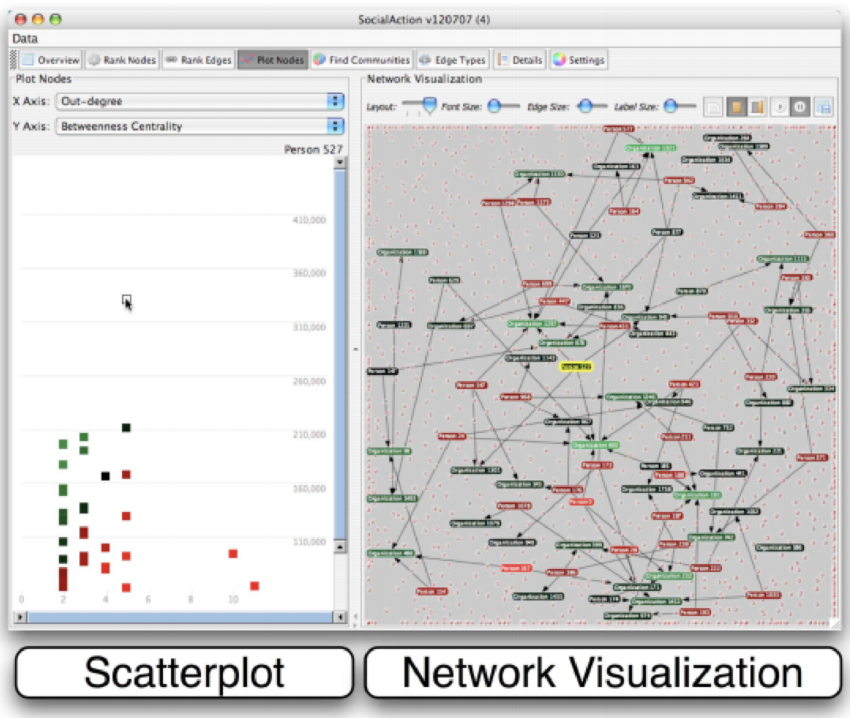


Figure 15: Coupled scatter plot with SNA metrics (Adam Perer & Ben Shneiderman, 2008)

to these outside groups. Coupled with the design are hues to highlight sentiment, opacity to encode activeness, size to exemplify influence, and glyphs for various types of users.

Similar projects such as Visual Backchannel (Dork, Gruen, Williamson, & Carpendale, 2010), Vox Civitas (Diakopoulos, Naaman, & Kivran-Swaine, 2010), and Chae et al.'s (2012) work centered around this connection between real-time trend analysis and spatio-temporal analysis. Visual Backchannel (Dork et al., 2010) incorporated both a stream/stacked graph, spiral graph, and image cloud to highlight current and most relevant topics, people, and images (respectively). Figure 17 shows example cases, where the topic stream provides both the 30 most relevant words associated with the search, but also the double encoded temporal information along the y-axis and the utilizations of hue, which includes a mechanism for searching/filtering. The People Spiral illuminates users of Visual Backchannel and their tweets on this current search. Lastly, the Image Cloud provides any images also related to a given search.

Vox Civitas (Diakopoulos et al., 2010) was tailored more toward a journalistic point of view and provides results that could be read and interpreted as one might see in news reports. This includes sentiment, a timeline of events surrounding the user-driven search, the tweets themselves, and keywords over time, which are shown in Figure 18. Chae et al. (2012), shown in Figure 19, focused more on the geographical data provided in Twitter. Coupled with topic modeling using Latent Dirichlet Allocation (Blei, Ng, & Jordan, 2003) and social media outlets, such as Facebook, Flickr, and YouTube, they provided plots/wordgrams to showcase unusual events during specific timeframes in social media, including case studies in the Chardon High School Shooting, the Occupy Wall Street protests, and the 2011 Virginia Earthquake.

1.5.1 Dynamic Twitter Network Analysis (DTNA)

DTNA (Dudas, 2013a, 2013b) is a novel, dynamic model that provides a smaller force-based graph computed every 8 to 12 seconds, depending on the latency of the data. The web-application is able to start with a specific geographical location (POI, city, state, country) within a specific bounding box around a given radius, and a username, hashtag, or keyword.

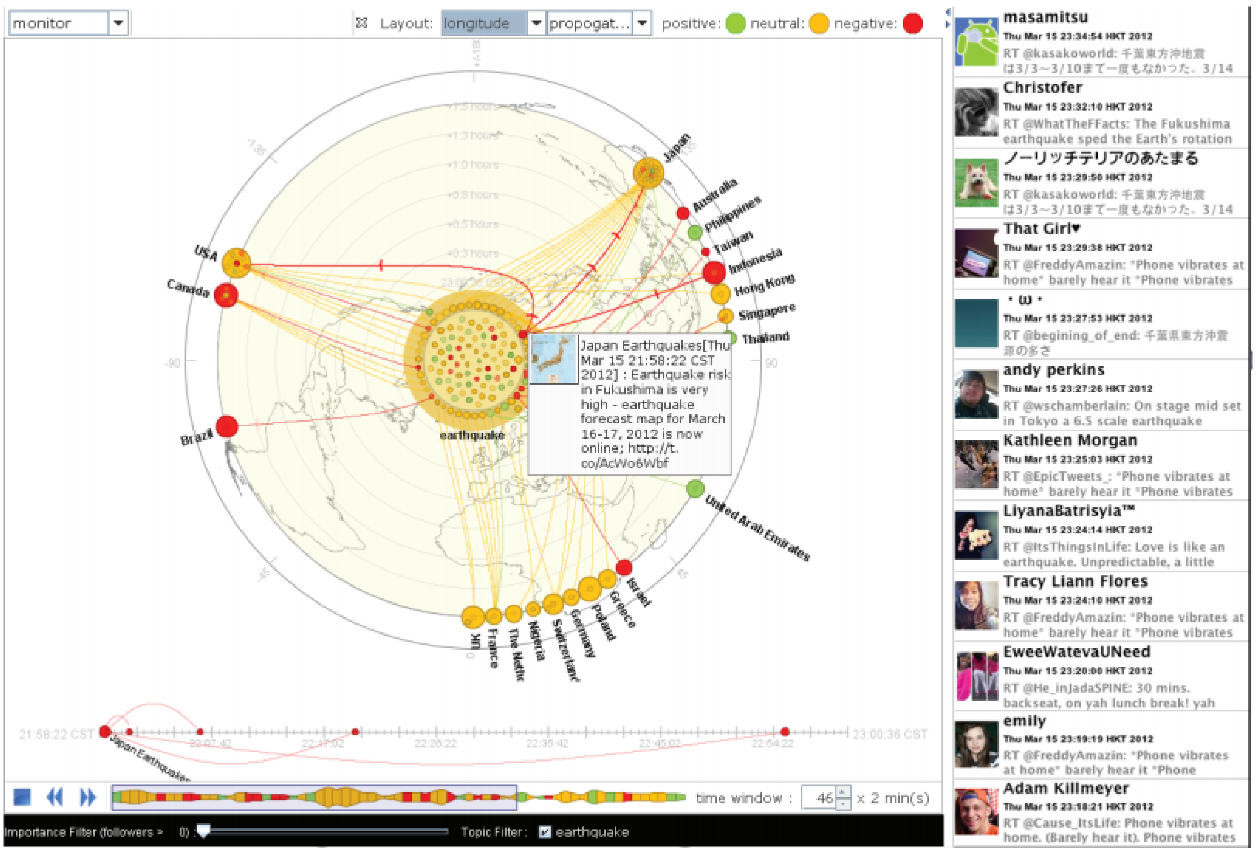


Figure 16: From Whisper (Cao et al., 2012), demonstrates Whisper's design including geographical, temporal, sentiment, and users

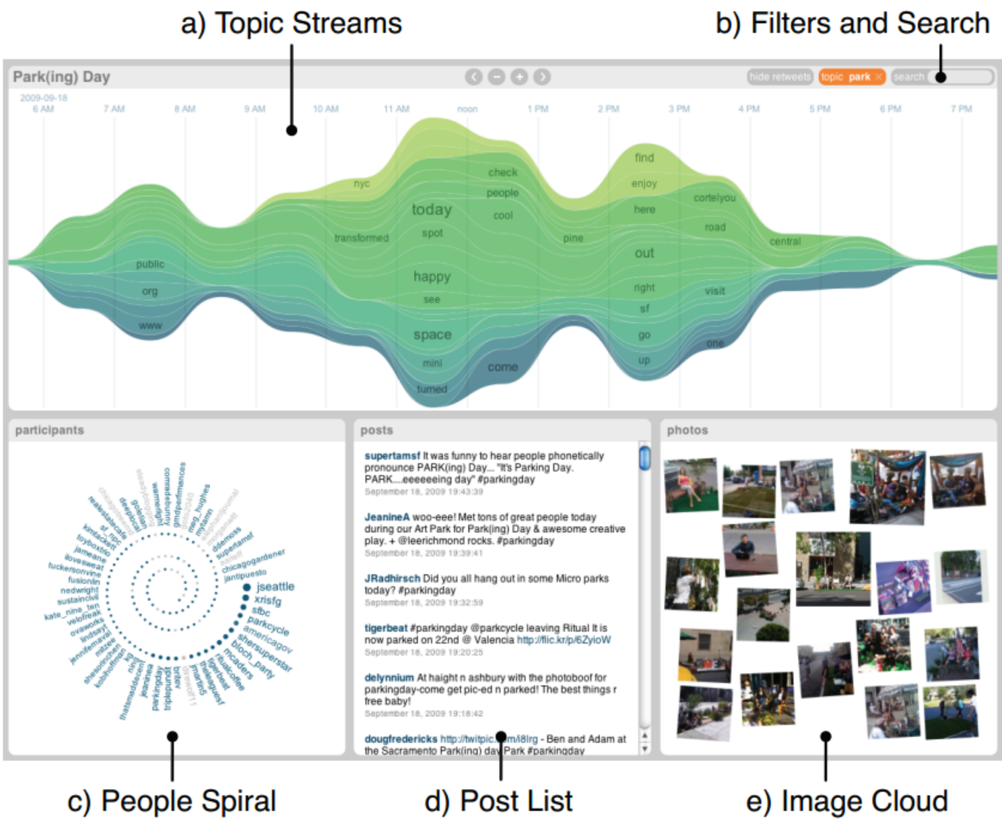


Figure 17: From Visual Backchannel (Dork et al., 2010), utilizing a Stream Graph and Spiral Chart, the UI provides trend knowledge based on topic and users

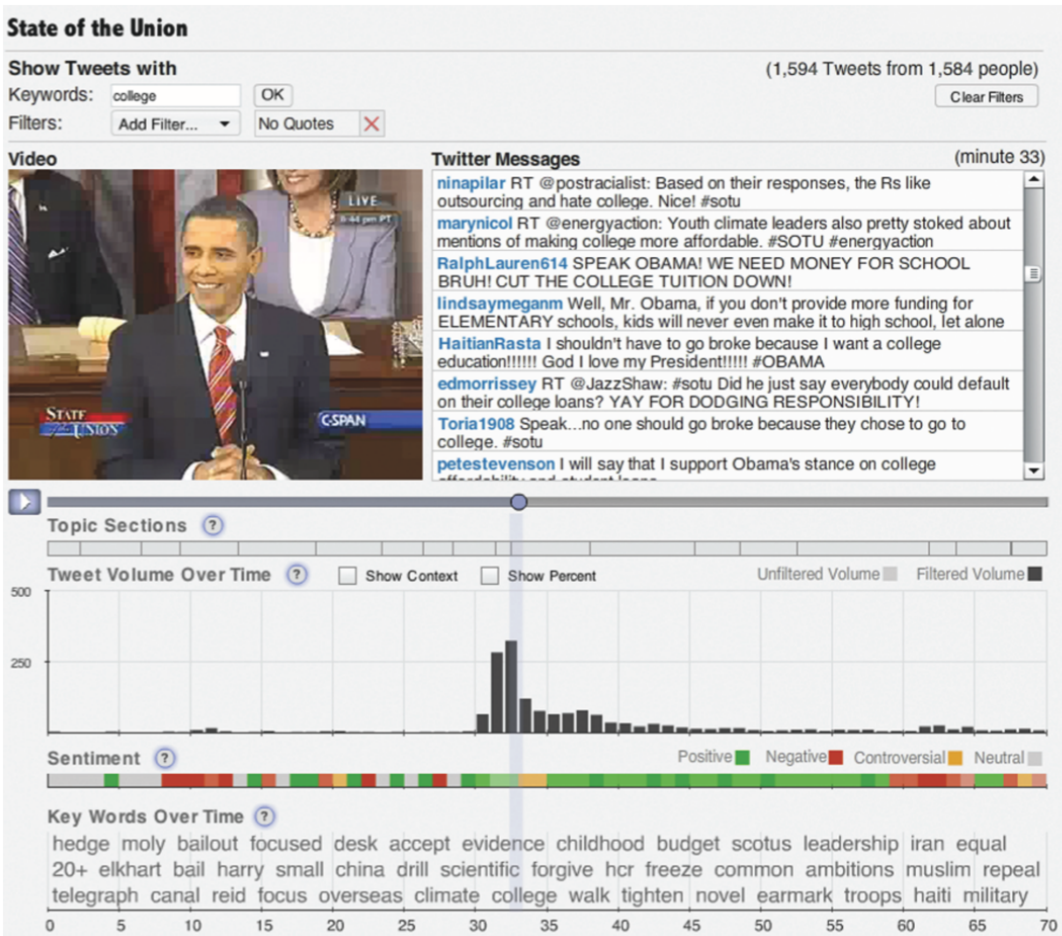


Figure 18: Journalistic layout and design for Twitter data (Diakopoulos et al., 2010)

Figure 19: From (Blei et al., 2003), applied LDA and spatial data via Twitter data

DTNA would then produce networks based on: 1) username → username; 2) username → hashtag; 3) hashtag → hashtag; and 4) username → URL. An example is listed in Table 1.

Once keyword(s) and area of interest are specified, the user is provided the D3.js (Bostock, Ogievetsky, & Heer, 2011) force-directed network visualization, Google Map, and Twitter feed, all of which continually updates as long as the website remains active. The Twitter data is saved to a MySQL database, including the timestamps of tweets. The network itself has built-in aesthetic saliencies that suggests other users, hashtags, or keywords that may provide better insight into the original search topic.

The graph for the tweets is created every time a query is requested from Twitter. The network, when visualized, is limited to only 100 nodes to limit the amount of attention needed

Table 1: Sample Tweet and Networks

James tweets: "Just saw @Dannie at the #Pitt #Library. Checked out http://bit.ly/U8AGsc"	
Username → Username:	James → Dannie
Username → Hashtag:	James → Pitt and James → Library
Hashtag → Hashtag:	Pitt → Library
Username → URL:	James → http://bit.ly/U8AGsc

by the user and reduces the search space to limit the cognitive load required by the user. The nodes themselves are selectable and will augment the original search query to include or remove the selected hashtag, username, or keyword. This is a measure to help limit the amount of noise in a given graph. A clustering algorithm (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008) is applied to the network and groupings are assigned and represented by color assignment. Based on log-based degree centrality, the size of the node reflects the number of connections each username or hashtag accompanies. Considering Twitter is a directed network, edges reflect the direction of the message. Figure 20 highlights both the groupings and degree centrality, with a bias towards increased inner-cluster strength versus out of cluster nodes. The importance of these saliency changes is to highlight nodes of interest or key players, which is an important feature in dark networks (Everton, 2012).

A mechanism was developed to encapsulate various datasets together (called projects) into a single visualization, thus allowing single or multiple participants to collect data from various vantage points as a means of understanding an event from different perspectives. This can then be outputted as a .gexf, .net, or .dl network type. To visualize this closer to real-time, a stream was developed in a social network analysis application called Gephi (Bastian, Heymann, & Jacomy, 2009), and uses a Streaming Network add-on to view the network as it evolves. Figure 21 below contains only 5 minutes of data collection using the username $\rightarrow$ hashtag connection. It was a grouping of both Syria and FSA search terms in Homs, Syria.

Along with the main DTNA web-application, a suite of additional web applications is available to analyze collected datasets for further investigation. This includes visualizing entire projects with what is called TweetViewer. TweetViewer for DTNA promotes the temporal data associated with the network by providing a timeline to allow a user the ability to specify at which point in time they are interested in viewing the network, or the network's evolution over time. As DTNA's network visualization evolves, edges reflect the direction of the message. Additionally, nodes that originated from the tweet are differentiated by augmenting and animating the nodes' borders as time progresses. Lastly, users can search based on the tweet to showcase nodes that include certain text. Figure 22 shows a network built using the hashtag "#FSA" in Homs, Syria, and is highlighted by tweets that include

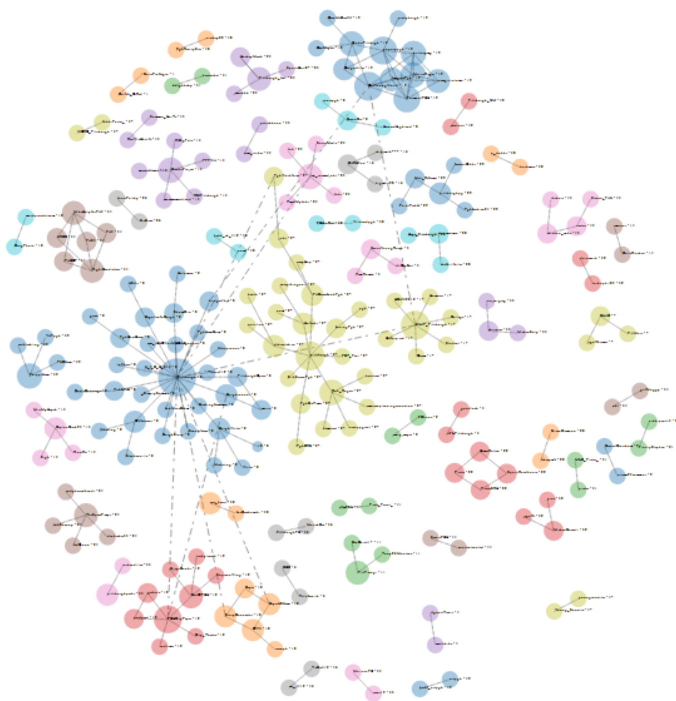


Figure 20: An example network visualization

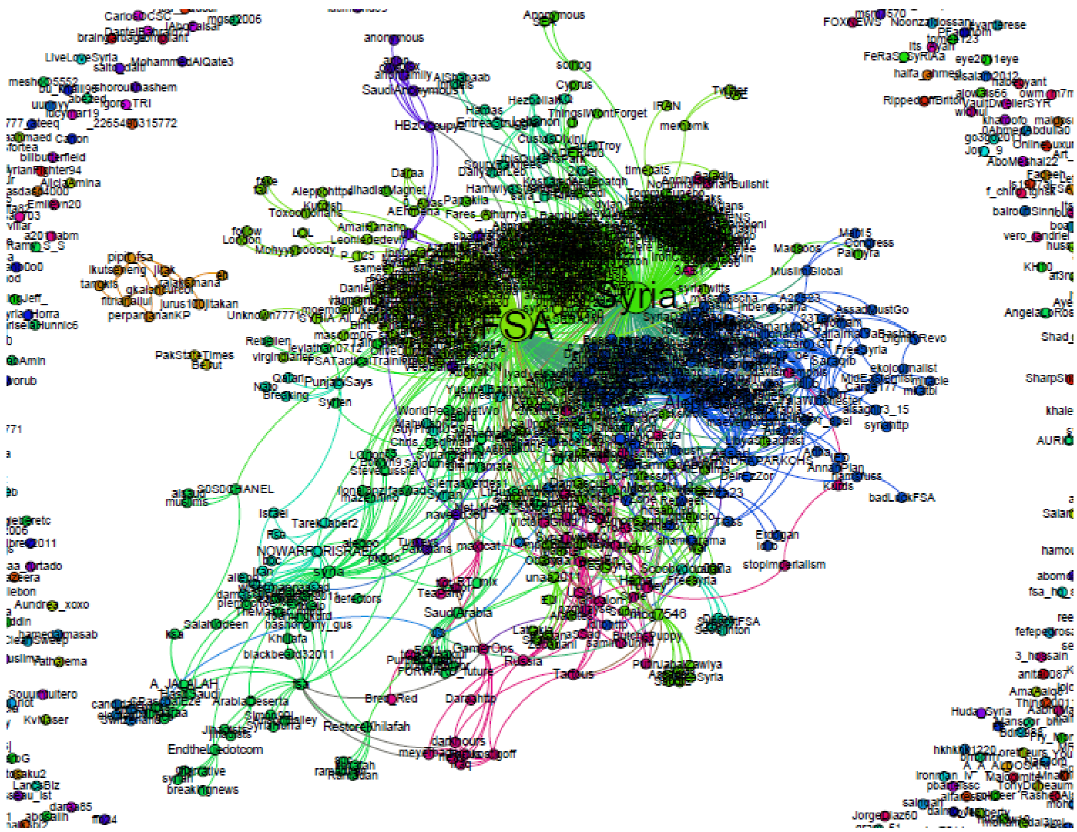


Figure 21: An example of streaming data from DTNA into Gephi (search terms: Syria and FSA)

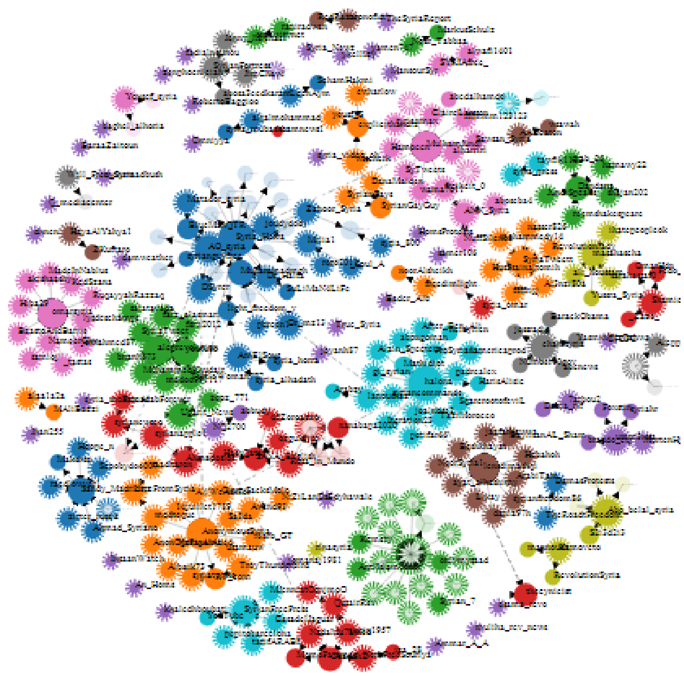


Figure 22: Searching TweetViewer for keyword "Syria"

the word "Syria."

Further analysis was put in place to extend the usability of these social media datasets to include work in sentiment analysis. This was implemented by using a sentiment lexicon (Nielsen, 2011), which includes word lists that are given a score from 5 to -5, based on their positive or negative connotations, respectively. This could be used to evaluate social movements or key terms and how these connections are influenced over time via sentiment lexicon. This does not have to be limited to strictly a global overview. It can be topic-based as well, allowing for searching of words or parts of words and connecting their positive or negative connections. The sentiment lexicon would be fixed around a circle, and the keywords would be placed in the center. Based on the number of occurrences of a keyword or phrase, in conjunction with a sentiment word, would determine the nodes' polarity towards the positive or negative side of the circle. Below are two instances of this analysis at two different timestamps, Figure 23.

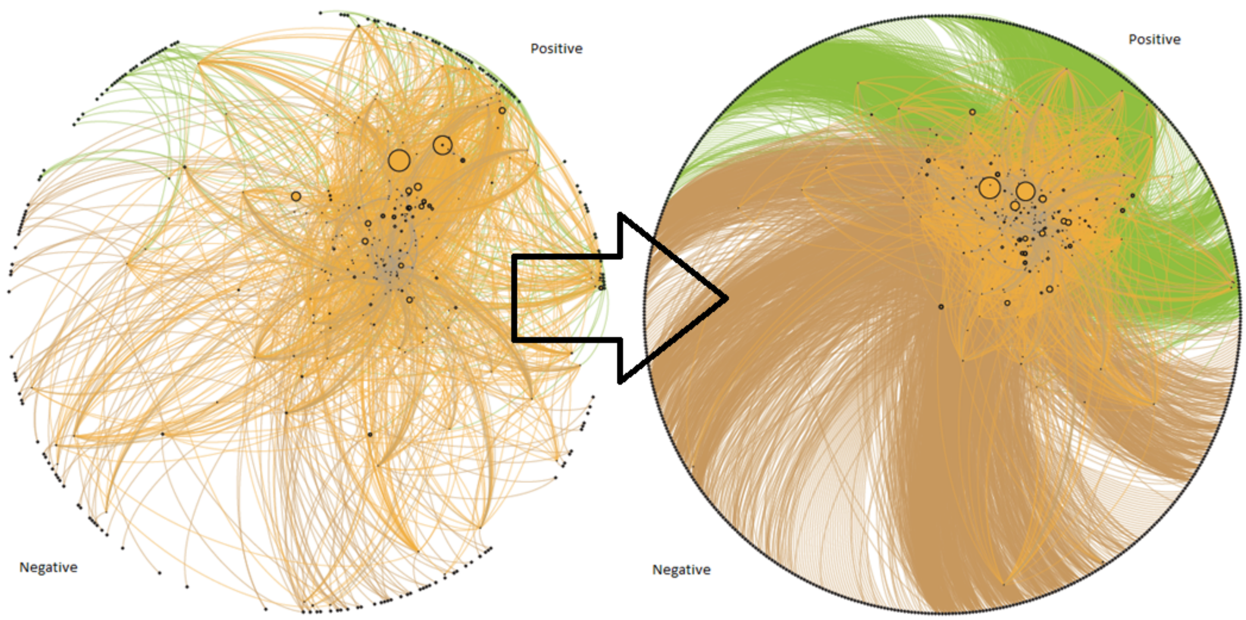


Figure 23: Sentiment visualization as the nodes shift from one side of the circle to the other side (from Negative to Positive)

1.5.2 Current Usage and Initial Findings

This project has been in development for three years, and at this point has collected roughly 1,325,534 tweets from 22 different countries including but not limited to Turkey, Iran, Syria, Thailand, Yemen, Somalia, Guatemala, Egypt, Philippines, Nigeria, Pakistan, and Russia using 42 different ISO 639-1 Code language codes. This geo-diverse data provides a plethora of user-content in a multitude of different locations, but the geographical granularity of this information is sparsely measurable. Figure 24 shows a Google Map with the positions of locations of interests for users of the system.

Provided are examples of the graphs being created through all projects within DTNA. Figure 25 shows a user → user network appended on a map based on geo-references. It contains 495,494 nodes and 577,289 edges. Figure 26 is a hashtag → hashtag search, which can be viewed more as taxonomy based on common words associated with searched keywords. This graph contains 33,301 nodes and 102,638 edges. Lastly, Figure 27 shows the same

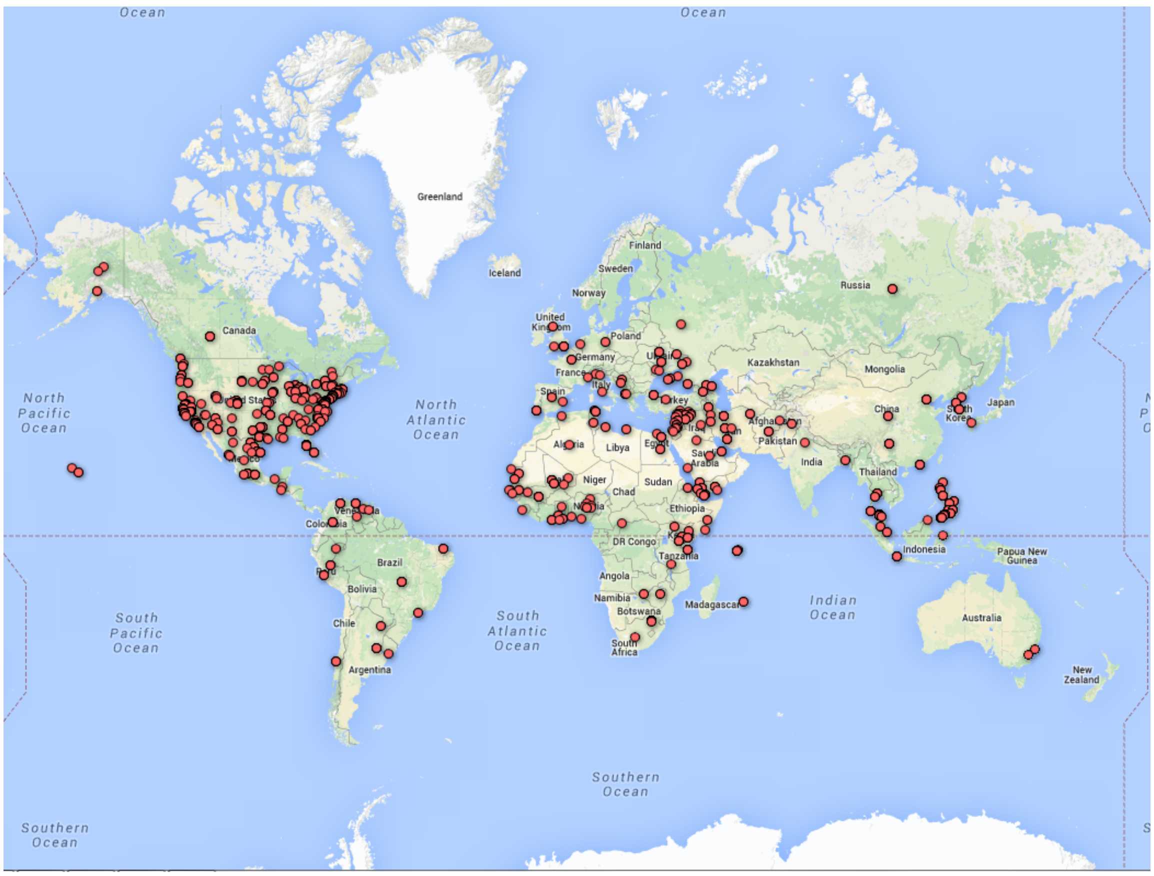


Figure 24: Map of DTNA data locations

network appended on a map showcasing within which proximity these networks are being searched.

Modus operando of the thesis work stems from the DTNA project, with the interest in connecting social media, specifically Twitter data, to a variety of networks in real-time. The focus of the study is on the connection between readability and usability of network graph structures, and then extends this paradigm to include added addition of visual aesthetic properties to increase the affordances of these graphs. This thesis (aided and informed by the Pilot Study) was developed to provide a better means of understanding the connection between the network graph and the application of visual aesthetics. The Pilot Study is provided in brief to inform the design of the main thesis.

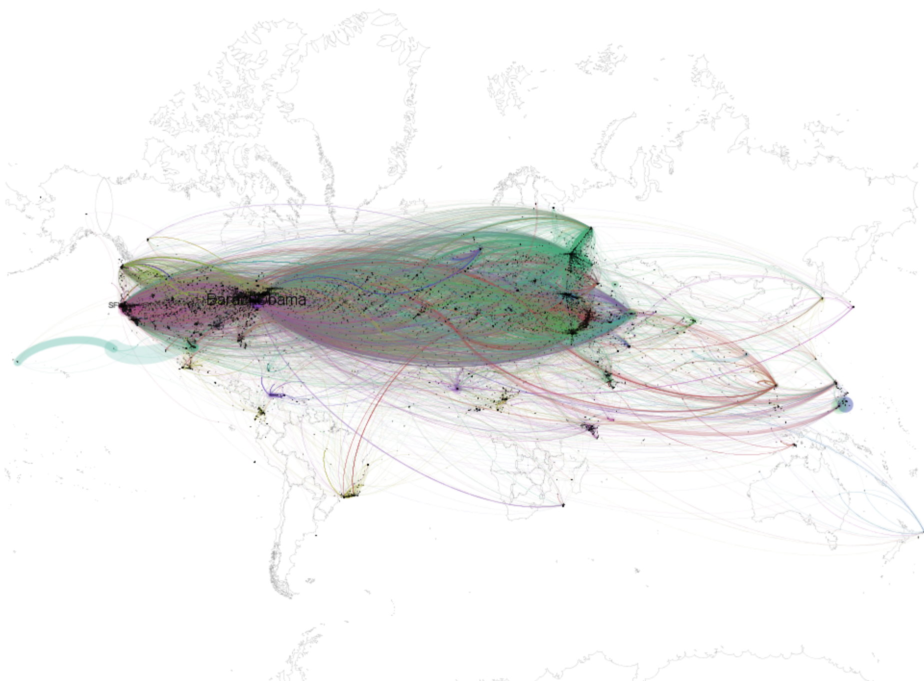


Figure 25: User → User connections mapped

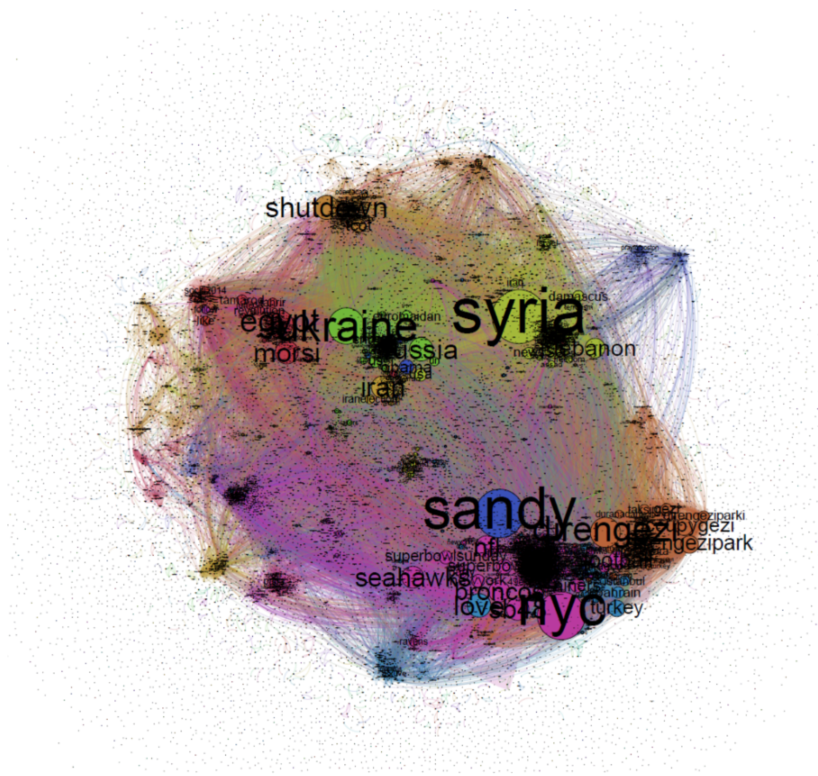


Figure 26: Hashtag → Hashtag network connections

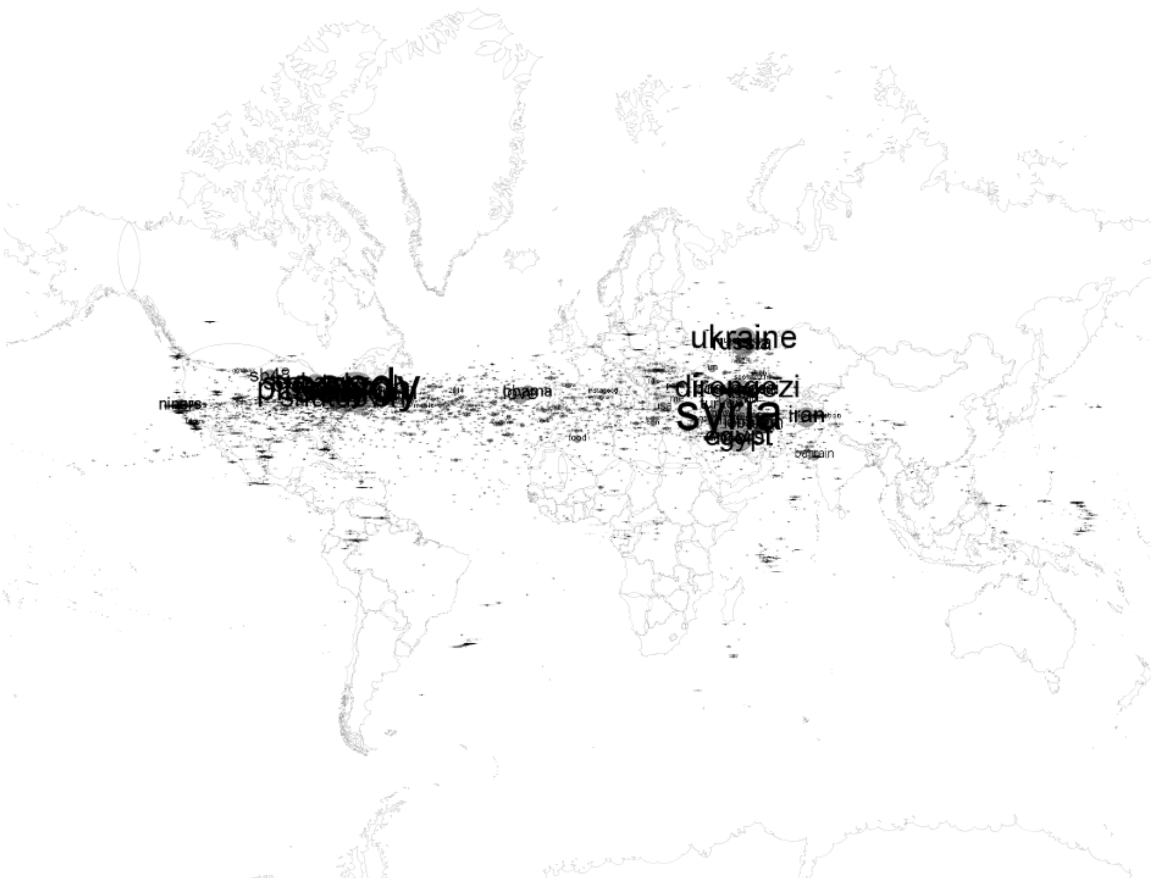


Figure 27: Hashtag → Hashtag connections mapped

2.0 PILOT STUDY

2.1 DEFINING THE VISUAL AESTHETICS

The visual aesthetics were derived from previous work and were broken up into two subcategories: visual encoding and visual styling. The visual encoding is a direct enhancement of the network graph by the underlying data from the network structure and can either enhance, such as degree centrality measure, or hinder, by using directional measure in a non-directional task. The visual styling of the graph is an aesthetic that is applied uniformly over the graph and not dependent on the graph structure or underlying topological structure and thus provides no encoded information. This styling may inherently provide encoding information, but this is only by coincidence and not a deliberate function of the designer.

2.1.1 Visual Aesthetics

2.1.1.1 Control (C) The control (Figure 28.1) for this experiment is very crucial to connecting and relating the topological structure of the network and the applied visual aesthetics. This is because the control does not contain the visual aesthetics and is the "natural" form of the network, where "natural" is understood as strictly providing readability in terms of the topological structure of the network with no enhancement based on the visual aesthetics. Finding the accuracy values, the eye-tracking movement, and the participant given perspective is countered by the control. Details on how the network were formed will be provided in the Experiment section, but understanding the role the control plays in this study was provided to prepare future conclusions and discussions.

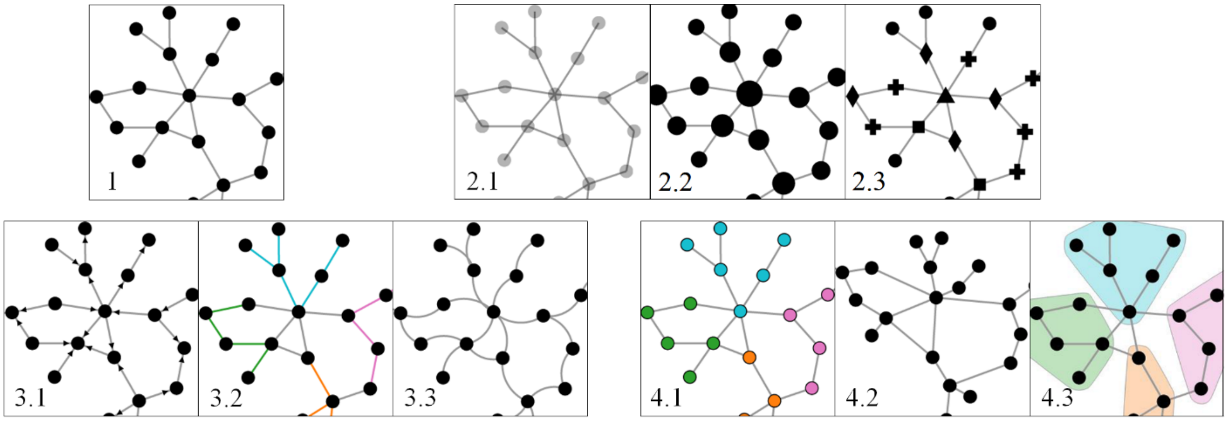


Figure 28: Shows an example network with the 10 aesthetics and their groupings: 1) Control (C); 2) Node Aesthetics (NA): 2.1) Opacity (NO), 2.2) Size (NS), 2.3) Glyphs (NG); 3) Link Aesthetics (LA): 3.1) Link Direction (LD), 3.2) Link Coloring (LCo), 3.3) Link Curves (LCu); and 4) Group Aesthetics (GA): 4.1) Group Coloring (GC), 4.2) Intra/Inter-Distance Bias (GDB), and 4.3) Convex Hulls (GCH)

2.1.1.2 Node Aesthetics (NA) The node aesthetics were designed to showcase both the visual encoding and visual styling within the graph. At this very local level, comparing the nodes would produce an optimal way to finding the most connected node(s) in the graph. Three visual aesthetics were designed based on Bertin's six classifications of these visual variables: size, value, texture, orientation, color, and shape (Bertin, 1983). Based on Bertin's work, they were then subdivided based on the problem set the task wanted to depict (finding the most connected node(s)), which were associative perception, selective perception, ordered perception, and quantitative perception. Size was used to highlight quantitative difference, varying based on radius (both degree and log-degree) and was classified as a visual encoding (Figure 28.2.2). Shape/Glyph provided an associative connection, where the use of a legend could allow participants to find like shapes that represented the most connected node, again a visual encoding (Figure 28.2.3). Finally, texture was used for both its selective and ordered quality to both distinguish itself from one another and provide a bottom-up ordering to most connected node(s). This was labeled a visual styling and was created by changing the opacity for each node to 0.3, showcasing the "texture" of each node based on their connections (Figure 28.2.1). Color, value, and orientation were omitted due to the redundancy in these four levels of perception.

2.1.1.3 Link Aesthetics (LA) The link aesthetics (Figure 28.3) were chosen based on how well they illustrate connectivity within the graph, both at the level of encoding and styling. Three separate visual aesthetics for links were depicted to show either: how information flows (Eades & Xuemin, 1989), common neighbors (Blondel et al., 2008), or reductions of clutters (using Lombardi drawing (Duncan, Eppstein, Goodrich, Kobourov, & Nöllenburg, 2011)). Considering the task, at its essence, is a link summation: how the links are exemplified may either help or hinder the user's ability to complete the task and for this, the main foci of this experiment will be on how the users traverse the graph to find the most connected node(s). Information flow (Figure 28.3.1), was designed to hinder the ability of the participant by visually encoding directionality in a non-directional task. Similarly to Rusu et al. (2011) and Jianu et al. (2009), common neighbors (Figure 28.3.2) and link curves (Figure 28.3.3) could be used to increase readability by decreasing the effects of edge crossing, where link coloring is a visual encoding constraint based on highly connected nodes and link curves is a visual styling based on Gestalt's continuation principle.

2.1.1.4 Group Aesthetics (GA) The group aesthetics (Figure 28.4) formed by these networks will all be tested using the same clustering algorithm, which is based on modularity optimization (Blondel et al., 2008). The reasoning for developing these visual aesthetics is that when grouping is involved with the task, the user will try to complete the task in a hierarchical manner, where a top-down approach can be utilized by users to make groups based on like items and see them as homogeneous groups. These groups, which are based on highly connected sets of nodes, can subdivide the problem set into more manageable cognitive chunks, especially when the graph becomes denser and less readable. These group visual aesthetics are based on Rock and Palmer (1990) and exemplified Gestalts grouping principles, where group coloring (Figure 28.4.1) is similar to group based on similarity, distance bias (Figure 28.4.2) is similar to proximity, and convex hull (Figure 28.4.3) is similar to group by

Table 2: Visual Aesthetics Breakdown

Visual Encoding (Task Related)	Visual Encoding (Not Task Related)	Visual Styling
Node Size (NS)	Link Direction (LD)	Node Opacity (NO)
Node Glyphs (NG)	Line Color (LCo)	Link Curve (LCu)
	Group Coloring (GC)	
	Group Intra/Inter-Distance Bias (GDB)	
	Group Convex hulls (GCH)	

closure. All of these were considered as a non-related visual encoding.

2.2 HYPOTHESIS

Based on the past literature in terms of network aesthetics, testing was first done to see if there was a significant difference between each topological structure of the network based on accuracy, as this would help in validating the initial design of the study.

- H1a: Under the same condition of network size and density, different network topological structuring will lead to significant differences in terms of accuracy (F1-Score), regardless of visual aesthetics and layout.

As per Xu's work (Xu et al., 2012), it is believed that the networks themselves will cause accuracy variation, and as the number of nodes increases (from 40 to 80), accuracy values will suffer.

- H1b: Increasing the number of nodes (from 40 to 80) will have a significant effect on the accuracy (F1-Score) of the task.

Considering the most highly cited issue with network graph readability typically results from the number of edge crossings (Purchase, 1997), that within each of these network sizes (40 nodes and 80 nodes), ones with more edge crossings will result in poorer accuracy values.

- H1c: Graphs that include more edge-crossings will have significant effects on the accuracy (F1-Score) of the given task.

Considering that task dependent visual encoding (NS, NG) should have a better means of overcoming these adverse conditions, such as poor readability of the topological structure of the network, they should outperform the other conditions. More specifically, based on Mackinlay's work (Mackinlay, 1986), NS should outperform NG. For the visual styling, the node based styling (NO) should be more superior than the link based styling (LCu). The worst conditions should be the visual encoding unrelated to the task, as in LD, LCo, GC, GDB, and GCH. This will be based on both accuracy and task completion time.

- H2a: Compared to the control, node based visual encoding (NS, NG) will have higher performance (accuracy and task completion time) than the control.
- H2a1: Specifically, node size (NS) will have higher performance (accuracy and task completion time) than node glyphs (NG).
- H2b: Compared to the control, visual encoding that is not task related (LD, LCo, GC, GDB, and GCH) will have lower accuracy levels than the control.
- H2c: Visual styling (NO and LCu) will not have a significant difference on the performance (accuracy and task completion time) of the task as the control.
- H2c1: Specifically, node opacity (NO) will have higher performance (accuracy and task completion time) than link curves (LCu), as the node styling is localized at task level.

Because the task and visual aesthetics can be linked when the visual encoding methods are applied, they will be seen as more visually advantageous to complete the task and thus have a higher utility ranking based on the qualitative information provided by the participants. This is based on the question, "The aesthetic helped in completing the task."

- H3a: Compared to the control, node based visual encoding (NS, NG) will be seen as more aesthetically positive (in terms of utility) in completing the task.
- H3b: Compared to the control, visual encoding that is not task related (LD, LCo, GC, GDB, and GCH) will be seen as less aesthetically positive (in terms of utility) in completing the task.

And specifically for the visual styling, based on Xu et al. (2012), a higher level of subjective visual approval should be seen when styling is applied.

- H3c: Compared to the control, visual styling (NO and LCu) will be seen as more aesthetically positive (in terms of utility) in completing the task.

Based on this node related task, the node aesthetics (NO, NG, and NS) should have a higher perceived utility than the link and group aesthetics.

- H3d: Compared to the control, node aesthetics (NO, NG, and NS) will be seen as more aesthetically positive (in terms of utility) in completing the task.

With the eye-tracking data, that variation of levels of aesthetics (nodes, links, and groups) will change the eye-movements of the participants. For node aesthetics (NA), based on this search task, NA will have more average saccade length and larger than average saccade length. Link aesthetics (LA) will confine eye movements to the 'skeleton' of the network and average saccade lengths will be more condensed. Group aesthetics (GA) will have the smallest average saccade lengths because they will be seen as sub-domain problems and analyzed as such, creating less eye-movement.

- RQ1: How will different forms of the aesthetic affect eye-movement? Specifically, can there be different eye-movement paths with different aesthetics (Visual Node Aesthetics (NA - NS, NG, NO) > Visual Link Aesthetics (LA - LD, LCo, LCu) > Visual Group Aesthetics (GA - GC, GDB, GCH)) in terms of average saccade lengths?

Also, tested is the idea that visual aesthetics can affect task focusing time (inverse task abandonment) (Cawthon & Moere, 2007), where task focusing time is defined as the amount of time the participant spends focused on the network to complete the task. This starts when the participant first views the network and ends when multiple saccades (10 consecutive) are outside the buffer area (60 pixels) of the network. Therefore, higher values of task-focusing time relates to higher levels of consistent focus on the network to complete the task. Based on this, it is believed that visual aesthetics that have a higher level of performance (accuracy and task completion time) will have a larger amount of task focusing time, as in they will focus more on the task based on their higher levels of performance. Also, visual aesthetics that have higher perceived utility will have a larger amount of task focusing time.

- H4a: Based on the conclusions of the H2 hypothesizes, visual aesthetics that have a significantly higher accuracy and task completion time will have a larger task focusing time.
- H4b: Based on the conclusions of the H3 hypothesizes, visual aesthetics that have a significantly higher perceived utility will have a larger task focusing time.

2.3 EXPERIMENT DESIGN

2.3.1 Graph Creation

Considering a need to map the 11 visual aesthetics to randomly created networks and, more importantly, determine the best means of testing out the bias of the topological structure of the networks themselves using a randomized blocking design, each visual aesthetic should have a corresponding network. Therefore, 11 randomly generated networks were created using the Erdös-Rènyi graph model (Erdös & Rényi, 1959; Gilbert, 1959). Erdös-Rènyi's model uses a probabilistic model to uniformly choose a random graph of size. Two sets were created to test a small network (n_s) and a large network (n_l). This was specified for vertices of 40 and probability .04 for edges ($n_s = 11\ graphs$) and vertices of 80 nodes and probability .0225 ($n_l = 11\ graphs$).

All 22 graphs utilized this same force-directed graph algorithm, and the final relaxed graph drawing was stored and presented in its static form for each participant.

Matching 1-to-1 for each visual aesthetic to each small and large network would still bias results based on the topological structure of the network. As mentioned, network readability can be affected by the number of nodes, number of edges, link density, edge-crossing, orthogonality, among other topological structures and heuristics. The visual aesthetics would need to be cycled through the networks, so if there was a bias it could be stated that this bias (positive or negative) would affect all the visual aesthetics evenly. Therefore, a randomized

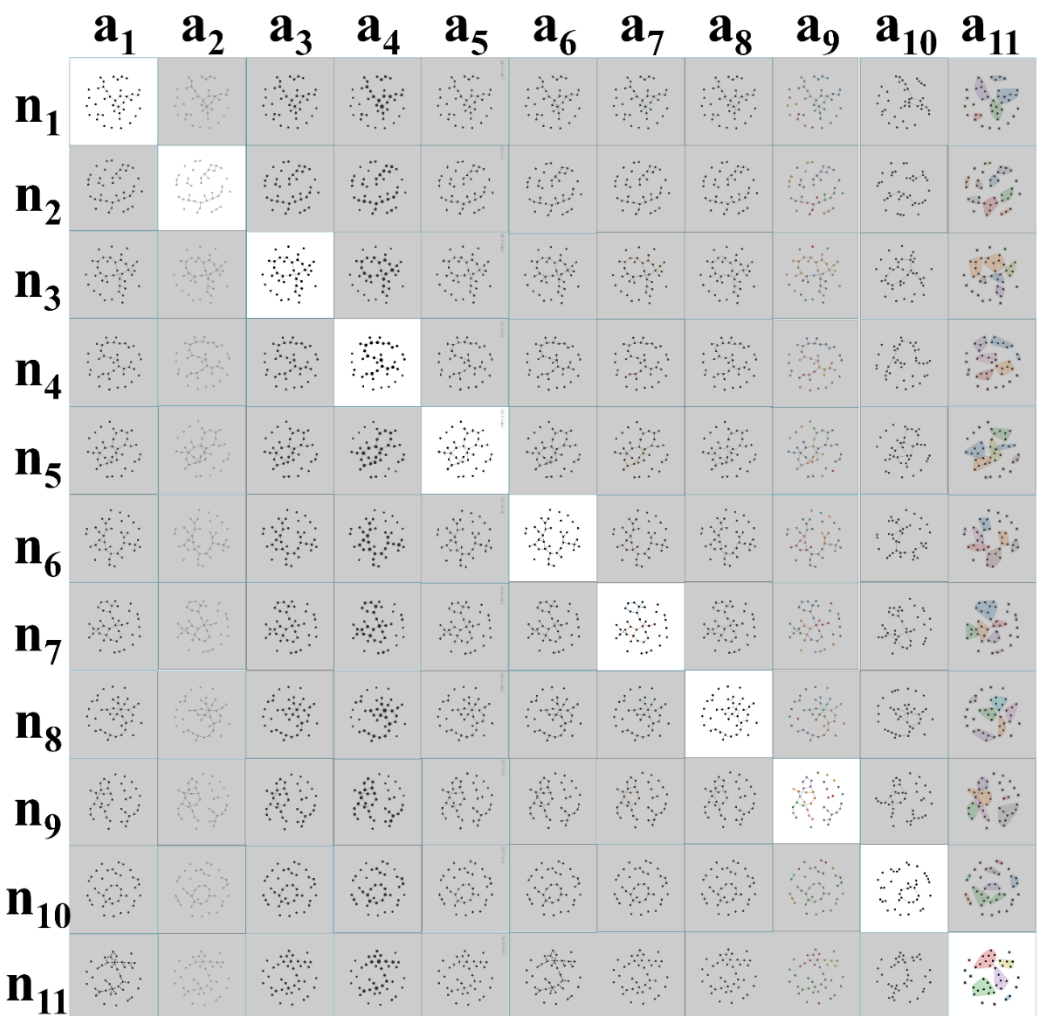


Figure 29: Visual aesthetics and corresponding small network (n_s) matrix. Highlight (area in white) showcases the trials for P_1

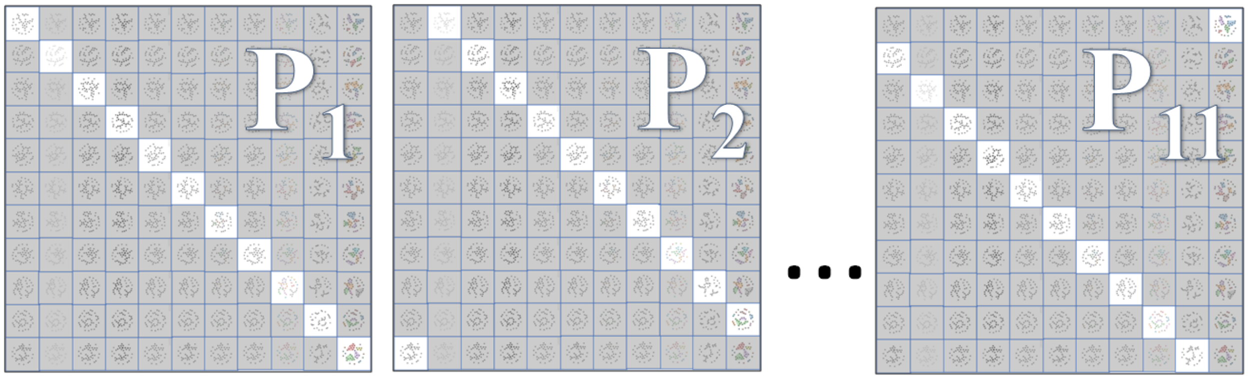


Figure 30: Shows highlighted areas for each participant (specifically P_1, P_2, and P_{11})

blocking design was used to mask the bias caused by the topological structures of the networks. Figure 29 shows the final blocking representation for the small networks utilized for testing, and would be the same design for the large network as well. Please note, the networks in Figure 29 are not shown in detail. To see the network to visual aesthetic relationship, see Figure 28, where a_1 to the control, a_2 to node opacity, and so on. To cycle participants through this blocking design, networks were selected along the diagonal of the matrix for each participant. This would match each visual aesthetic to a unique network. Figure 29 shows the blocking design with the highlighted area for participant 1 (P_1). Extending this out for all participants, Figure 30 shows the method participants were cycled through the testing design for participant 1 (P_1), participant 2 (P_2), and continuing to participant 11 (P_{11}).

As stated, with the utilization of eye-tracking, having a means to verify the results was imperative to validating the data collected from the eye-tracking software. With this in mind though, repeating the same network multiple times could allow for a learning bias to form based on this reoccurring network. Taking these two issues into account, it was decided to rotate the graphs by 0°, 90°, 180° and 270° degrees. This would create the redundancy needed and change the graph enough to remove the learning bias[1].

[1] In the Pilot Study, it was believed that the rotation of the network by 0°, 90°, 180° and 270° would not create a learning bias, but this assumption was not specifically tested. In the Thesis Experiment, this assumption was tested, and no significance was found in the rotation that would lead to a learning bias.

Therefore, the final design would then need to include:

$$\begin{array}{c} (11\ small\ networks + 11\ large\ networks) \\ x\ 4\ rotations\ (0^{\circ}, 90^{\circ}, 180^{\circ}, 270^{\circ}) \\ \hline 88\ total\ networks \end{array}$$

Out of these 88 networks, random selection was used to counterbalance each network.

2.3.2 Participants

Recruitment was handled locally on the University of Pittsburgh campus and required users to have 20/20 vision without the use of glasses or corrective lenses, as this would greatly hurt the ability to track eye-movement using the equipment and software (Tobii 1750 monitor and Clearview 2.7.1 software). The data was collected on a Dell OptiPlex 755 with 3326MB RAM, Intel(R) Core(TM)2 Duo CPU E8200 @ 2.66GHz (2 CPUs), and a ATI Radeon HD 2400 Pro. Data was collected from 15 University of Pittsburgh students (7 male, 8 female), between the ages of 18-29 years old, noting that one participant abstained from providing their date-of-birth, and in terms of experience with working with network graphs 9 had no experience, 2 had very limited or minimal experience, and 4 had experience. The participants were compensated for their time ($10/hr) and the average time to complete the experiment was 1 hour and 12 minutes.

Once the participants arrived, they were told about the experiment and its overall goal, which included signing a consent form with this information. For the first part of the experiment, they were asked to complete a task to establish reaction time to a static object on a screen of fluctuating size and color. This was done to get an initial baseline for the individual based on their reaction time, considering the objective of the task was the implementation in the visual aesthetics' and their "first impressions." This was a pre-test using Fitts' law experiment (Fitts, 1954) to calculate an index of performance for each participant. Testing these results for significance to remove participants due to their differential in times, no significance was found.

2.3.2.1 Breaking Mental Models Considering the majority of our participants would have limited knowledge of network graphs, and was the main focus of the experiment was connecting reactionary, subjective visual aesthetics properties and usability, a brief tutorial section of the experiment was used to explain very elementary foundation concepts in regards to network graphs and highlighting the degree centrality, as this task needed to be completed. The expectation for this was a consistent mental model of the graph structures and that knowing at least a fundamental level of information could increase the connection between the visual aesthetic and the usability.

As stated, participants with no or limited understanding of network graphs could attach previous intuition to the graph that were invalid based on the situation and their mental models. As two participants mentioned throughout the task, "It reminds me of connect-the-dots," this statement caused concern based on the mental models of the participants if applied to this task, which would greatly hinder the usability of the graph structure. Mental models can be very powerful tools in understanding user behavior, if understood prior to the experiment (Johnson-Laird, 1983; Johnson-Laird, Legrenzi, Girotto, Legrenzi, & Caverni, 1999), which was the case for this study. As shown by Kempton (1986), when participants were asked how thermostats worked or functioned, they discovered two distinct groups of user: 1) thermostats were analogous to car accelerators and heating/cooling a room would be done by over compensating by increasing or decreasing the heat/cold to "accelerate" this change; and 2) thermostats were analogous to switches where the thermostat was set, and if the room dropped below this temperature, the heat would turn on and vice-versa. Both cases, based on their mental models, would change the usability of the thermostat.

In the case of the car accelerators group, they would continually change the temperature throughout the day to compensate for heat or cold. In the case of the switches group, they only needed to change once or twice a day if necessary. Shown in this example, the usability of an object is dependent on our mental models of this object. If the participants connected the graph visualization to their mental models of "connect-the-dots," they could assume that all vertices should have a sequential ordering to one another and in extreme cases, when an image was not seen in the graph structure (which is seen in "connect-the-dots"), participants may linger on the self-imposing "connect-the-dots" image instead of the task asked of them.

This brief tutorial section ranged between 5-7 minutes, after which the participants were presented with a pre-survey that asked general questions about the user and his or her experience with network graphs, and included 2 example networks with 15 questions (6 for example one and 9 for example two) in regards to these elementary network foundation concepts. They were allowed to ask questions (no one required this) to help clarify points in these foundation concepts and if they were not sure of an answer, allowed to mark "N/A" and continue on (only 1 person put this for 1 question), and reviewing the results, only 4 total questions were missed (1 question by 4 participants) based on the entire 240 allotted questions (15 questions and 16 participants), producing a 98.33% accuracy level. The participants understood these elementary foundation concepts, thus increasing the network graph usability by having this foundation a priori.

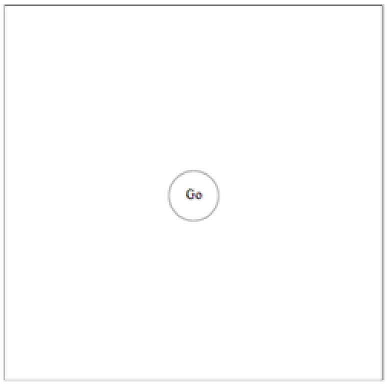
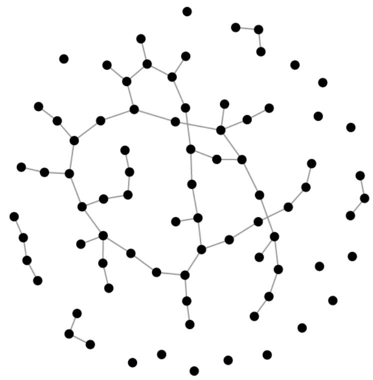
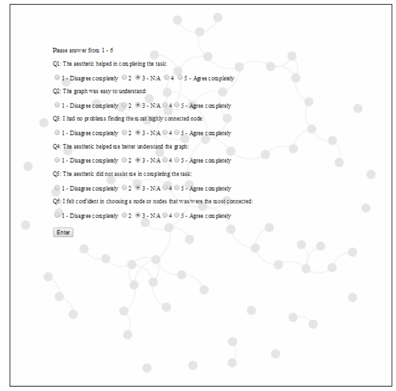
The subjects were then shown an example of all visual aesthetics on a single example network so they would understand their meaning throughout the study. It was stressed to the participants that even though the aesthetic may change, they were still asked to find the most connected node(s) in the network and this was consistent throughout. They were then shown a demonstration of the software which explained all three screens utilized throughout the study, and after a few demonstrations, they were allowed to try out the software on their own until they felt comfortable using it. Typically, the participants found the interface intuitive and would only require 2-3 minutes to feel comfortable with the software.

2.3.3 Interface Design

The main interface of the study utilized three screens, as shown in Table 3, including an initial countdown screen, the network itself, and an evaluation page afterwards. During the experiment, the screen was maximized to remove distracters from the analysis.

The initial countdown screen (Screen 1 from Table 3) was designed to draw the user's attention to the center of the screen. A small button was designed, which takes up .049% of the total allocated screen space for the experiment that would count down from 3 to 0 (or "Go") with a very slight animation change as the countdown would diminish. During the countdown phase, this button was un-clickable, but once it reached 0 or "Go," the participant

Table 3: The 3 screens that make up the interface.

		
Screen 1 (Warm-up Screen) Highlighting the starting button to move to the next network. The button and the word "Go" were enlarged only for clarity.	Screen 2 (Network Screen) The network and the screen where the user could interact with the network.	Screen 3 (Follow-up Screen) The questionnaire that would follow each network with a transparency to view the past network for better clarity in their response.

was then able to select it to move to the next network. Pohl (2009) did something similar to this by using a small cross in the center of the screen to set the focus of their experiment. This was expanded by including both the animation and selection (clicking the button) to focus gaze to the center of the screen and to have all participants start with the same eye-gaze location proximity. The intent with creating this type of button was to allow freedom in terms of moving to the next network shown, which allowed for breaks if necessary.

After selecting "Go," the participant was shown a static network graph (Screen 2 from Table 3). Considering the study looked at the "first-impression" of the visual aesthetics, once the network was viewed, the user was given 10 seconds to react to the network and select the most connected node(s). When interacting, to help with affordance and feedback, the outline of the node was changed from black to a dark blue coloring to show that the node had been selected. This would not indicate if they were correct or incorrect, just that they had selected a node and henceforth unable to deselect it. The black to dark blue change was only slight, so as to not distract but still indicate a "selected" node. The network interaction screen(s) was utilized in the accuracy, time, and eye-tracking portion of evaluation. Finally, after each network was shown, an evaluation screen (Screen 3 from Table 3) would be overlaid on top of the network with a 20% transparency on the background. This would allow the participant the ability to still read the wording of the questionnaire, but also provide them the network they were just shown to help with evaluating said network. The questionnaire was utilized to obtain the subjectivity measures in the final evaluation and was evaluated using a Likert scale from 1 to 5 (1 indicating complete disagreement and 5 indicating complete agreement to the question). The questions were as follows:

Q1: The aesthetic helped in completing the task.

Q2: The graph was easy to understand.

Q3: I had no problems finding the most highly connected node.

Q4: The aesthetic helped me better understand the graph.

Q5: The aesthetic did not assist me in completing the task.

Q6: I felt confident in choosing a node or nodes that was/were the most connected.

2.3.4 Accuracy Measure

In terms of accuracy measure, there was a need to encapsulate a means of understanding the relationship of the task (finding the most connected node or nodes) and various visual aesthetics. This brought up the need to balance both situations where a node was or was not the most connected, and the participant's ability to select or reject based on the signal and noise presented to them.

A signal-detection accuracy measure was utilized because it gave equal weighting to all interaction (Wickens et al., 2004). An F1-Score was used to better quantify this relationship, which was first introduced in terms of information retrieval systems and then extended to pattern recognition. This is a measure that combines both the precision (or positive predictive value, PPV) and the recall (or true positive rate, TPR).

In other words, in a selection task (as in selecting the most connected node) PPV would be equivalent to the number of correctly selected nodes that were the most connected, over the total number of selection, or the precision. The TPR would be equivalent to the number of correctly selected nodes that were the most connected, over the total number of conditional positive values. The harmonic mean of these two values was designated as the F1-Score or F-measure. The conditions for the F1-Score went as followed:

- T_p= subject selecting a node and correctly identified the most connected node(s)
- T_n= subject not selecting a node and correctly rejected the node as not the most connected node(s)
- F_p= subject selecting a node that was not the most connected node(s)
- F_n= subject not selecting a node and the node was the most connected

2.4 RESULTS

In terms of the analysis done on the data collected from the experiment, it was sectioned into three topics: 1) analysis of the topological structure of the network; 2) the analysis done strictly with interactions and how the visual aesthetics are both judged subjectively and in terms of their usability in completing the task; and 3) using the eye tracking software to determine average saccade lengths and task focusing time.

2.4.1 Topological

As specified, blocking design alternated the networks to visual aesthetics per participant to test this potentially confounding variable (the topological structure of the network). To showcase this necessity and to highlight current literature in terms of the stress of network formation algorithms versus the actual use of visual aesthetics, the topological structure of the network was tested to see if there were significant differences between the networks based on accuracy (F1-Score).

Using ANOVA, the F1-score was first validated between the two conditions of small and large networks ($p < .01$). Using ANOVA/Tukey Post-test ($p < .01$), it was determined that for both the small (1 network) and large (3 networks) networks there were significant values based on the F1-Score (Figure 31). This result showcases that the topological structure of the network itself provides its own initial bias and validates H1a. This was a condition

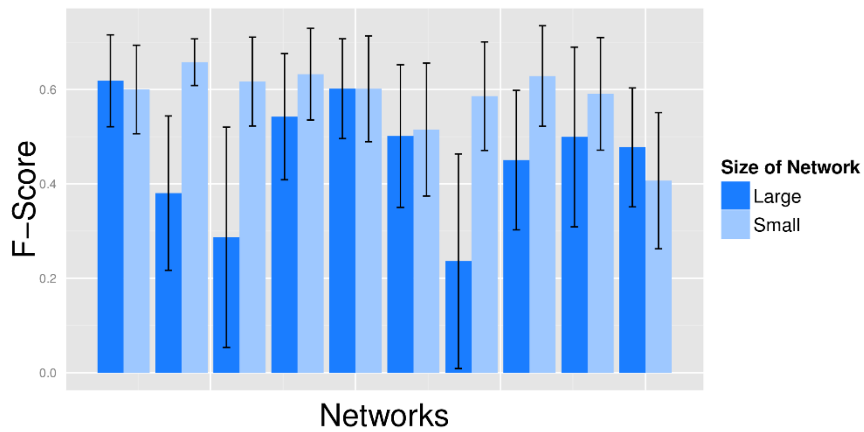


Figure 31: The results of network topological structure performance (based on accuracy) for both small (40 nodes) and large networks (80 nodes) and shows significance for small network 10 (p<.01) and large network 2, 3, and 7 (p<.01)

deemed important based on previous literature, but again was taken into account prior to testing to remove this bias.

To validate H1b and H1c, ANOVA with Tukey Post-test ($p < .001$) was used based on the number of nodes and the ratio of edge-crossing to number of edges, as both conditions could negatively affect readability of the networks and hinder the participants' ability to find the most connected node(s). It was found that the number of nodes did play a significant role in the accuracy measure (validating H1b - $p < .001$), but edge-crossing did not (not validating H1c).

Table 4 shows the F1-Scores, the number of nodes, the number of edges, and the edge crossing ratio (number of edge-crosses to number of edges) (ordered by F1-Score).

2.4.2 Accuracy vs. Time Measure

Evaluating the H2 hypotheses, the F1-score measure ran across the small network (n_s), large network (n_l), and then both networks. As mentioned, based on the time limit given to the participant, there was a potential for selecting the most connected node(s) and non-most

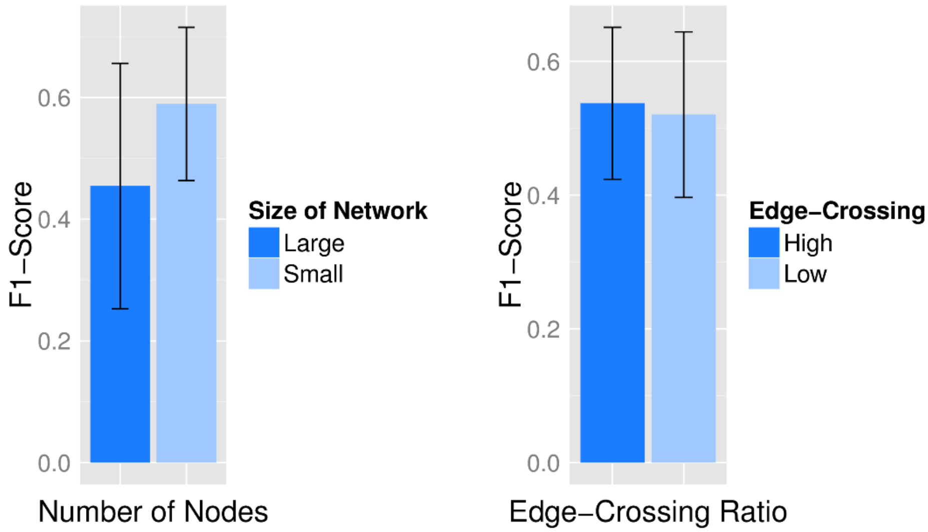


Figure 32: The results of network topological structure performance (based on accuracy - left - $p < .001$) on the small (40 nodes) and large networks (80 nodes) and (right) the ratio of edge-crossings (edge-crosses/edges).

Table 4: F1-Score, Number of Nodes, Number of Edges, and Edge-Crossing Ratio (Number of Edge-Crosses/Number of Edges) (ordered by F1-Score, only showing the ten lowest F1-Score values)

F1-Score	Number of Nodes	Numbers of Edges	Edge-Crossing Ratio
0.236	80	79	1.975
0.287	80	62	1.484
0.38	80	64	1.438
0.407	40	30	1.067
0.45	80	74	1.743
0.477	80	53	1.208
0.499	80	84	2.381
0.501	80	63	1.651
0.515	40	33	1.545
0.542	80	79	2.241

connected node(s), with the equal potential to miss the most connected node(s) and reject non-most connected node(s). Significance was found using ANOVA followed by Tukey HSD pair-wise. This is reported using $F(a, b) = c$, where a is the degrees of freedom, b the residuals, and c the F-value, similar in Netzel et al. (2014). In the case of accuracy (F1-Score), the small network (n_s) was not significant, but both the large network ($p < .0001$ - $F(10, 396) = 4.795$) and all networks ($p < .0001$ - $F(10, 763) = 3.179$) were significant. The Tukey HSD pair-wise for the large network showed significance ($p < .01$) NG-GCH, NG-GC, NG-LD, and NS-GCH, and for all networks ($p < .01$) NG-GCH and NG-GC. Overall, there was significant differences in the large and combined networks, but based on pair-wise comparison, this did not prove H2a as proposed in the hypothesis.

For our time variation, the 'first correct selection' was used, as it was possible to not select all of the most connected node(s). If the participant did not select any of the most connected nodes, that trial was given a max time of (10000 ms). Significance was found using ANOVA for small networks ($p < .001$ - $F(10, 424) = 2.798$), large networks ($p < .000\ 1$ - $F(10, 426) = 7.976$) and all networks ($p < .0001$ - $F(10, 860) = 7.817$). This was followed by Tukey HSD pair-wise for small network ($p < .05$) NS-GDB and NS-LCu, large network ($p < .01$) NS-all conditions except (NG - $p < 0.0832194$) and ($p < .05$) NG-GDB, NG-GCH, NG-GC, NG-Co, NG-LD, NG-LCo, and NG-LCu and all network ($p < .01$) NS-all conditions except (NG - $p < 0.069$) and ($p < .05$) NG-GDB, NG-GCH, NG-GC, NG-Co, NG-LD, and NG-LCu. This confirmed at least one visual encoding as significantly more optimal for finding the most connected node versus the control (C) in the shortest amount time. This also confirmed the H2a for time that the task related visual encoding outperformed the other aesthetics. H2a1 was not confirmed, as both NG and NS had a non-significant difference in both time and accuracy.

Even though H2b was shown as a non-significant result based on accuracy and time, Figure 33 provides insight into this hypothesis. Referring back to the top left-hand portion of Figure 33, low accuracy and high first selection time shows the visual encoding conditions (non-related to task) (LD, LCo, GC, GDB, and GCH) as a poor performance to the other conditions. Centered in the figures were both visual styling conditions (NO and LCu), showing little variation to the control. Finally, two of the conditions (NG, and NS), where

NS showed significance, were featured in the bottom-right section of the figures, illustrating a strong connection between visual encoding (related to task) and task completion.

Finally, for H2c there were similar results between the control and visual styling and based on no significance, prove H2c. H2c1 was not confirmed, based on no significance, but there was higher accuracy and lower time for LCu than NO. This is interesting based on the typically poor performance with curved lines (LCu).

2.4.3 Participant Feedback

In terms of statistical analysis, considering non-normalized Likert values were used to collect this measure, Kruskal-Wallis one-way ANOVA was used to find significance and Siegel and Castellan post-hoc for Kruskal-Wallis (McCrum-Gardner, 2008; Siegel & Castellan, 1988). All of the question responses were found to be significant in terms of visual aesthetics being utilized ($p < .01$). This was confirmed for both the small (40 nodes) and large (80 nodes), joined and separated.

For subjectivity measure, the pair-wise comparison was utilized, using two-tailed comparison between our control (no visual aesthetic) and the other visual aesthetics. Also compared were both the small and large networks separately and combined. Significance was found for these subjective measures ($p < .01$) and looking at the pair-wise/two-tailed post-hoc analysis, node size (NS), glyphs (NG), and opacity (NO) all were significant compared to all other conditions ($p < .01$); showcasing that the node visual aesthetics, were superior in terms of subjectivity when accomplishing degree centrality calculations, confirming H3a. No significance for the non-related visual encoding and visual styling was found, disproving H3b and H3c.

Figure 34 shows the results for one of the questions ("The aesthetic helped in completing the task.") between the small and large networks. This hypothesis was validated by comparing results for the predefined groups (nodes, links, and groups), which showed significance ($p < .01$) for the comparison between the control and the node aesthetics (NO, NG, and NS), as seen in Figure 34.

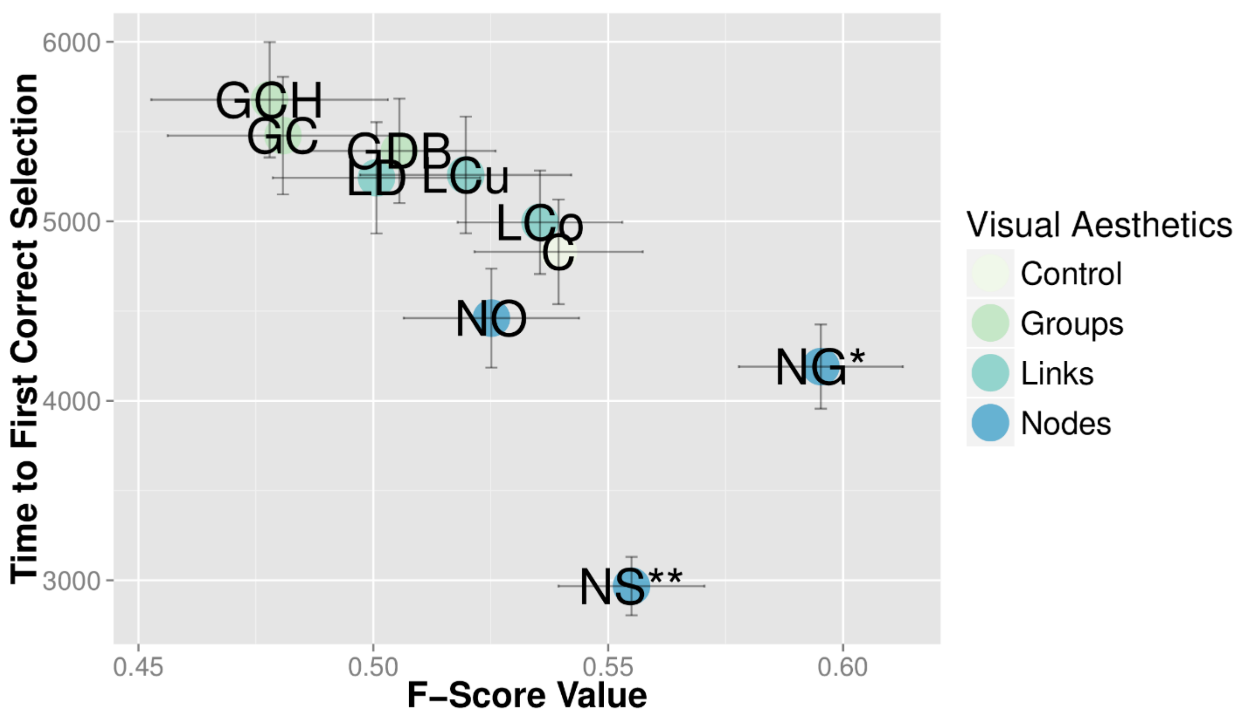


Figure 33: Results of both networks (n_s and n_l) in terms of F1-Score and time for first correct selection. Error bars represent standard error.

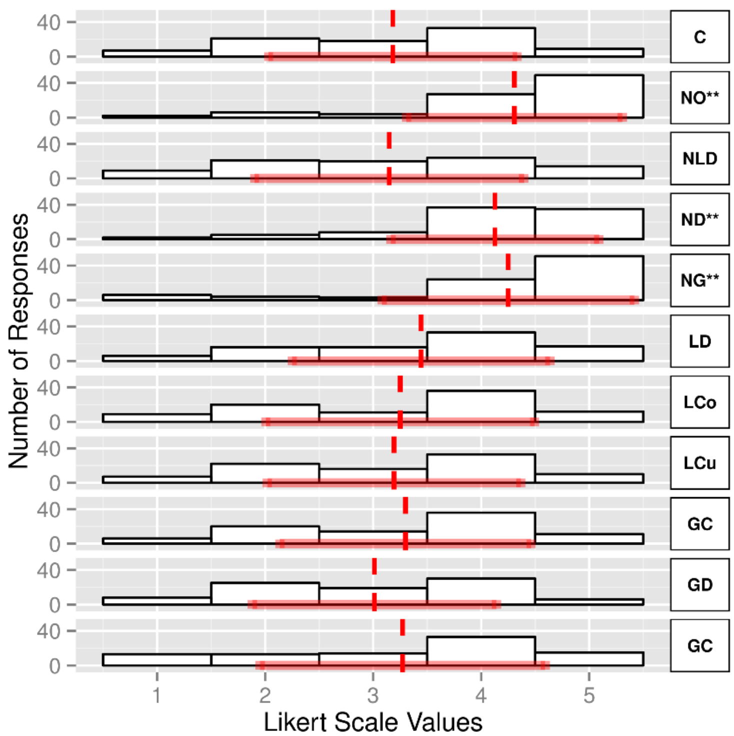


Figure 34: Response to "The aesthetic helped in completing the task." This is for both the small network (40 nodes) and the large networks (80 nodes). The vertical red line indicates the mean.

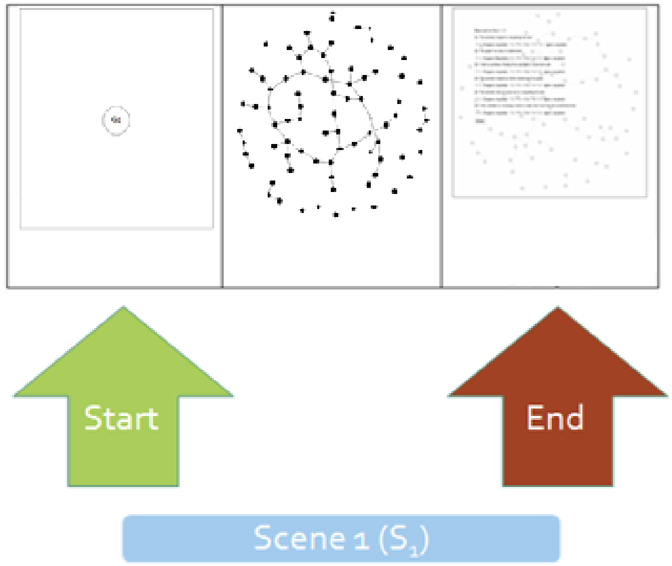


Figure 35: Example 'scene' for a given network

2.5 EYE-TRACKING

The eye-tracking methodology for this experiment was developed to not only collect eye-tracking data, but also provide a deterministic methodology for its analysis. The first step after the data was collected from the participants was to determine "scenes" based on user-defined areas of interest. These scenes were designated from when the participant was viewing and interacting with the network structures. The scene would start during the warm-up screen (Table 3 - Screen 1) and ended at the beginning of the evaluation screen (Table 3 - Screen 3) to capture all of eye-movements by the participant during the networking phase. Figure 35 illustrates an example network and designated scene. These designations could vary between 11 seconds up to 13 seconds, depending on the accuracy of the selection, which were both subject to human and application specific error.

This process was then repeated for all networks, in total 88 times (Pilot Study) and 36 times (Thesis Experiment) per participant. An example of a completed coding of designated scenes for a single participant can be found in Figure 36.

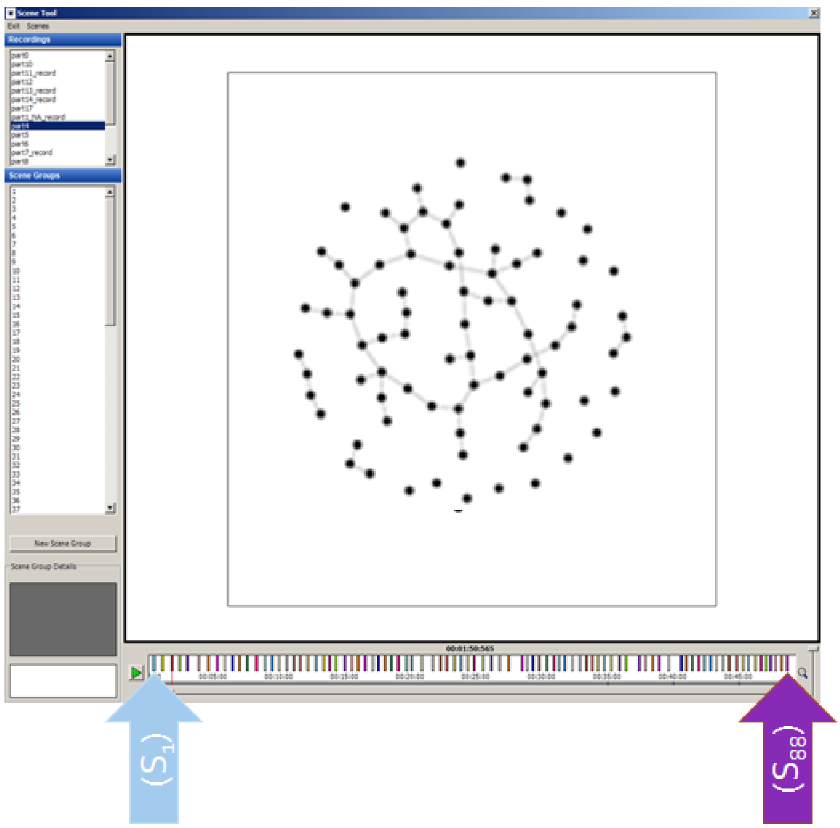


Figure 36: Completed coding for a single participant in Pilot Study

Once all 88 of the scenes were designated, the software produced three raw data formats: EFD (eye fixation data) file, EVD (event data) file, and AOI (areas of interest) file. The AOI file provided the start and stop times for each of the scenes that were designated. As specified, this would be a human defined window roughly 11-13 seconds, Figure 37. The EVD provided the events, as in mouse clicks and key presses, which were time stamped during the collection process. The EFD provided the raw data for the eye movements. To calculate an accurate start and end time for each scene or trial for participants, the following progression was used:

1. (Figure 37) Starting with the AOI file, each trial was encoded at the start of the warm screen, Table 3 - Screen 1, and ending with the follow-up screen, Table 3 - Screen 3.
2. (Figure 38) The user interface was designed to limit movement to the network screen, Table 3 - Screen 2, by requiring the participant to "click" on the "GO" button on the warm screen, Table 3 - Screen 1. Knowing this, the EVD file was used to find this exact moment this transition would occur (from warm-up screen to network screen), giving an accurate start time.
3. Knowing the start time, 10000 ms was added due to the fact that each trial was 10 seconds in length, giving an accurate end time.
4. With both the start and end times, the EFD data was parsed to find all eye fixations for each scene for each participant, which occurred every 20 ms.
5. (Figure 39) Within the EFD data, three levels of clarity were given for each fixation (left, right, and both). Only the "both" data was used in the eye-tracking analysis as it provided the least amount of error.
6. As specified by Poole et al. (2006), due to the sometimes chaotic eye-movements by participants, 60 ms was used for each fixation point. Therefore every 3 raw points (20 ms) were averaged to produce a 60 ms fixation point.
7. (Figure 40) Finally, the figures provided by the original Clearview software were compared to figures created using this data process (in the raw data form) to validate this methodology.

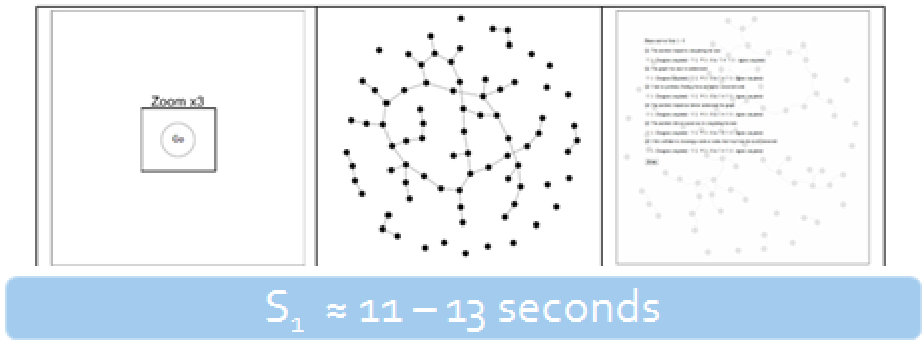


Figure 37: Example scene (S1), which was encoded for each scene and participant

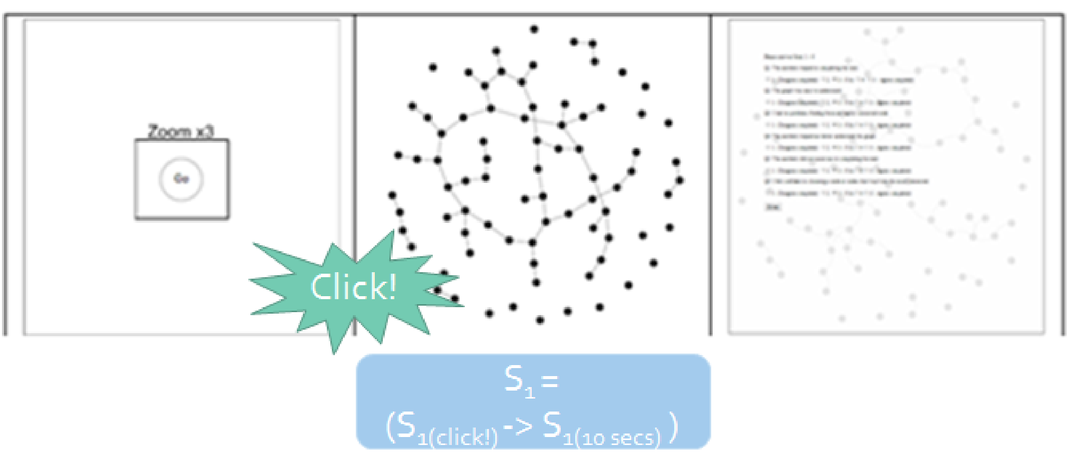


Figure 38: Redefining S1 using the event data given in EFD, where a "click" is used to transition to the network screen

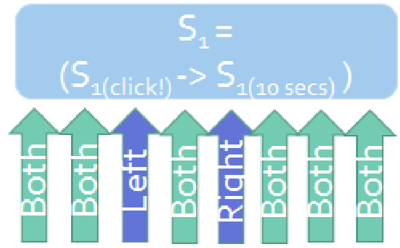


Figure 39: An example of the fixation data and how it was encoded (left, right, and both), where "both" provided the highest level of accuracy

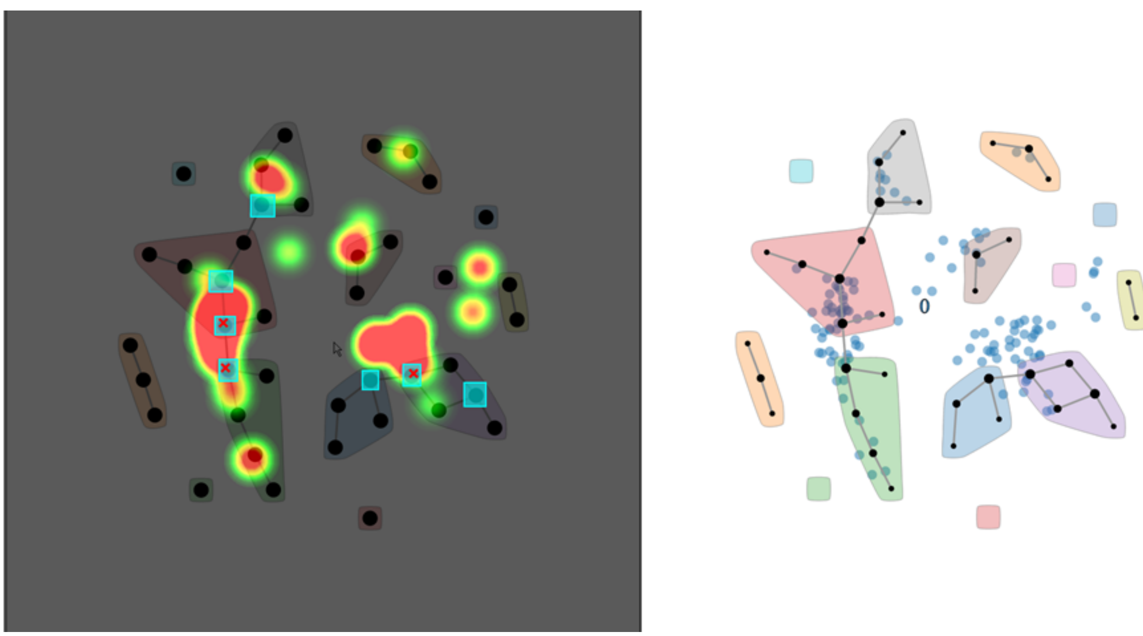


Figure 40: Comparing the results given in the Clearview software to data processed using the raw files

With this processed data created from the raw eye-tracking data sources (AOI, EFD, and EVD), the analysis was approached two different ways: 1) using average saccade length to dissect differences in how the visual aesthetics are parsed and utilized in the given task; and 2) the calculation of the task focusing time, which makes a connection between the visual aesthetics, the task, and user's ability to continue applying effort in completing the task.

2.5.1 Average Saccade Length

To calculate average saccade length for the graphs, the Euclidean distance between each set of three (60 ms) fixation points for each graph was found. They were then summed for each graph/visual aesthetic to determine a final average saccade length amount. Therefore, higher amounts show a more exhaustive search versus a more precise understanding of where to look in terms of the graph.

The interesting result is the comparison between the various levels (node aesthetics, link aesthetics, and group aesthetics), which showed that node aesthetics (NA) produced significantly higher saccade movements ($p < .01$) than the link aesthetics (LA) and group aesthetics (GA). This shows that for RQ1, aesthetics related to the task location would produce a higher exhaustive search. A significant difference was found for NA > LA and GA, as in the participant would need to scan (or did scan) all nodes for their aesthetic information to find and validate their responses.

2.5.2 Task Focusing Time (Inverse Task Abandonment)

Task focusing time relies on the participant's ability to focus on task-oriented directions. With eye-tracking data, a better understanding of this component can be gathered by defining an area in which the task could be accomplished (within the network structure) and not accomplished (the white space). This was computed by comparing the raw data source to the topological structure for each participant for each trial. A bounding box (a buffer area) was used around each node and link (roughly 60 pixels), where task focusing time starts when the participant first views the network and ends when multiple saccades (10 consecutive) are outside the buffer area (60 pixels) of the network. This would allow some random

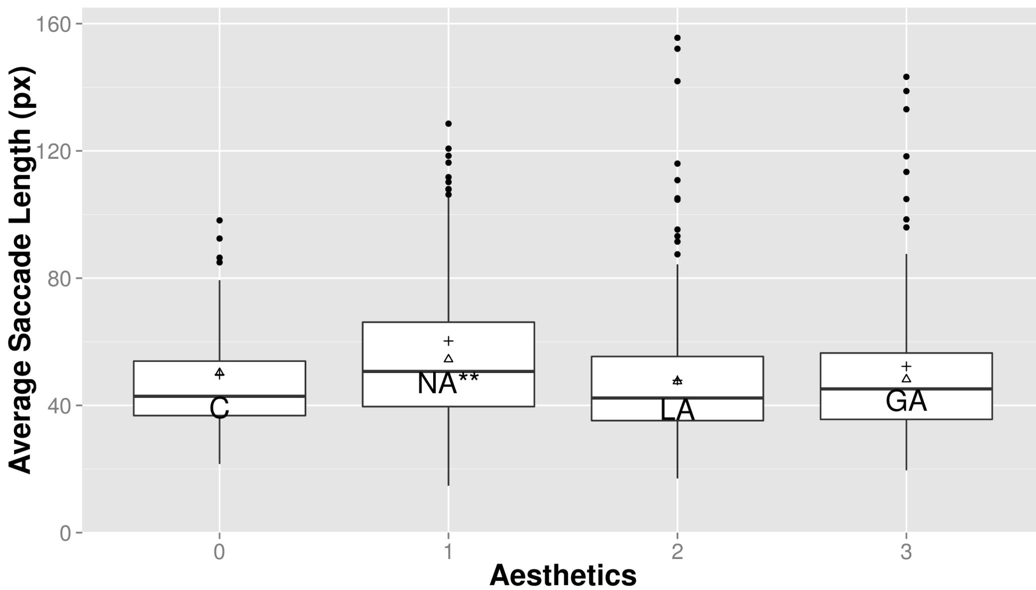


Figure 41: Average saccade length results showcasing each group of aesthetics (NA, LA, and GA) and their distribution versus time (px). The + indicates the mean of the large network (n_l) and Δ indicates the mean of the small network (n_s)

saccades, which have a tendency to occur naturally and could better classify the action on the participant's behalf. Also, this does not take into account the user resuming interest and going back to the task. The reason this study did not solicit this information directly from the participant was because either they would not know or assume they focused on the task throughout.

Once these values were calculated, the results were normalized based on the "time to first correct selection" values presented in Figure 33. This was because if the task was deemed "completed" by the participant, they would lose focus on the task. Figure 42 provides the results of this analysis for all network trials. Using ANOVA followed by Tukey HSD pair-wise, significance was found for all three node aesthetics (NA - $p < .01$) and LCo, LCu, and GC ($p < .05$), which proved both H4a and H4b. This provides evidence that visual aesthetics with high accuracy are perceived with higher utility and will keep the participant focused on that task.

2.6 CONCLUSION

In terms of the topological structure of the network and their effects on the accuracy of task:

- H1a: Confirmed. The results, seen in Figure 31, showed that the topological structure of the network affected the accuracy.
- H1b: Confirmed. The results, seen in Figure 32 - left, were also validated based on the number of nodes utilized in the graph (40 and 80 nodes).
- H1c: Not confirmed. The results, seen in Figure 32 - right, did not show a connection to the edge-crossing ratio (number of edge-crossing/number of edges) and the accuracy results.

For the accuracy (F1-Score) and time to completion, this was analyzed using a number of conditions based on the related visual encoding, non-related visual encoding, and visual styling.

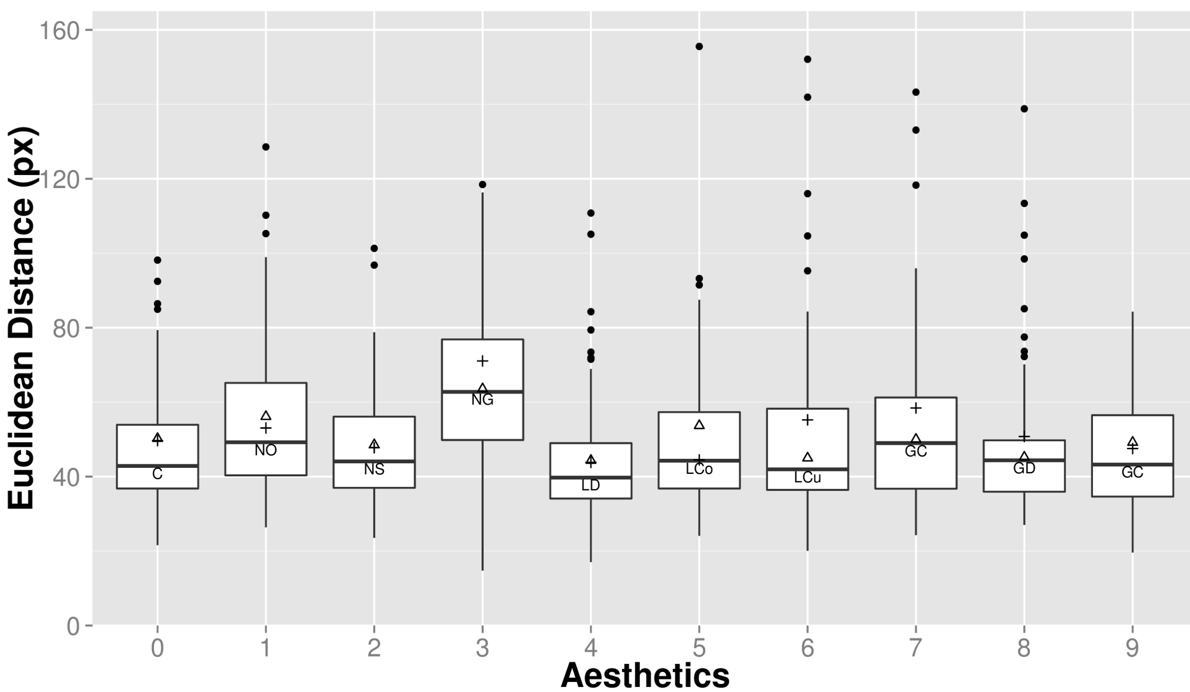


Figure 42: Task focusing time (inverse task abandonment) results showcasing each aesthetic and their distribution versus time (ms). The + indicates the mean of the large network (n_l)

- H2a: Confirmed, but only for node size (NS) condition when it came to time completion. The results, Figure 33, showed that there was a higher accuracy value and lower "time to first correct response" for both (NS and NG), but not significantly.
- H2b: Not confirmed. It is worth noting though, in Figure 33, there was a distinct pattern indicating that non-related visual encoding produced poor accuracy levels and increased time to completion.
- H2c: Confirmed. The results, Figure 33, based on no significant difference between NO and LCu and the control.

Provided was the qualitative evaluation of the perceived network utility of the graph.

- H3a/H3d: Confirmed. Results, Figure 34, showed node size (NS), glyphs (NG), and opacity (NO) were found significantly more optimal in terms of their perceived usability based on the qualitative data.
- H3b/H3c: Not confirmed. There was no significance for the non-related visual encoding and visual styling, Figure 34.

For the eye-tracking results, analysis was made of two types of usability criteria: average saccade length and task focusing time.

- RQ1: Confirmed. An increased average saccade length for the node aesthetics (NA) and the relationship NA > LA and GA, Figure 41.
- H4a/H4b: Confirmed. Based on the results, Figure 42, the higher accuracy and higher perceived utility visual aesthetics provide more focus by the participant.

2.7 MOVING FORWARD

Learning from the initial Pilot Study, a few changes were made to the design of the experiment, including an expanded hypothesis. Provided are expansions and changes to the Pilot Study and their subsequent design.

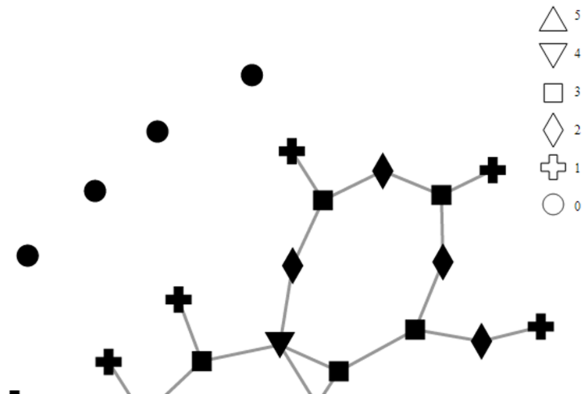


Figure 43: Example network using glyphs from the Pilot Study

2.7.1 Motivation/Incentive

As proposed by Ware (2002), to help with motivating individuals to provide more accurate results, an incentive (a $50 gift card to Amazon.com) was offered to the top two participants who completed the tasks with the highest accuracy value (F1-Score). By increasing the participants' willingness to maximize their selections, there would be an increase in the engagement for the user during the task (relative to the accuracy for these trials). This would then lend itself to increasing the accuracy levels and decreasing the error associated with the eye-tracking data, based on participants being more attentive to the task. This incentive was advertised on the experiment flyer so the participant knew of this incentive prior to testing, was included in the explanation of the experiment prior to testing, and required the participant to sign a waiver indicating their interest and specifics in the distribution of the incentive.

2.7.2 Visual Aesthetic Changes

One of the specialized visual encoding elements in the Pilot Study (the node glyphs) utilized a legend presented at the top right-hand portion of the screen, seen in Figure 43. This caused an issue in terms of defining when eye-movements should be recorded, based on the fact that all participants needed to:

1. Start their eye gaze at the center of the page (as most did for all graphs)
2. Move their eye gaze to the top right-hand portion of the screen
3. Establish the glyph (in the above example, the top-pointing triangle) that represented the most connected node
4. Return eye gaze to graph structure
5. Search the graph for this glyph (in the above example, the top-pointing triangle)

This was unique in terms of searching strategies employed by the rest of the visual encodings. To remedy this, glyphs were designed to be more intuitive based on representing the number of connections via their shape; this is shown in more detail in the Thesis Study.

Also, the color edges and the group node coloring were altered for the Thesis Study. Color edges, which performed poorly compared to the control, was altered to create a novel visual aesthetic using a color gradient, highlighting degree centrality based on link aesthetics. Also, the group node coloring was changed to provide visual encoding of the degree centrality, which would create novel, two encoding aesthetics (one for encoding the group membership and one for encoding the number of connections). Both of these will be described in more detail in Section 3.0 - Thesis Experiment.

2.7.3 Removed Visual Aesthetics

Four visual aesthetics were removed based on their poor or middling results in the Pilot Study and other extenuating reasons discussed. The first were the curved lines, as these were examined in-depth by Holten et al. (2010) and Xu et al. (2012), in which it was shown that curved edges either impaired readability or were of little consequence to readability. This was validated in the Pilot Study and removed because of this. Also, directed edges were removed based on their poor performance in both accuracy and time, and led to confusion on the given task of finding the most connected node(s) in a non-directed graph. Log-degree size was removed because it was a redundant examination of degree size, which produced worse results in terms of accuracy and time versus the degree size. Finally, the intra-inter distance bias was removed because even though the topological network structure was similar to other visual aesthetics, it varied in the structure out of necessity of the visual aesthetic. The only way to accomplish the intra-inter distance bias was to change the topological structure of the network, varying it from the other visual aesthetics.

2.7.4 Added Visual Aesthetic

One visual aesthetic was added based on the previous literature that was not analyzed in the Pilot Study. This aesthetic is tapered edges. Again, this will be described in more detail in Section 3.0 - Thesis Experiment.

2.7.5 Increasing the Number of Participants

The Pilot Study focused on having 22 participants. To gain better insight into the user's behavior and to accommodate the changes the visual aesthetics used, this was redone with 27 new participants. In terms of the literature, when it comes to the number of participants, only one reference specifies the number of participants needed to complete a suitable eye-tracking, usability study (Pernice & Nielsen, 2009). In this reference, they note that 20 participants are needed for adequate quantitative analysis, but for eye-tracking studies there should be upwards of 39 participants based on eye-tracking data (30 with high-quality eye-tracking data) when utilizing heatmaps as the main comparative medium. To meet this criterion, the Thesis Experiment is not solely using heatmaps for the analysis but specific gaze plots that are created by raw data of the eye-tracking system to provide better clarity. Also, each network is seen multiple times (2x) by each participant, providing a more concrete understanding of their behaviors as comprised of the 20 participants needed for quantitative analysis and 30 (high quality) needed for an adequate eye-tracking study. To further this point, a new section is added to the study to provide qualitative data based on the eye-tracking data and participants' feedback to their own eye-tracking replays, called retrospective talk-aloud interview. The qualitative analysis will be discussed in more detail in Section 3.0 - Thesis Experiment.

3.0 THESIS EXPERIMENT

After the Pilot Study was concluded, changes and updates were made to solidify the study agenda to better understand how visual aesthetics could alter a person's ability to complete a given task. The Thesis Experiment refined the visual aesthetics to focus on how variations at the node, link, and group level and how they could redefine the strategies utilized, reduced the amount of trials to collect more meaningful data, incorporated more participants to the study, and introduced qualitative analysis for further understanding. There is overlap between this study and the previous one (Pilot Study), but primary notes about each of the experiment components will be reintroduced to help with navigating the overall design and finding.

3.1 DEFINING THE VISUAL AESTHETICS

Similarly to the Pilot Study, the visual aesthetics were developed to provide both encoding and styling variations. The number of visual aesthetics were reduced to provide better focus on the comparative nature of the visual aesthetics. In the Pilot Study, there were visual aesthetics labeled "Not Task Related." This was effectively removed because they did not provide any added benefit for the given task. Each level of the design (node, link, and group) provided both visual encoding and visual style. The node was provided three visual encodings to help distinguish which encoding (based on color, shape, and size) would produce better results in terms of accuracy and time. To expand on this, participants were able to provide qualitative feedback for each of the visual aesthetics, including both their interaction in selecting the most connected node or nodes and determining their approach or strategies

Table 5: Visual aesthetic updates (*italicized* - removed and **bold** - changed or reclassified)

Visual Encoding (Task Related)	Visual Encoding (Not Task Related)	Visual Styling
Node Size (NS)	*Link Direction (LD)*	Node Opacity (NO)
Node Glyphs (NG)	*Line Color (LCo)*	*Link Curve (LCu)*
Node Color (NC)	*Group Coloring (GC)*	**Link Tapered (LT)**
Link Gradient (LG)	*Group Distance Bias (GDB)*	**Group Convex Hulls (GCH)**
Group Coloring (GC)		

Table 6: Visual aesthetics for the Thesis Experiment

Visual Encoding	Visual Styling
Node Size (NS)	Node Opacity (NO)
Node Glyphs (NG)	Link Tapered (LTa)
Node Color (NC)	Group Convex Hulls (GCH)
Link Gradient (LG)	
Group Coloring (GC)	

when accomplishing the task. Table 5 is an updated version of Table 2, where *italicized* items have been removed and **bold** items were added or reclassified. Table 6 is the the final table.

3.1.1 Visual Aesthetics

This resulted in producing nine visual aesthetics, both representing all levels of network design (node, link, and group) and visual encoding and styling. Provided are detailed descriptions for each of these levels and aesthetics. Figure 44 showcases all nine visual aesthetics, utilizing the same network structure for each. This highlights the overarching theme of the visual aesthetics, as they can be viewed as filters to augment both the readability and utility of the topological network structure. Also, it should be noted that all visual aesthetics were created and designed prior to the development of the network graphs as to not bias their design based on the network topology provided.

The control for this experiment follows the same constraints of the Pilot Study as a

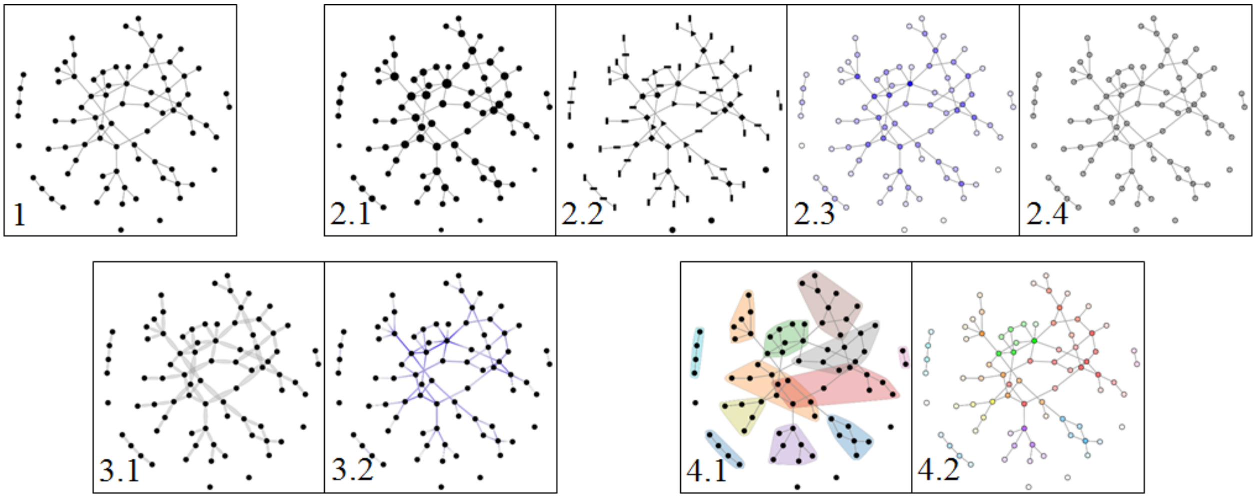


Figure 44: Shows an example network with the 9 visual aesthetics and their groupings: 1) Control; 2) Node Aesthetics: 2.1) Size, 2.2) Glyphs, 2.3) Color Saturation, 2.4) Opacity; 3) Link Aesthetics: 3.1) Link Tapered, 3.2) Link Gradient; and 4) Group Aesthetics: 4.1) Convex Hulls, and 4.2) Group Coloring and Node Saturation

crucial piece for connecting and relating the topological structure of the network and the applied visual aesthetics. This is because the control does not contain visual aesthetics and is limited in providing increased readability and affordance to network aesthetics by methods previously presented in regards to minimizing edge overlap, orthogonality, and collinearity. These same network aesthetics are also applied to all networks within this study so the control is used as the baseline to measure the visual aesthetics' effects on accuracy, time, subjective measures, and eye-tracking data based on the topological structure of the network.

3.1.1.1 Node Aesthetics (NA) The node aesthetics (NA) were designed to optimize variation between values to maximize the information presented to the participant. The only node aesthetic that does not follow this maximized information is the node opacity (NO) since this is a type of style and does not allow such distinction.

The first NA, node size (NS), uses a linear increase of diameter based on the number of connections entering or leaving the node, or similarly, degree centrality. This was maximized in terms of differentiating from one another by increasing the diameter (versus only the area) to create nodes that varied in size by $\pi\left(\frac{n}{2}\right)^2$.

Algorithm 1 (Node Size)
$G = (V, E)$, s source, t target, v vertex where $s \in V$ and $t \in V$ $NS = \lim_{v\to\infty} \pi\left(\frac{1}{2}\right)^2, \pi\left(1\right)^2, \pi\left(\frac{3}{2}\right)^2, \ldots, \pi\left(\frac{v}{2}\right)^2$

Creating glyphs provided its own unique set of complexities. Glyphs can have a perceived quality based on the creator's or viewer's mental model of their design and therefore typically represent nominal functions (Ware, 2012). Due to this, the glyphs were designed to adhere to objects that already share their mental mapping in mathematical roots by using polygons to represent the number of connections. This altered the previous view of glyphs from the Pilot Study, as these were listed only as associated connections, but now providing quantitative perception (each point of the polygon means an increase in value).

In Figure 45, three rows are shown. The top row shows the polygons and their subsequent numerical value (bottom row). As there are no non-gons, mono-gons, or di-gons, glyphs for 0, 1, and 2 were made to look similar to their numerical counter parts (0 and 1). For di-gons, these are 1-dimensional objects with a slight increase in a 2nd dimension to provide depth. The other polygons present their like previously understood polygon (triangle, square, pentagon, hexagon, etc.). As the number of connections increased, there became an apparent issue with the polygons with five or more points as they become indistinguishable from one another. An algorithm was put together to create objects with greater than five connections:

Algorithm 2 (Node Glyphs)
$G = (V, E)$, s source, t target, v vertex where $s \in V$ and $t \in V$

$\theta = \frac{2\pi}{deg(v)}$

$r = \frac{\sqrt{sizeOfObject}}{2}$

$c = amountOfCurvature$

$for\ all\ i = 0, 1, 2, \ldots, n\ in\ (deg(v) > 5)$

$\quad x\,(i) = (r * cos(i * \theta))$

$\quad y\,(i) = (r * sin(i * \theta))$

$\quad if\ (i\ ==\ 1)$

$\quad\quad dx\ =\ x[i]\ -\ x[i-1]$

$\quad\quad dy\ =\ y[i]\ -\ y[i-1]$

$\quad\quad dr\ =\ \sqrt{(dx * dx\ +\ dy * dy)} * c$

$\quad\quad path = moveto([i-1], y[i-1])$

$\quad\quad path = elliptical\ Arc(dr, dr, \text{“0 1,0”}, x[i], y[i])$

$\quad else\ if\ (i\ >\ 1)$

$\quad\quad dx\ =\ x[i]\ -\ x[i-1]$

$\quad\quad dy\ =\ y[i]\ -\ y[i-1]$

$\quad\quad dr\ =\ \sqrt{(dx * dx\ +\ dy * dy)} * c$

$\quad\quad path = lineto(x[i-1], y[i-1])$

$\quad\quad path = elliptical\ Arc(dr, dr, \text{“0 1,0”}, x[i], y[i])$

$\quad if\ (i+1\ ==\ n)$

$\quad\quad path = lineto(x[i], y[i])$

$\quad\quad path = elliptical\ Arc(dr, dr, \text{“0 1,0”}, x[0], y[0])$

$\quad\quad path = closepath()$

$end\ for$

$return\ path$

This algorithm produced a slight curve to the edges of the polygons (greater than 5 points), increasing their readability and differential to other polygons. In Figure 45, the middle row showcases these polygons. It should be noted that after the networks were computed, the degree centrality ranged from four to nine, so it is believed this correction

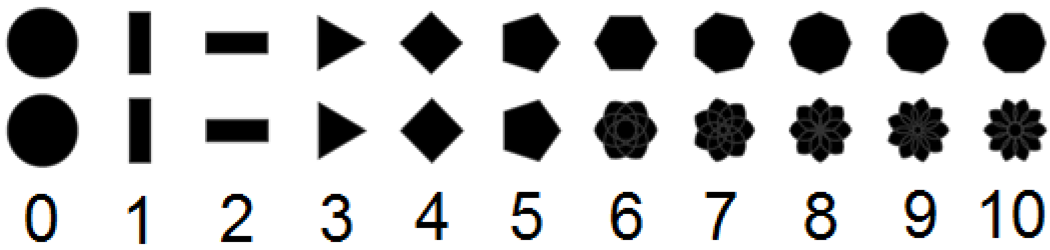


Figure 45: Glyph representation for polygons and their quantitative counter parts. (Top row) shows the initial design and (middle row) illustrates the “point featuring” of the same polygons.

was relevant.

Introduced in this experiment is a third alternative to showcasing degree centrality measure, this being the range of lightness or brightness of a given hue. Hues range from a value of 0° to 360°, where 0° represents red primary, 120° green primary, 240° blue primary, and 360° completes the cycle back to red primary. The hue 246 (indigo) and saturation 100% was used with a variation of the lightness based on the number of connections. Because the lightness was varied, any hue could be utilized and still provide the same range values. To calculate the lightness, the following algorithm was produced:

Algorithm 3 (Node Color)
$G = (V, E)$, s source, t target, v vertex where $s \in V$ and $t \in V$
$for\ all\ 0, 1, 2, \ldots, v\ \in\ vertex$
$\quad lightness = \frac{deg(v)}{max(deg(V))}*100$
$\quad lightness = \frac{lightness}{2}$
$\quad lightness = 100 - lightness$
$\quad return\ HSL\,(246,\ 100,\ lightness)$
$end\ for$

This algorithm would maximize the lightness value for each network graph (individually), where contrary, the lightness could be maximized based on the highest degree centrality measure for all networks (combined). The second option would minimize the distinction between the number of connections in situations where the network graph had a degree centrality less than the maximum degree centrality for graphs. As in:

1. The lightness range was set from 50% to 100%. For each network, the minimal number of connections in the graph would receive 100% and the maximum would receive 50%. .
2. The lightness range was set from 50% to 100%. For all networks, the minimal number of connections for all graphs would receive 100% and the maximum for all graphs would receive 50%

Again, to maximize the differences between the number of connections in a graph, option (1) was chosen.

Lastly, the node opacity (NO) was reintroduced from the Pilot Study. This is considered to be node visual styling. The only difference between the control and the node opacity is that the opacity was set to .3 and again, showcases the 'texture' of each node based on their connections.

3.1.1.2 Link Aesthetics There were two link aesthetics created for this experiment: 1) link tapered (LTa) and 2) link gradient (LG). Link tapered was inspired by the works of Netzel et al. (2014) and Holten et al. (2011). It should be strongly noted that for both the experiments, each researcher was focusing on directed graphs and both found tapered links to be superior for directed graphs. The networks utilized in this experiment were created as undirected graphs, so the design needed to be augmented by removing directed information. Contrary to these works, the start of the tapered edge was moved in between the nodes and would go in both directions (to each respective node). After this was done, there appeared to be a visual artifact based on the sharp tapered edge at the starting location. It would be best to think about putting two triangles together, which are the mirror image of one another, example: . The peak in the middle could cause issues with readability, as per suggested in Ware's (2002) "when edge cross at acute angles, they will be more likely to cause visual confusion." To remedy this, curved edges were used to soften the edge meet up, so they resembled the equidistant comets found in Netzel et al. (2014). Also, similarly to the curved links in the Pilot Study, link tapered would be a visual styling based on Gestalt's continuation principle. The following algorithm was used to construct these:

Algorithm 4 (Link Tapered)
$G = (V, E)$, s source, t target, v vertex where $s \in V$ and $t \in V$
$v_s = (x_s,\ y_s)\ and\ v_t = (x_t,\ y_t)$
$dx = \ x_s - \ x_t$
$dy = \ y_s - \ y_t$
$dr = \sqrt{dx^2\ *\ dy^2}$
$path = moveto\,(x_s, y_s)$
$path = elliptical\ arc\,(dr, dr, \text{“}0\ 0, 1\text{”}, x_t, y_t)$
$path = elliptical\ arc\,(-dr, -dr, \text{“}0\ 0, 1\text{”}, x_s, y_s)$

Even though the tapered link incorporates ideas from tapered edges, equidistant comets, and Lombardi drawing (Duncan et al., 2011), the naming was kept to tapered links, as based on the original design.

The other link aesthetic, called link gradient, was designed to incorporate similar visual encoding as the node visual encoding, but moving the degree measurements to the links connected to the node versus the node itself.

Algorithm 5 (Link Gradient)
$G = (V, E)$, s source, t target, v vertex where $s \in V$ and $t \in V$
$gradient = \left[v_s = \ 100 - \left(\frac{\deg(v_s)}{max(\deg(V))} {*}50\right),\ v_t = 100 - \left(\frac{\deg(v_t)}{max(\deg(V))} {*}50\right)\right]$
$if\ (\deg(v_s) > \ \deg(v_t))$
$\quad return\ linearGradient\,([gradient\,[v_s]],\ \ [gradient\,[v_t]])$
$if\ (\deg(v_s) < \deg(v_t))$
$\quad return\ linearGradient\,([gradient\,[v_t]],\ \ [gradient\,[v_s]])$
$if\ (\deg(v_s) == \deg(v_t))$
$\quad return\ HSL\left(246,\ 100,\ 100 - \left(\frac{\deg(v)}{max(\deg(V))} {*}50\right)\right)$

The link gradient is proposed as a novel means of visually encoding links based on data typically represented at the node level. Because this study is focused on the creation and understanding of networks in short, finite amounts of time, the link gradient takes advantage of the natural structure of the network and limits the scope of the eye-movement to only highly connected areas of the graph. Similar arguments in regards to this were brought from Selassie et al. (2011) work, where the analogy of a "two-way highway" was used to discuss links arriving and leaving a node and color coding them accordingly. As the nature of this task is unidirectional, a "one-way street" analogy could be more applicable for the link gradient. Also, the link gradient increases the data-ink ratio of the graph and adheres heavily with the ideals of Gestalt principles of continuation. If the link gradient improves the readability of the network graph, then additional attributed information can be adhered to the nodes themselves without the conflictions brought about by applying two different pre-attentive attributes or feature maps to a single object.

3.1.1.3 Group Aesthetics Similarly to the group aesthetics in the Pilot Study, groups were formed using the same clustering algorithm modularity optimization (Blondel et al., 2008). Based on gaze-replays in the Pilot Study, there was a difference in how participants approached the group aesthetics versus the node/link aesthetics. Participants tended to search for connected nodes within the community structure, to their benefit or detriment. Unfortunately, the modeling of the eye-tracking (the saccade movements) did not support this enough to warrant acknowledgement within the study.

To rectify this idea, group aesthetics were again used with the two variations: convex hulls and two-encoding method, but with the inclusion of qualitative feedback from the user to validate this top-down approach. The convex hull is seen in this experiment as a visual styling, as the visual aesthetic is not related to the task, but applied ubiquitously as the other visual styling. The algorithm was not designed for this study and utilized the "hull geom" created natively in D3.js (Bostock et al., 2011).

The two-encoding method (called two-encoding coloring), encodes both the group membership and the degree centrality, similarly to node coloring (NC - Algorithm 3). Again the idea being that these groups, which are based on highly connected sets of nodes, could subdivide the problem set into more manageable cognitive chunks, especially when the graph becomes denser and less readable.

3.1.1.4 Side-by-Side Comparisons Provided is a side-by-side comparison (control vs. visual aesthetics) of all visual aesthetics.

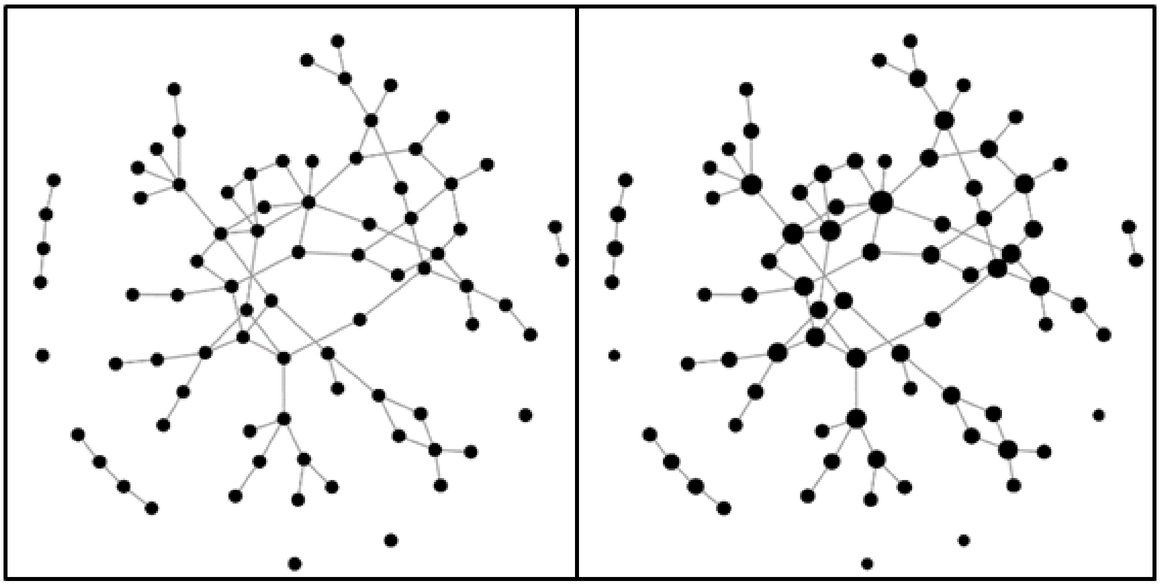


Figure 46: Side-by-side: Control and Node Size (NS)

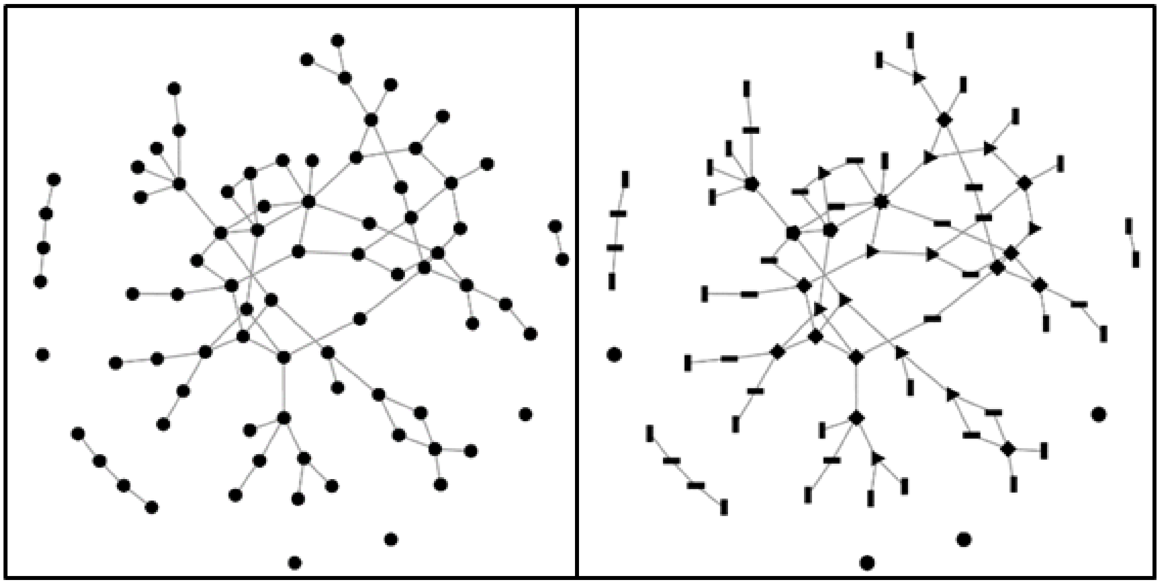


Figure 47: Side-by-side: Control and Node Glyphs (NG)

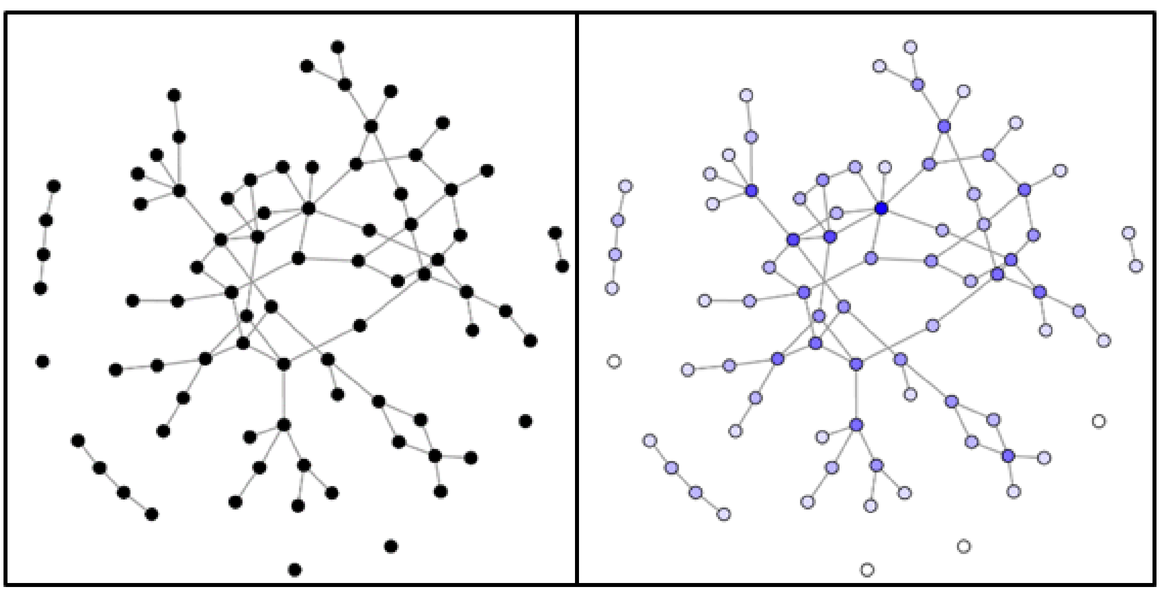


Figure 48: Side-by-side: Control and Node Color (NCo)

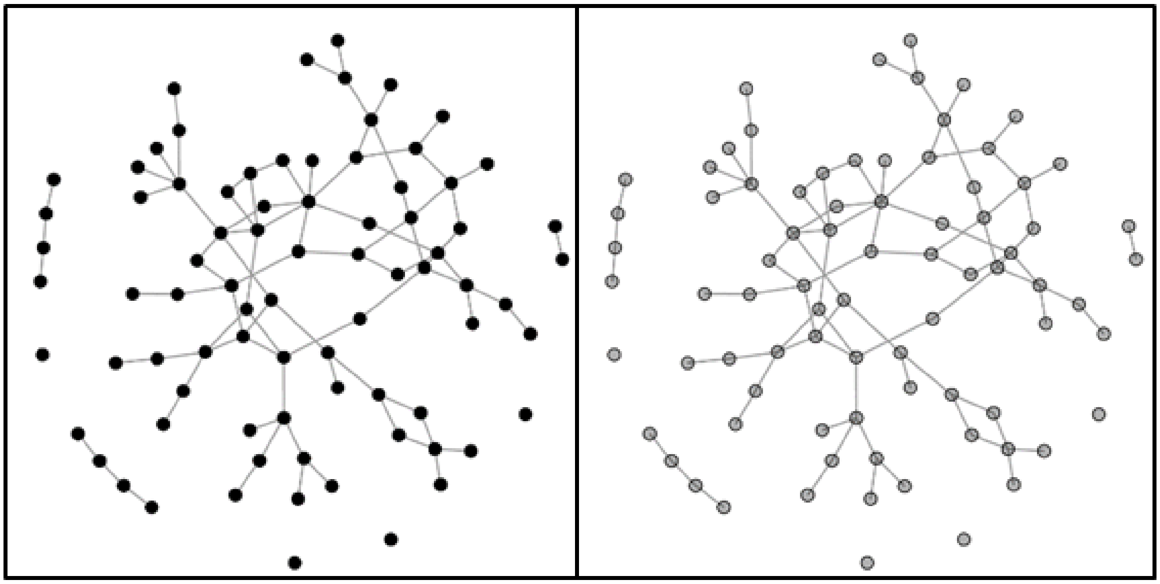


Figure 49: Side-by-side: Control and Node Opacity (NO)

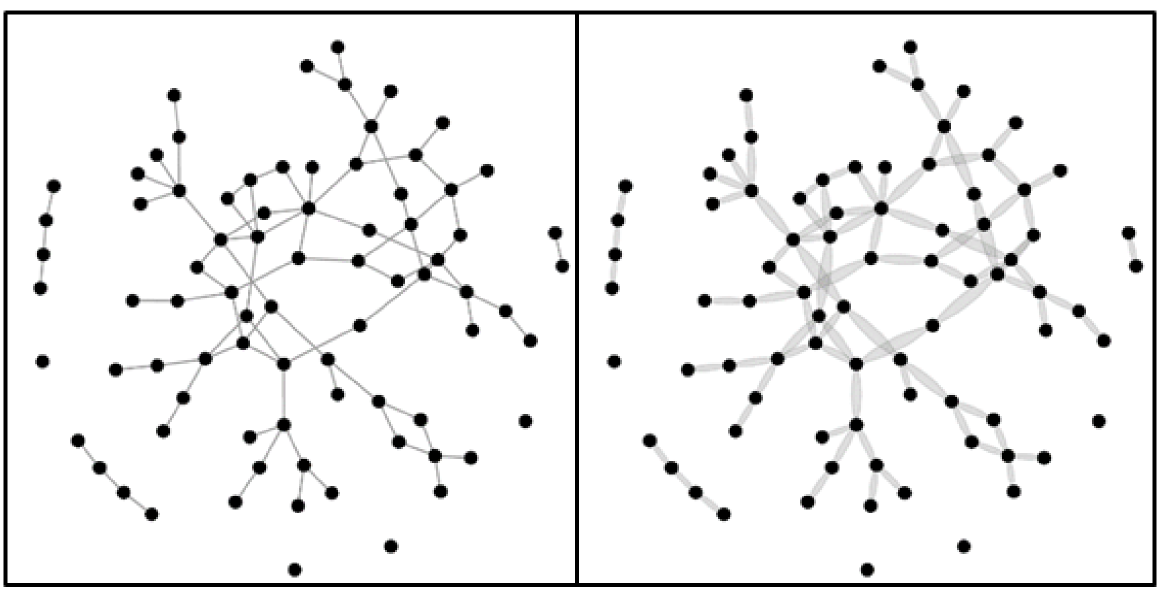


Figure 50: Side-by-side: Control and Link Tapered (LTa)

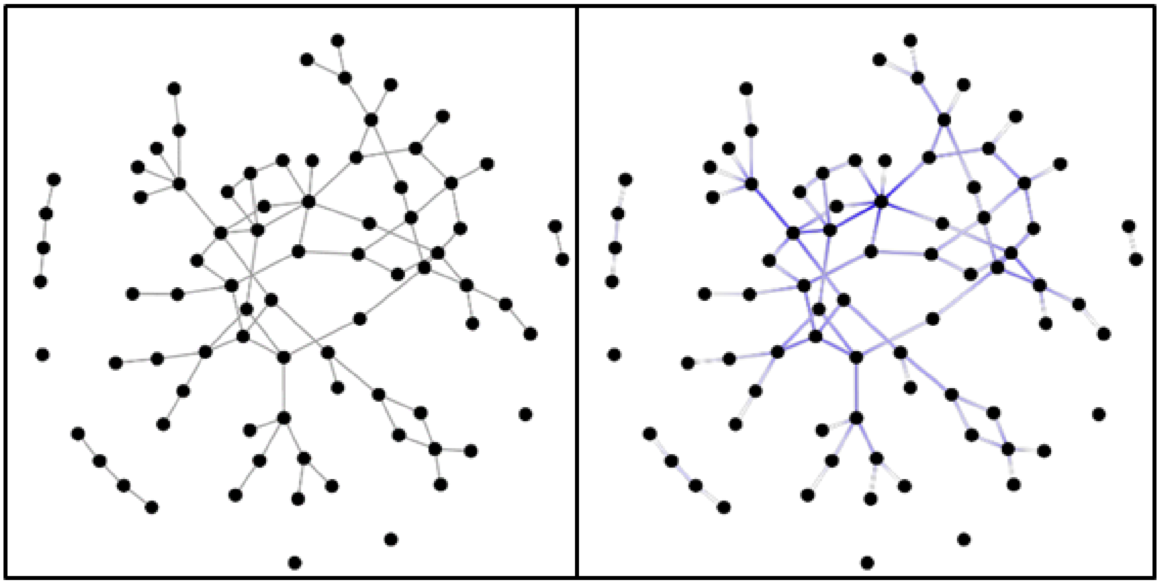


Figure 51: Side-by-side: Control and Link Gradient (LG)

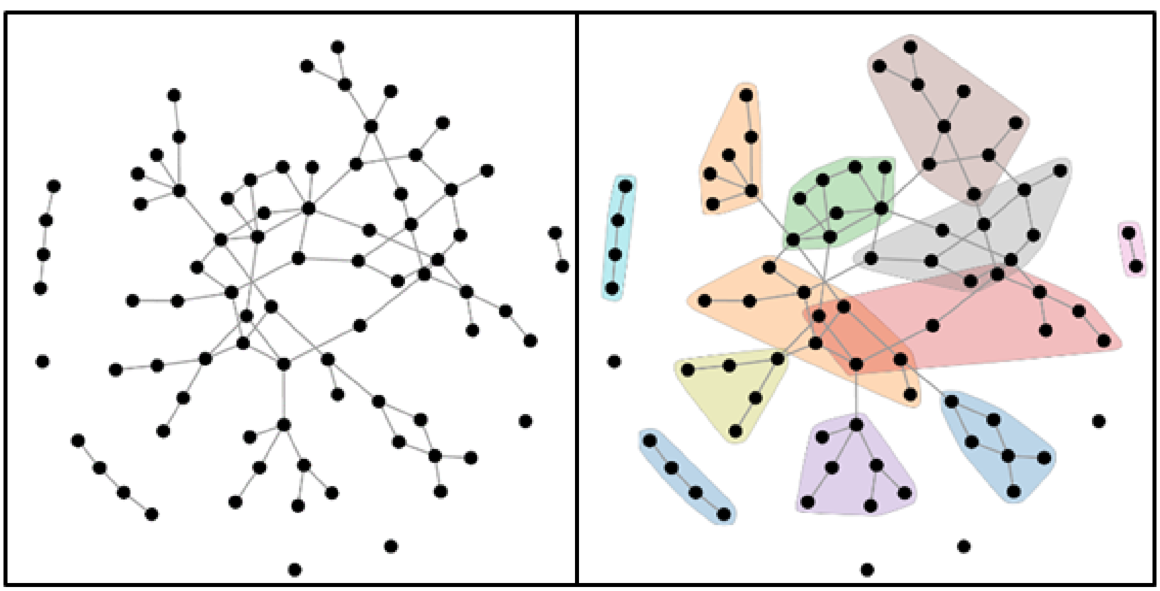


Figure 52: Side-by-side: Control and Group Convex Hull (GCh)

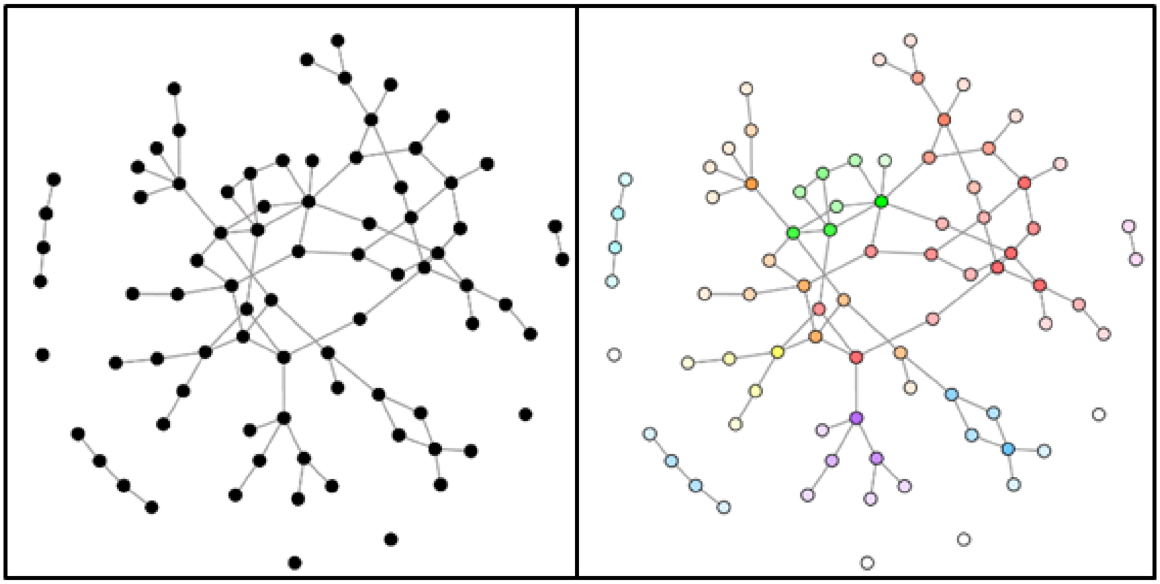


Figure 53: Side-by-side: Control and Group Two-Encoding Coloring (GCo)

3.2 GRAPH CREATION

The same randomized blocking design was incorporated into the Thesis Experiment. Therefore 9 randomly generated networks were created using the Erdös-Rènyi graph model (Erdös & Rényi, 1959; Gilbert, 1959). Erdös-Rènyi's model uses a probabilistic model to uniformly choose a random graph of size. Two sets were created to test a small network (n_s) and a large network (n_l). The only difference between the Pilot Study and the Thesis Experiment was that probability of edges was increased for both the small and large network, specifically for the small networks, vertices of 40 and probability .07 for edges ($n_s = 9$ $graphs$); and the large networks, vertices of 80 nodes and probability .029 ($n_l = 9$ $graphs$). The formation of the networks was handled using D3's (Bostock et al., 2011) force-directed graph, which incorporates both a Verlet integration with simple graph constraints (Jakobsen, 2001) and balanced using Dwyer's layout constraints (Dwyer, 2009). Figure 54 and Figure 55 showcase the randomized blocking design used in the Thesis Experiment, for small and large networks, respectively.

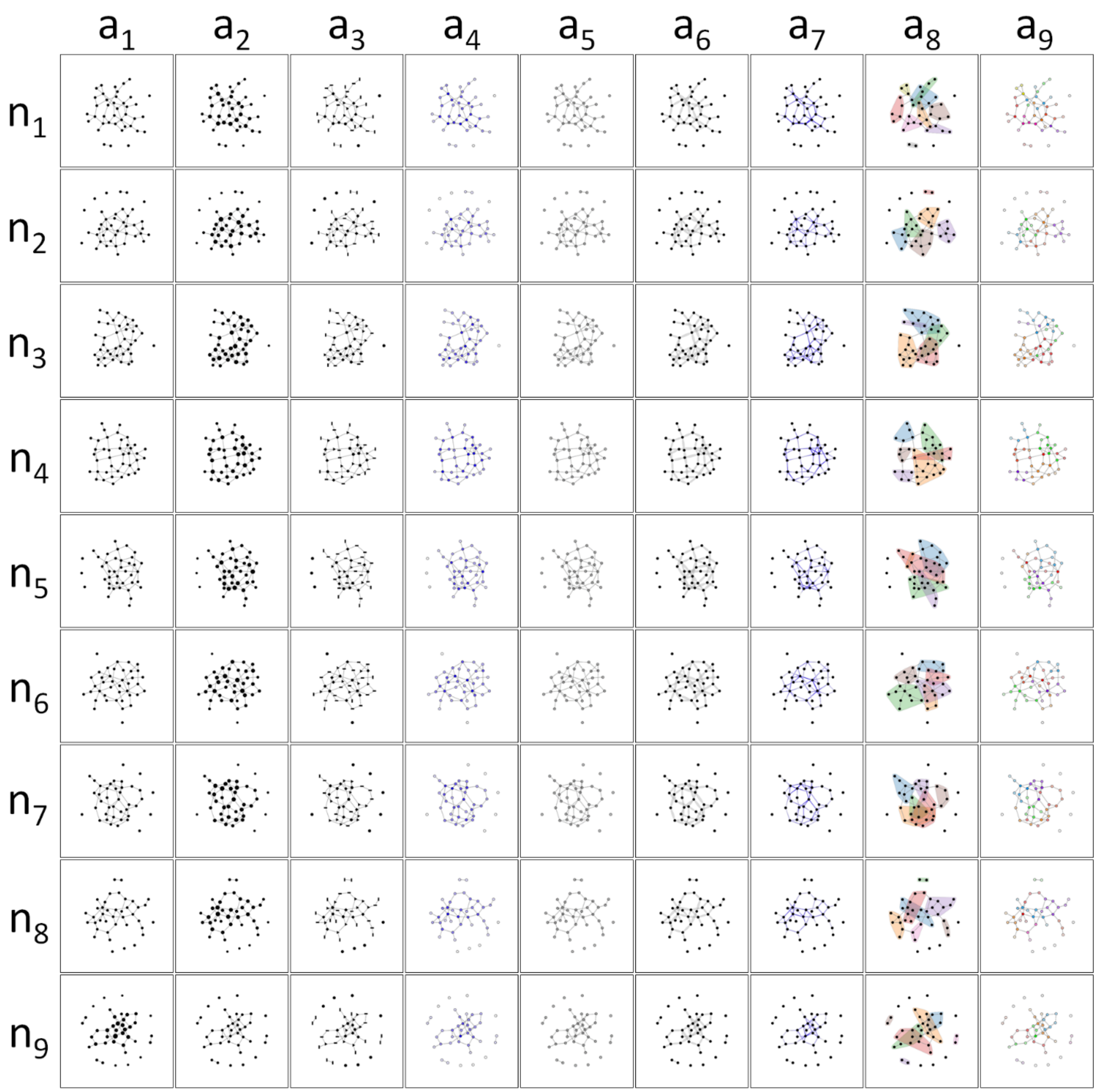


Figure 54: All aesthetics and network graphs for the small networks in the Thesis Experiment

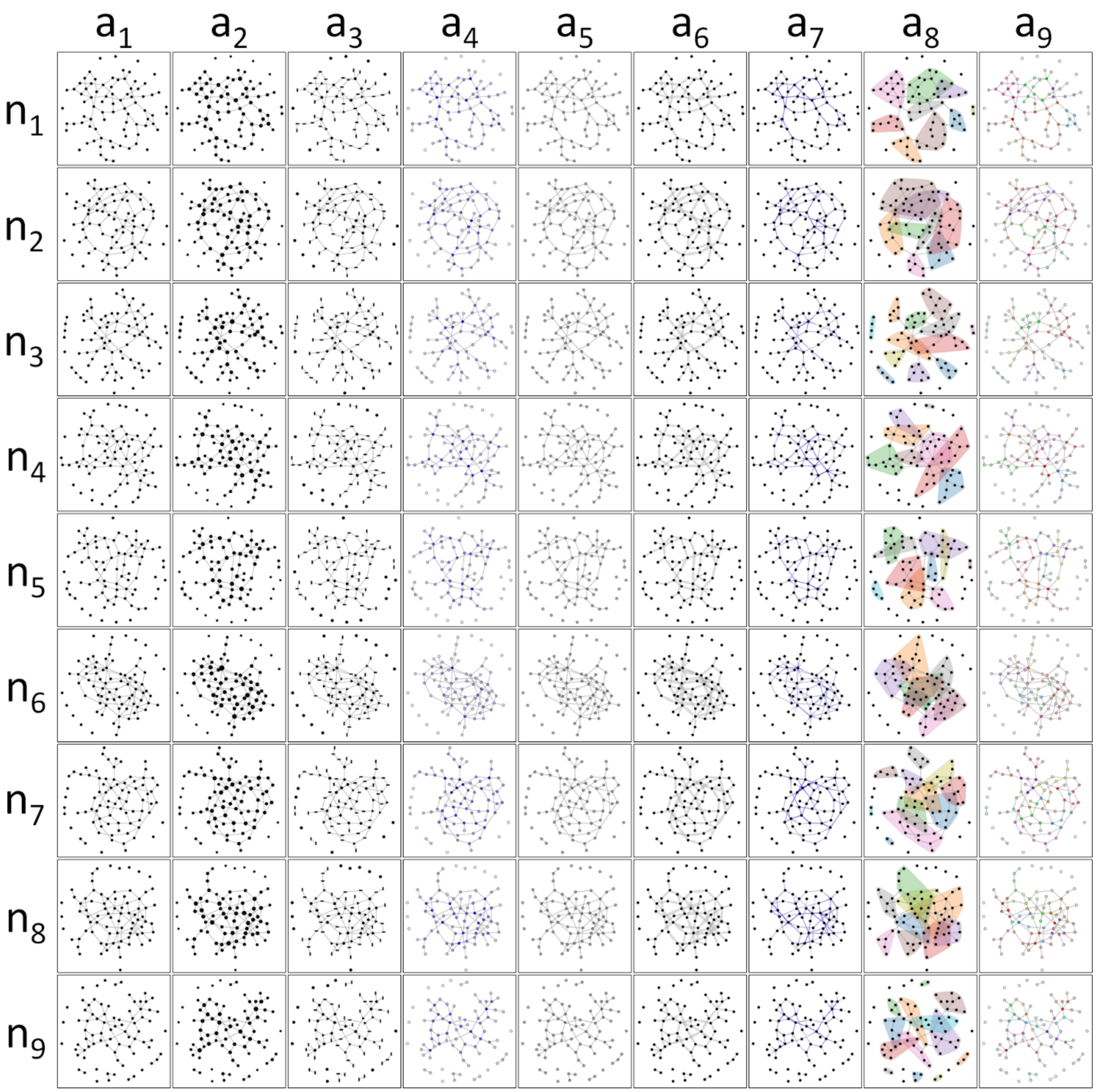


Figure 55: All aesthetics and network graphs for the large networks in the Thesis Experiment

3.3 HYPOTHESIS

The focus of this work is the coordination of the topological structures of the network and visual aesthetics applied to these network. The hypothesis will be focusing on how participants find utility with the given visual aesthetic, perceive their affordances, and how the appearance of the visual aesthetics can help or hinder their overall utility. The hope of the project is to bring focus to the application of visual aesthetics to network graphs and how visual aesthetics can alter their meaning (both positively and negatively). This will be accomplished by evaluating their overall utility, both in accuracy and time, to showcase which of the visual aesthetics can provide better utility than the control. Once this is established, identifying how the perceived affordance in both utility and confidence, and the subjective beauty of the graph coincide with these results. From there, a metric derived from the Pilot Study (Task Focusing Time) will be compared to the results from our subjective measure of beauty provided by the participants.

At this point, all that is known is which visual aesthetic outperforms the other visual aesthetics. This is where the application of eye-tracking data will bring to light how the participants utilized these visual aesthetics. This will be done with the average saccade lengths as a means of interpreting exploration, the use of ground-truth derived from the topological structure of the network graph, and how search patterns vary based on the visual aesthetics. Finally, the qualitative feedback received by the participant, where search strategies are discussed and further exploration on short term memory are brought to light based on saccadic suppression. The hypothesis will then emphasis:

- Utility of the visual aesthetics
- Affordances by perceived utility and confidence of the visual aesthetics
- Correlating subjective beauty to task focusing time
- Exploration, user behavior, and strategies given to visual aesthetics
- Edge-crossing vs. degree centrality: variation by visual aesthetic
- Short term memory implications in change blindness and network complexity

3.3.1 Utility of the Visual Aesthetics

Consistent to the Pilot Study, both the accuracy and time is evaluated for each visual aesthetics. Based on the Pilot Study, the visual encoding aesthetics (NS, NG, NC, LG, and GCo) should outperform the visual styling (NO, LTa, and GCh) and the control (C). The control is used to evaluate the node visual encoding (NS, NG, NC), the link visual encoding (LG), the group visual encoding (GCo), and the visual styling (NO, LTa, and GCh). Specifically for the visual styling, the expectation brought by the Pilot Study is reevaluated; that being the visual styling will have similar results to the control and node styling will be superior to link and group styling.

- H1a: Compared to the control, node based visual encoding (NS, NG, NC) will have higher performance (accuracy and task completion time) than the control.
- H1b: Compared to the control, link based visual encoding (LG) will have higher performance (accuracy and task completion time) than the control.
- H1c: Compared to the control, group based visual encoding (GCo) will have higher performance (accuracy and task completion time) than the control.
- H1d: Visual styling (NO, LTa, and GCh) will not have a significant difference on the performance (accuracy and task completion time) of the task as the control.
- H1e: Specifically, node opacity (NO) will have higher performance (accuracy and task completion time) than link tapered (LT) and group convex hull (GCh), as the node styling is localized at task level.

Node visual encoding (NS, NG, NC) were designed to represent ratio or ordinal data sets. Based off work by Mackinlay (1986), it is expected that NC > NS > NG in terms of accuracy (larger values mean higher accuracy) and NC < NS < NG for time (lower values mean less time to completion). This was further examined by Lucille Nowell both in her dissertation (L. T. Nowell, 1997) and later work (L. Nowell, Schulman, & Hix, 2002), where based off quantitative data, the following relationships for her work and the works of Christ (1975), Mackinlay (already mentioned), and Cleveland (1984) are presented Table 7.

Based off these works, it is expected that: NC > NS > NG in terms of accuracy and NC < NS < NG for time.

Table 7: Reconstructed from Nowell's work (Nowell et al., 2002), edited to include a total column based on quantitative data, where the lower number indicates a better result in their respective experiments.

	Time	Error	Christ's Search Time	Mackinlay	Cleveland	Total
Color	1	1	1	2	2	7
Shape	2	2	3	3	3	13
Size	3	3	2	1	1	10

- H2a: In terms of accuracy, the comparisons of shape, size, and color is expected to be: NC > NS > NG (higher accuracy value)
- H2b: In terms of completion time, the comparisons of shape, size, and color is expected to be: NC < NS < NG (lower completion time)

3.3.2 Affordances by perceived utility and confidence of the visual aesthetics

One major difference between the Thesis Experiment and the Pilot Study was the questionnaire screen, as the questions themselves were changed to address the issues of perceived affordance, the graphs' subjective attractiveness, and the confidence associated with visual aesthetics.

Though the model of affordance as defined by Colin Ware, in terms of the utilization of Gibson's Affordance Theory (Gibson, 1977), is "perceivable possibility of action (Ware, 2012), page 19" and later "... cognitive constraints relating to the user's understanding of the data space (Ware, 2012), page 356." Even in the original work, Gibson makes note that affordances could be seen as a double-edged sword, both in enhancing the system's usability and hurting the system's usability, if done recklessly. That stated, the recruitment of this study was not limited to just experts in network graphs, so this becomes a much more divided question as it is, "relating to the user's understanding of the data space." All participants subjected to this study will be told that the visual aesthetics are a means of understanding the topological graph structure, but never deliberately told that certain visual aesthetics will enhance their search or hinder their search. This could cause a clear divide in the participants who have experience with network graph and those who do not. This value is obtained using two follow-up questions:

- "The aesthetic helped in completing the task."
- "I felt confident in choosing a node or nodes that was/were the most connected."

This idea was examined in the Pilot Study, but limited to a single question ("The aesthetic helped in completing the task"), but it was believed that this needed to be expanded on with the Thesis Experiment to better evaluate affordance. In the Pilot Study, the participants' recalled the node aesthetics (both encoding and styling) as more helpful to the control, link aesthetics, and group aesthetics.

- H3a: Compared to the control, node based visual aesthetics (NS, NG, NC, and NO) will have higher rate in affordance than the control, regardless of experience. For both:
  - "The aesthetic helped in completing the task."
  - "I felt confident in choosing a node or nodes that was/were the most connected."

For the other two visual encoding (TG, GCo), it is believed that these visual aesthetics will also have a higher degree of affordance than the control. This will be limited to, "The aesthetic helped in completing the task," as they will be evaluated as a superior form versus the control, though they will not be seen as a means that provides confidence to the given task.

- H3b: Compared to the control, link gradient (LG) and group coloring (GCo) will have higher rate in affordance than the control, regardless of experience. For just:
  - "The aesthetic helped in completing the task."

The visual styling will not appear to have a higher affordance level versus the control, as they do not provide additional guidance to the participant in completing the task.

- H3c: Compared to the control, link tapered (LTa) and group convex hull (GCh) will the same affordance level as the control.

3.3.3 Correlating Subjective Beauty to Task Focusing Time

This measure was tested in the Pilot Study, but the task focusing time (inverse task abandonment) was compared for accuracy and perceived affordance of the task, and not the subjective beauty of the given aesthetics. This idea of task abandonment centers on work by Cawthon and Moere (2007) and again later by Moere et al. (2012), specifying that the "latency in task abandonment .. (is) correlated to a visualization's perceived beauty (Moere et al., 2012)." Within both of Moere's work, there was a lack of eye-tracking to calculate this measure. This was altered slightly to fit more with the ideas of "task abandonment," but using it as a means to calculate task focusing time. Task focusing time was therefore defined as the amount of time that the participant spends focused on the network to complete the task. This was done for both the small and large networks separately, as the number of nodes may influence the amount of time spent focusing on the graph.

- H4a: Results of subjective, perceived view of beauty for each aesthetic will be correlated with the larger amounts (in terms of time) of task focusing time for small graphs.
- H4b: Results of subjective, perceived view of beauty for each aesthetic will be correlated with the larger amounts (in terms of time) of task focusing time for large graphs.

3.3.4 Exploration, User Behavior, and Strategies Given to Visual Aesthetics

One of the first utilizations of the eye-tracking is applied to two different means of understanding of the participant's exploration and strategic means of completing the task. Previous work by Netzel et al. (2014) analyzed average saccade length based on their various link aesthetics, where average saccade length based on long scans were seen as exploratory and short scans as confusion. Based on the research questions posed in the Pilot Study, a slight variation between node aesthetics versus the control, link aesthetics, and group aesthetics was addressed, but with limited understanding of the participant's intent. It was hypothesized that the reason for the increase in searching was due to the node aesthetics providing a more "shotgun" search, as the task solution was not regulated to the network itself. Review of the gaze replays of the Pilot Study showed that eye-movements tended to use the links to traverse the network, even when node visual encoding was utilized. Contrary to this approach, the link gradient should limit the eye-movement, as the traversal of the network uses the link as its main pathway. The link gradient was designed to promote directional viewing to specific areas.

- H5a: Compared to the control, node aesthetics (NS, NG, NC, NO) will have longer average saccade lengths.
- H5b: Compared to the control, node aesthetics (NS, NG, NC, NO) will more fixation.
- H5c: Compared to the control, link gradient (LG) will have shorter average saccade lengths

With the qualitative feedback from the retrospective think-aloud (RTA) protocol, the amount of exploration (saccade lengths) should result in more in-depth or more numerous strategies in solving the task. As this is qualitative feedback, a research question will be devised to compare these results.

- RQ1: Based on the RTA, more strategies will be developed based on the amount of exploration (longer saccade lengths) utilized by the participants.

3.3.5 Edge-Crossing vs. Degree Centrality: Variation by Visual Aesthetic

A focal point for much of the past literature in network graph visualizations are the issues caused by edge-crossing (Purchase, 2002; Purchase et al., 2002; Ware et al., 2002). Edge-crossing itself can be limited, but in some situations, unavoidable. The variation of the visual aesthetics and their effects on accuracy have already been addressed, but looking beyond this is a need to compare the visual aesthetics against the edge-crossing, to see if a certain visual aesthetics is utilized, that the edge-crossing effect could be minimized. It is not believed that visual aesthetics alone can completely remove this bias, but could greatly reduce its emphasis in the network visualization domain. This will be addressed by two types of points of interest (POI): 1) the nodes with the largest degree centrality (the target of this task) and 2) areas with edge-crossing. The node visual encoding can create a situation where traversal of the network is moot, because solution can be found independent of the topological structure. It is hypothesized that the node visual encoding (NS, NG, NC) will have a superior positive influence on highlighting the degree centrality POI and minimize the edge-crossing POIs.

- H6a: Compared to the control, the node visual encoding (NS, NG, NC) will be able to highlight the degree centrality POIs.
- H6b: Compared to the control, the node visual encoding (NS, NG, NC) will be able to reduce the distraction caused by the edge-crossing POIs.

The link visual encoding (LG) should be able to highlight the degree centrality POIs, but still requires a need to use the edge to traverse the network structure. So for the LG, edge-crossing POIs will still have similar results to the control.

- H6c: Compared to the control, the link visual encoding (LG) will be able to highlight the degree centrality POIs.

3.3.6 Change Blindness Based on Network Complexity

One main thought that was discussed in the analysis of the Pilot Study that needed to be addressed in the Thesis Experiment was if change blindness would/could occur through the phases of the trial. Change blindness being an inability of the observer to detect change (even large changes) in a visual scene (Wickens et al., 2004); specifically in the Pilot Study, in which the rotation of the graphs ($0°, 90°, 180°, 270°$) and corresponding visual aesthetics would have an adverse effect on the viewer of the graph in terms of working or short-term memorization. This idea has been addressed before in a subtopic called difference maps, but has not been tested to provide conclusive results that change blindness is a necessary issue caused by network graphs (Archambault et al., 2011).

However, for this specific project, change blindness had more to do with the possibility of participants learning from previous network graphs because the visual aesthetics were kept constant per network graph and the only variation was the rotation. This is tested in the RTA section of the study and validated based on the number of selections and correct selections, as it would be expected that if change blindness does not occur, the accuracy will increase to a perfect mark (F1-Score = 1) and time to completion will decrease with trails. More details about the RTA are presented in Section - 3.4.8. Participants are split into three groups, each group seeing all the same visual aesthetics in the same order, but with different network graphs. Group 1 will see a complex graph (as determined by its poor accuracy level in the Pilot Study), group 2, a mid-complex graph, and group 3, a simpler graph. Three possible outcomes are predicted. This can be verified by accuracy values for each participant:

- H7a: Participants will not be able to tell the same graph was used multiple times in a row for all three groups (visual aesthetics affecting change blindness)
- H7b: Participant will be able to tell the same graph was used multiple times in a row, but only certain groups (network complexity affecting change blindness)
- H7c: Participants will be able to tell the same graph was used multiple times in a row for all groups (no change blindness)

3.4 EXPERIMENTAL DESIGN

In terms of the design of the experiment, there are a few similarities to the Pilot Study, but all parts are discussed in more detail in this section. The layout of the experiment can be seen in Figure 56, where the width of the section generally defines the amount of time relative to one another.

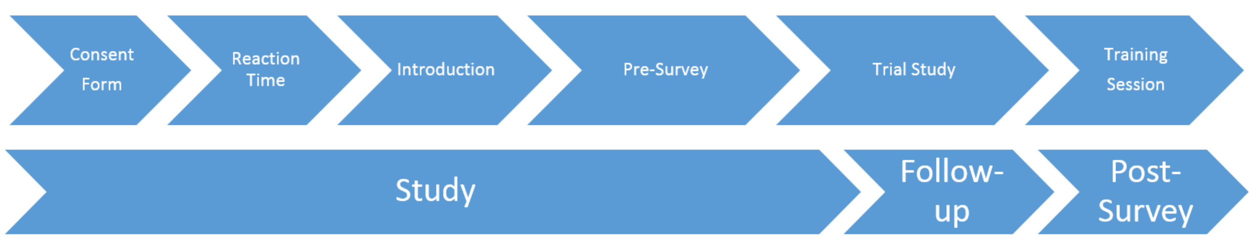


Figure 56: Thesis Experiment (first half - top and second half - bottom)

3.4.1 Consent Form/Incentive

When the participant arrived, they were greeted and provided a consent form for the study. They were informed of the requirements (University of Pittsburgh student, 18 years old or older, and 20/20 vision without the need of glasses or contacts) and that they would be paid $10 for their time. Once this was completed, a second form was utilized to allow the participant to opt into the incentive mentioned in Section - 2.7.1 (Motivation/Incentive), which all participants did. It was noted to the participant that they would receive payment for the study, regardless of accuracy scores and that the (Motivation/Incentive) should be seen as an inducement for their time and efforts.

3.4.2 Reaction Time

A reaction test was utilized to collect preliminary data on the participant based on their interaction with a Fitt's law interactive website (Wichary, 2005). This was only used to collect data on the participant's ability to select objects on a screen in coordination with moving the mouse to various designated parts of the screen. Four experiments are given on the website, where the first was used as a demo, the second as training, and the third (Figure 57) and fourth (Figure 58) as data. The data compared all participants for outliers. These outliers may have needed to be removed from the study based on their ability or inability to react quickly enough to objects on a screen, as the study focused on rapid reaction to visual aesthetics' properties on a given network graph. Of the 189 total objects selected by the participants, 27 participants x (3 objects in exp 3 + 4 objects in exp 4), only 6 outliers were recorded. No significance was found for the participants, as in no participant was seen as significantly better or worse than the other participants, which meant all participants were kept in the study, and there was no need to bias or weigh results.

3.4.3 Introduction

Participants were provided a short tutorial on the basics of network graphs, which included using four actors interacting with one another and how this would transcribe to the node-

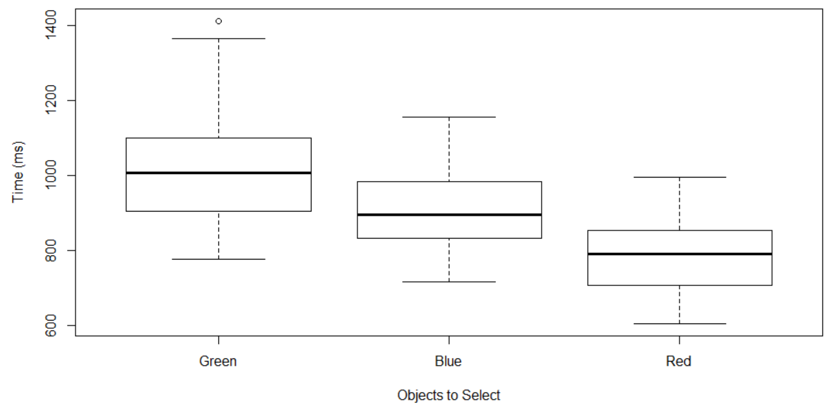


Figure 57: Results of Fitt's law test for experiment 3

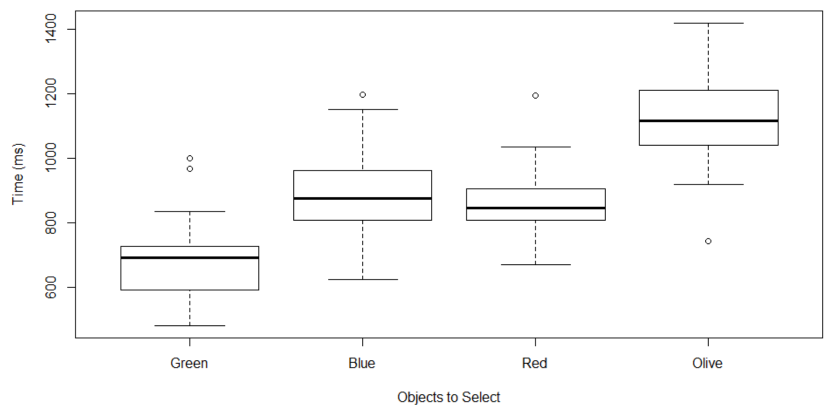


Figure 58: Results of Fitt's law test for experiment 4

link or network diagram. The participants were told some basic terminology, such as: nodes, links, communities, paths, and given task for the study (finding the most connected node or nodes in a given network).

3.4.4 Pre-Survey

Once this was completed, a pre-survey (Appendix A) was given to the participant. This requested minimal information about the participant, including gender, date of birth, and experience with using networks or node-link diagrams. This also included an example network, created using Microsoft Visio, with eight questions in regards to this network. The example network and questions were used to validate the participant's understanding of a network graph, again at very high level.

3.4.4.1 Participants 27 participants[1] were recruited on the University of Pittsburgh campus, ranging in disciplines from information science, engineering, statistics, computer science, and humanities. This included 13 males and 14 females, 11 self-reported being an expert or knowledgeable with network graphs, node-link diagrams, or force-directed graphs, and 16 self-reported having no prior knowledge of this subject matter. The average age of the participant was $\mu = 25.5$ and $\sigma = 9.24$, with one participant abstaining from providing this information.

3.4.4.2 Results of Pre-Survey Example Graph The pre-survey (Appendix A) contained an example network graph and 7 questions about the graph. The participants were allowed to ask questions for clarification purposes (3 participants needed to do this). Out of 189 questions of the pre-survey example (7 questions x 27 participants), only 4 questions were missed, resulting in a 97.8% accuracy level. Therefore, it was felt that the participants, regardless of experience, understood a basic network graph and had enough high-level knowledge to complete the task.

3.4.5 Trial Study

In the trial study, participants were able to try the software utilized in the experiment. This included all the screens presented in Table 3. They were shown how to use the software

[1] The study was approved by the Institutional Review Board at the University of Pittsburgh and consent forms were collected from all participants

and allowed to try the software on their own as much as they needed to feel comfortable with the interface. The networks that were shown were from the previous experiment and no aesthetics were applied to the graphs. The timing utilized in the actual study for the network (10 seconds) was used as well in the trial version, but the participants were not told the exact amount, only that they had limited time that was kept consistent throughout the experiment. Also, during the trial phase, they were allowed to ask clarification questions about the interface because this time was meant to prepare the participants for the study.

3.4.6 Aesthetics Training Session/Eye-Tracking Calibration

Finally, the participant was shown a "slideshow-like" example of aesthetics being utilized in the experiment. The aesthetics were shown one at a time. During the control aesthetics (the first aesthetic shown to the participants), the participants were told which node had the most connections in the graph. In Figure 59, all nine aesthetics are provided and they were shown to the participants from right to left, top to bottom. The control aesthetic provides the solution of the most connected node in the graph, as this was pointed out to the participant during this phase. As the aesthetics were shown, they were not told how the visual aesthetic highlighted the most connected node(s), only for what level the aesthetic was designed (node, link, or group level). To emphasize this point, here is a detailed progression of the training session:

1. Participant was shown the control aesthetic
2. The most connected node was pointed out to the participant
3. The node size (NS) aesthetic was shown to the participant (without mentioning that the size was indicative of the number of connections)
4. Step 3 is repeated for the node glyphs (NG), node coloring (NC), node opacity (NO), link tapered (LTa), link gradient (LG), group convex hull (GCh), and group two-encoding coloring (GCo) (again, without telling the participants the visual aesthetics affordances in regards to the task).
5. Participant was allowed to ask clarification questions about the visual aesthetics.

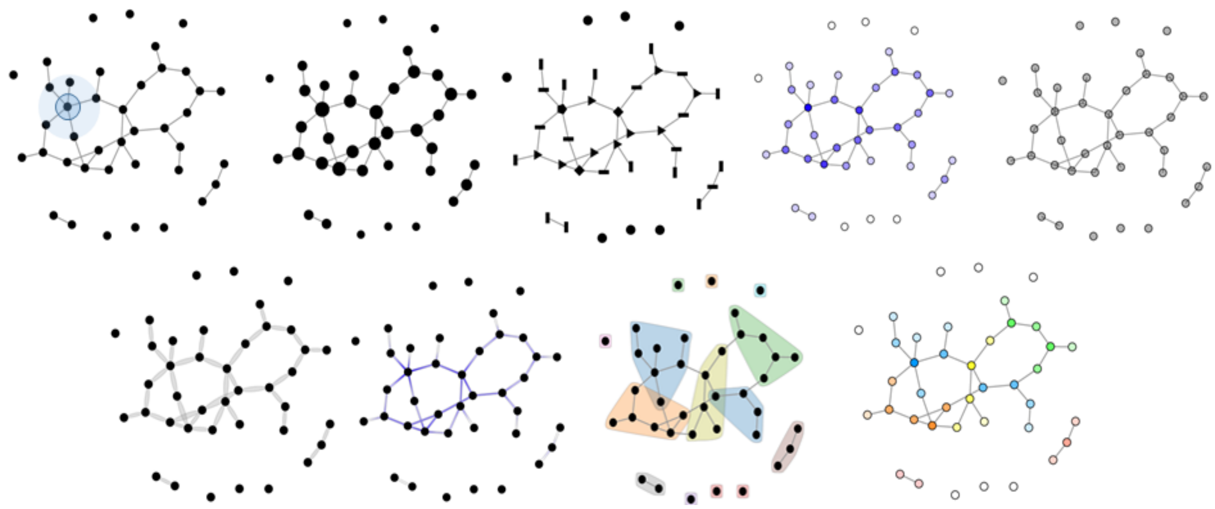


Figure 59: Aesthetics training session provided to the participants, where the most connected node in the graph is highlighted in blue for the control, as this node was noted to the participant at the start of the training session.

a. If the participant asked if the visual aesthetic provided a certain affordance (example: size indicates number of connections), they were told to keep in mind that the visual aesthetics were designed to help clarify the readability of the network graph. They were not told the actually meaning of the affordance.

The lack of clarification of the affordance of the visual aesthetic was crucial for the experiment, as this was one of the focal points of the experiment.

After this was concluded, the eye-tracking software was calibrated to the participant and the main part of experiment commenced.

3.4.7 Primary Experiment

The primary experiment was very comparable to the Pilot Study, where networks and visual aesthetics were selected using the randomized blocking method. Figure 60 shows an example of this selection per participant, as of participant 1, 2,...,9, where P_1 is participant 1, P_2 participant 2, and P_9 is participant 9.

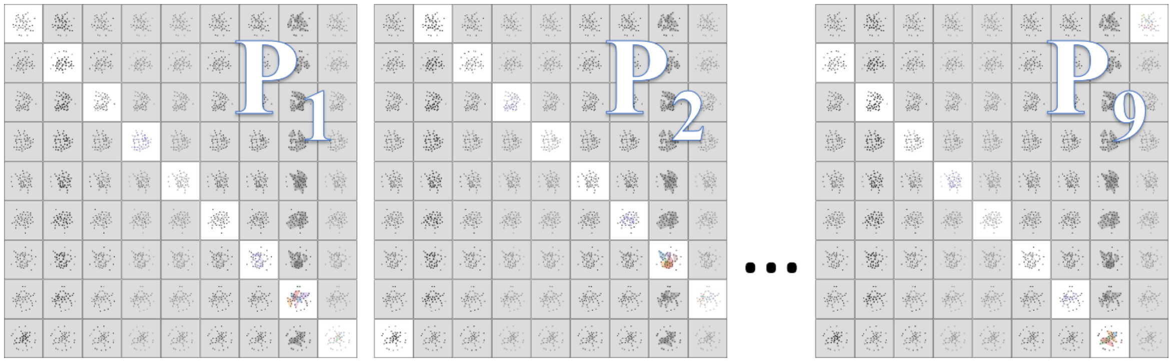


Figure 60: Shows highlighted areas for each participant (specifically P_1, P_2, and P_9)

Once the block was selected (based on the number of the participant), the selected networks were randomized for both the small and large network, and the two rotations $(0°, 90°)$, as a single group to counterbalance the networks. This would result in the following:

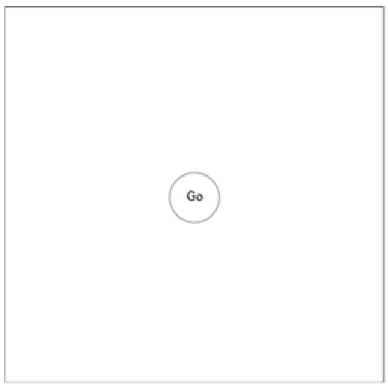
$$\begin{array}{c} (9\ small\ networks + 9\ large\ networks) \\ x\ 2\ rotations\ (0°, 90°) \\ \hline 36\ total\ networks\ per\ participant \end{array}$$

3.4.7.1 Interface Design Virtually the same interface was used in the Thesis Experiment as was used in the Pilot Study. The only differences were found on the "Follow-up" screen, which provide: 1) three questions instead of six, 2) a white background for the questions to improve readability, and 3) slight a change to the text to evenly space out the response from the participants (1 - 5 responses), Table 8.

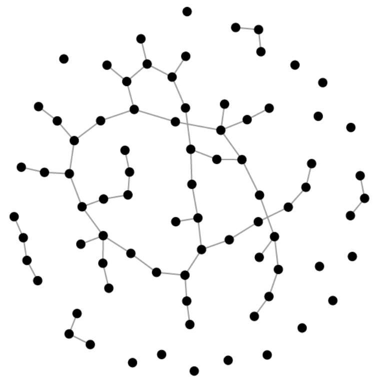
The three new questions presented were:

1. The aesthetic helped in completing the task.
2. I thought the graph was aesthetically pleasing to look at.
3. I felt confident in choosing a node or nodes that was/were the most connected.

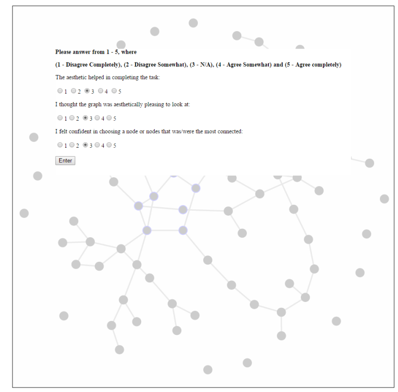
Table 8: The 3 screens that make up the interface for the Thesis Experiment.



Screen 1 (Warm-up Screen) Highlighting the starting button to move to the next network. The button and the word "Go" were enlarged only for clarity.



Screen 2 (Network Screen) The network and the screen where the user could interact with the network.



Screen 3 (Follow-up Screen) The questionnaire that would follow each network with a transparency to view the past network for better clarity in their response.

3.4.7.2 Accuracy Measure The same accuracy measure was used as in the Pilot Study, so an abridged version of the accuracy measure is presented.

A signal-detection accuracy measure was implemented because it gave equal weighting to all interactions (Wickens et al., 2004). To better quantify this relationship, a F1-Score was used. This is a measure that combines both the precision (or positive predictive value, PPV) and the recall (or true positive rate, TPR). In other words, in a selection task (as in selecting the most connected node), PPV would be equivalent to the number of correctly selected nodes that were the most connected, over the total number of selection, or the precision. The TPR would be equivalent to the number of correctly selected nodes that were the most connected, over the total number of conditional positive values. The harmonic mean of these two values was designated as the F1-Score or F-measure. The conditions for the F1-Score went as followed:

- T_p= subject selecting a node and correctly identified the most connected node(s)
- T_n= subject not selecting a node and correctly rejected the node as not the most connected node(s)

- F_p= subject selecting a node that was not the most connected node(s)
- F_n= subject not selecting a node and the node was the most connected

3.4.8 Follow-up Experiment/Retrospective Think-Aloud Interview

A major change to the experiment was the inclusion of a retrospective think-aloud (RTA) interview. The term RTA was first coined in Hyrskykari et al. (2008) work, where it was determined that RTA was found to be significantly better than conventional think-aloud protocol (CTA). CTA is commonly used in usability testing by having the participant think-aloud or talk-aloud as they are completing the task. This method can help in better understanding strategies utilized by the participant to complete the task and can help combine both the intent of the user and the eye-movements they are exhibiting. A shortcoming of this approach is the unnatural means of collecting this data, as typical interaction with most systems, people tend not to explain how to complete a task while actually completing the task, at the same time. RTA uses the eye-tracking results in gaze-replays, and allows the participant the ability to see their eye-movements and add commentary based on reliving this interaction. To utilize this RTA, a second set of eye-tracking data was collected after the primary experiment. The RTA was used to:

1. Gain a better understanding of the participant's actions or strategies in completing the task.
2. See if the aesthetics and rotation would be enough of a change, that the participants themselves (and verified by the data collected) did not notice they were being subjected to the same network structure on repeated measures.

This second point is worth mentioning, as this part of the study was not used in the accuracy/time calculation, so there was no need to use the same randomized blocking design as in the main study. This was only meant to see if 1) they noticed the same network structure, and 2) how they completed the task. By this point in the study, the participants would have had visual contact with each of the aesthetics, once at the training phase, and four times in the study itself. This was meant to only provide the strategies in this specific case (visual aesthetic changes). How this was constructed was by utilizing the network accuracy

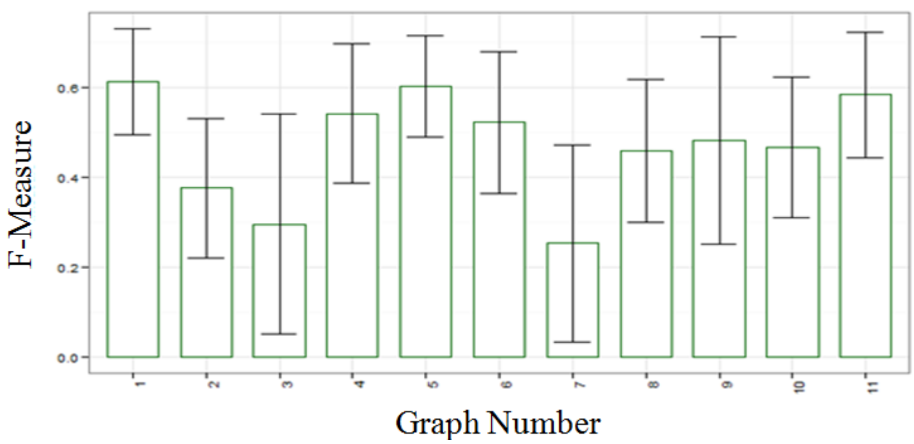


Figure 61: The results of network topology performance for large networks (80 nodes) and shows significance for network 2, 3, and 7 (p < .01)

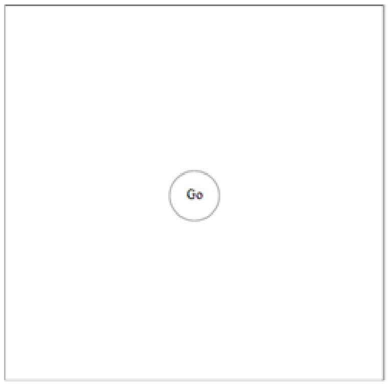
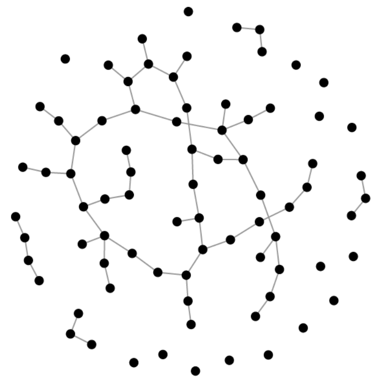
results from the Pilot Study. Figure 61 shows the results for large graphs and their accuracy measures. From these results, graph 7 (the lowest accuracy measure), graph 10 (the mean accuracy measure), and graph 1 (the highest accuracy measure) were split between the 27 participants (1 group with graph 9, 1 group with graph 7, and 1 group with graph 1).

The participants would see the same graph nine consecutive times using only the "Warm-up Screen" and "Network Screen," but the graphs would 1) be rotated each time by 90° and have a different visual aesthetic (they would go in the same order: C, NS, NG, NC, NO, LTa, LG, GCh, GCo). Table 9 showcases the screens used in the RTA and Figure 62 illustrates the three networks used in RTA, where participants (1-9) would see n_7, participants (10-18) would see n_9, and participants (19-27) would see n_1.

The RTA took roughly 2 minutes to complete and afterwards, the participants were asked:

1. Did they notice that the visual aesthetics changed, but the network structure did not?
   a. After they responded, they were shown a direct example of this rotation to confirm their answers
   b. The accuracy measures would be utilized to confirm their answers

Table 9: The interface used in the retrospective think-aloud protocol.

	
Screen 1 (Warm-up Screen) Highlighting the starting button to move to the next network. The button and the word "Go" were enlarged only for clarity.	Screen 2 (Network Screen) The network and the screen where the user could interact with the network.

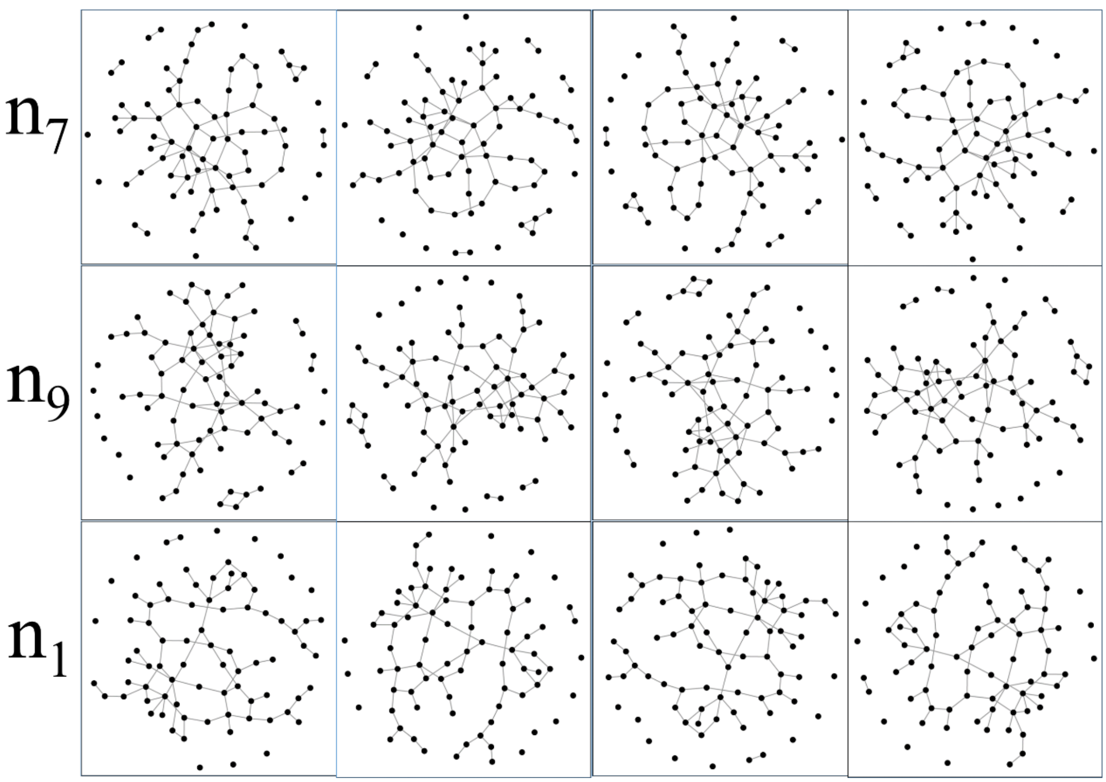


Figure 62: Illustrates the three networks utilized, rotated: $0°, 90°, 180°, 270°$

2. (While watching each visual aesthetic) how they completed the task and further clarification of strategies used based on the visual aesthetics.

This was only utilized if they did not provide this information during question 2. The RTA was used to help resolve the hypothesis related to "Utility of the Visual Aesthetics" and "Change Blindness Based on the Network Complexity."

3.4.9 Post-Survey

Finally, a post-survey was provided to the participants (Appendix B). The questions used a 1-5 Likert scale and were used to evaluate the design as a whole. Also, participants would be able to provide feedback in their own words in regards to the study. Overall, the participants found the task both easy to complete (μ = 3.89 and σ = .993) and easy to understand (μ = 4.89 and σ = .314). Participants would recommend this interface to other people (μ = 4.07 and σ = .766), as the found the visualizations to be visually pleasing (μ = 3.81 and σ = .818). Lastly, even though they were limited in the amount of time given to them for the task (10000 ms), they still thought this sufficient in completing the task (μ = 3.70 and σ = 1.41).

Out of the 7 comments left in regards to the study, 6 made positive remarks, including "Cool study :)", "enjoyed the study", "That is a great study. I like so much the project in general", "Interesting study. Lots of good questions here. I wonder how aesthetics relate to network complexity and if different aesthetics do different things.", "Its [sic] very interesting!", and "Well coordinated and explained/very interesting task, made me think." The other comment was in regards to the visual aesthetics themselves and was reported in the RTA.

4.0 RESULTS, DISCUSSION, AND CONCLUSION

4.1 UTILITY OF THE VISUAL AESTHETICS

To evaluate the H1(a-d) hypotheses, the analysis took advantage of the randomized blocking design by including all nine aesthetics across both the small network (n_s) and large network (n_l) for the 27 participants. Similar to the Pilot Study, time limiting factors enforced the equal potential for participants to select both the most connected node(s) and non-most connected node(s), with the equal potential to miss the most connected node(s) and reject non-most connected node(s). Statistical significance was tested based on ANOVA using a linear model with blocking on the network graphs, followed by Tukey HSD pairwise test. This is reported using $F(a,b) = c$, where a is the degrees of freedom, b the residuals, and c the F value, similar in Netzel et al. (2014). This was first done across the F1-Score (accuracy), which produced significances for the aesthetics $(p < .001 - F\,(8, 804) = 15.390)$.

4.1.1 F1-Score (Accuracy)

At the accuracy level, evaluating H1a looked at comparing the control to the node based visual encoding (NS, NG, NC). This is where the Tukey HSD pair-wise test was utilized only on the factor of the control (C) versus all the visual aesthetics. Based on this, the H1a was confirmed for all node visual encodings, with (NG, NC - $p < 0.01$) and (NS - $p < 0.1$). H1b (accuracy) was also confirmed based on factor (C) vs. (LG), which produce significances (LG - $p < 0.1$). H1c (accuracy) was not confirmed due to lack of significance compared to the control. H1d was confirmed based on the lack of significance compared to the control, which was expected from the Pilot Study. Finally, H1e was not confirmed as no significance was found.

4.1.2 Time to First Correct Selection (Time)

For time variation, "first correct selection" was used because it was possible to not select all of the most connected node(s). If the participant did not select any of the most connected nodes, that trial was given a maximum time of 10000 ms. Again, the Tukey HSD was utilized for confident levels of .01, .05, and .1. For H1a, this was confirmed for all node visual encodings, with (NG, NC - $p < 0.01$) and (NS - $p < 0.05$). H1b and H1c were rejected because no significance was found. H1d was confirmed, as the visual styling (NO, LTa, and GCh) had no significance. H1e was not significant, establishing (NO) was not superior to (GCh and LT) based on time. Figure 63 provides both the accuracy and time dimensions for all aesthetics and all graphs, Figure 64 for large graphs, and Figure 65 for small graphs.

4.1.3 Color, Shape, and Size

Based on previous work, two hypotheses were proposed: NC > NS > NG in terms of accuracy and NC < NS < NG for time. For the accuracy, it can be seen in Figure 63 that this relationship does not hold true, as NG > NC > NS is clearly the dominant sequence for this ordinal data set. As mentioned in the Section 3.1.1.2- Node Aesthetics (NA), the node aesthetics were developed to maximize the information or data given to the participant. This relationship is discussed in more detail through the qualitative analysis to gain a better understanding why this occurs. To strengthen this argument, Tukey HSD was used to compare the difference of means of each of the node aesthetics (NS, NG, and NC). For the confidence levels of .05 (Figure 66), node glyphs are significantly more accurate than the node size and color. Reducing this confidence level to .15 (Figure 67), node color shows significance to node size. The statistical significance is then given as " > " for the ($p < 0.05$) and " $\geq$ " for the $p < 0.15 : NG > NC \geq NS$. For the time dimension, the same relationship is shown as NG < NC < NS, but without having any statistical significance. With this result, the focus of the discussion is limited to just the accuracy level of analysis.

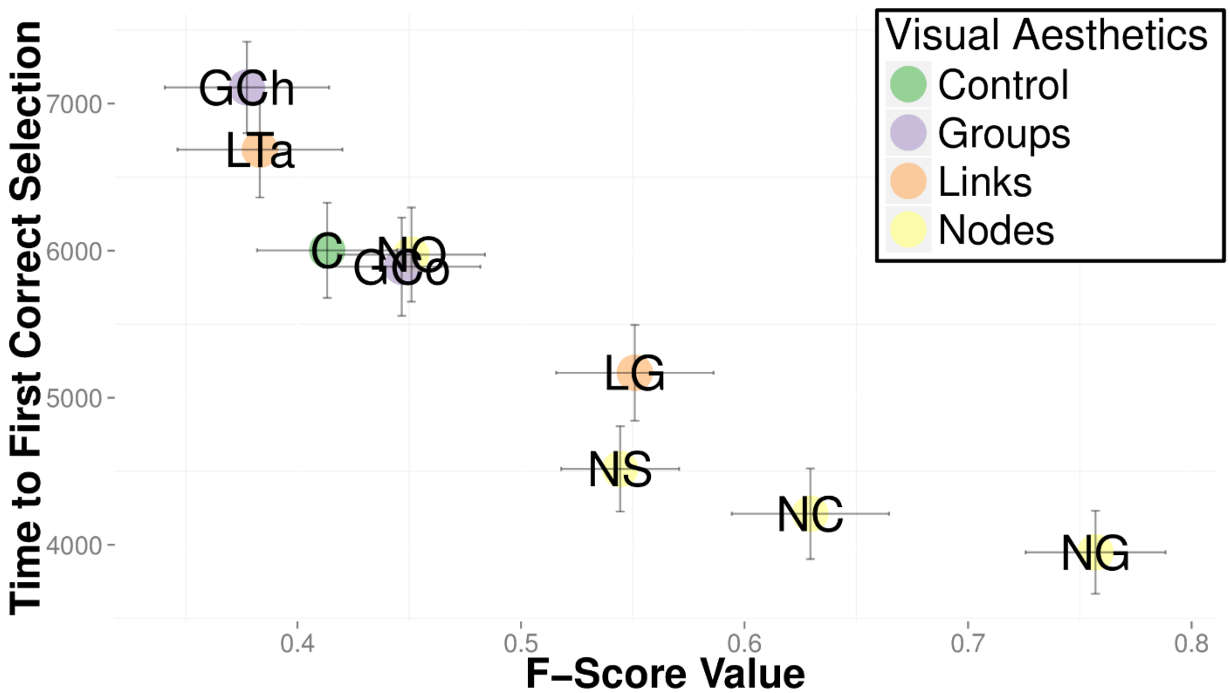


Figure 63: (All Networks) Task performance measured based on accuracy (F-Score) and time, for all aesthetics

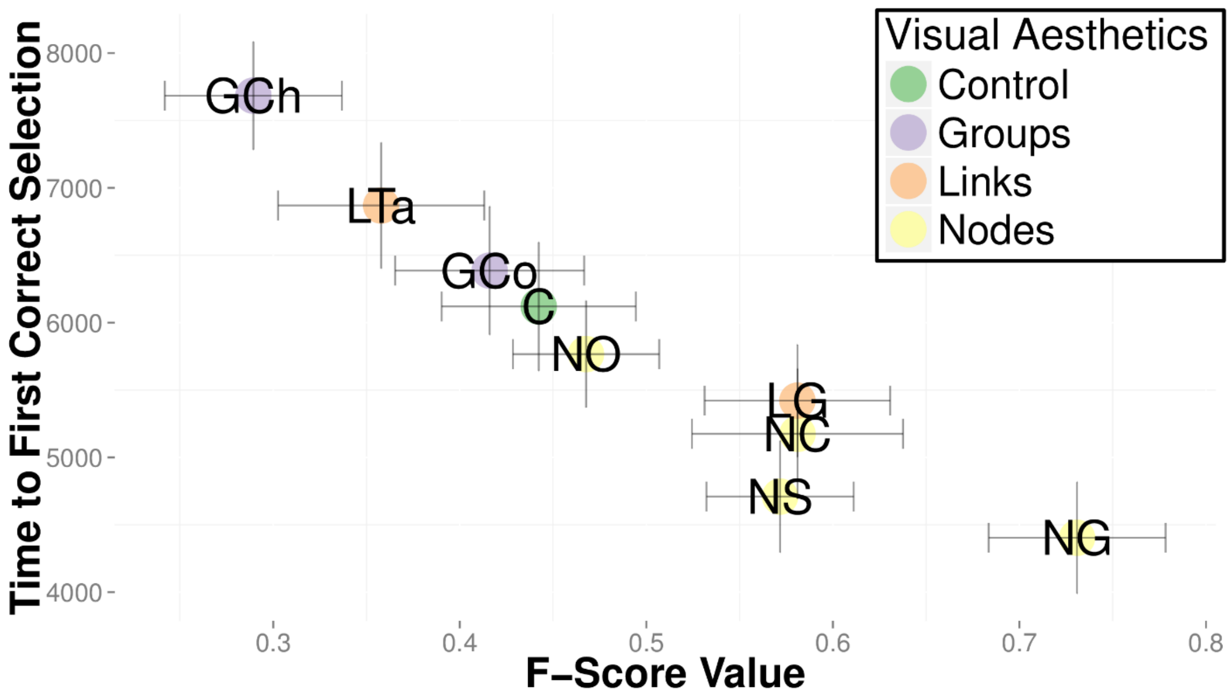


Figure 64: (Large Networks) Task performance measured based on accuracy (F-Score) and time, for all aesthetics

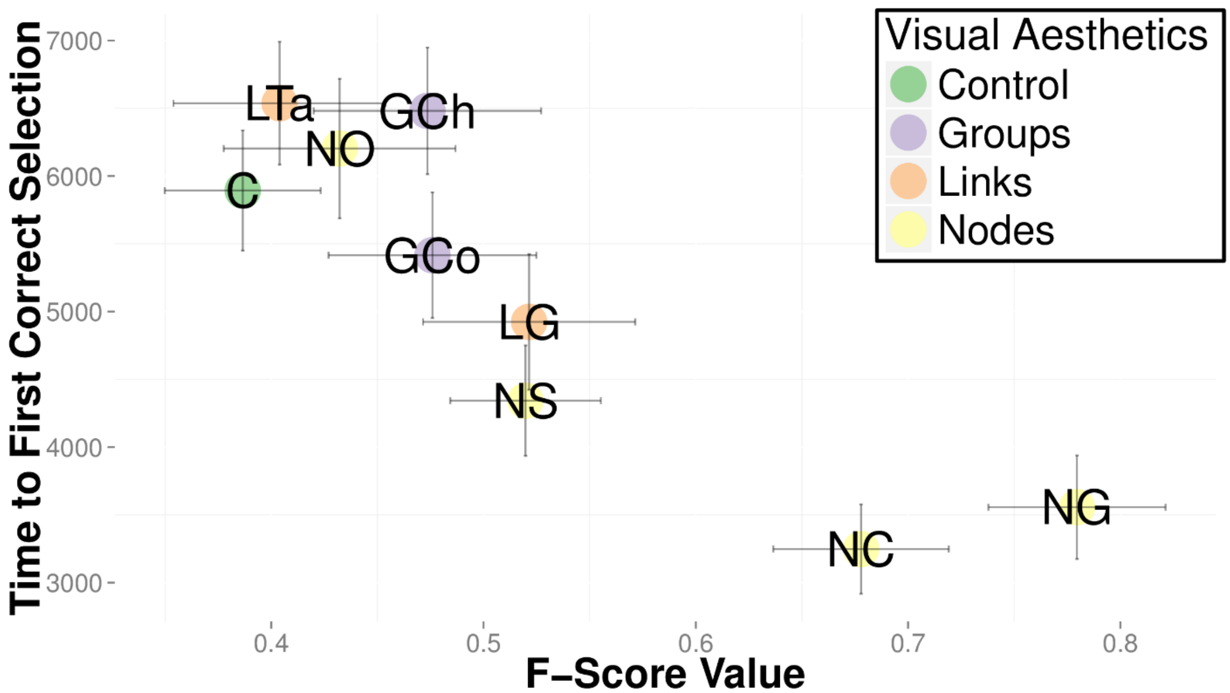


Figure 65: (Small Networks) Task performance measured based on accuracy (F-Score) and time, for all aesthetics

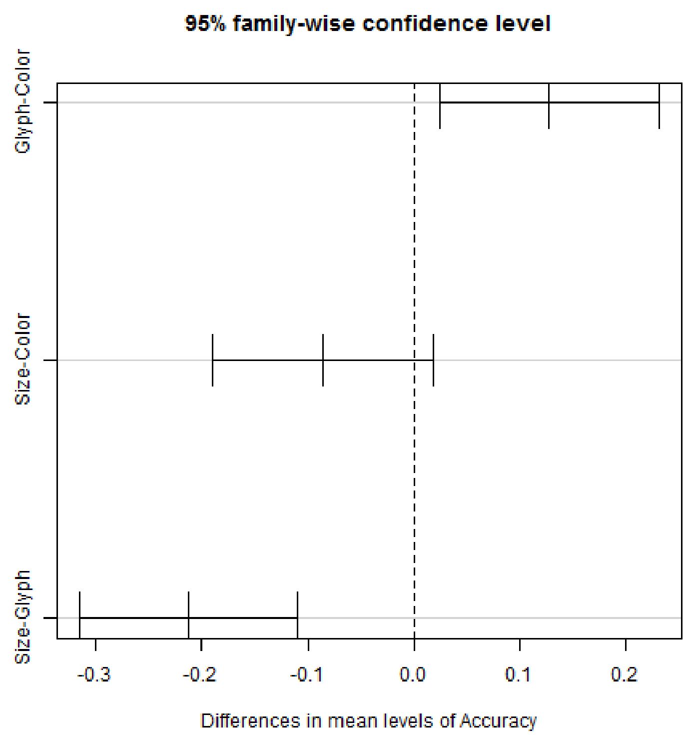


Figure 66: Tukey HSD one-tailed test for confidence level of 95%

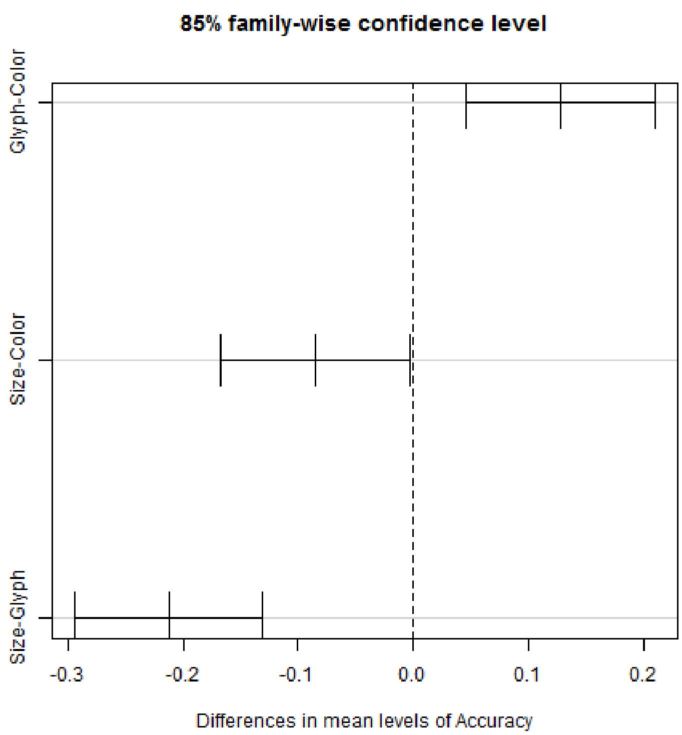


Figure 67: Tukey HSD one-tailed test for confidence level of 85%

4.2 AFFORDANCES BY PERCEIVED UTILITY AND CONFIDENCE OF THE VISUAL AESTHETICS

Affordance for the visual aesthetics is judged by two means: 1) the perceived utility (Q1: "The aesthetic helped in completing task"); and 2) the confidence of the participant (Q3: "I felt confident in choosing a node or nodes that was/were the most connected"). The same statistical analysis was used in the Pilot Study based on non-normalized Likert values. That meant using Kruskal-Wallis one-way ANOVA to show significance overall and Siegel and Castellan post-hoc for Kruskal-Wallis (McCrum-Gardner, 2008; Siegel & Castellan, 1988) for the conditions (visual aesthetics). Both the questions (Q1 and Q3) were significant in terms of visual aesthetics being utilized ($p < .01$). Each question is addressed separately based on hypotheses (H3) for this section.

4.2.1 "The aesthetic helped in completing the task."

Each visual aesthetic was designed to provide a new means of interpreting the network graphs being analyzed. Hypotheses H3a and H3b specified that the majority of the visual aesthetics would be perceived as significant versus the control. To test this, Kruskal-Wallis was utilized producing significant results ($\chi^2 = 237.3481$, $df = 8$, $p < .001$). Using a Siegel and Castellan suggested two-tailed pairwise test for ($p < .01$) produces significance for (NS, NG, NC, NO, LG, and GCo), proving the first part of H3a, H3b, and H3c, which specified that LTa and GCh, based on being stylings unrelated to the node, would not have significant values compared to the control. Figure 68 provides a histogram of the given responses, where 1 indicates, "Disagree with Completely," 3 indicates, "Neutral Opinion," and 5 indicates "Agree with Completely."

4.2.2 "I felt confident in choosing a node or nodes that was/were the most connected."

The same analysis was done for Q3 ("I felt confident in choosing a node or nodes that was/were the most connected"), which showed significance using the Kruskal-Wallis rank

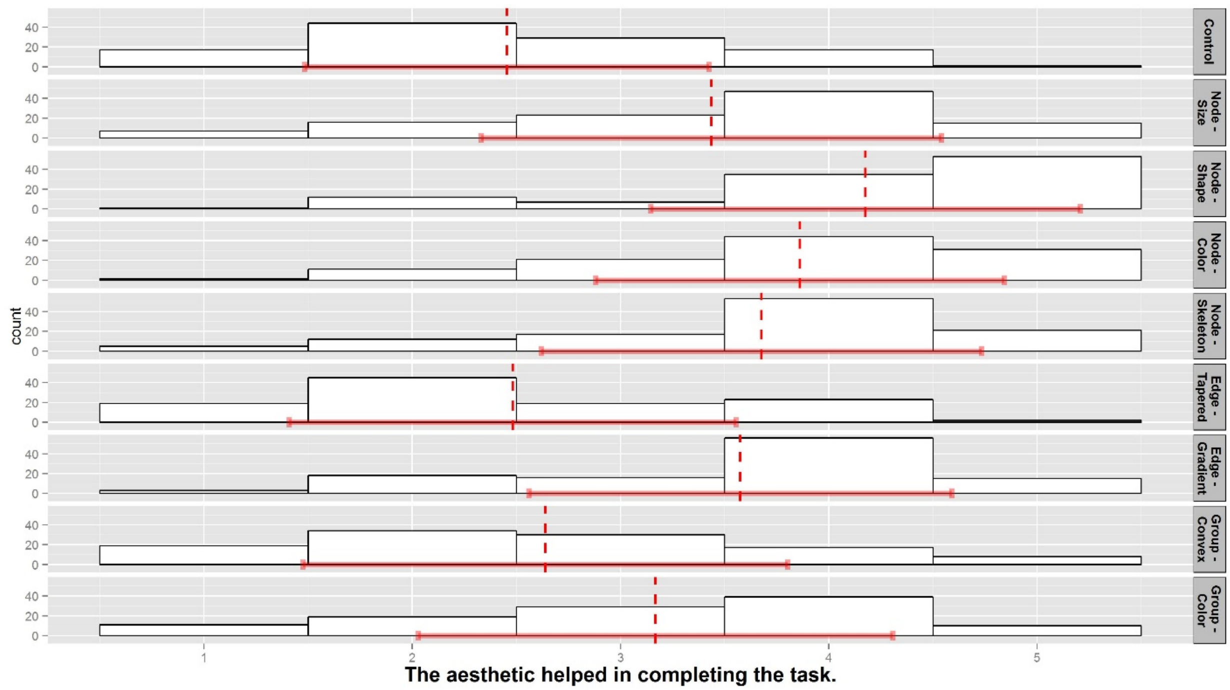


Figure 68: Results of the follow-up question, "The aesthetic helped in completing the task." The vertical red line indicates the mean and the horizontal red line indicates the standard deviation.

sum ($\chi^2 = 142.5345,\ df = 8,\ p < .001$). Using the multiple comparison test with ($p < .01$) produces significance for (NG, NC, NO), which proves H3a for conditions (NG, NC, NO), which meant the node size (NS) was disproven for this affordance marker. Figure 69 illustrates this histogram of the given responses, where 1 indicates, “Disagree with Completely,” 3 indicates, “Neutral Opinion,” and 5 indicates, “Agree with Completely.”

4.3 CORRELATING SUBJECTIVE BEAUTY TO TASK FOCUSING TIME

Correlating the relative and subjective beauty of the visual aesthetics and the task focusing time for each aesthetics requires two measures: 1) the participant feedback for Q2, “I thought the graph was aesthetically pleasing to look at,” and 2) eye-tracking results and calculation of the task focusing time.

4.3.1 Subjective Beauty

For the participant feedback, the Kruskal-Wallis rank sum results as ($\chi^2 = 145.0557,\ df = 8,\ p < .001$), which showed significance for the visual aesthetics. Using the multiple comparison test with ($p < .01$) produces significance for (NG, NC, NO, LG, GCh, GCo), which meant that both the node size and link tapered equal or near the same level as the control. Figure 70 provides insight into specific values of this question using a histogram of the given responses, where 1 indicates, “Disagree with Completely,” 3 indicates, “Neutral Opinion,” and 5 indicates, “Agree with Completely.”

4.3.2 Task Focusing Time

The task focusing time was computed similarly to the Pilot Study. Again, the raw eye-tracking data was utilized, and for the three levels of clarity given for each fixation (left, right, and both), only data that was at levels “both” was used to decrease error. These points were averaged for each three data points to provide 60 ms worth of data, which can stem from chaotic eye movements within this 60 ms range (Poole & Ball, 2006). On average,

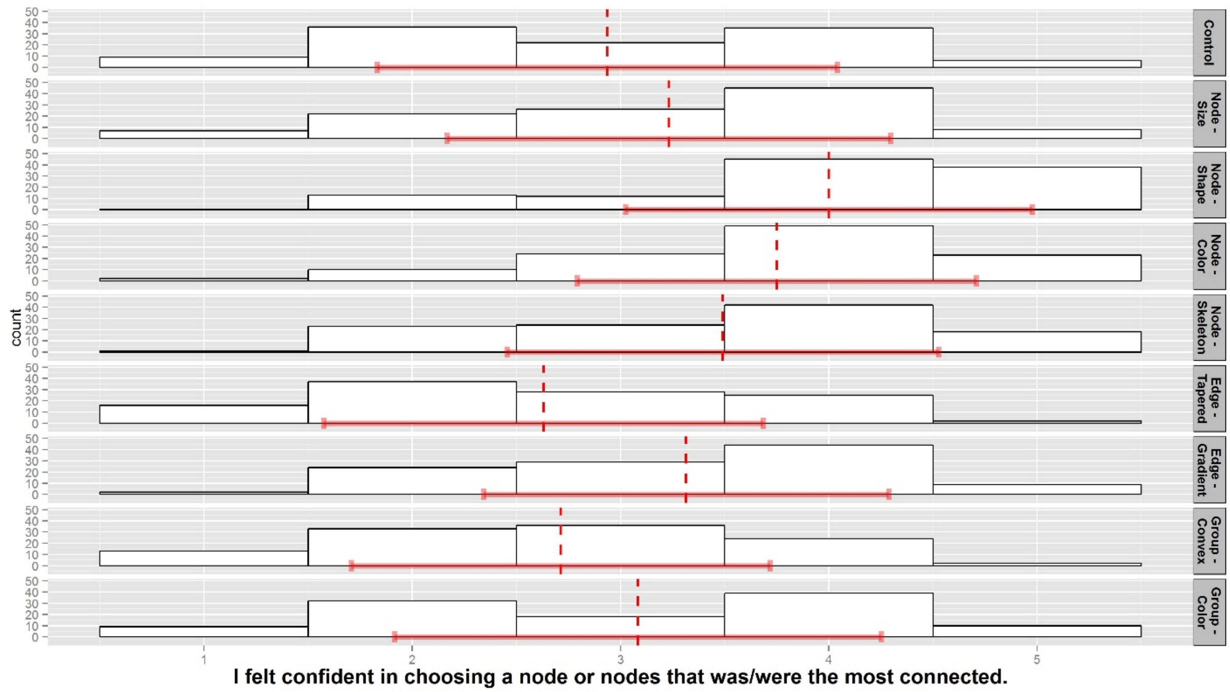


Figure 69: Results of the follow-up question, "I felt confident in choosing a node or nodes that was/were the most connected." The vertical red line indicates the mean and the horizontal red line indicates the standard deviation.

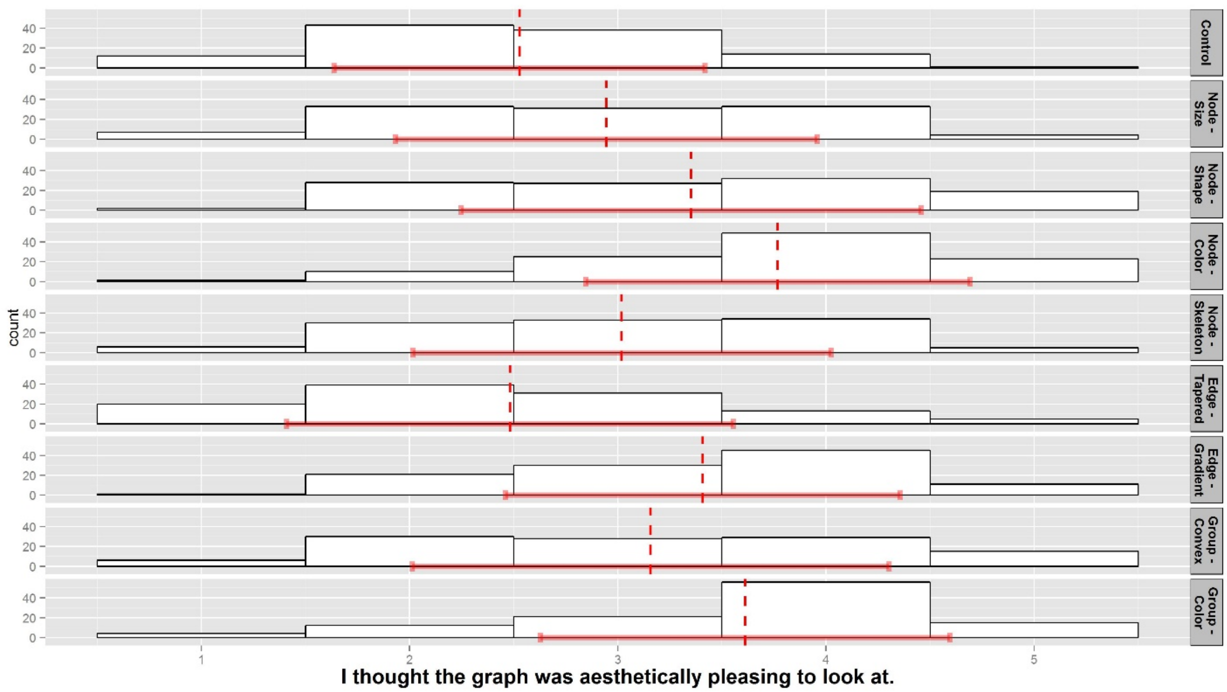


Figure 70: Results of the follow-up question, "I thought the graph was aesthetically pleasing to look at." The vertical red line indicates the mean and the horizontal red line indicates the standard deviation.

this would result in 156.67 data points per trial, per participant, and 152,280 total data points used for the analysis.

$$156.67\ data\ points * 36\ trials * 27\ participants = 152280$$

The analysis used the library package called Shapely (Gillies, Bierbaum, Lautaportti, & Tonnhofer, n.d.), which utilizes the geometric engine, open source library (GEOS), typically used for spatial analysis. Considering points/lines are used to create the network structure, it works for planar geometric object analysis as well. A buffer was placed around each node and link, creating a dilation of 5 pixels around each spatial object. An exhaustive algorithm was developed to:

1. Collect eye-tracking dataset for a specific participant, looking at a specific network ($\mu = 470$ per participant, per trial).
2. Create averaged saccade values for 60 ms of data ($\mu \approx 156.6-6$ per participant, per trial).
    a. Each of these data points is represented by an x, y value.
3. Create a network object for the corresponding trail.
4. Dilate a buffer of 5 pixels on the nodes and links.
5. Run each average saccade point within this buffered network.
6. If the average saccade point is within the buffered, this is repeated from step 5.
7. If not, this is labeled a miss, and if 10 consecutive "misses" occur, record this time.
    a. Where time is relative to the number of average saccade points that pass step 6.
8. This is then repeated for each trial, for each participant.

This creates a dataset of 972 values for participants, where 486 is used for the small networks and 486 for the large networks. During this calculation, both the response for Q2 ("I thought the graph was aesthetically pleasing to look at.") and the task focusing time for each trial were compared. In this experiment, only 964 of the 972 were calculated due to issues with the 8 eye-tracking datasets. These 8 were replaced with averages to ensure there was no missing data during analysis. Finally, considering the networks themselves (based on the topological structure) could vary both the mean task focusing time and the subjective beauty, groups of 9 were averaged together to remove this bias to utilize the randomized block design seen in Figure 44, reducing this total of 972 to 54 observations. This was first graphed (Figure 71) to view what type correlation is observed for all networks ($df = 52$). A positive, but fairly weak correlation is found for all graphs, suggesting that if a graph is very pleasing to look at, more time will be spent looking at the given graph. A Pearson's correlation was used to provide more insight, which finds that Pearson correlation is 0.249 and t-test provides a $p < .03$, confirming a weak, positive correlation between the two variables.

This was tested using separate plots for the small and large network. Figure 72 shows a positive correlation for both the small and large networks, albeit for the small graphs, a much greater increase. Using the Pearson analysis for the small networks, a much higher correlation coefficient of 0.4019 and t-test $p < .018$ was found, providing strong evidence that for smaller networks, as the subjective beauty of the graph increases, so does the amount of time focusing on them. However, the larger network graphs had a minor correlation and confidence based on its correlation coefficient of 0.1537 and $p < .222$.

These results make sense because both the small and large graphs used the 10000 ms timer, though the larger graph required more area to be searched, indicating that the 10000 ms was perhaps more appropriate for the small graphs. The smaller graph, which incorporates 482 observations, provides sufficiently strong evidence that an aesthetically pleasing design will result in more time spent in effort on its discovery.

4.4 EXPLORATION, USER BEHAVIOR, AND STRATEGIES GIVEN TO VISUAL AESTHETICS

Evaluating the exploration and user behavior presented itself to be a daunting task because trying to connect user action to eye-movement to talk-aloud protocol can leave a great amount of room for error. This was completed for this study by focusing on the eye-movements, both with the participant RTA and independently of explanation. The eye-tracking analysis was broken into three parts: fixation (when a participant focuses their

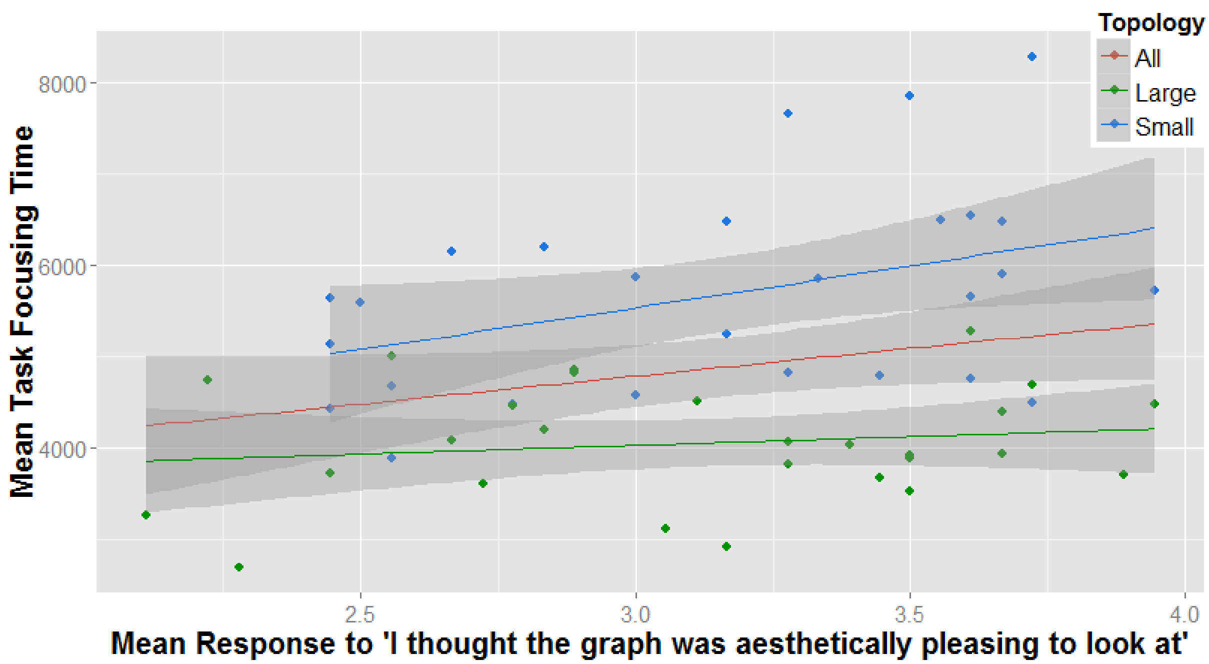


Figure 71: Subjective beauty vs. task focusing time with a linear regression line and 95% confidence region for all networks

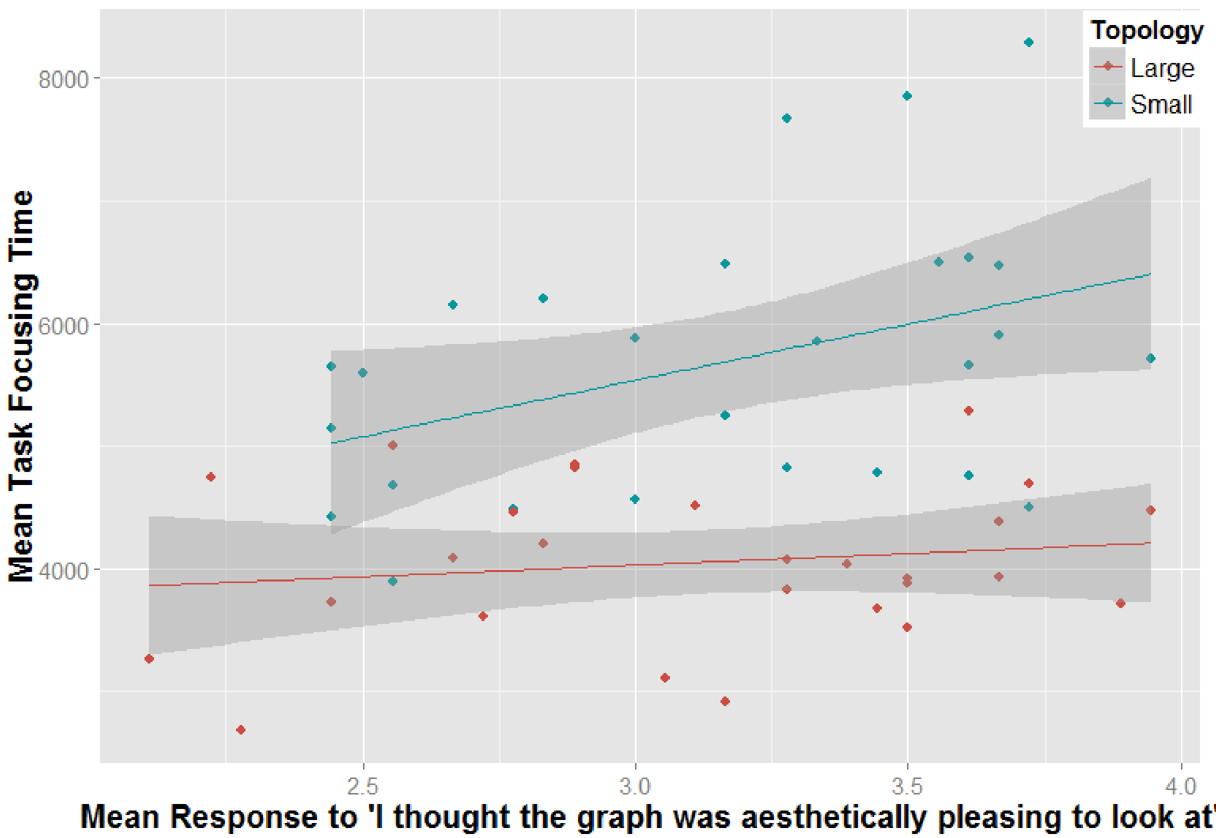


Figure 72: Both the small (red) and large (green) networks of the subjective beauty vs. task focusing time with a linear regression (with respective colors) and 95% confidence region.

eye on a specific part of the screen for a prolonged amount of time) and saccade (short eye movements that connect between fixations) (Poole & Ball, 2006). The algorithm utilized to capture these specific values is as follows:

1. Retrieve a set of eye-tracking data for a given participant and a given trial, producing a sequence of x, y coordinates for the 10000 ms span.
2. Increment each of these x, y coordinates from 0 ms to 10000 ms.
3. Set the first x1, y1 position and iterate to the next x2, y2 position.
   a. If x2, y2 is within a buffer (30 pixels in diameter) of x1, y1, label as fixation.
   b. If x2, y2 is outside a buffer (30 pixels in diameter) of x1, y1, label as saccade.
4. Repeat process for all data, up to 10000 ms.
5. Repeat for all trials for a given participant.
6. Repeat for all participants.

However, not all the eye-tracking data could be utilized for all participants, as some may have had issues with tracking caused by eye-tracking apparatus, software, or just simply blinking. The 10000 ms span was subdivided into 100 x (100 ms) segments and averaged, providing a thorough amount of data points, and removal of error, as previously specified. During the calculation of the fixation, saccade, and scanpath time, the average saccade and scanpath distances were also computed.

As per Netzel et al. (2014), the average saccade length and average fixation duration were plotted, Figure 73, where average saccade length can indicate more exploratory action (long saccade length) and average fixation duration can show more in depth searches (longer fixation duration). It should be noted in Figure 73 that the average number of fixations equates to amount of time, as each fixation point equals a 20 ms segment in eye-tracking data. Significance was found for the average fixation duration ($p < .01 - F\,(8, 891) = \ 8.371$) and average saccade length ($p < .05 - F\,(8, 891) = \ 12.6\ $). Though addressing the H5a and H5b, the only significance was found for the node size and group convex hull, only validating the H5b claim that a node visual encoding (NS) would provide more fixation duration than the control.

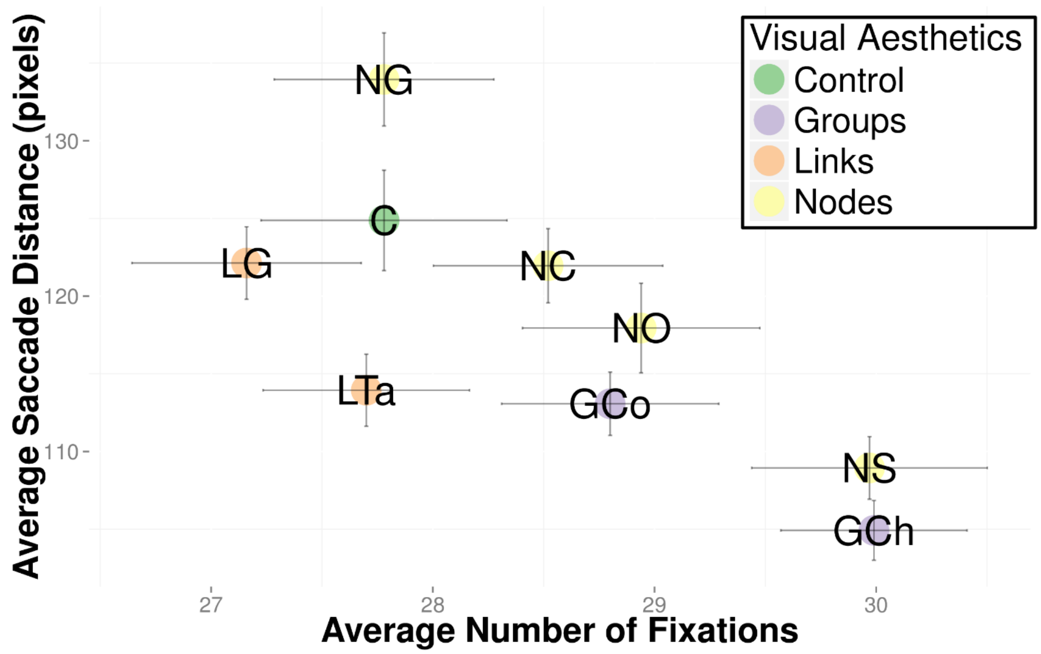


Figure 73: Plot of average saccade distance (pixels) and average number of fixations (time)

This was further investigated by plotting the average number of fixations and saccade versus each of the time increments, seen in Figure 74 and Figure 75. Each of these plots showcased their respective averages and each of the visual aesthetics, including a regression line with a 95% confidence region and a horizontal line of the mean of all values for all visual aesthetics. It should be noted that for each graph, the y-axis was scaled .5 above the mean and 1.5 below uniformly for each graph. This was done to highlight the regression line. Considering these figures included a time scale, they are more complete in terms of how each of these factors played a part in the visual aesthetics analysis. Figure 74 (Fixations) indicates that when the regression line is above the mean line, it is a less efficient search (Goldberg & Kotval, 1999). Figure 75 (Saccades) indicates that when the regression line is above the mean line, it is seen more as a search behavior (Goldberg & Kotval, 1999).

Based on these results and the RTA, this chart indicates both the strategies utilized by the participants and how the fixations and saccades could be interpreted. The descriptions were put together using the terminology and strategies utilized by the participants. Also,

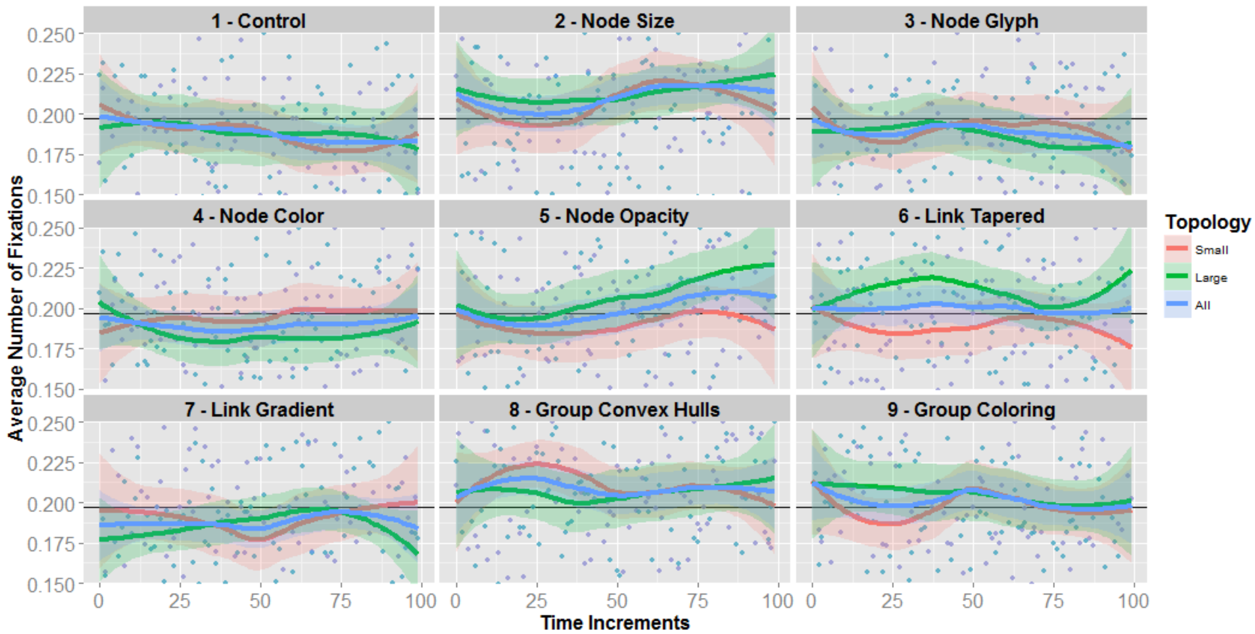


Figure 74: Average number of fixations on a time scale with a regression line. The solid, horizontal black line indicates the mean value of the number of fixations in total.

featured is an example network with corresponding eye-tracking data from the participants. This is a variation of a heatmap, where an affinity propagation algorithm is used to cluster points together (Frey & Dueck, 2007), this is referred to as the affinity propagation heatmap (APH). This is done to concatenate all eye-tracking data from multiple sources (participants) into areas of high interest.

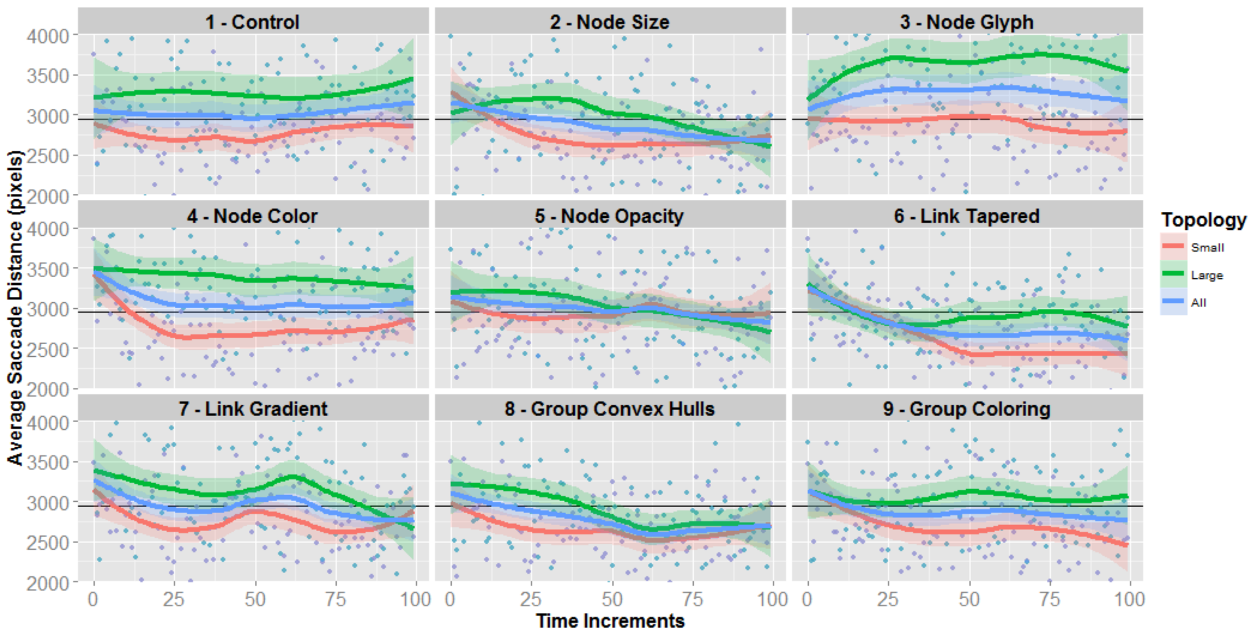


Figure 75: Average number of saccades on a time scale with a regression line. The solid, horizontal black line indicates the mean value of the number of saccades in total.

4.4.1 Control

The majority of the comments showed that the denser parts of the network were the main focal points. This denseness was described by both the cluster of edges and the cluster of nodes. Also, because the nodes with fewer connections were towards the outside of the graph, most of their focus was on the center structure. Some of the supplied comments specified that they simply counted the lines to find the most connected node, not using any of the network structure as guidance. Based on the eye-tracking data, participants found little guidance in the structure as their larger saccades and fewer fixations indicate.

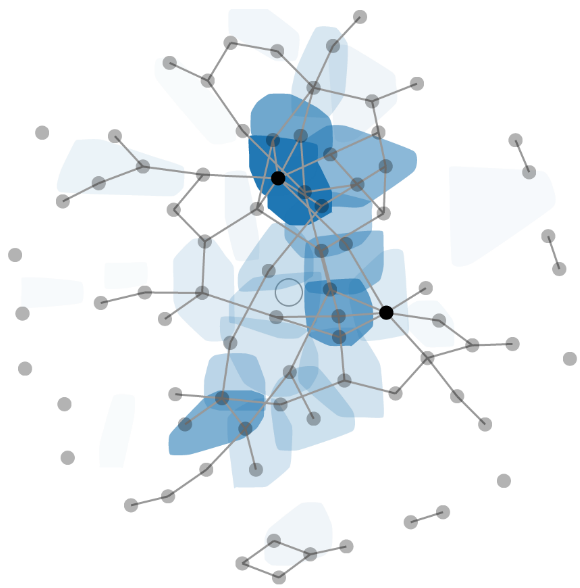


Figure 76: (Control - No Visual Aesthetics) Graph shows affinity propagation heatmap based on 9 participants.

4.4.2 Node Size

The node size was very difficult for participants to make sense of, even though most participants could figure out that the larger size of the node meant more connections. Most participants did not like the node size, as it would cause confusion based on more surface area being added to the overall structure (a poor use of "data-ink ratio"). The high number of fixations seem to stem from participants struggling to compare nodes by their given size once a high density area was found. This was exacerbated when two nodes of similar size were a far distance apart. The node size, in terms of short term memorization, seem to cause a large amount of issues because it was just too difficult to compare. One participant specifically commented, "This got worse as the number of connections increased," which certainly would cause a major issue with the given task.

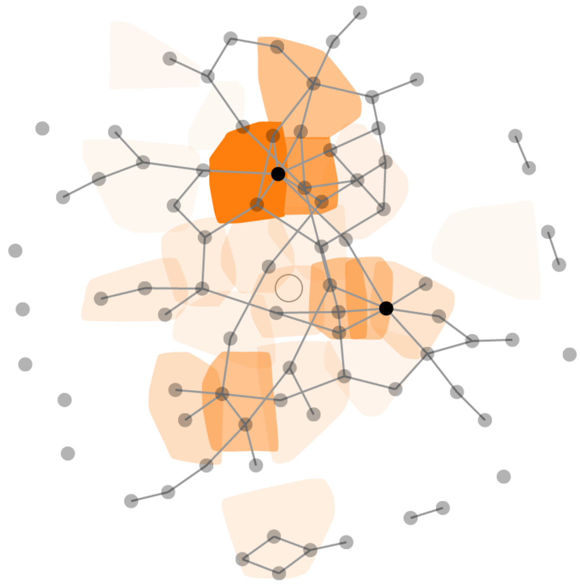


Figure 77: (Node Size) Graph shows affinity propagation heatmap based on 9 participants.

4.4.3 Node Glyphs

Node Glyphs were by far the favorite in terms of trying to complete the task. Multiple participants provided unique strategies for the node glyphs, including: 1) number of points on the shape indicates the number of connections; 2) find the unique shapes; 3) always search for the 6-sided object, referred to as the star or flower; 4) find the shape that is least represented; and 5) find the dense areas and select the unique shape. Some of these strategies overlap with one another, but do have a slight uniqueness to them. Indicated by the heatmap, Figure 78, which showed very large scanpaths, participants were asked why they tended to search large areas of the graphs and they specified that: 1) they needed to search the whole graph to figure out what the glyphs meant; 2) they were searching for unique objects; or 3) if they thought they had found the most connected node, the shapes made it easier to compare at further distances (easier for them to internalize).

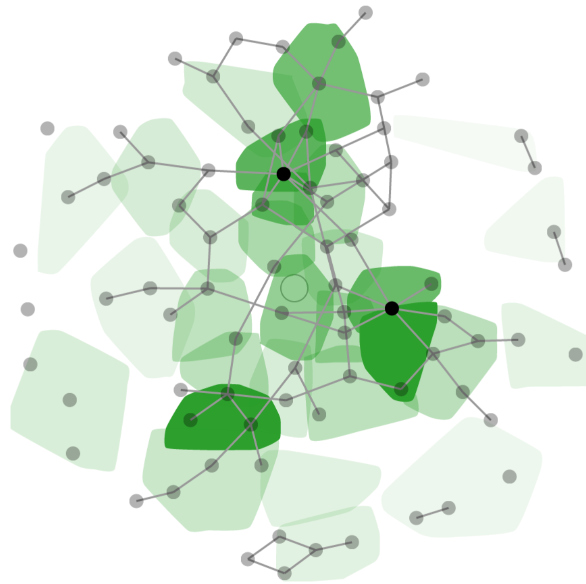


Figure 78: (Node Glyphs) Graph shows affinity propagation heatmap based on 9 participants.

4.4.4 Node Coloring

Most participants were able to deduce that the darker node (in terms of lightness of color) was indicative of more connections; however, not all participants saw this connection and these participants saw the node coloring as more distracting than helpful. With the node size, comparison between connections of 5, 6, and 7, was very difficult due to the lack of differences in the nodes' appearances. Of the 27 participants, 10 mention this being their favorite in terms of visual appearance (visually pleasing), and as mentioned before, led to them focusing on the task longer than the other visual aesthetics. Based on the scanpath and saccade, this continuous searching showed that the participant would use large scanpath through much of the search and finally smaller saccades to select answers.

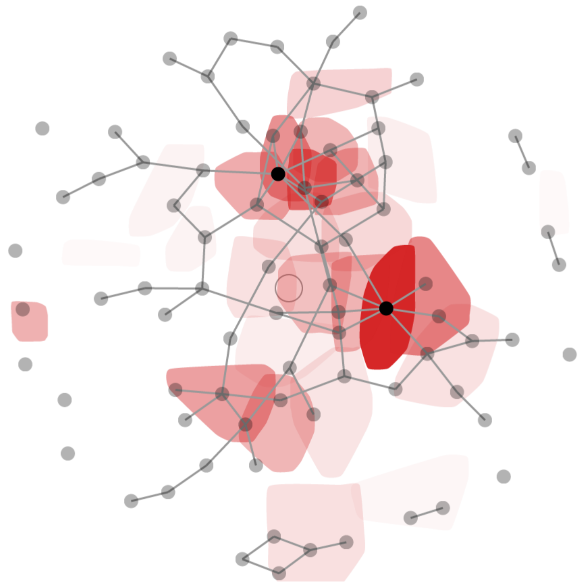


Figure 79: (Node Coloring) Graph shows affinity propagation heatmap based on 9 participants.

4.4.5 Node Opacity

Node opacity functioned and acted very similar to the control, though participants thought it to be more aesthetically appealing. The strategies that developed from the node opacity were very unique, in that participants: 1) used the nodes as targets, and selected nodes that were more filled in; and 2) treated the node like a pie chart and found nodes that were "cut into more pieces." Participants also noted that they liked that the node was more "accessible," helping greatly in discerning if the link was connected to the node, or just passing by. Based on the fixation, saccade, and scanpath, the node opacity was not very unique, again with similarities to control.

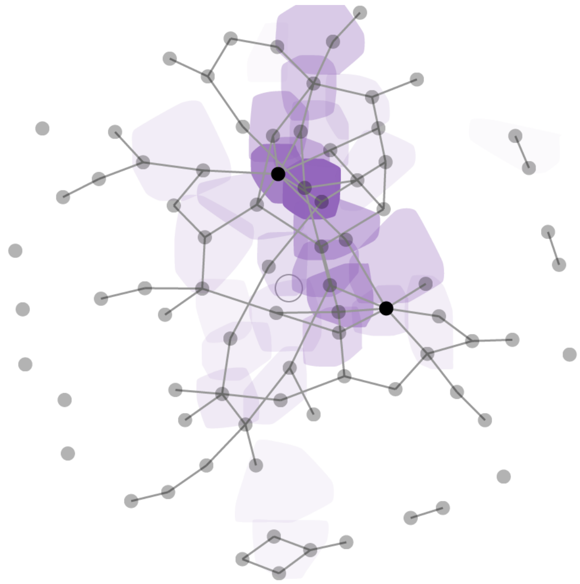


Figure 80: (Node Opacity) Graph shows affinity propagation heatmap based on 9 participants.

4.4.6 Link Tapered

By far this visual aesthetic was least liked aesthetically and in terms of completing the task. Participants had a very difficult time developing strategies, as the visual aesthetic itself did not lend itself to exploration. Edge overlap became a major issue and was exacerbated by its design. Participants were unsure what the size of the link indicated, if anything. The link gradient had a greater amount of fixation and saccades, but it seemed to be due to the fact that it was hard to navigate. Even in Figure 81, there is a lack of gaze on one of the most connected nodes. Overall, trying to incorporate a directed graph aesthetic into a non-directed graph task worked poorly in its overall effectiveness.

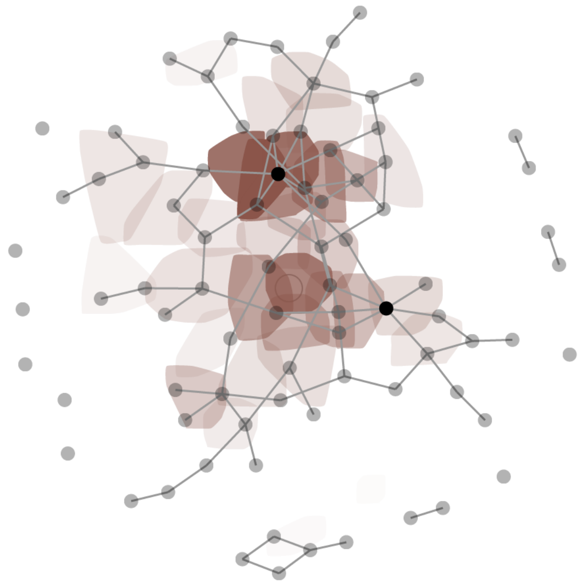


Figure 81: (Link Tapered) Graph shows affinity propagation heatmap based on 9 participants.

4.4.7 Link Gradient

After the node coloring and group coloring, this was the third most appealing visual aesthetic (as told during the RTA). This was a very unique visual aesthetic in regards to the RTA, as most liked it visually, but were unsure what it was trying to showcase. Most participants would focus on the darker lines, as seen in Figure 82, which would limit their searching to very specific areas. Based on the average saccade distance (pixels) and average number of fixation, this did not showcase a good design for exploration, even though it was one of the better in terms of accuracy. Overall, the participants liked the link gradient, but they did not know what to do with it.

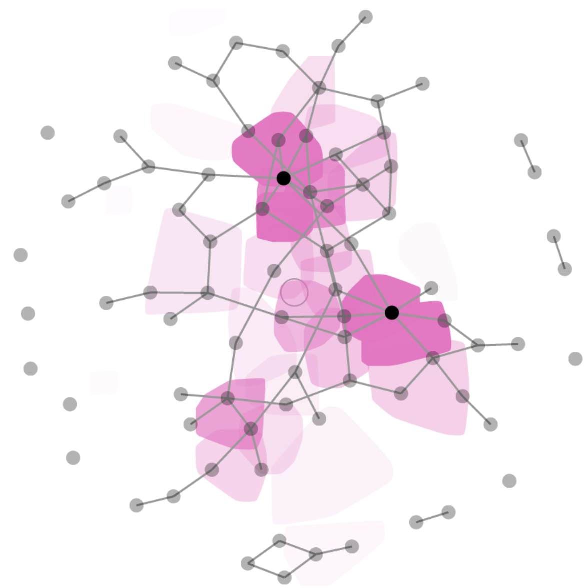


Figure 82: (Link Gradient) Graph shows affinity propagation heatmap based on 9 participants.

4.4.8 Group Convex Hulls

The convex hull was seen as one of the more visually pleasing graphs, but also one that was very difficult to utilize. A number of strategies were developed to help in completing the task: 1) the most connected node will fall in areas where the convex hull overlaps; 2) find the most connected node in the group and find the most connected node in the graph; 3) nodes that are on the outskirts of the group are more likely to be the most connected node, as they tend to connect both inside and outside the group; 4) the most connected node tends to be in the larger groups; and 5) nodes that tend to be the most central to the group are more than likely the most connected node. As it can be seen, some of these strategies contradict one another, and some are just artifacts of the clustering and convex hulls. The most interesting strategy (find the most connected node in the group and find the most connected node in the graph) was one of the main driving forces of the convex design, which was to create smaller, more manageable sub-problems. This was mentioned by 5 of 27 participants, so this tended to be a small characteristic of the convex hull. The creation of sub-domains by the

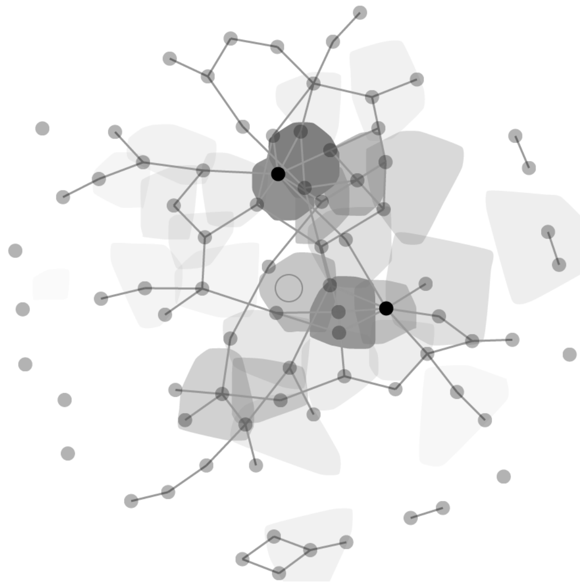


Figure 83: (Group Convex Hulls) Graph shows affinity propagation heatmap based on 9 participants.

convex hull is very much supported by the numerous fixation and low scanpaths. Searching was done within the convex hull and then migrated to the next community or group. Even though the convex hull aesthetic was not the most accurate strategy, it does showcase that participants will use the groupings to create more manageable sub-problems, and if provided more time, could have an increased utility.

4.4.9 Group Coloring

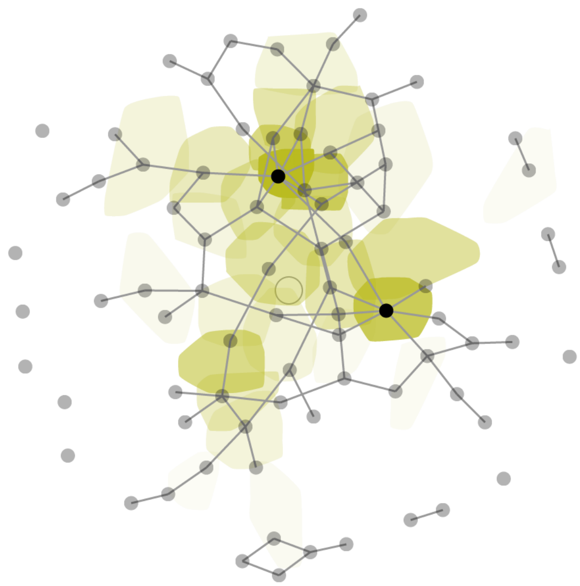


Figure 84: (Group Coloring) Graph shows affinity propagation heatmap based on 9 participants.

Group coloring was the second most aesthetically pleasing, next to the node coloring. However, similar to the convex hull, this was not seen as a visual aesthetic that could help with the given task. The majority of the strategies were the same as the node coloring - find the darkest node based on coloring. This became more difficult though based on its use of multiple colors. In this regards, even though it showed the same groups/communities as the convex hulls, participants remarked that it was not as helpful for "parsing the network" in smaller problems because there was less definition between the given groups. That being said, the increase in the saccades was very similar to the convex hulls. Multiple participants (5 of the 27) stated that they would look for the dark color, as in red or orange, when searching because these would tend to standout. The two-encoding function of the group coloring tended to cause more problems than help with the given task.

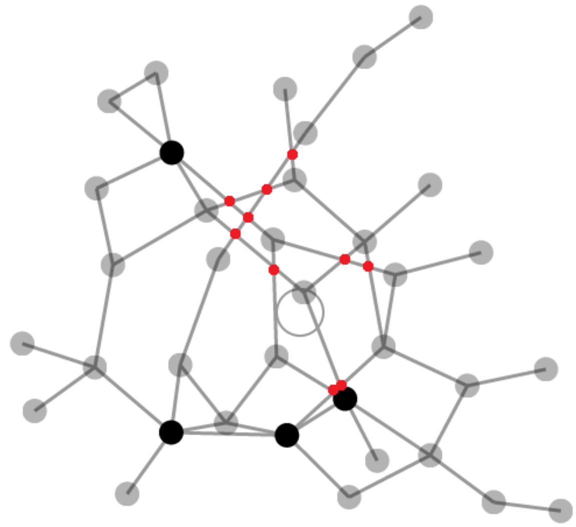


Figure 85: An example network, where degree centrality POIs are highlighted nodes and the edge-crossing POI are highlighted using a red dot

4.5 EDGE-CROSSING VS. DEGREE CENTRALITY: VARIATION BY VISUAL AESTHETIC

This final, data-intensive hypothesis is one of the most fundamental parts of this work; that being if any of the visual aesthetics could overcome a previously seen pitfall of network graph readability in edge-crossing. This problem was addressed by defining two types of points of interest (POI): 1) the nodes with the largest degree centrality (the target of this task) and 2) areas with edge-crossing. Figure 85 illustrates this by highlighting the most connected nodes, degree centrality POI, and red dots on where edges cross, edge-crossing POI.

The algorithm for this metric required two lists, one for the degree centrality POI and the other the edge-crossing POIs. This was done by:

1. Iterating through all nodes, and creating a vector of all nodes with the highest degree centrality and their (x, y) position.

   a. Label these points as degree centrality POIs.

2. Iterating through all edges, find if two edges overlap, and where this occurred, the (x, y) position.

    a. Label these points as edge-crossing POIs.

Once these two vectors were created for each graph:

1. Retrieve a set of eye-tracking data for a given participant and a given trial, producing a sequence of x, y coordinates for the 10000 ms span.
2. Increment each of these x, y coordinates from 0 ms to 10000 ms.
3. For each gaze point, create a buffer of 10 pixels in radius.
4. Iterate through the degree centrality POIs.

    a. If gaze within a degree centrality POIs buffer, label as degree centrality POIs.

5. Iterate through the edge-crossing POIs.

    a. If gaze within an edge-crossing POIs buffer, label as edge-crossing POIs.

6. Repeat process for all data, up to 10000 ms.
7. Repeat for all trials for a given participant.
8. Repeat for all participants.

The same averaging was done to produce 100 x (100 ms) segments, providing enough data for analysis and removal of errors. This was then graphed, Figure 86, and overlaid on top of one another. In the figure, temporal variation for degree centrality targets (red curve) and the edge-crossing (green curve) POIs, for all visual aesthetics. The black horizontal line indicates the average overall POIs for control. . This would illustrate visual aesthetics that are better or worse than the control. Because it is on a time scale, this figure also showcases when the participant would see either the edge-crossing or the most connected node(s). It should be noted that in the figure there were more edge-crosses than most connected nodes, and these were not normalized in the figure.

Some of the more noteworthy findings from this are: 1) node glyphs focused attention on the most connected nodes and limited issues with the edge-crossing; 2) there was much more edge-crossing viewed in the convex hull than all other visual aesthetics; and 3) similarity of

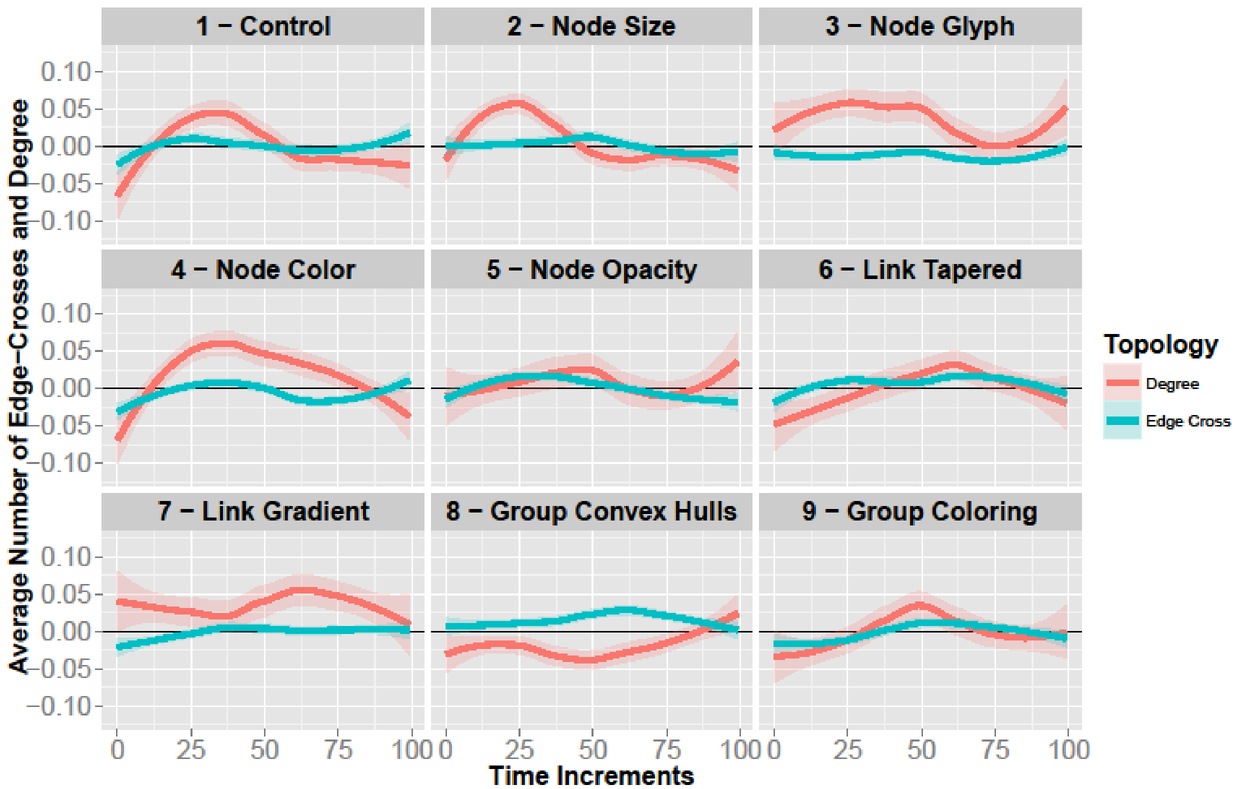


Figure 86: Temporal variation degree centrality targets (red curve) and the edge-crossing (green curve) POIs, for all visual aesthetics. The black horizontal line indicates the average overall POIs for control.

the patterns for the control and node visual aesthetics, which suggests similar search patterns for both.

To help evaluate this measure, similar analysis was used as that done for the accuracy and time values. Both the degree centrality POIs and edge-crossing POIs were found to be significant ($p < .001 - F(8, 891) = 12.48$) and ($p < .001 - F(8, 891) = 20.14$), respectively. Using Tukey HSD pair-wise test on the factor of the control (C) versus all the visual aesthetics was completed and showed that for both degree centrality POI and edge-crossing POI, node glyphs show significance versus the control ($p < 0.01$), confirming H6a at least for node glyphs. Also worth mentioning was how well link gradients did for the participants focusing on the most connected node(s), significance ($p < 0.01$).

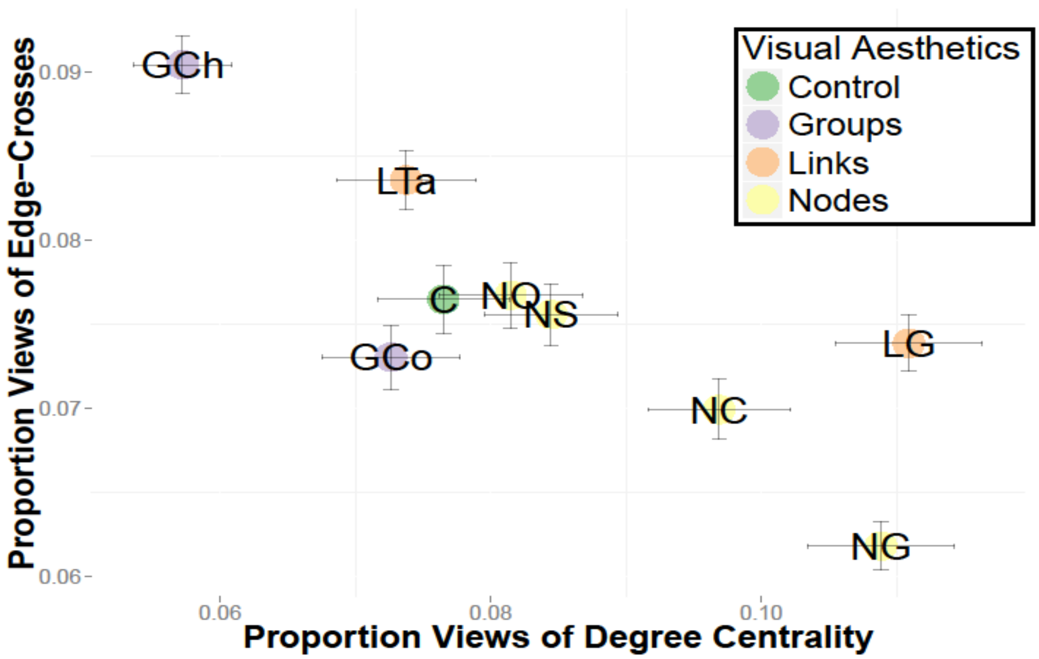


Figure 87: Average view of Degree POIs vs. average view of the edge-crossing POIs plotted for all aesthetics.

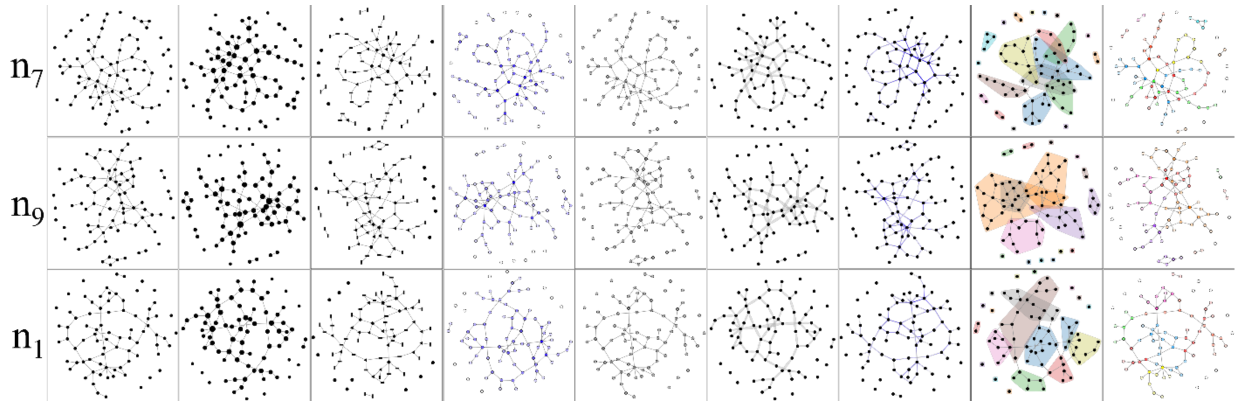


Figure 88: The three graphs utilized in RTA, as they would have been seen by the participants.

4.6 CHANGE BLINDNESS BASED ON NETWORK COMPLEXITY

This last area of research focuses more on the working memory associated with visual aesthetics and topological structure of the graphs. The 27 participants were split into 3 groups (9 per group) during the RTA portion of the study. All participants would see the same network graph repeated 9 times, with all visual aesthetics being represented. The only difference between each of the graphs would be: 1) the visual aesthetic applied; and 2) the graph rotated by 90°. Based on the previous experiment, one group would view the worst graph (in terms of accuracy), one group would view the mean average graph (in terms of accuracy), and one group would see the best graph (in terms of accuracy) from the large network graphs, Figure 88.

The hypothesis specified three conditions:

1. H7a: Participants will not be able to tell the same graph was used multiple times in a row for all three groups (visual aesthetics affecting change blindness).
2. H7b: Participant will be able to tell the same graph was used multiple times in a row, but only certain groups (network complexity affecting change blindness).
3. H7c: Participants will be able to tell the same graph was used multiple times in a row

Table 10: Responses by the participants in regards to noticing the consistent graph structure utilized in the RTA. These are color coded with least accurate network in the Pilot Study (Top), the mean average network in the Pilot Study (Middle), and the most accurate network in the Pilot Study (Bottom)

1	2	3	4	5	6	7	8	9
No	No	No	No	No	No	Yes	No	No
10	11	12	13	14	15	16	17	18
No	No	No	No	No	Yes	No	No	No
19	20	21	22	23	24	25	26	27
Yes	No	No	Yes	Yes	Yes	Yes	Yes	No

for all groups (no change blindness).

This was addressed by utilizing the RTA, where the first question asked the participant, "Of these last 9 graphs all featured the 9 aesthetics used in this experiment, did you happen to notice that the graph used had the same structure throughout the last observations?" After they responded, they were shown an example of the rotating networks, and asked to clarify their answers, if need be. The breakdown for their response can be found in Table 10.

Based on their responses, the hypothesis H7b seems to be the most applicable, where the network complexity dictated the adverse effects that occur with change blindness. This was further investigated by calculating the participants T_p, F_p, and F_n scores. As a reminder, here is the signal-detection break down:

- T_p= subject selecting a node and correctly identified the most connected node(s)
- T_n= subject not selecting a node and correctly rejected the node as not the most connected node(s)
- F_p= subject selecting a node that was not the most connected node(s)
- F_n= subject not selecting a node and the node was the most connected

Based on the assumption that if there was a learning bias unknown to the users, the T_p scores would peak at the maximum amount and continue to the end of the RTA, or at least relatively close to that maximum value. The first set (group 1), Figure 89, showed very poor

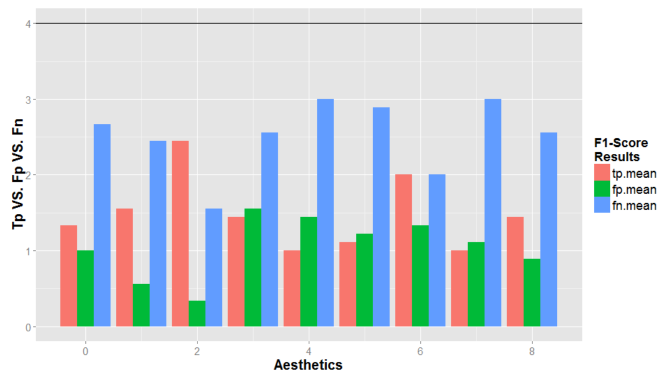


Figure 89: Group 1 results for the least accurate graph from the Pilot Study. The horizontal line at (4) indicates the optimal value, as there were 4 most connected nodes.

results throughout and replicated scores similar to the Pilot Study results. The second set (group 2), Figure 90, had similar results to the Pilot Study. However, group 3, Figure 91, had a consistent result (except for the last visual aesthetic), but this could be characterized as an artifact of a missed selection or just late response by human error.

Therefore, H7b was confirmed, which means if there was change blindness occurring between each of the graphs, it would be caused more so by the network complexity and not the visual aesthetics themselves.

4.7 SUMMARY OF RESULTS

4.7.1 Utility of the Visual Aesthetics

The utility of the visual aesthetics were tested based on their accuracy and time. Five hypotheses were developed to compare node based visual encoding (NS, NG, NC), link based visual encoding (LG), group based visual encoding (GCo), and visual styling (NO, LTa, and GCh) versus the control. Node opacity (NO) will have higher performance (accuracy and task completion time) than link tapered (LT) and group convex hull (GCh). For H1a, this

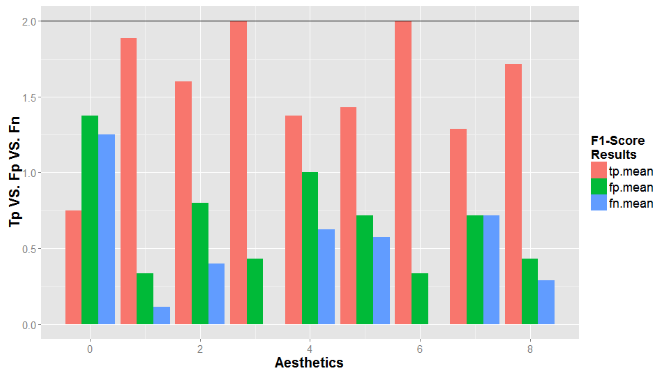


Figure 90: Group 2 results for the mean accurate graph from the Pilot Study. The horizontal line at (2) indicates the optimal value, as there were 2 most connected nodes.

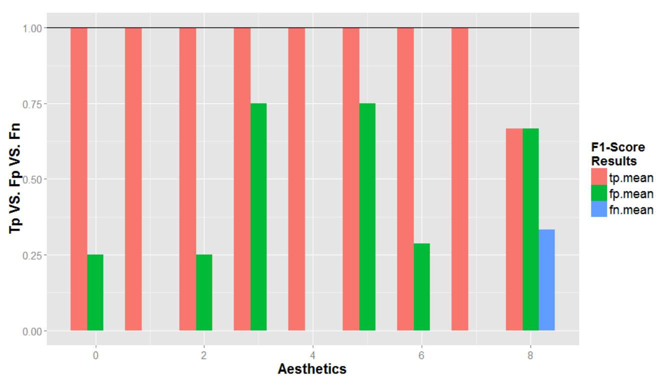


Figure 91: Group 3 results for the most accurate graph from the Pilot Study. The horizontal line at (1) indicates the optimal value, as there was 1 most connected node.

was confirmed for all node visual encodings, with (NG, NC - $p < 0.01$) and (NS - $p < 0.05$). H1b and H1c were rejected, as no significance was found. H1d was confirmed, as the visual styling (NO, LTa, and GCh) had no significance. H1e was not significant, establishing, based on time, (NO) was not superior to (GCh and LT). These results can be seen in Figure 63.

A secondary set of hypotheses were developed to test the node size, shape, and color. This is discussed in much of the past literature in regards to visual aesthetics. Two hypotheses were proposed: NC > NS > NG in terms of accuracy and NC < NS < NG for time. It was shown that for a confidence level of 95%, node glyphs were superior to node size and node coloring, and at an 85% confidence level, node coloring was superior to node size in terms of accuracy, Figure 67. Though this relationship was again shown for the task completion time, no significance was found.

Glyphs were previously thought as a means of showing nominal data and not ratio or interval data, which was not the case for this given experiment. The reason being that it seems the network graph itself provided context for the glyphs, providing a new means to utilize shapes in graph visualizations. Also, the novel link gradient was also shown to have improved accuracy levels, which means that local level information could be transferred to the links that connect these point. This should be considered moving forward without other measures, such as betweenness and closeness measure, as both rely on the links to provide these measures. This would allow for more information to be encoded at the node level, which could include attribute data of the node or a redundancy measure, such as the glyphs to maximize their effects.

4.7.2 Affordances by Perceived Utility and Confidence of the Visual Aesthetics

Affordance was derived for two follow-up questions given to the participants after each network graph trial in the experiment. These were:

- “The aesthetic helped in completing the task.”
- “I felt confident in choosing a node or nodes that was/were the most connected.”

The first question, “The aesthetic helped in completing the task,” was created to elicit a direct comparison of the visual aesthetic and their affordance on the task. The second question, “I felt confident in choosing a node or nodes that was/were the most connected,” was asked to show to what degree the visual aesthetic affected the affordance. Three hypotheses were developed which compared the control to the node encoding and node styling (NS, NG, NC, and NO), the link and group encoding (LG) and (GCo), and the visual styling (NO, LTa, and GCh). The node encoding was thought to be superior to the control for both questions, the link and group encoding superior to the control only for the question, “The aesthetic helped in completing the task,” and the visual styling evaluated as the same as the control. It was found that there was significance for (NS, NG, NC, NO, LG, and GCo), proving the first part of H3a and H3b. H3c was also confirmed based on not having any significance compared to the control. These results can be seen in Figure 68.

These results show that the actual utility of the visual aesthetics have strong correlation to the perceived affordance for the visual aesthetics. It is worth mentioning again that the study did not limit itself to only network graph, node-link, or force-directed graph experts, but was open to people who may not have had any experience at all with these types of visualizations. Showing a high correlation between the affordance and utility provides confidence in the usability of the network graph and the visual aesthetics applied to them, that these visual aesthetics can make the network graph more accessible to even novices of the social network domain.

4.7.3 Correlating Subjective Beauty to Task Focusing Time

It was seen before that subjective beauty or visually appealing graphs lead to an increased amount of time focused on them when given a task. This idea was tested only by the amount of participation utilized by the subjects in completing the task, neglecting the fact that this testing was not done with the actual engagement of the participants, as they could become either disinterested in the task or have difficulty in completing the task, which could account for these continued views. This topic directly tested the eye-gazes as a function of task focusing time to the subjective beauty of the given visual aesthetic. This was designed to show correlation to the task focusing time and subjective beauty provided by participant feedback for, “I thought the graph was aesthetically pleasing to look at.” The two hypotheses looked at this correlation for both the small and large network separately.

It was found that for both types of network, there was a positive correlation between the task focusing time and the subjective beauty of the aesthetic. However, the small network had a much higher degree of confidence based on a correlation coefficient of 0.4019 and t-test $p < .018$, where the large graph had a correlation coefficient of 0.1537 and $p < .222$. Figure 72 provides graphical results of the testing.

As mentioned, the results of this part of the study validates a previously thought connection between the amount of time and effort spent by the participant and its correlation to the participant's perception of the visual aesthetic or styling. This validation shows that when designing graphics or visualizations, the adage "maximize the data-ink ratio" may not be as applicable as originally thought, as the addition of the visual aesthetics and styling could entice the observer to maximize the time spent on their exploration and understanding.

4.7.4 Exploration, User Behavior, and Strategies Given to Visual Aesthetics

With eye-tracking studies in the domain of usability studies or testing, a large amount of literature has been discussed in terms of the perceptual meaning of specific eye-movements, specifically in regards to fixation and saccades. Three hypotheses were designed to provide insight in the average saccade lengths and the number of fixations, where it was believed that node visual aesthetics would have multiple fixations and longer saccades. Also, because of its design, the link gradient would have shorter saccades because it would limit searching to a small domain of the graph. Also a research question was developed to connect the fixation and saccades, the affinity propagation heatmaps, and the RTA feedback, and where there were connections between these parts of the study. This was deemed a research question, as the feedback from the RTA was not correlated between the participants, only recanted from the conversation in regards to participant strategies in the RTA.

There was a slight verification of the hypothesis as only the node size and group convex hull showed significance in terms of the average number of fixations, proving part of the H3a. The RQ did provide much further insight into visual aesthetics and the resulting eye-tracking data. Specifically with the larger saccades used in glyph, it was directly connected to the amount of exploration utilized to validate the meaning of the glyph. There were shorter amounts of saccades in the link gradient, which confirmed to some degree that it provided a more focused search. The shorter saccades and numerous fixations in the group convex hull also showcased that the task was subdivided into smaller problems, based on the grouping; however, this was not the case as much for the group coloring, as the grouping was not as apparent. Lastly, the high symmetry between the control and all node visual aesthetics suggested that when limited to node searching, patterns would emulate one another in their completion.

4.7.5 Edge-Crossing vs. Degree Centrality: Variation by Visual Aesthetic

Based of node degree centrality POIs and edge-crossing POIs, it was deduced that some of the visual aesthetics could overcome issues typically derived from the edge-crossing found in graphs. This was found to be more so in the node visual encoding (NS, NG, and NC) than the other visual aesthetics, and that for the degree centrality POIs, the link gradient would produce higher results than the control based on its focused search.

Figure 86 provided the best means of analyzing these results, as it overlaid both the node degree centrality POIs and edge-crossing POIs in a temporal setting. Based on the analysis, it showed significance that only the node glyphs both promoted the node degree centrality POIs and minimized the edge-crossing POIs, which is a very important finding for future design implication. Also confirmed was the maximization of degree centrality POIs created by the link gradient, which provided insight into its ability to limit the eye-movements to target at hand.

4.7.6 Change Blindness Based on Network Complexity

Though this study was not designed specifically to test working memory concerns caused by change blindness in graph structures, it did lend itself in providing a strong glimpse at whether network complexity or the visual aesthetics would cause such an issue. Table 10 showed the verbal results of the RTA focused on noticing the small changes in the graph structure over repeated measures. This was also confirmed using the F1-scores related to accuracy for the interaction session in the RTA. Both showed that the network complexity caused more of an issue when it came to working memory and selective attention.

4.8 CONCLUSION

In this dissertation, both human interaction and perception were analyzed using both a usability study and the eye-tracking system with a focus on how visual aesthetics affect both the readability, affordance, and usability of a network graph based on finding the most connected node(s) in a graph. A novel methodology was developed that allows researchers the ability to remove a common confounding variable in a network structure bias by the use of randomized blocking design. It was shown in both experiments that for a localized task, such as network degree centrality, participants would prefer and utilize more successful node visual aesthetics, though the novel link gradients did provide higher utility similar to the node encoding.

Based on the eye-tracking data, a method was developed to compute fixations, saccades, scanpaths, task focusing time, and degree centrality/edge-crossing foci which were used to analyze the participant engagement quantitatively versus more traditional subjective measures. These results were incorporated with the feedback collected during the retrospective think-aloud protocol and the affinity propagation heatmaps to provide detailed strategies used by the participants and how results varied based on visual aesthetics interpretive meaning. Also, this showed that when given the same network with different visual aesthetics, their eye-movements to search these networks changed based on the level of visual aesthetics (node, link, and group). Considering these networks did not change throughout the experiment, the visual aesthetics could be seen as a filter applied to these networks, increasing or decreasing accuracy based on visual aesthetics and independent of the topological structure of the network.

The two major results specific to the visual aesthetics were the utility of the node glyphs and link gradients. The glyph is typically utilized in nominal data, but seemed to perform very well when given context in the network graph, creating more strategies based on their ease of remembrance versus the color and size of the nodes based on their specific uniqueness in their design. The promise of link gradient showed that data typically associated for node encoding, could be moved to the link encoding to allow for further encoding of attribute data at the node level or to create a redundancy with node encoding.

Lastly, it was examined how visual aesthetics and network complexity could affect the abilities of the working memory, caused my change blindness. This may not necessarily provide a new domain to network graph visualizations, as difference maps have been utilized for years. It does show at which level this effect occurs and which situations may be more relevant to apply these difference maps in future projects.

It is believed that this work does not have to be limited to only a social-scope but could also be beneficial to other network visualizations, such as computer networks, cartography, protein-to-protein interactions, etc. Within the domain of those particular fields, the application of visual aesthetics can be an incredibly important component to highlight structural or meta-data values. Also, when network aesthetics cannot be utilized (as in a map, where location cannot be changed) the addition of visual aesthetics become much more important to help in terms of readability.

4.9 LIMITATION OF STUDY

One of the larger limitations of this thesis is use in a single task when evaluating a network graph. Works by Lee et al. (2006), showcase a multitude of topological-based, attribute-based, and browsing tasks that can be utilized, but this thesis used only the one task. This was done very deliberately, focusing the discussion on the visual aesthetics and providing a simplified task, therefore making this experiment more approachable to novices.

The network variation could have also been expanded to include small networks, large networks, and variations on the density of these two types of graphs. This would provide more clarity into how complexity would alter behavior and strategies for a more diverse data set.

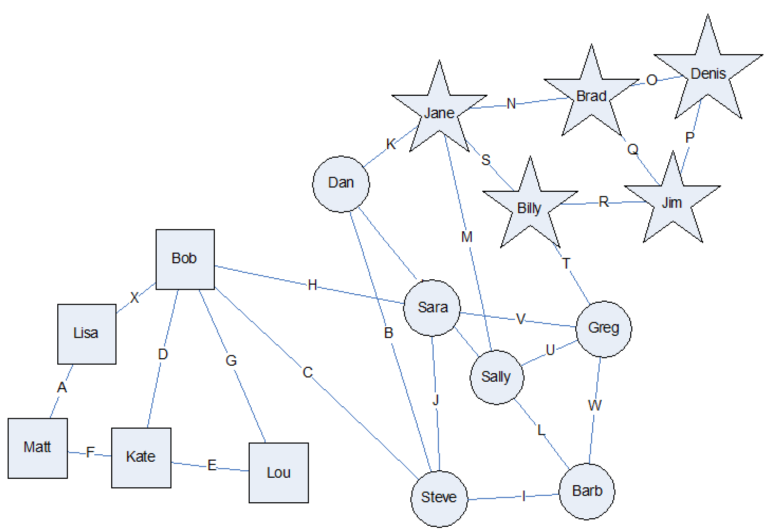
Finally, only a handful of visual aesthetics were presented in this study. Though there were 14 variations between the two studies, there is almost a limitless amount of creative means of injecting both styling and encoding within a network structure. Further, studies like this could include a new set of stylings and encodings that may build off these results or perhaps present novel means of creating new ways of reading and interpreting network graphs.

APPENDIX A

PRE-SURVEY

1. Gender: Male or Female
2. Date of birth:
3. Do you have experience using network graph, force-based graph, or node-link diagrams?
4. Have you ever used network graph, force-based graph, or node-link diagrams software/ visualization software before? If so, please list them (Note: you do not need to be proficient in using them, just have some experience).

For the following questions, the network is representing communication (talking / conversing) between each participant in the graph and the different glyphs / shapes represents groups / communities / cliques. If you are unsure of the answer, please mark it with "N/A" and proceed to the next question. Use Graph 1 for all the following questions:



1. Using Graph 1, does Jane communicate directly to Greg? Yes or No
   a. If yes, via which link (just list the label)? __________
2. Using Graph 1, how many connections does Bob have? __________
3. Using Graph 1, what is the shortest path between Kate and Billy (just list the label(s))?

   ____
4. Using Graph 1, is there a path between Matt and Steve? Yes or No
5. Using Graph 1, who communicates the most in this graph (could be more than one)?

   ____
6. Using Graph 1, are Steve and Bob in the same group? Yes or No
7. Using Graph 1, are Kate and Matt in the same group? Yes or No
8. Using Graph 1, are Sara and Billy in the same group? Yes or No

APPENDIX B

POST-SURVEY

Please answer using the 5-pt scale, where (1 - Very much disagree/very disappointing/highly negative) and (5 - very much agree/very enjoyable/highly positive):

1. I found the task easy to complete:

   1 - Very much disagree   2   3 - No Preference   4   5 - Very much agree

2. I found the visualization visually appealing:

   1 - Very much disagree   2   3 - No Preference   4   5 - Very much agree

3. The aesthetic all seemed unique to me:

   1 - Very much disagree   2   3 - No Preference   4   5 - Very much agree

4. I would recommend this interface to other people:

   1 - Very much disagree   2   3 - No Preference   4   5 - Very much agree

5. I found tasks straightforward to complete:

   1 - Very much disagree   2   3 - No Preference   4   5 - Very much agree

6. The task itself was easy to understand:

   1 - Very much disagree   2   3 - No Preference   4   5 - Very much agree

7. I found the time allotted to me was enough to complete the task:

   1 - Very much disagree   2   3 - No Preference   4   5 - Very much agree

8. I had trouble with differentiating the aesthetic properties:

   1 - Very much disagree   2   3 - No Preference   4   5 - Very much agree

9. I felt confident in selecting nodes:

   1 - Very much disagree   2   3 - No Preference   4   5 - Very much agree

10. I did not care for the aesthetic properties:

1 - Very much disagree   2   3 - No Preference   4   5 - Very much agree

Feel free to add comments about the design of the visualization, the task, or the project in general.

References

Ahn, Y. Y., Han, S., Kwak, H., Moon, S., & Jeong, H. (2007). Analysis of topological characteristics of huge online social networking services. In (pp. 835–844). ACM.

Andrienko, N., & Andrienko, G. (2006). *Exploratory analysis of spatial and temporal data.* Springer.

Archambault, D., Purchase, H. C., & Pinaud, B. (2011). Difference map readability for dynamic graphs. In *Graph drawing* (pp. 50–61). Springer.

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *science*, *286*(5439), 509–512.

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks.

Bennett, C., Ryall, J., Spalteholz, L., & Gooch, A. (2007). The Aesthetics of Graph Visualization. In *Computational aesthetics* (pp. 57–64).

Bertin, J. (1983). Semiology of graphics: diagrams, networks, maps.

Blanchard, A., & Horan, T. (1998). Virtual communities and social capital. *Social science computer review*, *16*(3), 293–307.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *the Journal of machine Learning research*, *3*, 993–1022.

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), P10008.

Borgatti, S. P. (2006). Identifying sets of key players in a social network. *Computational & Mathematical Organization Theory*, *12*(1), 21–34.

Bostock, M., Ogievetsky, V., & Heer, J. (2011). D Data-Driven Documents. *Visualization* *and Computer Graphics, IEEE Transactions on*, *17*(12), 2301–2309.

Brandes, U., & Nick, B. (2011). Asymmetric relations in longitudinal social networks. *Visualization and Computer Graphics, IEEE Transactions on*, *17*(12), 2283–2290.

Bruns, A. (2011). How Long Is a Tweet? Mapping Dynamic Conversation Networks on Twitter Using Gawk and Gephi.

Burch, M., Konevtsova, N., Heinrich, J., Hoeferlin, M., & Weiskopf, D. (2011). Evaluation of traditional, orthogonal, and radial tree diagrams by an eye tracking study. *Visualization and Computer Graphics, IEEE Transactions on*, *17*(12), 2440–2448.

Cao, N., Lin, Y.-R., Sun, X., Lazer, D., Liu, S., & Qu, H. (2012). Whisper: Tracing the spatiotemporal process of information diffusion in real time. *Visualization and Computer Graphics, IEEE Transactions on*, *18*(12), 2649–2658.

Carr, D. B., Littlefield, R. J., Nicholson, W. L., & Littlefield, J. S. (1987). Scatterplot matrix techniques for large N. *Journal of the American Statistical Association*, *82*(398), 424–436.

Cawthon, N., & Moere, A. V. (2007). The effect of aesthetic on the usability of data visualization. In *Information visualization, 2007. iv'07. 11th international conference* (pp. 637–648). IEEE.

Chae, J., Thom, D., Bosch, H., Jang, Y., Maciejewski, R., Ebert, D. S., & Ertl, T. (2012). Spatiotemporal Social Media Analytics for Abnormal Event Detection and Examination using Seasonal-Trend Decomposition.

Cheong, M., & Lee, V. C. S. (2011). A microblogging-based approach to terrorism informatics: Exploration and chronicling civilian sentiment and response to terrorism events via Twitter. *Information Systems Frontiers*, *13*(1), 45–59.

Christ, R. E. (1975). Review and analysis of color coding research for visual displays. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *17*(6), 542–570.

Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, *79*(387), 531–554.

Conati, C., & Merten, C. (2007). Eye-tracking for user modeling in exploratory learning environments: An empirical evaluation. *Knowledge-Based Systems*, *20*(6), 557–574.

Diakopoulos, N., Naaman, M., & Kivran-Swaine, F. (2010). Diamonds in the rough: Social media visual analytics for journalistic inquiry. In *Proceedings of the ieee symposium on visual analytics science and technology.*

Dickerson, M., Eppstein, D., Goodrich, M. T., & Meng, J. Y. (2005). Confluent drawings: Visualizing non-planar diagrams in a planar way. *J. Graph Algorithms Appl*, *9*(1), 31–52.

Donath, J., & Boyd, D. (2004). Public displays of connection. *bt technology Journal*, *22*(4), 71–82.

Dork, M., Gruen, D., Williamson, C., & Carpendale, S. (2010). A visual backchannel for large-scale events. *Visualization and Computer Graphics, IEEE Transactions on*, *16*(6), 1129–1138.

Douglas, M., Brian, R., & Arthur, M. (2005). *Cognitive Psychology.* John Wiley & Sons.

Dudas, P. M. (2013a). *Cooperative, Dynamic Twitter Parsing and Visualization for Dark Network Analysis.* Fort Worth, TX. Retrieved from https://www.ideals.illinois.edu/bitstream/handle/2142/39982/295.pdf?sequence=2

Dudas, P. M. (2013b). *Dynamic Twitter Network Analysis for Dark Network Analysis.* West Point, New York.

Duncan, C. A., Eppstein, D., Goodrich, M. T., Kobourov, S. G., & Nöllenburg, M. (2011). Lombardi drawings of graphs. In *Graph drawing* (pp. 195–207). Springer.

Dwyer, T. (2009). Scalable, versatile and simple constrained graph layout. In *Computer graphics forum* (Vol. 28, pp. 991–998). Wiley Online Library.

Dwyer, T., Riche, N. H., Marriott, K., & Mears, C. (2013). Edge Compression Techniques for Visualization of Dense Directed Graphs. *IEEE transactions on visualization and computer graphics*, *19*(12), 2596–2605.

Eades, P., & Xuemin, L. (1989). How to draw a directed graph. In *Visual languages, 1989., ieee workshop on* (pp. 13–17). IEEE.

Ehmke, C., & Wilson, S. (2007). Identifying web usability problems from eye-tracking data. In *Proceedings of the 21st british hci group annual conference on people and computers: Hci... but not as we know it-volume 1* (pp. 119–128). British Computer Society.

Erdös, P., & Rényi, A. (1959). On random graphs. *Publicationes Mathematicae Debrecen*, *6*, 290–297.

Ersoy, O., Hurter, C., Paulovich, F. V., Cantareiro, G., & Telea, A. (2011). Skeleton-based edge bundling for graph visualization. *Visualization and Computer Graphics, IEEE Transactions on*, *17*(12), 2364–2373.

Everton, S. F. (2012). Network Topology, Key Players, and Terrorist Network. *Connections*, *32*(1), 12–19.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology*, *47*(6), 381.

Frey, B. J., & Dueck, D. (2007). Clustering by passing messages between data points. *science*, *315*(5814), 972–976.

Garey, M. R., & Johnson, D. S. (1983). Crossing number is NP-complete. *SIAM Journal on Algebraic Discrete Methods*, *4*(3), 312–316.

Giancoli, D. C. (2008). *Physics for scientists and engineers with modern physics.* Pearson Education.

Gibson, J. J. (1977). The theory of affordances. *Hilldale, USA*.

Gilbert, E. N. (1959). Random graphs. *The Annals of Mathematical Statistics*, *30*(4), 1141–1144.

Gillies, S., Bierbaum, A., Lautaportti, K., & Tonnhofer, O. (n.d.). Shapely. *GIS-Python Lab*.

Goldberg, J. H., & Kotval, X. P. (1999). Computer interface evaluation using eye movements: methods and constructs. *International Journal of Industrial Ergonomics*, *24*(6), 631–645.

Goldberg, J. H., Stimson, M. J., Lewenstein, M., Scott, N., & Wichansky, A. M. (2002). Eye tracking in web search tasks: design implications. In *Proceedings of the 2002 symposium on eye tracking research & applications* (pp. 51–58). ACM.

Görg, C., Pohl, M., Qeli, E., & Xu, K. (2007). Visual representations. In *Human-centered visualization environments* (pp. 163–230). Springer.

Gou, L., & Zhang, X. (2011). TreeNetViz: Revealing Patterns of Networks over Tree Structures. *Visualization and Computer Graphics, IEEE Transactions on*, *17*(12), 2449–2458.

Gou, L., Zhang, X. L., Luo, A., & Anderson, P. F. (2012). SocialNetSense: Supporting Sensemaking of Social and Structural Features in Networks with Interactive Visualization.

Granka, L. A., Joachims, T., & Gay, G. (2004). Eye-tracking analysis of user behavior in WWW search. In *Proceedings of the 27th annual international acm sigir conference on research and development in information retrieval* (pp. 478–479). ACM.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. *ACM SIGKDD Explorations Newsletter*, *11*(1), 10–18.

Heer, J., Bostock, M., & Ogievetsky, V. (2010). A tour through the visualization zoo. *Commun. ACM*, *53*(6), 59–67.

Heer, J., & Boyd, D. (2005). Vizster: Visualizing online social networks.

Holten, D., Isenberg, P., Fekete, J.-D., & Van Wijk, J. (2010). Performance Evaluation of Tapered, Curved, and Animated Directed-Edge Representations in Node-Link Graphs.

Holten, D., Isenberg, P., Van Wijk, J. J., & Fekete, J.-D. (2011). An extended evaluation of the readability of tapered, animated, and textured directed-edge representations in node-link graphs. In *Pacific visualization symposium (pacificvis), 2011 ieee* (pp. 195–202). IEEE.

Huang, W. (2007). Using eye tracking to investigate graph layout effects. In *Visualization, 2007. apvis'07. 2007 6th international asia-pacific symposium on* (pp. 97–100). IEEE.

Huang, W., & Eades, P. (2005). How people read graphs. In *proceedings of the 2005 asia-pacific symposium on information visualisation-volume 45* (pp. 51–58). Australian Computer Society, Inc.

Huang, W., Hong, S.-H., & Eades, P. (2006). Layout effects on sociogram perception. In *Graph drawing* (pp. 262–273). Springer.

Hyrskykari, A., Ovaska, S., Majaranta, P., Räihä, K.-J., & Lehtinen, M. (2008). Gaze path stimulation in retrospective think-aloud. *Journal of Eye Movement Research*, *2*(4), 1–18.

Jakobsen, T. (2001). Advanced character physics. In *Game developers conference* (pp. 383–401).

Jianu, R., Rusu, A., Fabian, A. J., & Laidlaw, D. H. (2009). A coloring solution to the edge crossing problem. In *Information visualisation, 2009 13th international conference* (pp. 691–696). IEEE.

Jianu, R., Rusu, A., Hu, Y., & Taggart, D. (2014). How to display group information on node-link diagrams: an evaluation. *Visualization and Computer Graphics, IEEE Transactions on*, *20*(11), 1530–1541.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness* (No. 6). Harvard University Press.

Johnson-Laird, P. N., Legrenzi, P., Girotto, V., Legrenzi, M. S., & Caverni, J.-P. (1999). Naive probability: a mental model theory of extensional reasoning. *Psychological review*, *106*(1), 62.

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive psychology*, *8*(4), 441–480.

Kempe, D., Kleinberg, J., & Tardos, É. (2003). Maximizing the spread of influence through a social network. In *Proceedings of the ninth acm sigkdd international conference on knowledge discovery and data mining* (pp. 137–146). ACM.

Kempton, W. (1986). Two Theories of Home Heat Control*. *Cognitive Science*, *10*(1), 75–90.

Kohonen, T. (1984). Self-organization and associative memory Springer. *New York Berlin Heideberg*.

Krzywinski, M., Birol, I., Jones, S. J. M., & Marra, M. A. (2011). Hive plotsrational approach to visualizing networks. *Briefings in Bioinformatics*.

Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? In (pp. 591–600). ACM.

Lee, B., Plaisant, C., Parr, C. S., Fekete, J.-D., & Henry, N. (2006). Task taxonomy for graph visualization. In *Proceedings of the 2006 avi workshop on beyond time and errors: novel evaluation methods for information visualization* (pp. 1–5). ACM.

Lotan, G., Graeff, E., Ananny, M., Gaffney, D., Pearce, I., & Boyd, D. (2011). The revolutions were tweeted: Information flows during the 2011 Tunisian and Egyptian revolutions. *International Journal of Communication*, *5*, 1375–1405.

Mackinlay, J. (1986). Automating the design of graphical presentations of relational information. *Acm Transactions On Graphics (Tog)*, *5*(2), 110–141.

Marr, D. (1982). Vision: A computational investigation into the human representation and processing of visual information, Henry Holt and Co. *Inc., New York, NY*, 2.

Marriott, K., Purchase, H., Wybrow, M., & Goncu, C. (2012). Memorability of visual features in network diagrams. *Visualization and Computer Graphics, IEEE Transactions on*, *18*(12), 2477–2485.

Matsuo, Y., Hamasaki, M., Nakamura, Y., Nishimura, T., Hasida, K., Takeda, H., ... Ishizuka, M. (2006). Spinning multiple social networks for semantic web. In (Vol. 21, p. 1381). Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999.

McCrum-Gardner, E. (2008). Which is the correct statistical test to use? *British Journal of Oral and Maxillofacial Surgery*, *46*(1), 38–41.

McGuffin, M. J., & Jurisica, I. (2009). Interaction techniques for selecting and manipulating subgraphs in network visualizations. *Visualization and Computer Graphics, IEEE Transactions on*, *15*(6), 937–944.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology*, 415–444.

Meyer, M., Munzner, T., & Pfister, H. (2009). MizBee: a multiscale synteny browser. *Visualization and Computer Graphics, IEEE Transactions on*, *15*(6), 897–904.

Moere, A. V., Tomitsch, M., Wimmer, C., Christoph, B., & Grechenig, T. (2012). Evaluating the effect of style in information visualization. *Visualization and Computer Graphics, IEEE Transactions on*, *18*(12), 2739–2748.

Moreno, J. L. (1960). The sociometry reader.

Netzel, R., Burch, M., & Weiskopf, D. (2014). Comparative Eye Tracking Study on Node-Link Visualizations of Trajectories.

Newman, M. (2010). *Networks: an introduction.* Oxford University Press, Inc.

Nielsen, F. Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *Arxiv preprint arXiv:1103.2903*.

Nowell, L., Schulman, R., & Hix, D. (2002). Graphical encoding for information visualization: an empirical study. In *Information visualization, 2002. infovis 2002. ieee symposium* *on* (pp. 43–50). IEEE.

Nowell, L. T. (1997). Graphical encoding for information visualization: using icon color, shape, and size to convey nominal and quantitative data.

Perer, A., & Shneiderman, B. (2008). Integrating statistics and visualization: case studies of gaining clarity during exploratory data analysis. In (pp. 265–274). ACM.

Pernice, K., & Nielsen, J. (2009). How to conduct eyetracking studies. *Nielsen Norman Group*.

Pohl, M., Schmitt, M., & Diehl, S. (2009). Comparing the readability of graph layouts using eyetracking and task-oriented analysis. In *Proceedings of the fifth eurographics conference on computational aesthetics in graphics, visualization and imaging* (pp. 49–56). Eurographics Association.

Poole, A., & Ball, L. J. (2006). Eye tracking in HCI and usability research. *Encyclopedia of Human-Computer Interaction, C. Ghaoui (ed.)*.

Poole, A., Ball, L. J., & Phillips, P. (2005). In search of salience: A response-time and eye-movement analysis of bookmark recognition. In *People and computers xviiidesign for life* (pp. 363–378). Springer.

Purchase, H. (1997). Which aesthetic has the greatest effect on human understanding? In *Graph drawing* (pp. 248–261). Springer.

Purchase, H. (2000). Effective information visualisation: a study of graph drawing aesthetics and algorithms. *Interacting with computers*, *13*(2), 147–162.

Purchase, H. (2002). Metrics for graph drawing aesthetics. *Journal of Visual Languages & Computing*, *13*(5), 501–516.

Purchase, H., Carrington, D., & Allder, J.-A. (2002). Empirical evaluation of aesthetics-based graph layout. *Empirical Software Engineering*, *7*(3), 233–255.

Raab, J., & Milward, H. B. (2003). Dark networks as problems. *Journal of Public Administration Research and Theory*, *13*(4), 413–439.

Rock, I., & Palmer, S. (1990). The legacy of Gestalt psychology. *Scientific American*, *263*(6), 84–90.

Rusu, A., Fabian, A. J., & Jianu, R. (2011). Using the gestalt principle of closure to alleviate the edge crossing problem in graph drawings. In *Information visualisation (iv), 2011*

15th international conference on (pp. 488–493). IEEE.

Saffrey, P., & Purchase, H. (2008). The mental map versus static aesthetic compromise in dynamic graphs: a user study. In *Proceedings of the ninth conference on australasian user interface-volume 76* (pp. 85–93). Australian Computer Society, Inc.

Schiessl, M., Duda, S., Thölke, A., & Fischer, R. (2003). Eye tracking and its application in usability and media research. *MMI-interaktiv Journal*, *6*, 41–50.

Segel, E., & Heer, J. (2010). Narrative visualization: Telling stories with data. *Visualization and Computer Graphics, IEEE Transactions on*, *16*(6), 1139–1148.

Selassie, D., Heller, B., & Heer, J. (2011). Divided edge bundling for directional network data. *Visualization and Computer Graphics, IEEE Transactions on*, *17*(12), 2354–2363.

Shah, D. V., Kwak, N., & Holbert, R. L. (2001). " Connecting" and" disconnecting" with civic life: Patterns of Internet use and the production of social capital. *Political Communication*, *18*(2), 141–162.

Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. In *Visual languages, 1996. proceedings., ieee symposium on* (pp. 336–343). IEEE.

Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences (McGraw-Hill, New York).*

Treisman, A. (1985). Preattentive processing in vision. *Computer vision, graphics, and image processing*, *31*(2), 156–177.

Tufte, E. R. (2006). *Beautiful evidence* (Vol. 1). Graphics Press Cheshire, CT.

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting elections with twitter: What 140 characters reveal about political sentiment. In (pp. 178–185).

Turner, M. (2006). *The artful mind: cognitive science and the riddle of human creativity.* Oxford University Press.

Tutte, W. T. (1963). How to draw a graph. *Proc. London Math. Soc*, *13*(3), 743–768.

Wang, C., & Shen, H.-W. (2011). Information theory in scientific visualization. *Entropy*, *13*(1), 254–273.

Ware, C. (2012). *Information visualization: perception for design.* Elsevier.

Ware, C., Purchase, H., Colpoys, L., & McGill, M. (2002). Cognitive measurements of graph aesthetics. *Information visualization*, *1*(2), 103–110.

Wellman, B., Haase, A. Q., Witte, J., & Hampton, K. (2001). Does the Internet increase, decrease, or supplement social capital? Social networks, participation, and community commitment. *American behavioral scientist*, *45*(3), 436–455.

Whitehouse, K. (2014). *Give me a beak! Street squawks over Twitter user numbers.*

Wichary, M. (2005). *An interactive presentation of Fittss Law.* (Vol. 2014) (No. 01-10). Retrieved from http://fww.few.vu.nl/hci/interactive/fitts/

Wickens, C. D., Gordon, S. E., & Liu, Y. (2004). An introduction to human factors engineering.

Wikipedia. (2014). *Twitter* (Vol. 2014) (No. 20-May-2014). Wikipedia,The Free Encyclopedia. Retrieved from http://en.wikipedia.org/w/index.php?title=Twitter{&}oldid=608646356

Xu, K., Rooney, C., Passmore, P., Ham, D.-H., & Nguyen, P. H. (2012). A user study on curved edges in graph visualization. *Visualization and Computer Graphics, IEEE Transactions on*, *18*(12), 2449–2456.

Yardi, S., & Boyd, D. (2010). Dynamic debates: An analysis of group polarization over time on twitter. *Bulletin of Science, Technology & Society*, *30*(5), 316–327.

Zinsmaier, M., Brandes, U., Deussen, O., & Strobelt, H. (2012). Interactive Level-of-Detail Rendering of Large Graphs. *Visualization and Computer Graphics, IEEE Transactions on*, *18*(12), 2486–2495.