The Evaluation of Information Visualization Techniques

Using Eye Tracking

by

Qing Liu

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved July 2015 by the
Graduate Supervisory Committee:

Ann McKenna, Chair
Jennifer Bekki
Nancy Cooke

ARIZONA STATE UNIVERSITY

August 2015

ABSTRACT

Node-link diagrams are widely used to visualize the relational structure of real world datasets. As identical data can be visualized in infinite ways by simply changing the spatial arrangement of the nodes, one of the important research topics of the graph drawing community is to visualize the data in the way that can facilitate people's comprehension. The last three decades have witnessed the growth of algorithms for automatic visualization. However, despite the popularity of node-link diagrams and the enthusiasm in improving computational efficiency, little is known about how people read these graphs and what factors (layout, size, density, etc.) have impact on their effectiveness (the usability aspect of the graph, e.g., are they easy to understand?). This thesis is comprehensive research to investigate the factors that affect people's understanding of node-link diagrams using eye-tracking methods. Three experiments were conducted, including 1) a pilot study with 22 participants to explore the layout and size effect; 2) an eye tracking experiment with 43 participants to investigate the layout, size and density effect on people's graph comprehension using abstract node-link diagram and generic tasks; and 3) an eye tracking experiment with the same participants to investigate the same effects using a real visualization analytic application. Results showed that participants' spatial reasoning ability had significant impact on people's graph reading performance. Layout, size, and density were all found to be significant effects under different task circumstances. The applicability of the eye tracking methods on visualization evaluation has been confirmed by providing detailed evidence that demonstrates the cognitive process of participants' graph reading behavior.

DEDICATION

I dedicate this dissertation to my Mom, Dad, and Xin. Thank you for being there for me, I love you all.

ACKNOWLEDGMENTS

I sincerely thank Dr. Ann McKenna for being an incredible advisor, mentor, and friend. Your patience, flexibility, and faith in me during the dissertation process enabled me to attend to life while also earning my Ph.D. Thank you, Ann!

I also would like to thank my committee members Dr. Nancy Cooke and Dr. Jennifer Bekki. Your academic support and input are greatly appreciated. Thank you!

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

Figure Page

CHAPTER 1

INTRODUCTION

Graphs are defined as "a set of vertices and set of edges that connect the vertices (Battista, Eades, Tamassia, & Tollis, 1998)." In the context of mathematics and computer science, a graph is a formal mathematical representation of a network. Graphs are widely used to model the relational structure of real world data, such as social networks, computer networks, partial orders, and algebraic geometry, etc. Graphs are usually visualized as node-link diagrams. One challenge of the graph drawing community is to take a set of nodes and their relationships as input and automatically visualize them to optimize both efficiency (e.g. how much time it costs to produce/visualize the graph?) and effectiveness (e.g., how much cognitive and perceptive resources are required for viewers to understand the embedded information correctly?). In the past three decades, most efforts have been placed on improving computational efficiency of the automatic graph layout algorithms, whereas relatively little attention has been given to understanding how people actually read the graphs.

In order to ensure the readability of the node-link diagrams, many aesthetic criteria (rules to layout the graphs) have been proposed (Bennett, Ryall, Spalteholz, & Gooch, 2007; Coleman & Parker, 1996). These criteria are used as quality measures of the graphs. However, in the absence of empirical support, as most of these criteria are proposed based on intuition of the designers, their relationship with better comprehensive performance cannot be guaranteed. Furthermore, as most criteria are mutually exclusive (Purchase, Carrington, & Allder, 2002), their relative importance needs to be ranked for designers to make compromises among them to achieve both good readability and efficiency. In addition, empirical evidence has revealed that useful layouts for certain application domains obey different aesthetic criteria (Purchase et al., 2002). There is a need to validate the criteria both on abstract graphs and in a real application context.

Eye tracking is a research method to measure the individual's eye movements. It provides information on where a person is looking at a specific time and how their eyes shift from one place to another (Rayner, 1998). In visualization evaluation research, eye tracking methodology can help uncover the subtle cognitive processes that are otherwise hard to observe using traditional usability measurements (Goldberg & Helfman, 2011).

The focus of this dissertation research is to empirically investigate how people read, interpret, and respond to different graph layouts, sizes and densities using eye-tracking methodology. By using eye-tracking methods, the study is grounded in data that relates directly to cognitive processing for how users understand the visualizations across task contexts, graph sizes and densities, rather than based on self-report, or perception data. Contributions of this study include: 1) determining which layout should be used and what properties (edge crossing frequency, crossing angle, etc.) should be optimized given certain tasks and context; 2) understanding the cognitive process of people's graph-reading behavior; 3) suggestions for designers to choose appropriate algorithms to visualize a node-link diagram for a certain applications.

The remainder of this document is structured as follows. The “background and related” work section provides an overview of the research literature on graph theory and eye tracking methods. This section also introduces and justifies the choice of eye tracking methodology to evaluate graph visualizations. The “pilot study” section describes a pilot eye tracking study to investigate the layout effect on graph reading performance. Based on the lessons learned from this pilot study, two formal eye tracking experiments were structured to investigate the layout, size and density effect on people's graph comprehension using abstract node-link diagrams (“experiment with abstract graphs”) and a real application (“experiment with DIA2”), respectively. Finally, the findings and suggestions for future research are discussed in the “discussion and conclusion” section.

CHAPTER 2

BACKGROUND AND RELATED WORK

The purpose of this section is to provide the theoretical and methodological background for understanding the rest of this document. The chapter begins with a review of what is known about the aesthetic criteria for laying graphs out in the context of graph theory. The need to validate these criteria for graph layouts, which is the main focus of this study, is also discussed in this section. Also included is an extensive discussion of using eye tracking as a research method for visualization evaluation, including an introduction to eye tracking technology, a review on previous eye tracking research in the Human-Computer-Interaction (HCI) field and a discussion of commonly used eye movement metrics.

Graph Theory

In the context of mathematics and computer science, the study of graphs is called Graph Theory.

Graph and node-link diagram. A great number of real-world data sets have relational structures that consist of entities and the relationships between them (e.g., social networks, citation networks, communication networks, and neural networks, etc.). Graphs are usually visualized as node-link diagrams for an easier understanding of the embedded information.

A visualization of a graph is only useful when the node-link diagram is readable and can convey the underlying information effectively. One issue with graph drawing is that the same data set can be visualized in infinite ways by simply changing the spatial arrangement of the nodes (see Figure 1 as an example of using different layouts to visualize identical information). Empirical studies have shown that spatial layout has an effect on people's comprehension of the graphs (Purchase, 1997; Purchase, Cohen, & James, 1996; Ware, Purchase, Colpoys, & McGill, 2002).The two main concerns of the graph drawing community are: 1) the computational *efficiency* of constructing geometric representations of abstract graphs; and 2) the *effectiveness* of conveying the underlying information to viewers (Purchase, McGill, & Colpoys, 2001).

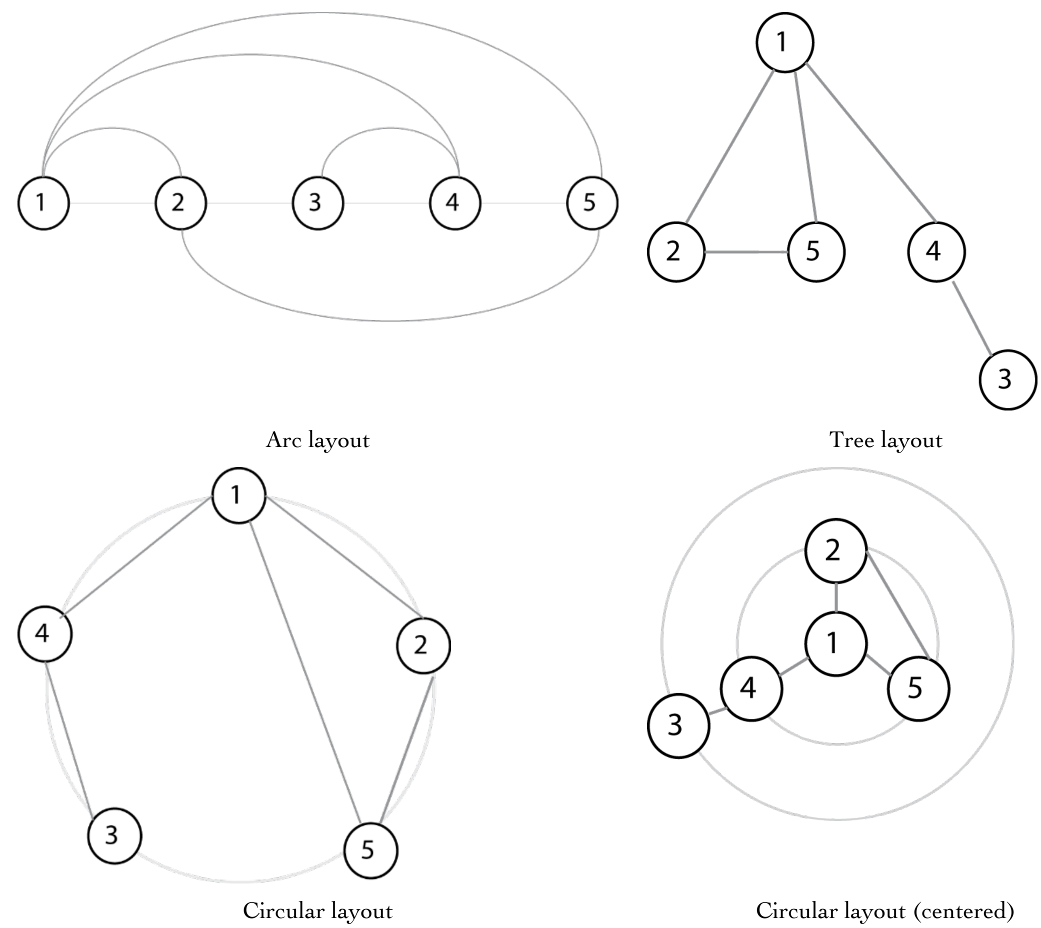


Figure 1. Different layouts representing identical information.

Aesthetic criteria. Aesthetic criteria (rules for laying out graphs) have been widely used as quality measures to evaluate the "goodness" of a visualization (W. Huang, 2013). In the past three decades, a variety of aesthetic criteria have been proposed with an assumption that they will improve the readability and understanding of graphs. Table 1 summarized the most accepted aesthetic criteria.

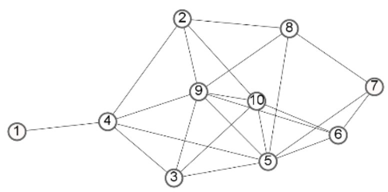
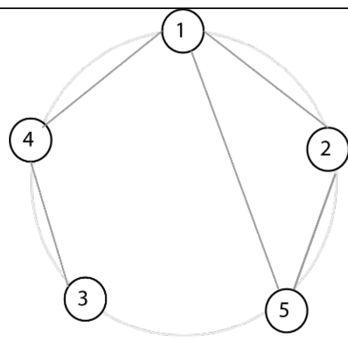
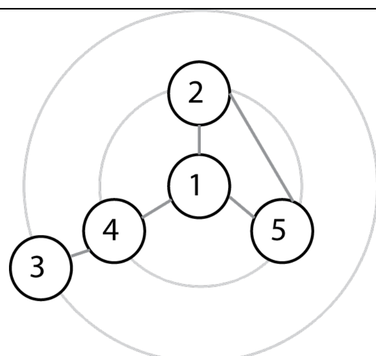
Table 1. Commonly accepted aesthetic criteria for graph visualization.

Concerns	**Brief description**	**Proposed by**
Crosses	The number of edge crossings in the drawing should be minimized	(Reingold & Tilford, 1981)
Bends	The total number of bends in polyline edges should be minimized	(Tamassia, 1987)
Angles	The minimum angle between edges extending from a node should be maximized	(Coleman & Parker, 1996; Gutwenger & Mutzel, 1998)
Orthogonally	Fix nodes and edges to an orthogonal grid	(Papakostas & Tollis, 2000; Tamassia, 1987)
Symmetry	Where possible, a symmetrical view of the graph should be displayed	(Gansner & North, 1998)

By far the most accepted aesthetic criteria for graph visualization is to *minimize the number of edge crossings* (Davidson & Harel, 1996; Purchase, 2002; Tamassia, Di Battista, & Batini, 1988; Taylor & Rodgers, 2005). Because of its great impact on graph comprehension, extensive validation of this criterion has been carried out to understand user preference and optimize user performance. With the same idea in mind, it is believed that *minimizing the number of edge bends* can also benefit the effectiveness of graph visualization. Edges with sharp bends are more difficult to follow, as they are more likely to be perceived as two separate objects. If several edges are connected to a same node, they should be spaced at even angles around the node. This leads to the criterion to *maximize minimum edge angles* between all edges of a node (Purchase, 2002; Taylor & Rodgers, 2005). The side effect of these criteria is that the nodes with more connections, which are more likely to be the important nodes, may be put closer to the center. To satisfy these criteria, graph designers may have to make compromises with competing structural criteria. Purchase (2002) proposed to *maximize edge orthogonality* by placing the edges and edge segments to match the lines of an imaginary grid. It is believed that this can reduce edge crossings and maximize the angles. Besides the spatial relationships between nodes and edges, the overall graph layout is also an important factor of aesthetics. *Maximizing global symmetry* is the one of the most widely studied criteria.

Most of the criteria could be mutually exclusive (optimizing one criterion would result in suboptimal performance of another criterion). The graph drawing community had developed many layout algorithms to meet the requirements of both computational efficiency and optimizing certain aesthetic criteria. See Table 2 for a review of widely used layouts.

Table 2. Examples of widely used layouts.

Layout	Sample graph	Description	Reference
Force-directed		Forces [1] are assigned among the edges and nodes based on their relative positions. These forces are then used to simulate the motion of edges and nodes or to minimize their energy.	(Fruchterman & Reingold, 1991)
Circular		All nodes are places on a circle.	Scott (2000)
Concentric		All nodes are laid on circumference of circles in a way that their distances from the center exactly reflect their centrality[2] levels.	(Brandes, Kenis, & Wagner, 2003)

Criteria validation. There are two main issues with these aesthetic criteria. First, they were primarily proposed based on the intuition of algorithm designers. Thus, their

[1] Forces are assigned among the set of edges and the set of nodes. Typically, spring-like attractive forces based on Hooke's law are used to attract pairs of endpoints of the graph's edges towards each other; repulsive forces like those of electrically charged particles based on Coulomb's law are used to separate all pairs of nodes.

[2] Degree centrality, the nodes with high degree (connections) are likely to be at the intuitive center.

positive effect on human graph comprehension is not guaranteed, and empirical studies are needed to support the assumptions embedded in the criteria. Second, the literature suggests that the establishing of aesthetics should be based on empirical evidence and the theories of how people read graphs (W. Huang, 2013; Purchase, Pilcher, & Plimmer, 2012). Appendix A is a summary of studies on criteria validation. Studies are described based on their independent variable, dependent variable, number of subjects, stimuli, tasks, and conclusion.

Purchase is the pioneer of investigating the cognitive measure of the aesthetics of graph drawing. Her study in the mid-1990’s showed that human performance with node-link diagrams is negatively correlated with the number of edge crossing and bends (Purchase et al., 1996). By comparing the relative importance of five commonly accepted aesthetic criteria (bends, cross, angles, orthogonality, symmetry), Purchase and her colleagues found that edge crossing was the most prominent affecting factor, followed by edge bends and symmetry (Purchase, 1997). This finding coincided with the results of Körner et al.'s research, which argued that the reasoning stage of human's graph comprehension process was only affected by the visual property of edge crossing (Körner, 2011). Empirical evidence also revealed that the impact of edge crossings varies under different experimental settings: Huang's research on layout effect found that the edge crossing just affects human performance on path-finding tasks but not node-locating tasks (Huang, 2007b). In another study, Huang et al. found that edge crossing does not inhibit human performance when the crossing angle is large (Huang, Eades, & Hong, 2009). Ware et al. (Ware et al., 2002) found that path continuity also has an important impact on graph comprehension.

Beside the research on individual aesthetics, there is substantial research focused on the effect of layouts on graph comprehension. Pohl et al. (Pohl, Schmitt, & Diehl, 2009) conducted a study to look into the readability of three kinds of commonly used layouts: force-directed, hierarchical, and orthogonal and found that force-directed out-performed the other two layouts on almost every task. Burch et al. (Burch, Konevtsova, Heinrich, Hoeferlin, & Weiskopf, 2011) conducted a study to investigate the readability of layouts depicting hierarchical structures (traditional, orthogonal, and radial tree). The results showed that the traditional and orthogonal tree layouts significantly outperformed radial tree layouts.

Although previous empirical studies confirmed the strong influence of graph layout on the readability of node-link diagrams, a comprehensive analysis and comparisons of different layouts in terms of performance, quality, and relative importance is still lacking in the literature.

Specifically, the diagrams used in these prior studies usually have sparse nodes and low density. As a node-link diagram is inherently a space-inefficient representation that suffers from scalability problems for large datasets (Burch et al., 2013; Ghoniem, Fekete, & Castagliola), it is insufficient to investigate the readability of graphs without considering the influence of the size and density of the graph. In this dissertation research, graph size and density were investigated as independent variables.

Furthermore, as it is often not possible to optimize across multiple criteria; instead, compromises usually have to be made. It is important to understand the relative important priorities of aesthetic criteria. Purchase initiated the studies in this direction by prioritizing five aesthetics (edge crossing, number of bends, symmetry, maximizing the minimum angle between neighboring edges, and maximizing orthogonality); however, the knowledge gained from her study is useful only for a specific task (node locating). This dissertation research included more factors (layout, size, density and task types) to guarantee the generalizability of the results when the method or task is varied.

Finally, as Bennett et al. (Bennett et al., 2007) pointed out, semantics and tasks are as important as structure when creating graphs. The same factors were not only evaluated using abstract graphs and generic tasks, but also using a real application context.

Cognitive model on graph comprehension. Several cognitive models have been developed and compared for graph comprehension during the last two decades. Top-down models break the process into theoretical sub-processes that are pre-determined by an individual's viewing strategies. In contrast, bottom-up models provide implications to higher cognitive processes based on observed behaviors. Lohse (1992) proposed a top-down model to predict the time needed to complete common graphical tasks. Carpenter and Shah (1998) modeled the graph comprehension process as a series of internal sub-processes, including encoding, pattern interpretation, and integration. In line with Carpenter and Shah, Bojko and Stephenson (2005) proposed a model of visual search which includes two stages: 1) deployment of attention and 2) target processing. In the first stage, people have to focus their attention on the targets. The 'noticeability' of the target, in other words, the layout of the graph will affect the efficiency and effectiveness of people's comprehension. In the second stage, people need to make sense of the target and relate it to their goal of information searching. At this stage, the efficiency of the task is affected by the understandability of the target. In this dissertation research, Bojko and

Stephenson's model was used to define metrics to investigate the cognitive process of people's comprehension of node-link diagrams. Specifically, the graph comprehension process has been divided into two sub-stages: target searching and target comprehension.

Eye Tracking as a Research Method

The objectives of this section are threefold. First, the basics of eye tracking technology are introduced. Second, previous visualization evaluation studies that used eye-tracking methods are reviewed, and the reason to use eye tracking in this study is presented. Third, the advantages of a range of different eye movement metrics are illustrated with reference to state-of-the-art visualization evaluation research.

Eye tracking technology. Since the first use of eye tracking technology in reading research over 100 years ago (Rayner, Pollatsek, Ashby, & Clifton, 2012), many different methods have been used to track eye movements. Early historical methods are invasive, requiring electrodes to be mounted on the skin around the eye or wearing large contacts that covered the cornea. Most modern eye tracking systems are camera based, using video images of the eye to determine where a person is looking. This method relies on determining properties of the eye, such as iris-sclera boundary and corneal reflections (Duchowski, 2007).

The remote tracking system implemented in these methods uses infrared light directed to the eye to create strong reflections in target eye features in order to easily track them. Then, camera-based trackers compute the gaze direction of the eye along with head position by determining the relationship between the corneal reflection and pupil. (See Figure 2 as an illustration of corneal reflection and bright pupil as seen in the infrared camera image).

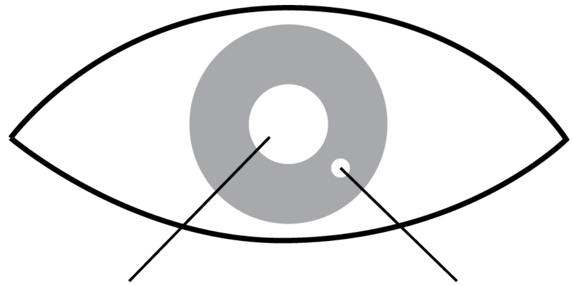


Figure 2. Bright pupil and corneal reflection as seen in the infrared camera image (adapted from Poole & Ball, 2006).

During analysis, eye movements are typically divided into fixations and saccades. A *fixation* is when the eye gaze pauses in a certain spot. A *saccade* is when it moves to another position. The resulting series of fixations and saccades is called a *scan-path*. Most information from the eye is made available during a fixation, but not during a saccade. In eye tracking, the fixation is the most important point of data.

Why use eye tracking in visualization evaluation. When people read graphs, their eye movements are not random, but guided toward interesting and informative regions (Buswell, 1935; Yarbus, Haigh, & Rigss, 1967). Just and Carpenter (1976) formulated one of the strongest validations for the power of eye tracking—the Mind-Eye Hypothesis. They argued that what people are looking at indicates what they are thinking about. During the 1980s, the Eye-Mind Hypothesis was often questioned in light of covert attention—the attention to something that one is not looking at, which people often do. However, in experimental scenarios when participants are looking at a visual stimulus, the eye-mind hypothesis usually holds true: people usually pay attention to and think about what they are looking at (Goldberg & Tang, 2011).

Using eye tracking can have several benefits for visualization evaluation research: first, other than traditional measurements such as response time and accuracy, sequences of fixations (scan-path) provided by eye movement data can help inform the strategy used by the viewer (Goldberg & Helfman, 2011). Understanding the difference between several design alternatives is valuable for improving the visualization to maximize efficiency. Second, eye tracking is sensitive enough to detect whether the predictable influence came from certain design features. Comparing the differences in strategies used by viewers between trials with and without error can provide diagnostic information for designers.

In sum, regarding the applicability of eye tracking method in this study, this study considered: 1) sometimes fixations do not necessarily translate into a conscious cognitive process; 2) fixations can be interpreted in different ways depending on the context and objective of the study; 3) during the processing of a visual scene, participants will move their eyes to relevant features in that scene; and 4) eye tracking data should not be used alone, but with self-report and interview data.

Previous eye tracking research. Eye tracking has been successfully used in psychology and education for many years. However, very few studies are available in the graph evaluation field. In 2005, Huang and Eades conducted the first eye tracking study to investigate how people read relational node-link graphs by looking into the effect of different layouts (Huang & Eades, 2005). The analysis of eye tacking data led them to the discovery of the geo-metric tendency in path-finding tasks of node-link diagram. Huang's research started the promising use of eye tracking in graph evaluation; however, by looking across available studies, we found that eye tracking data were mostly used for qualitative analysis in the form of heat-map visualization and video of eye movement trajectory, but rarely for statistical analysis and quantitative findings. As an attempt to fully understand the use of eye tracking data, Burch et al. calculated a transition matrix of the Areas of Interest (AOIs) to trace the attention distribution of the participants (Burch et al., 2011). However, the definition of AOI was somewhat arbitrary and may expose their study to questionable methodology. The only identified study that used eye-tracking data to extract quantitative conclusions is Körner's research on hierarchical graphs. In this work, the number of fixations in different phases is used to identify distinct stages of comprehension (Körner, 2011). In line with Körner's research, the same idea of using eye tracking data to distinguish different comprehension stages has been used in this dissertation research. However, unlike the bottom-up model adopted by Körner, in this research, participants' fixation data has been used to statistically support the two predefined sub-stages of graph comprehension. In order to support this top-down discovery path, other than the number of fixations, more metrics (e.g., time till first fixation on target, duration between target locating to task completion) have been included to provide quantitative evidence.

Eye movement metrics. There are over ten different kinds of eye movements (e.g., saccades, smooth pursuit, mergence, vestibular, and physiological nystagmus, etc.). Duchowski (2007) suggests that, based on the functionality of the eye movement, only fixation, smooth pursuits and saccades need be modeled to gain insights into the overt localization of visual attention. When the eyes focus on a point it is called a fixation (the duration varies from 100~600 milliseconds depending on the filter algorithm adopted), and saccades are the movements between these fixations. The duration of a fixation is usually an indication of information processing. Table 3 gives a review of commonly used eye movement metrics.

Table 3. Commonly used eye movement metrics.

Metrics	**Meaning / Interpretation**	**Reference**
Number of fixations	A higher number of fixations indicates less efficient search	(Goldberg & Kotval, 1999)
Fixation duration	A longer fixation duration indicates difficulty in extracting information, or it means that the object is more engaging in some way	(Just & Carpenter, 1976)
Time to first fixation	Faster time to first-fixation on an object or area means that it has better attention-getting properties	(Byrne, Anderson, Douglass, & Matessa, 1999)
Gaze	Gaze is usually the sum of all fixation durations within a prescribed area. Gaze is best used to compare attention distributed between targets.	(Mello-Thoms, Nodine, & Kundel, 2002)
Number of saccades	More saccades indicate more searching	(Goldberg & Kotval, 1999)
Saccade amplitude	Larger saccades[3] indicates more meaningful cues, as attention is drawn from distance	(Goldberg, Stimson, Lewenstein, Scott, & Wichansky, 2002)
Regressive saccades	Regressions indicate the presence of less meaningful cues.	(Sibert, Gokturk, & Lavine, 2000)
Scan-path duration	A longer-lasting scan-path indicates less efficient scanning	(Goldberg & Kotval, 1999)
Scan-path length	A longer scan-path indicates less efficient searching	(Goldberg et al., 2002)
Scan-path direction	Scan-path direction can determine a subject's search strategy	(Aaltonen, Hyrskykari, & Räihä, 1998)

[3] The amplitude of a saccade is the angular distance the eye travels during the movement.

CHAPTER 3

PILOT STUDY

This study served as a pilot for the following two formal, follow-up studies on graph layout, size and density effect on participants' graph-comprehension using abstract graphs and with a real application respectively (see Chapters 4 & 5). The objectives of this pilot study were two-fold: first, set the attribute value (reasonable graph size, path length, etc.) for the formal experiments by investigating the layout (force-directed, circular, and random) and size effect on participants' graph comprehension; and second, validate the applicability of the eye tracking method on graph evaluation. Accuracy, completion time, and eye movement of the participants were recorded as they conducted node-locating and path-finding tasks.

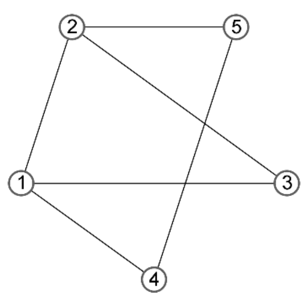
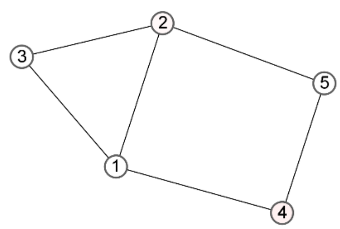
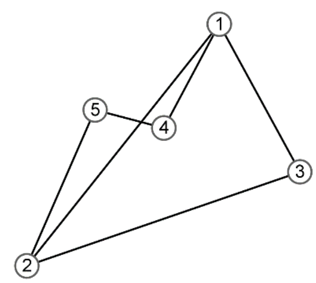
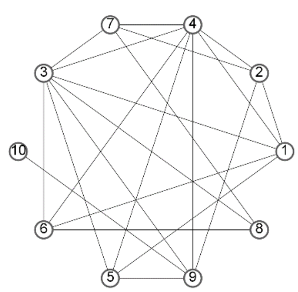
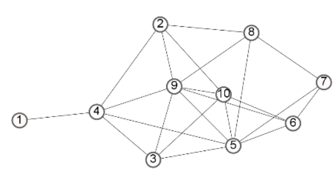
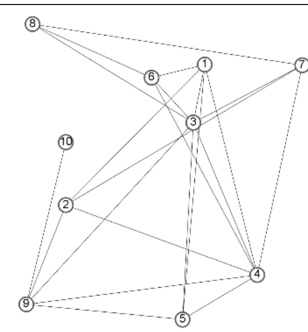
Method

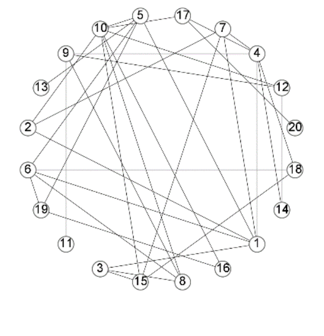
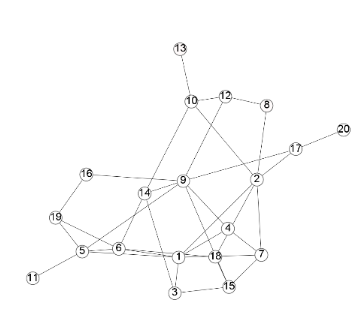
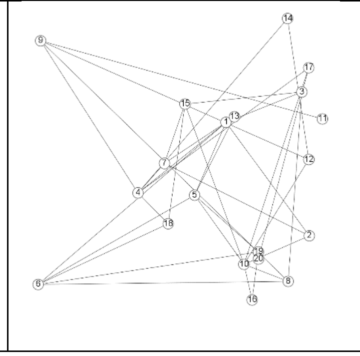
Participants. 22 (14 males and 8 females) volunteers were recruited from peers in similar graduate programs at the researcher's institution. They were graduate students from various fields, including applied psychology, computer science and engineering. All of the participants had normal vision and were regular computer users.

Apparatus. An EyeTech VT2 eye tracking system running QuickCapture software was used to collect eye movement data at a sampling rate of 35 Hz. Stimuli were presented on a 17-inch diagonally measured 1280×1024 display. Pupil position and corneal reflection from both eyes were used to locate participants' gaze positions on the screen. Heat-map and scan-path visualizations of participants' eye movement were produced using the open source software OGAMA (Open Gaze And Mouse Analyzer) for analysis purposes. Statistical analysis was conducted using Microsoft Excel and SPSS.

Stimuli. Graphs were generated using the open source software Gephi. Circular, force-directed and random layout algorithms were used to produce graphs in sizes of 5 nodes, 10 nodes, and 20 nodes (see Table 4). Nodes were labeled by numbers. Graphs with the same size illustrated the same underlying data structure. By using graphs that are equally unfamiliar and meaningless to users, this study focused on the abstract characteristics of graphs.

Table 4. Stimuli used in pilot study.

	Circular	Force-directed	Random
5			
10			

	Circular	Force-directed	Random
20			

Tasks. Participants were asked to conduct two generic tasks of graph comprehension: node-locating and path-finding. These tasks were chosen because of their frequency of user in real-world applications.

1. Node-locating: find the nodes with the highest number of connections.

2. Path-finding: find the shortest path between two highlighted nodes.

Experimental Design. The study followed a repeated-measures design.

The factors of interest include:

- Layout: force-directed, circular, and random;
- Graph size: graphs were created at three size levels to include 5, 10 and 20 nodes;
- Tasks: node-locating (find the nodes with the highest number of connections) Path-finding (find the shortest path between two highlighted nodes).

Each participant performed one trial of each layout type, each diagram size, and each task type, resulting in 18 trials in total.

Procedure. Participants were first introduced to the basic characteristics of node-link diagrams. Then, they were given time to ask questions and practice with two warm-up tasks (one for the node-locating and one for the path-finding task), during which they were instructed to put emphasis on correctly answering because the focus of the study was participants' graph reading strategy. A 16-point calibration was carried out every 3 trials to guarantee the reliability of the data. Participants were informed about the tasks before they saw the stimuli. They conducted one node-locating and one path-finding task on each stimuli. The same set stimuli including 18 node-link diagrams were presented to all participants. Time recording was started as soon as they saw the stimuli on the screen and ended when they gave answers verbally. Answers were recorded manually by the experiment coordinator. The whole experiment took about 30 minutes per participant.

Results & Discussion

Across all tasks, accuracy did not differ by gender, $F(1, 394) = 2.76$, $p = .10$. As illustrated in Figure 3, the path-finding task had a significantly higher accuracy compared to node-locating task. Since node-locating and path-finding are essentially two different kinds of cognitive tasks, the following analysis presents the findings in node-locating task and path-finding task separately.

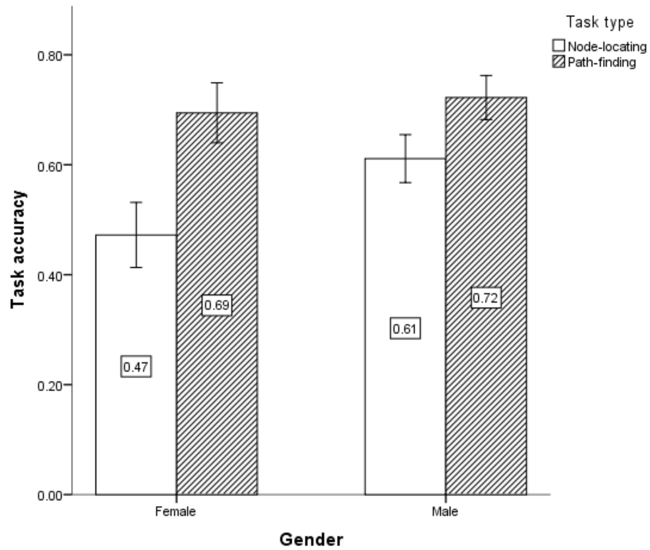


Figure 3. Task accuracy (percentage of correct answers) by gender and task type. Error bars indicate standard errors[4].

Node-locating task. Table 5 summarizes the mean and standard deviation of node-locating tasks. A two-way ANOVA was conducted to investigate the effect of graph size and layout on task accuracy and completion time. Results (see Table 6 for a summary of results of ANOVA analysis) show that both graph size and layout had a significant effect on task accuracy, $F(2, 189) = 33.01$, $p < .05$ and $F(2, 189) = 9.42$, $p < .05$ respectively. The larger the graph was, the lower the task accuracy participants achieved. The interaction between graph layout and size also affect task accuracy significantly, $F(4, 189) = 5.12$, $p < .05$). Figure 4 illustrates that at size 20, circular layout graphs had significantly higher accuracy than graphs in the other two layouts. Further analysis of participants' eye movement revealed that force-directed and random layouts had more back-and-forth checking on dense areas, whereas circular layouts provided participants with a visual "circle" to guide their eyes to compare the number of connections between

[4] Error bars used in graphs of this document indicate standard errors.

nodes. As a result, the participants' attention distributed on the circular layout graph was more 'organized' and in a 'circular' form, whereas more scattered with force-directed and random layout (see Figure 5 and Figure 7).

Table 5. Mean and standard deviation of task accuracy and time of node-locating tasks.

Layout		Accuracy Mean	Accuracy Std. Deviation	Time (s) Mean	Time (s) Std. Deviation
Circular	5	.86	.35	8.09	3.64
	10	.64	.49	19.73	9.58
	20	.68	.48	27.82	11.81
	Total	.73	.45	18.55	12.07
Force-directed	5	.82	.39	8.50	5.77
	10	.64	.49	19.00	8.68
	20	.14	.35	34.50	18.68
	Total	.53	.50	20.67	16.24
Random	5	.86	.35	8.55	8.61
	10	.41	.50	20.23	12.04
	20	.00	.00	48.86	37.35
	Total	.42	.50	25.88	28.51
Total	5	.85	.36	8.38	6.25
	10	.56	.50	19.65	10.05
	20	.27	.45	37.06	26.21
	Total	.56	.50	21.70	20.32

Table 6. Summarized results of ANOVA analysis of node-locating tasks. Significance level α = .05. Effect size calculated using partial eta squared (η^2).

	Dependent Variable							
	Accuracy				Time (s)			
Source	df	F	Sig.	η^2	df	F	Sig.	η^2
Layout	2	9.42	<.05	0.09	2	3.65	0.03	0.04
Size	2	33.01	<.05	0.26	2	53.55	<.05	0.36
Layout * Size	4	5.12	<.05	0.10	4	3.14	0.02	0.06
Error	189				189			
Total	198				198			
Corrected Total	197				197			

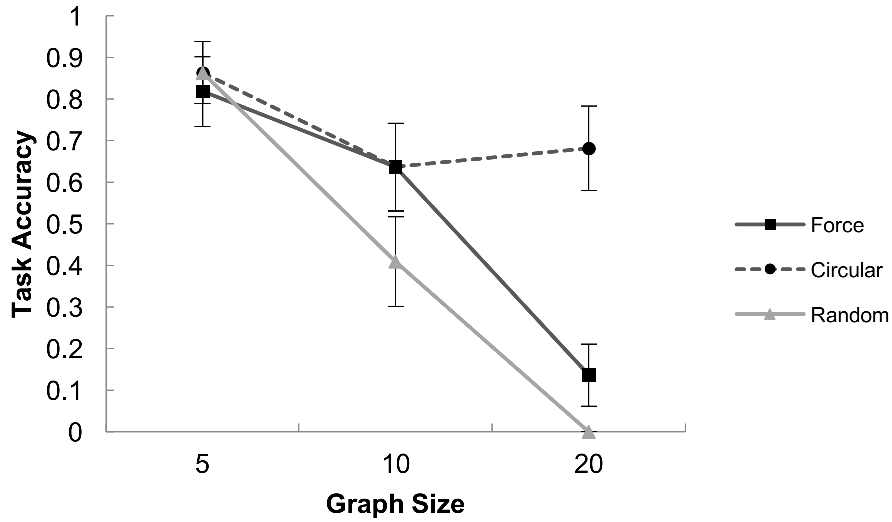
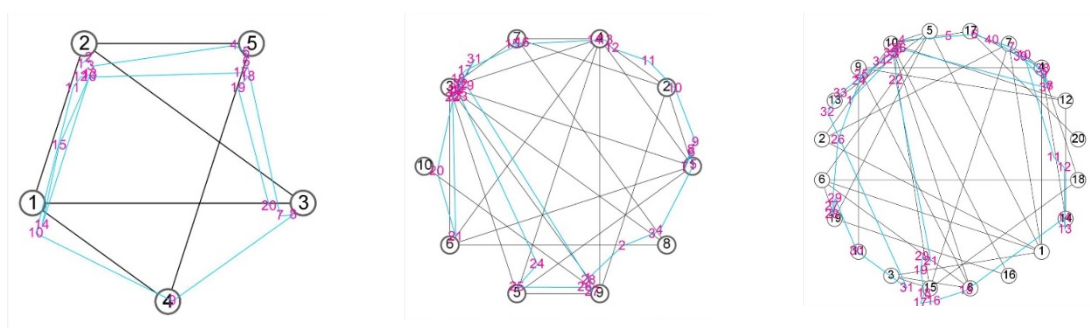
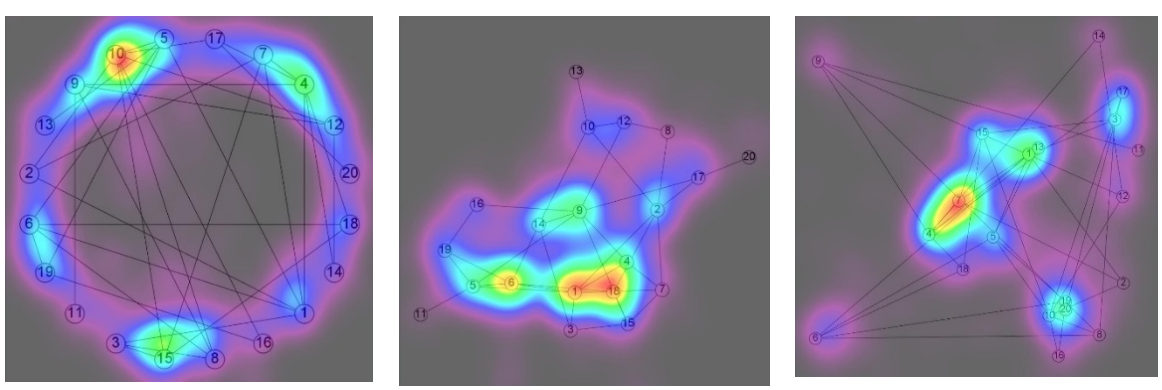


Figure 4. Graph size and layout effect on task accuracy of node-locating task.



Size 5 Size 10 Size 20

Figure 5. Scan-path visualizations of one participant' eye movements for node-locating tasks with circular layout graphs at three sizes. Pink number indicates the viewing sequence; blue lines represented the link between two successive fixations. Note the guide effect (circular scan-path) of the layout.



Circular layout Force-directed layout Random layout

Figure 6. Heat-map visualization of participants' aggregated eye movement of node-locating task with three layouts. Warmer color indicates more fixations in that area. Note back and forth movements confirming eye movements with force-directed and random layout graphs.

As illustrated in Figure 8, size, layout, and their interaction also had significant effect on completion time of node-locating task, $F(2, 189) = 53.55$, $p < .05$,$F(2, 189) = 3.65$, $p < .05$, and $F(4, 189) = 3.14$, $p < .05$. The layout effect was significant at graph size of 20 nodes. Random layout that didn't optimize the readability of the graphs at all had significantly longer time-on-task than graphs in force-directed and circular layout.

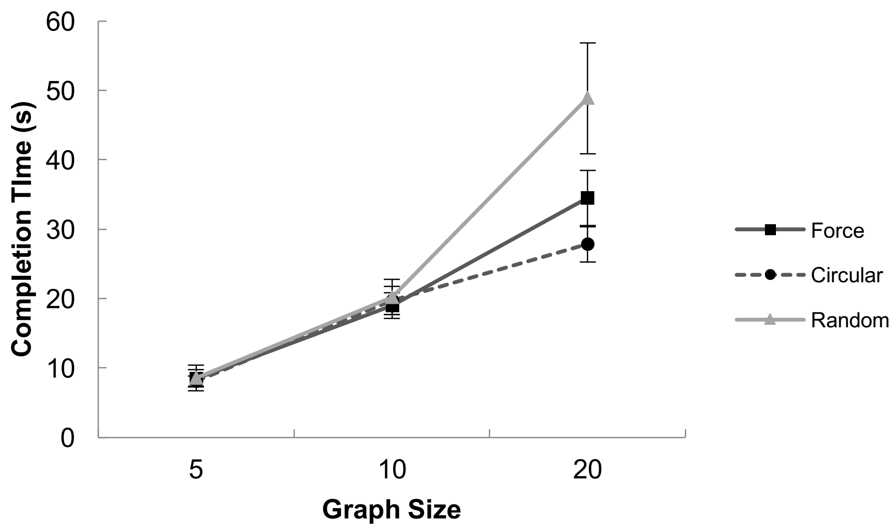


Figure 7.Graph size and layout effect on completion time of node-locating task.

Path-finding task. Table 7 summarizes the mean and standard deviation of path-finding tasks' accuracy and completion time. Two-way ANOVA (see Table 8 for the summary of the results) shows that size, layout and their interaction also had significant effect on accuracy of the path-finding tasks, $F(2, 189) = 66.18$, $p < .05$, $F(2, 189) = 86.20$, $p < .05$, $F(4, 189) = 22.88$, $p < .05$ respectively. Participants answered all path-finding tasks correctly with graphs in size 5. However, as graph size increased, the accuracy of random layout dropped dramatically and ended with no participants answering correctly for graph size20 (see Figure 8). Scan-path visualization of participants' eye movement data of path-finding tasks with the random layout shows frequent over-checking that indicated confusion with graphs (see *Figure 9*).

Table 7. Mean and standard deviation of task accuracy and completion time of path-finding tasks.

		Time (s)		Accuracy	
Layout		Mean	Std. Deviation	Mean	Std. Deviation
Circular	5	8.50	3.64	1.00	0.00
	10	35.91	18.96	0.86	0.35
	20	25.95	18.85	0.95	0.21
	Total	23.45	19.12	0.94	0.24
Force	5	7.82	8.30	1.00	0.00
	10	35.86	31.75	0.50	0.51
	20	16.95	6.32	0.95	0.21
	Total	20.21	22.34	0.82	0.39
Random	5	8.36	3.80	1.00	0.00
	10	27.91	18.71	0.14	0.35
	20	74.77	36.94	0.00	0.00
	Total	37.02	36.70	0.38	0.49
Total	5	8.23	5.59	1.00	0.00
	10	33.23	23.86	0.50	0.50
	20	39.23	34.98	0.64	0.48
	Total	26.89	27.98	0.71	0.45

Table 8. Summarized results of ANOVA analysis on path-finding tasks.

	Dependent Variable							
	Time (s)				Accuracy			
Source	df	F	Sig.	η^2	df	F	Sig.	η^2
Layout	2	13.21	0.00	0.12	2	86.20	0.00	0.48
Size	2	44.94	0.00	0.32	2	66.18	0.00	0.41
Layout * Size	4	20.80	0.00	0.31	4	28.88	0.00	0.38
Error	189				189			
Total	198				198			
Corrected Total	197				197			

In regard to completion time of the path-finding task, Figure 10 revealed that the layout effect is significant at graph size 20. Graphs that have been optimized using either the force-directed or circular layout algorithm had a significantly shorter time-on-task than the graphs that were randomly produced.

Circular layouts had significantly higher accuracy than force-directed layouts at size 10 but not 20. Interestingly, the accuracy increased for force-directed layouts as graph size increased from 10 to 20. As both end nodes (node 8 and 13, randomly selected) of this task were located in the upper-right corner of the graph, the shortest path between the end nodes was quite straightforward. The heat-map visualization revealed that participants' fixations almost exclusively focused on the upper-right corner and did not cross the graph, resulting in a shorter response time and higher accuracy for the path-finding task in this example. These same findings were confirmed during our interviews with the participants after the experiment: they found it is surprisingly easy to get the answer for this task because the two highlighted nodes are located close and somewhat isolated from other "messed-up" neighbors (see Figure 11).

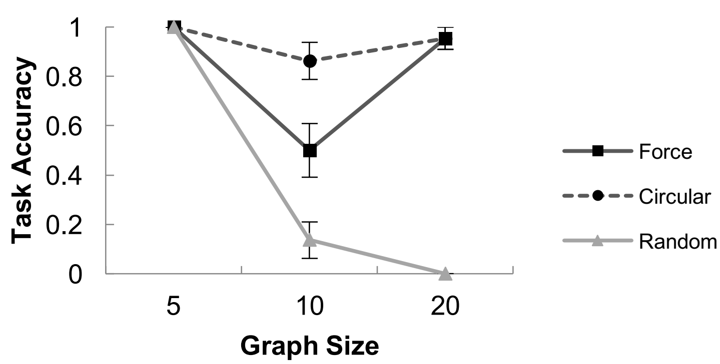
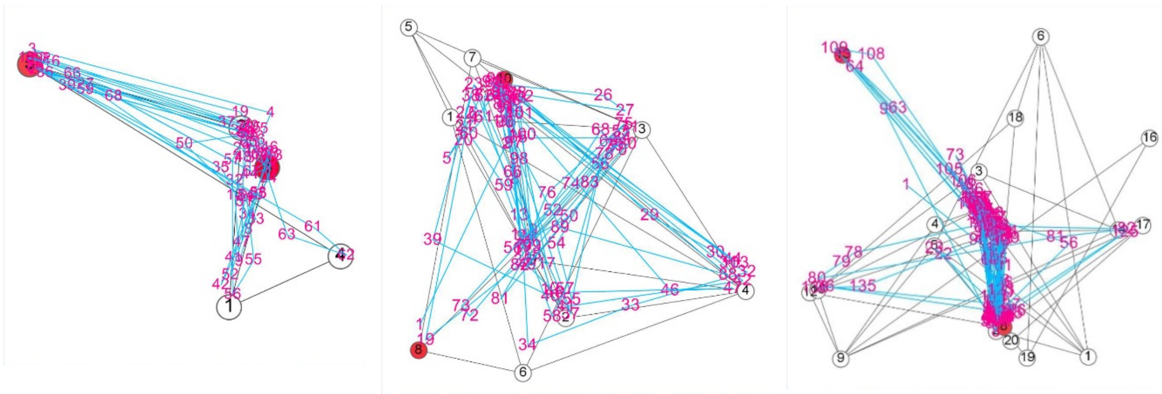


Figure 8. Graph size and layout effect on accuracy of path-finding task.



Size 5 Size 10 Size 20

Figure 9. Scan-path visualization of participants' eye movement on path-finding tasks with random layout at three sizes. Node the frequent back-and-forth checking behavior.

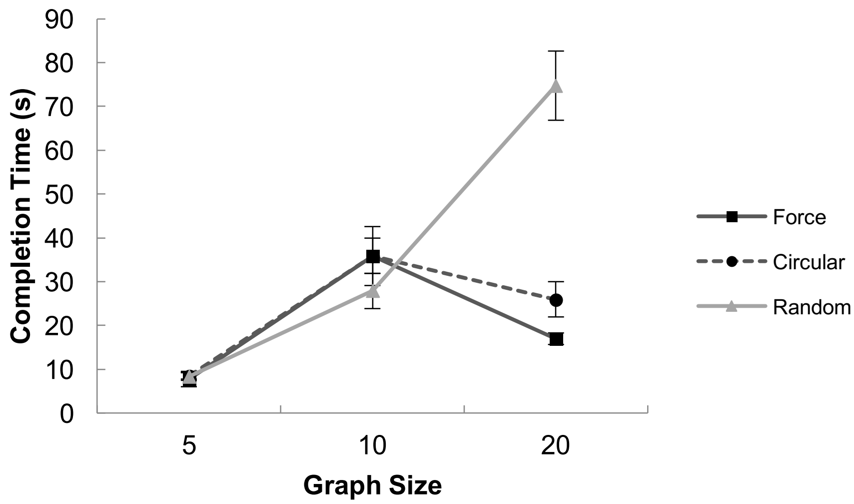
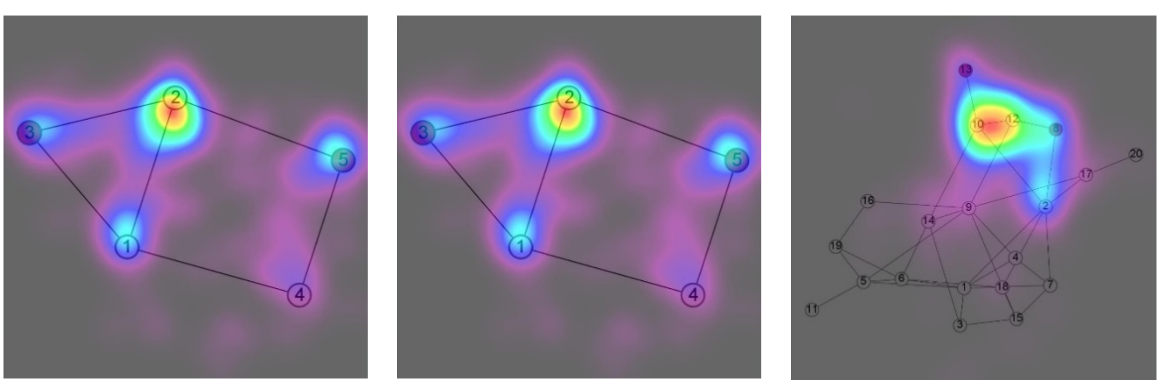


Figure 10. Graph size and layout effect on completion time of path-finding task.



Size 5 Size 10 Size 20

Figure 11. Heat-map visualization of participants' eye movement of path-finding tasks with graphs at three sizes. Note the centralized fixations of size20 graph.

Summary

The size and layout effect were confirmed by this pilot study. Participants' performance on tasks dropped as graph size increased. Compared to random layouts, participants performed better on those layouts that were optimized using some aesthetic criteria such as minimizing edge crossing and node overlapping. Specifically, for node-locating tasks, circular layout graph had significant higher task accuracy and short task completion time; for path-finding tasks, circular layout and force-directed layout performed equally well. However, the high accuracy with circular layouts on node-locating tasks was achieved when the graph was relatively sparse (density = .2). Although the graphs had identical density at same size level, density was not investigated as an independent variable and was not controlled strictly across three size levels. Thus, the same size and layout effect was investigated at different density levels (.2 and .4 respectively) in the following chapter (see Chapter 4).

The visualizations (heat-map & scan-path) of participants' eye movement have been used as qualitative support for the task accuracy and completion time findings. They provide insights from a cognitive perspective to answer 'how' questions arising from behavior observation. More metrics were included in the eye tracking analysis in the following experiments (see Chapter 4 & 5) to deepen our understanding of people's graph reading behavior.

CHAPTER 4

EXPERIMENT WITH ABSTRACT GRAPHS

This experiment was designed to investigate not only the layout effect but also the effect of graph density, as scalability is always a concern with node-link diagrams.

Based on the experience of the pilot study, several adjustments were made to the experimental design:

1. The three layouts (force-directed, circular, and concentric) investigated in the experiment were chosen not only because they are the most commonly used layouts for visualizing relational structure, but also because they are the exact layout used by Deep Insight Anytime Anywhere (DIA2), the application used as a task context in next experiment (see Chapter 5). A comparison was conducted to understand the layout effect on both abstract graphs and a real application—visualizations with contextual information. The random layout as a control group was dropped because the existence of layout effect has already been shown in the pilot study.
2. Density was included as an independent variable. Two density levels (0.2 and 0.4) were investigated in the experiment.
3. Spatial ability has been shown to relate to people's comprehension of graphs (Lohse, 1997). During the pilot study, it was noticed that participants who mentioned that they are good at video games (related to spatial ability) tended to have better task performance. Thus, a spatial reasoning test was added in the experiment to provide data to analyze the

relationship between people's spatial reasoning ability and their performance on node-link graph reading.

4. In the pilot study, the same diagram was used for both node-locating task and path-finding task if they were in the same layout and size. In this experiment, different graphs were used for two types of tasks to avoid the learning effect.
5. Eye tracking data from the pilot showed patterns of different information-processing stages. Instead of highlighting the end-nodes for path-finding tasks, which was the situation in pilot study, the numbers of the designated end-nodes were read to participants before they saw the stimuli in order to better investigate their graph reading strategies.
6. Participants were instructed that there was only one correct answer and that this experiment focused on the right answer rather than quick responses in an effort to minimize random answers.
7. Unlike the pilot study, think-aloud protocols were not encouraged to yield a more strictly controlled experiment.

Methods

Participants. 43 participants (36 males, 7 females) were recruited through ASU Polytechnic Psychology Subject Pool. They were students enrolled in PSY101 (introductory psychology) on ASU Polytechnic campus. Participants had diverse majors, including Engineering, Business, and Professional Flight. Figure 12 shows the distribution of participants' majors. Participants' age ranges from 18 to 35 with a mean of 20.81 (SD = 4.02). They received one credit for PSY 101 course for their participation.

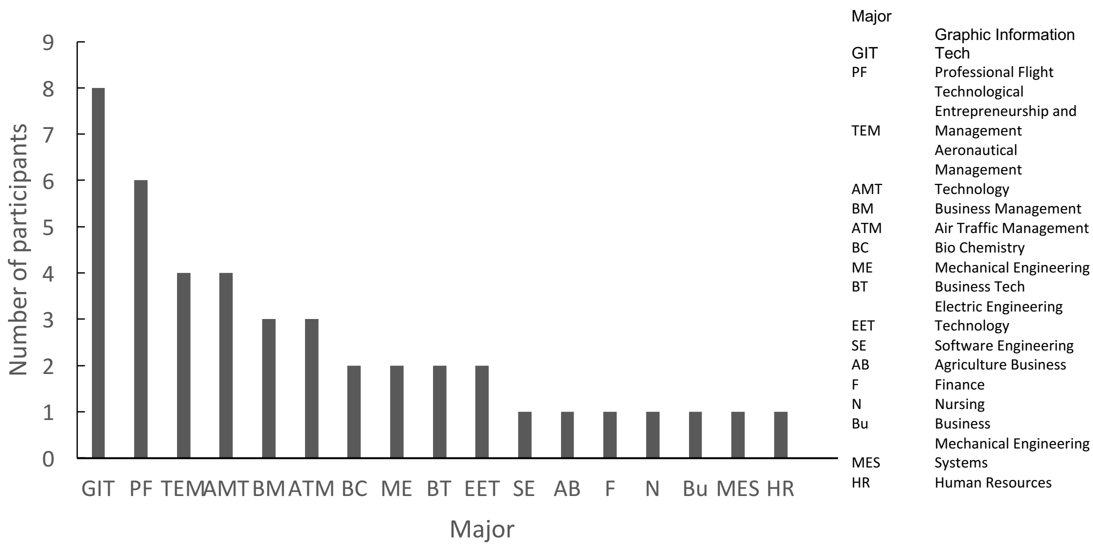
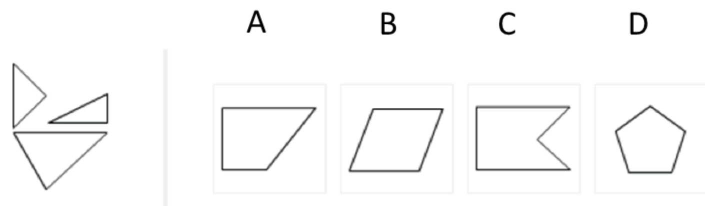
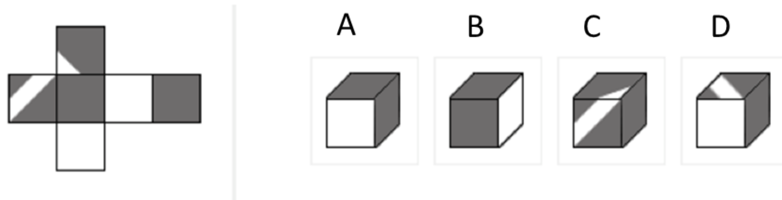
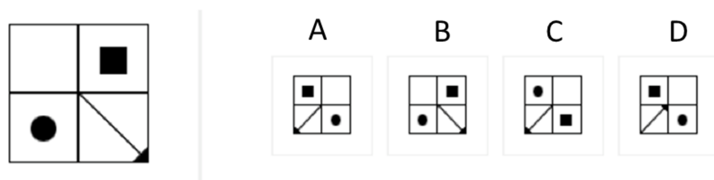
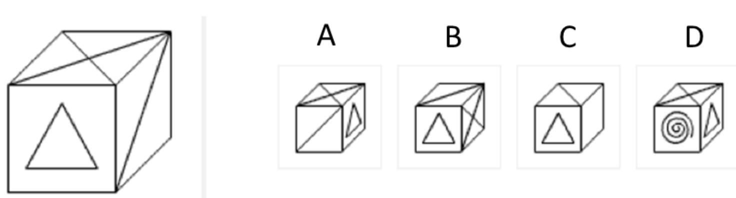


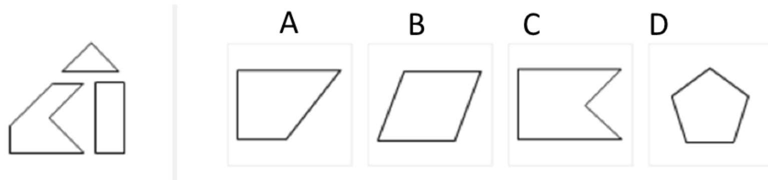
Figure 12. Major distribution of experiment participants.

Apparatus. An Eyetech VT2 model eye tracker, running QuickCapture software was used to capture participants' eye movement at a sampling rate of approximate 35 Hz. Stimuli were displayed using a Dell 20.8 inch screen with 1600 × 1200 resolution. Another screen, which was a duplication of the stimuli-display screen, was used for experiment coordinator to control the experiment process. Eye tracking data was analyzed and visualized using the combination of open source software Ogama, Excel and SPSS.

Graph Stimuli. A standard spatial reasoning test (adapted from jobtestprep.co.uk) that includes 5 tasks was used to test participants' spatial reasoning ability. These 5 tasks were chosen to include the popular task types (see Table 9). Participants were asked to choose one from the four shapes on the right side which was the ‘best match’ with the shape on the left side after spatial re-arrangement (e.g. combine, fold up, mirror, etc.)

Table 9. Stimuli used for spatial reasoning test.

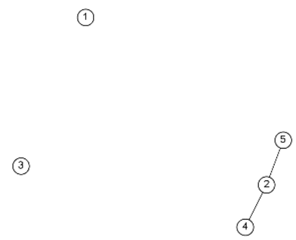
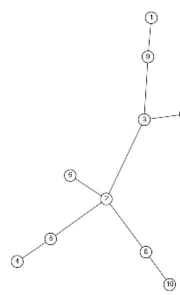
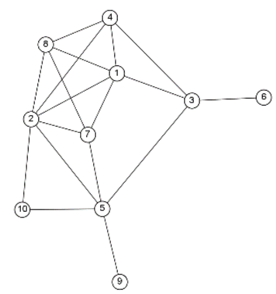
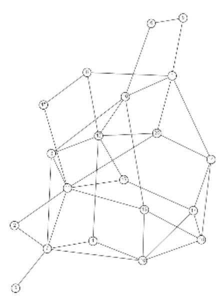
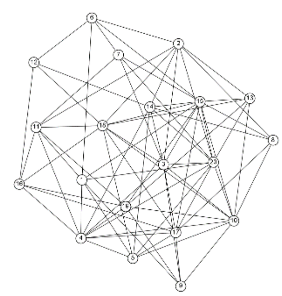
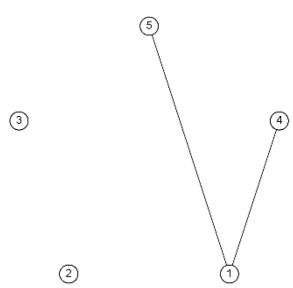
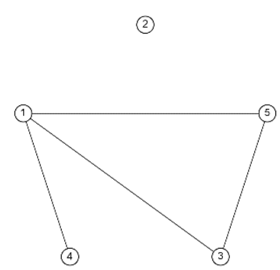
Stimuli	Category
	Organizing two dimensional shapes
	Perspectives
	Mirror images
	Spatial reasoning cubes

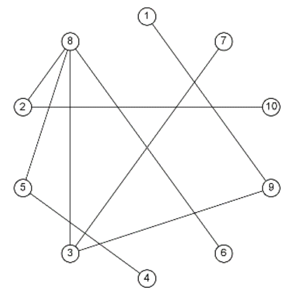
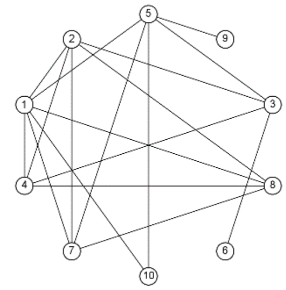
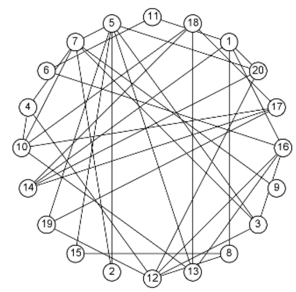
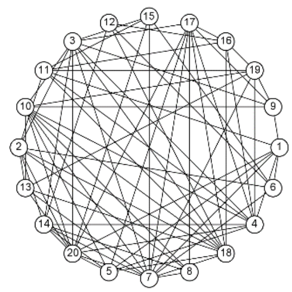
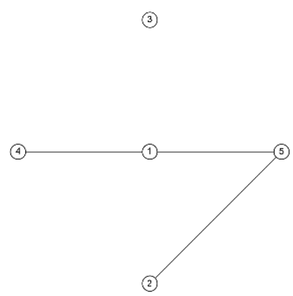
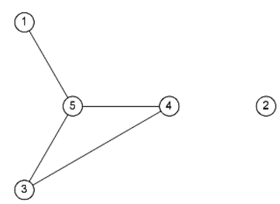
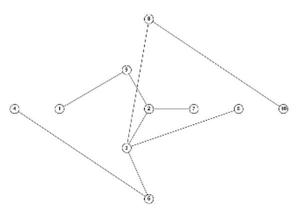
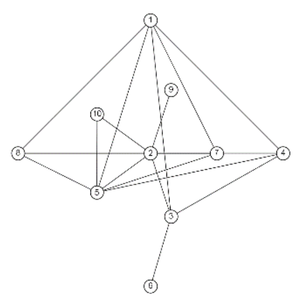
Stimuli	Category
	Organizing two dimensional shapes

Thirty-six node-link diagrams were produced using open source software Gephi. There were several considerations when creating these graphs: first, the Erdos-Renyi G (n, m) model was used as graph generator to control the size and density of the produced graphs. Graphs included 3 graph sizes (5, 10, and 20) and two densities (0.2, 0.4); second, each graph size and density level had 3 layouts, including force-directed, circular and concentric generated using ForceAtlas, Circular Layout and Concentric Layout algorithms respectively; third, graphs with the same size and density illustrated the same underlying data structure; fourth, graphs were drawn with black line on white background to ensure their readability. Also, unnecessary overlapping (other than inherited from density) was minimized using Noverlap layout algorithm.

Table 10 shows the produced graphs for the node-locating task and Table 11 shows those generated for the path-finding task.

Table 10. Stimuli produced using Gephi for node-locating tasks.

Layout	Size	.2 density	.4 density
Force-directed	5		
	10		
	20		
Circular	5		

Layout	Size	.2 density	.4 density
	10		
	20		
Concentric	5		
	10		

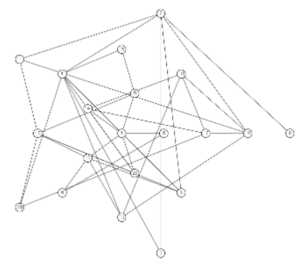
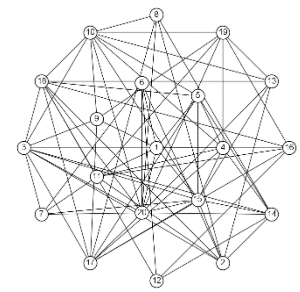
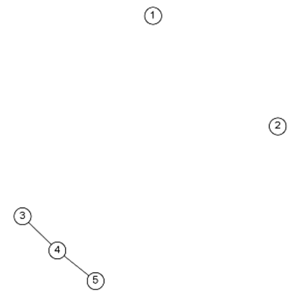
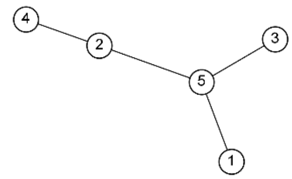
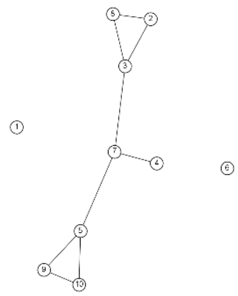
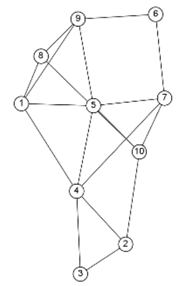
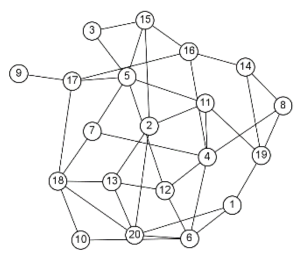
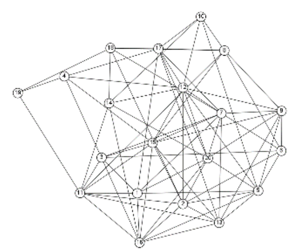
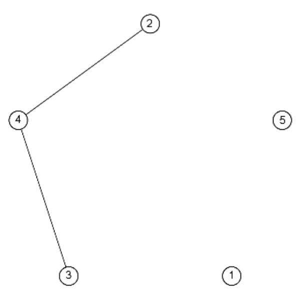
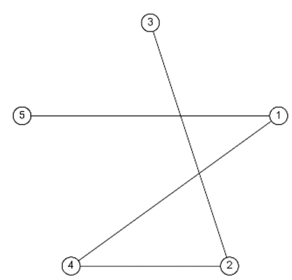
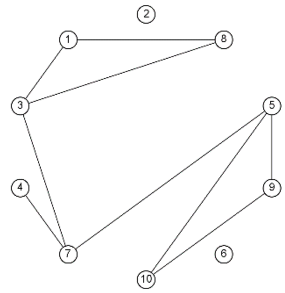
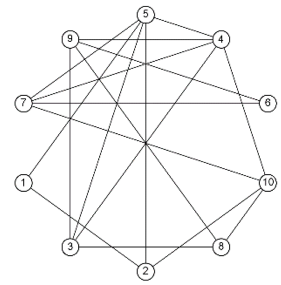
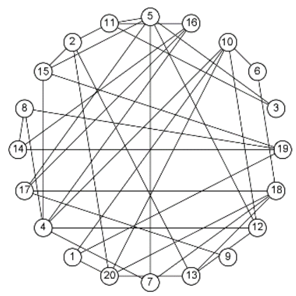
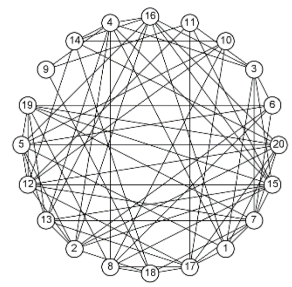
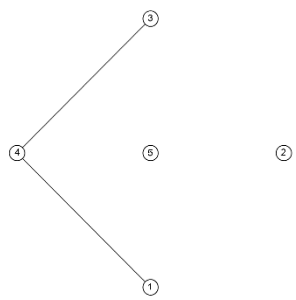
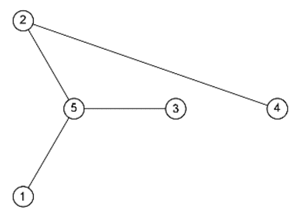
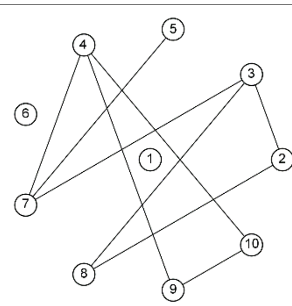
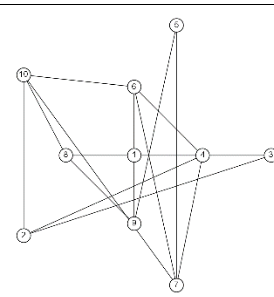
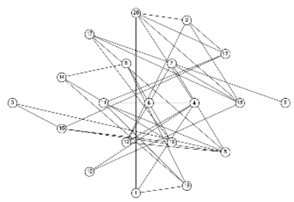
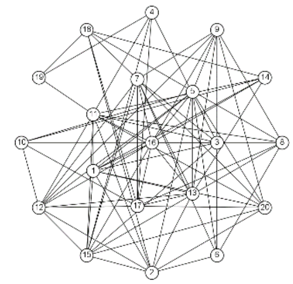
Layout	Size	.2 density	.4 density
	20		

Table 11. Stimuli produced using Gephi for path-finding tasks.

Layout	Size	.2 density	.4 density
Force-directed	5		
	10		
	20		
Circular	5		

Layout	Size	.2 density	.4 density
	10		
	20		
Concentric	5		
	10		

Layout	Size	.2 density	.4 density
	20		

Tasks. There were two kinds of cognitive tasks: 1) the node-locating task (N task) asked participants to find the most connected node in the graph; 2) the path-finding (P task) asked participants to locate the shortest path between two designated nodes (note here that the shortest path is the one that goes through the least number of nodes).

Procedure. Participants were first welcomed and the experimenter explained the purpose of the study and the basic mechanism of the eye tracker. They were given time to read the consent form and ask questions. After addressing concerns and questions about the study (if any), participants' demographic information (age, gender, and major) was collected. Then a demo was presented to show how to conduct the spatial reasoning test, N task and P task. They had two practice trials with the spatial reasoning test and one practice trial each for the N task and P task. Participants were instructed to focus on finding the correct answer rather than quickly answering, and they were reminded that there was only one correct answer for spatial reasoning tests and N tasks. Then, 5 spatial reasoning test tasks were conducted. The task was completed as participants verbally gave their answers. Task completion time and participants' answers were recorded for further analysis. Eye tracking data were not collected for the spatial reasoning test.

The main experiment included 36 counterbalanced trials in total (18 each for N task and P task), and was divided into 2 blocks (18 N tasks followed by 18 P tasks). Participants were reminded to constrain their head movement during tasks in order to get reliable eye tracking data. The number of the two designated nodes for P tasks was read to participants (instruction: please find the shortest path between node * and *) before they saw the stimuli. Time recording was started as soon as they saw the stimuli on the screen and ended when they gave answers verbally. Answer and completion time were recorded manually by the coordinator. Eye tracking data were recorded using Eyetech VT2 eye tracker. A 16-point calibration was conducted every 6 trails to ensure the quality of the eye tracking data. A short interview was conducted after participants completed the tasks. For each participant, the experiment took around 40 minutes on average to complete.

Experimental design. The experiment was a repeated-measures design. The study presented 3 graph layouts (force-directed, circular, concentric) × 3 graph sizes (5, 10, 20 nodes) × 2 densities (0.2, 0.4) × 2 task types (node-locating, path-finding) to each participant. Each participant performed one trial of each graph layout, size, density, and two task types, resulting in 36 trials in total.

Results

Spatial Reasoning Test. 42 (Male=35, Female=7) participants' data were included in this analysis. One participant's data was excluded because of missing data.

Figure 13 is the histogram of participants' task accuracy. The spatial reasoning test proved to be a challenge for participants: no one answered all five tasks correctly.

Only one participant gave correct answers to four tasks. Most of the participants had 40% (two tasks) accuracy.

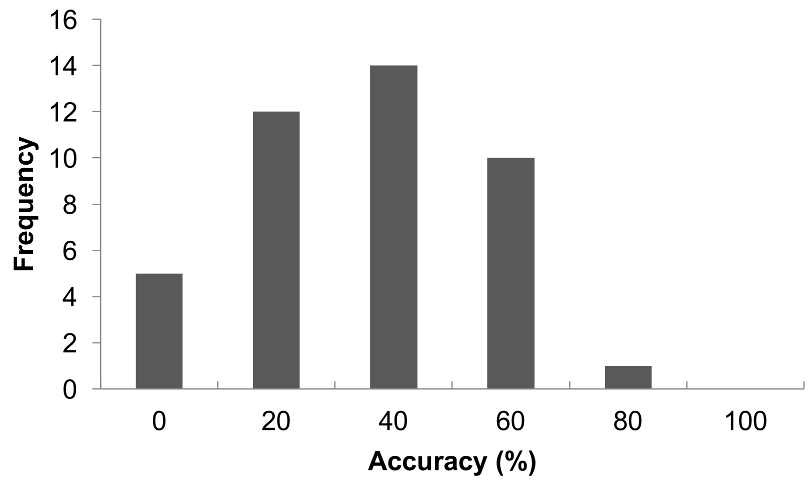


Figure 13. Histogram of participants' task accuracy of spatial reasoning test.

No statistically significant difference was found between male and female on task accuracy (see Figure 14).

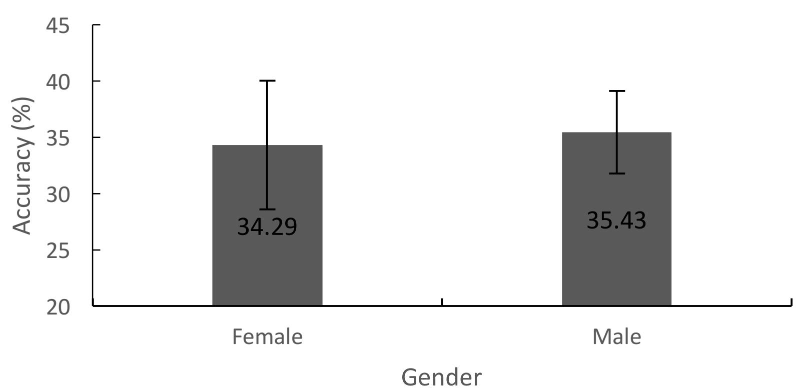


Figure 14 . Task accuracy of female and male participants for spatial reasoning test. Note, other than noted specifically, error bars indicate Standard Error. Significance level α = .05.

Figure 15 and Figure 16 show the task accuracy and completion time by task respectively. Interestingly, participants had the shortest completion time and lowest accuracy on Task 3. By further analysis of their answers, it was noticed that most participants chose B (see Table 9. Stimuli used for spatial reasoning test Table 9), which is identical to the test shape as their answer. However, the intention of this task is to find the mirror shape, which is A. Most participants overlooked the original intention and misunderstood it as a straightforward test (find the identical shape), resulting in quick but wrong answers.

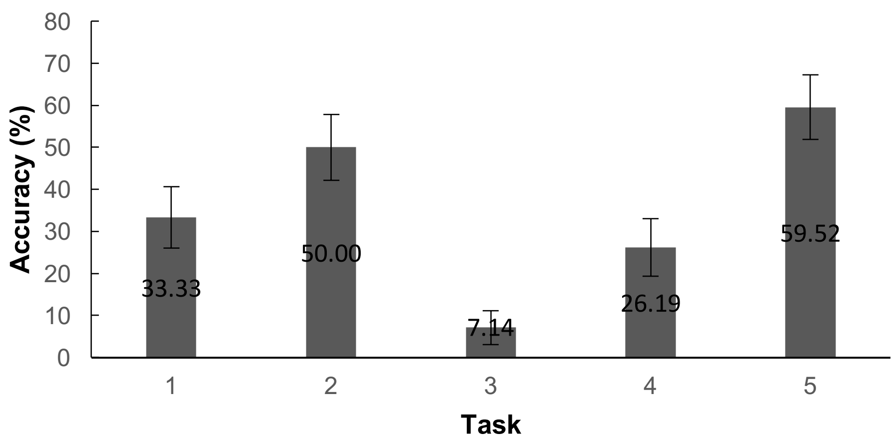


Figure 15. Participants' task accuracy on every spatial reasoning tasks.

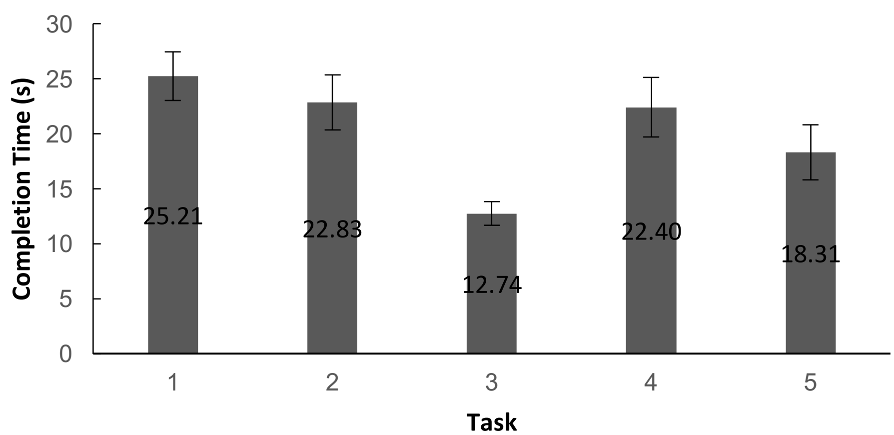


Figure 16. Participants' task completion time of every spatial reasoning task.

Node-locating and Path-finding task. Participant S01 and S36 had an average task completion time of 40.92 (s) and 39.56 (s) respectively, which is substantially higher than the average time of all participants (*M* = 13.44, *SD* = 17.50). Their data were excluded for the analysis of task completion time. Two data points that were higher than 100 seconds were considered outliers, and have been replaced with the average completion time for that task (e.g., a 113 seconds completion time has been replaced with the average time of task 1 which is 26.72).

The 41 participants' data (35 male, 6 female) used for analysis have an average task completion time (across tasks) of 11.98 (*SD* = 12.62) seconds.

As illustrated in 17, significant correlation was found between participants' task accuracy and their performance on the spatial reasoning test (measured by the number of correct answers) for P tasks (Pearson's $r = .081$, $p = .028$, $\alpha = .05$), but not for N tasks ($r = .063$, $p = .87$). As demonstrated in Figure 18, the correlation between participants' task completion time and spatial reasoning performance was significant for both N tasks ($r = -.134$, $p < .001$) and P tasks ($r = -.076$, $p = .039$). Participants who performed better on the spatial reasoning test tended to have a higher task accuracy and shorter task completion time.

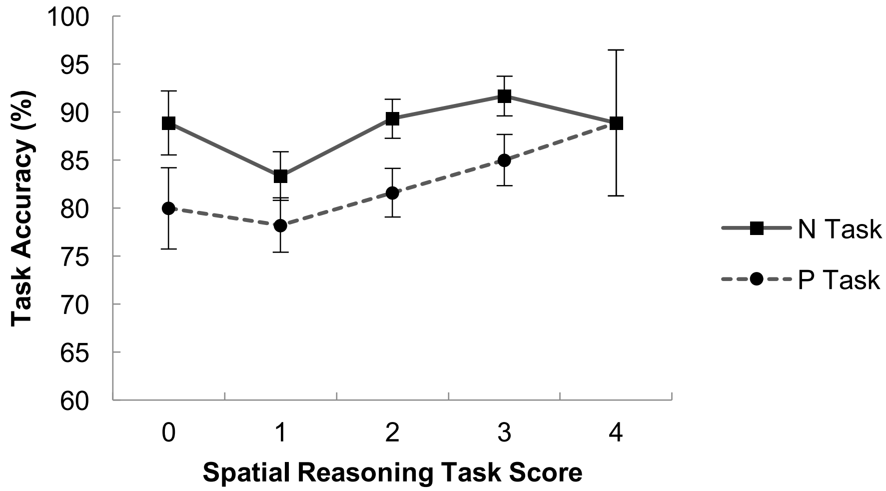


Figure 17. *Participants' task accuracy on N and P tasks as a function of spatial reasoning test score.*

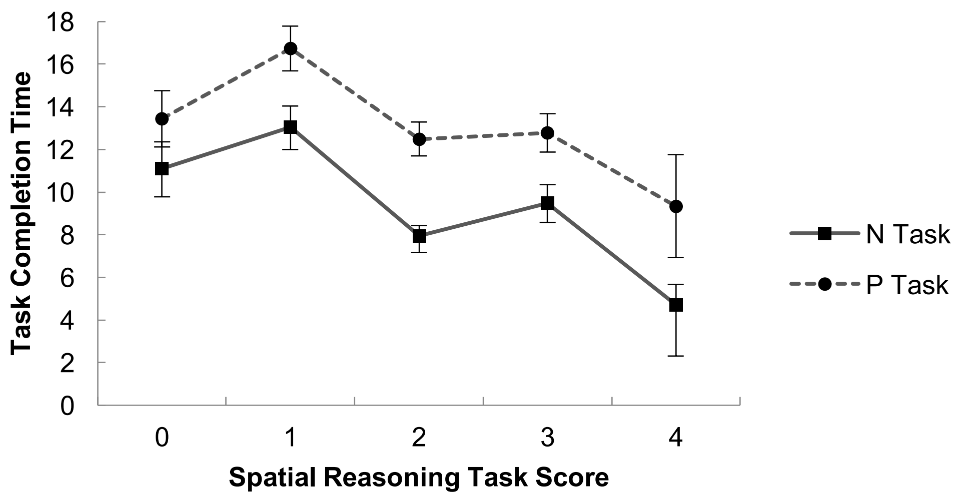


Figure 18. *Participants' task completion time (s) on N and P task as a function of spatial reasoning test score.*

Since node-locating and path-finding tasks were essentially different cognitive tasks, the following analysis was divided into analysis on node-locating tasks and path-finding tasks accordingly.

Task accuracy and completion time for the node-locating task. Figure 19 and Figure 20 provide an overview of accuracy and completion time of node-locating tasks. Table 12 summarizes the mean and standard deviation of task accuracy and completion time of node-locating tasks.

Table 12. Mean and standard deviation of task accuracy and completion time of node-locating tasks.

				Accuracy		Time (s)	
Size			Task No.	Mean	SD	Mean	SD
5.00	.20	Circular	15	1.00	0.00	1.88	0.93
		Concentric	6	0.88	0.33	4.37	4.48
		Force	11	0.98	0.16	2.39	1.41
		Total		0.95	0.22	2.88	2.94
	.40	Circular	18	0.95	0.22	2.78	1.42
		Concentric	17	1.00	0.00	3.00	1.96
		Force	14	0.93	0.26	4.15	3.50
		Total		0.96	0.20	3.31	2.51
	Total	Circular		0.98	0.16	2.33	1.28
		Concentric		0.94	0.24	3.68	3.50
		Force		0.95	0.22	3.27	2.79
		Total		0.96	0.21	3.09	2.74
10.00	.20	Circular	9	1.00	0.00	4.68	3.14
		Concentric	12	1.00	0.00	5.07	3.10
		Force	1	1.00	0.00	6.93	7.49
		Total		1.00	0.00	5.56	5.07
	.40	Circular	13	0.98	0.16	8.56	5.03
		Concentric	4	0.83	0.38	10.93	5.62
		Force	5	1.00	0.00	7.80	4.26
		Total		0.93	0.25	9.10	5.14
	Total	Circular		0.99	0.11	6.62	4.60
		Concentric		0.91	0.28	8.00	5.39
		Force		1.00	0.00	7.37	6.07
		Total		0.97	0.18	7.33	5.39
20.00	.20	Circular	7	0.76	0.43	18.02	10.35
		Concentric	10	0.98	0.16	14.24	10.78
		Force	2	0.80	0.40	14.71	8.03
		Total		0.85	0.36	15.66	9.86
	.40	Circular	16	0.46	0.50	24.30	11.37
		Concentric	3	0.76	0.43	20.68	14.67

Size			Task No.	Accuracy Mean	Accuracy SD	Time (s) Mean	Time (s) SD
		Force	8	0.56	0.50	21.28	13.50
		Total		0.59	0.49	22.08	13.24
	Total	Circular		0.61	0.49	21.16	11.26
		Concentric		0.87	0.34	17.46	13.19
		Force		0.68	0.47	17.99	11.52
		Total		0.72	0.45	18.87	12.08
Total	.20	Circular		0.92	0.27	8.19	9.41
		Concentric		0.95	0.22	7.89	8.26
		Force		0.93	0.26	8.01	8.14
		Total		0.93	0.25	8.03	8.60
	.40	Circular		0.80	0.40	11.88	11.61
		Concentric		0.86	0.35	11.53	11.61
		Force		0.83	0.38	11.08	11.16
		Total		0.83	0.38	11.50	11.43
	Total	Circular		0.86	0.35	10.04	10.71
		Concentric		0.91	0.29	9.71	10.22
		Force		0.88	0.33	9.54	9.87
		Total		0.88	0.32	9.76	10.26

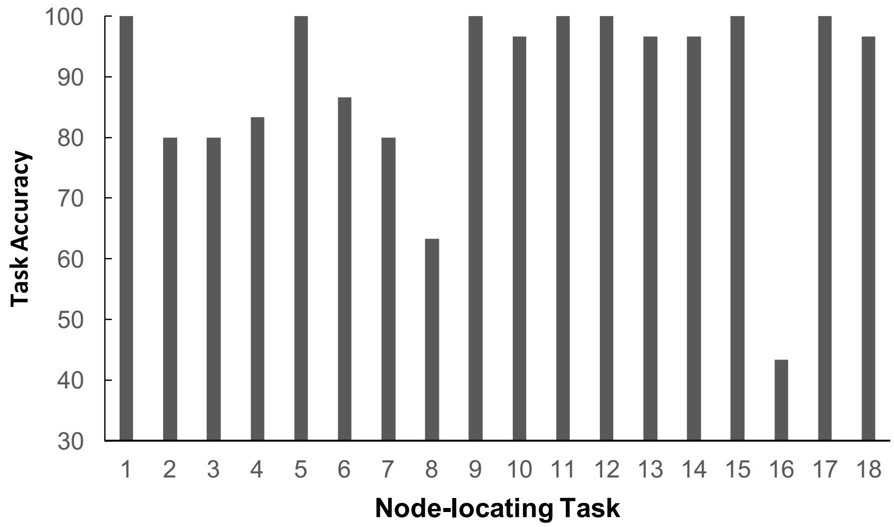


Figure 19. Task accuracy of 18 node-locating tasks.

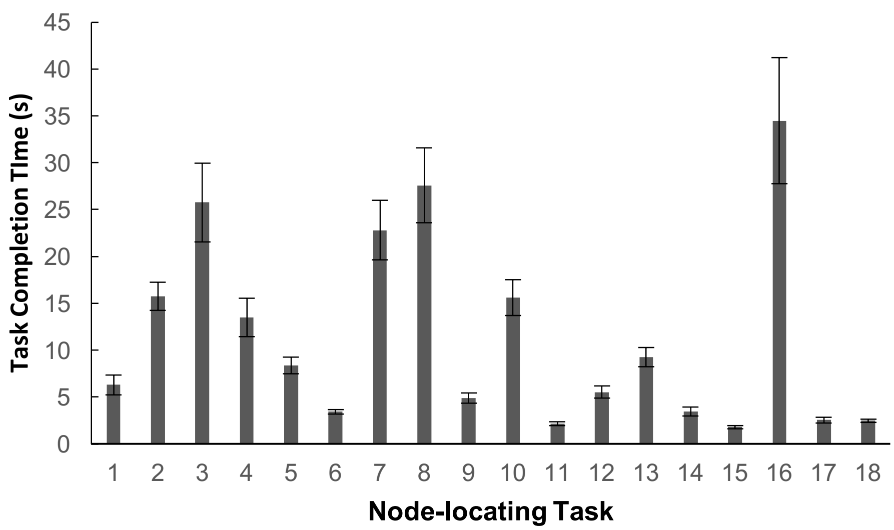


Figure 20. Task completion time of 18 node-locating tasks.

A three-way ANOVA was conducted to explore the size, density, layout, and their interaction effect on participants' task performance measured by accuracy and completion time. Participants' performance on spatial reasoning test was also included as a covariate variable. As illustrated in Table 13, participants' score on the spatial reasoning test (designated as SR Score in the table) had significant effect on task completion time $F(1, 719) = 25.95$, $p < .05$) but not task accuracy. Graph size and density were found to have significant effect on both task completion time $F(2, 719) = 303.64$, $p < .05$, $F(1, 719) = 41.01$, $p < .05$, and accuracy $F(2, 719) = 58.23$, $p < .05$, $F(1, 719) = 23.71$, $p < .05$. The graph size had a larger effect size measured as partial eta squared[5] (.46 and .14 for task completion time and accuracy respectively), compared to the small effect size of graph density (.05 and .03). The interaction between size and density was also significant for both task completion time $F(2, 719) = 10.24$, $p < .05$, and task accuracy $F(2, 719) = 13.41$, $p < .05$. Layout as a main effect was not significant for both task accuracy and task completion time. However, the interaction between graph size and layout was found to be significant for both task completion time $F(4, 719) = 3.62$, $p < .05$, and task accuracy $F(4, 719) = 8.98$, $p < .05$.

Table 13. Summary results of the ANOVA analysis on node-locating tasks.

Source	Dependent Variable					
	Time (s)			Accuracy		
	F	Sig.	η^2	F	Sig.	η^2
SR Score	25.95	.000*	0.04	3.73	0.05	0.01
Size	303.64	.000*	0.46	58.23	.000*	0.14
Density	41.01	.000*	0.05	23.70	.000*	0.03
Layout	0.29	0.75	0.00	1.79	0.17	0.01
Size * Density	10.24	.000*	0.03	13.41	.000*	0.04
Size * Layout	3.62	.006*	0.02	8.98	.000*	0.05
Density * Layout	0.13	0.87	0.00	0.21	0.81	0.00

[5] The interpretation of partial eta squared was according to: .02~ .13 small effects; .13~ .26 medium effects; >.26 large effect.

Source	Dependent Variable					
	Time (s)			Accuracy		
Size * Density * Layout	1.62	0.17	0.01	2.33	0.06	0.01

As illustrated in Figure 21, at size 10, the task accuracy of concentric layout graphs was lower than force-directed and circular layout. However, this was not the case for graphs of size 20: as graph size enlarged from 10 to 20, accuracy of circular and force-directed layout dropped dramatically, whereas concentric layout has a more consistent performance, which is demonstrated by a significantly higher accuracy.

All three layouts had almost the same task completion time at different sizes with one exception: at size 20, participants took more time to complete the task with the graphs in circular layout than the other two layouts (see Figure 22).

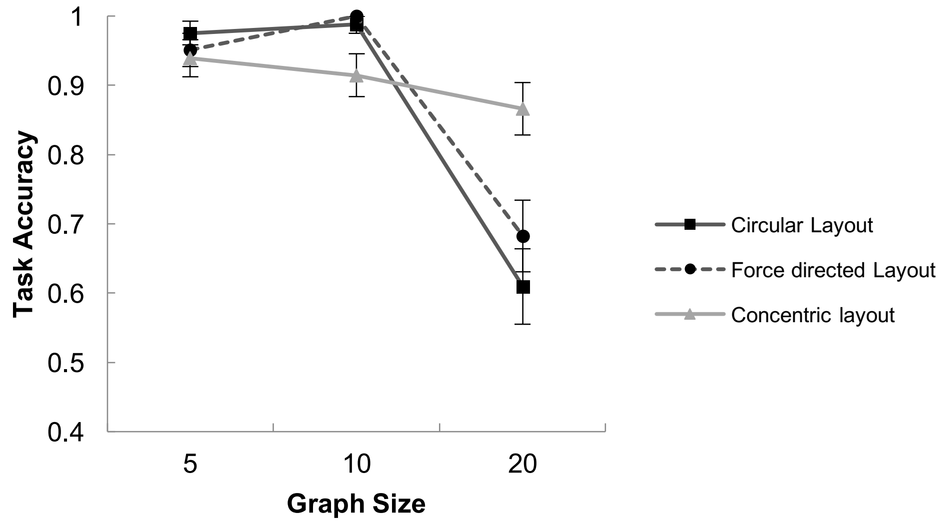


Figure 21. Layout, size, and their interaction effects on task accuracy.

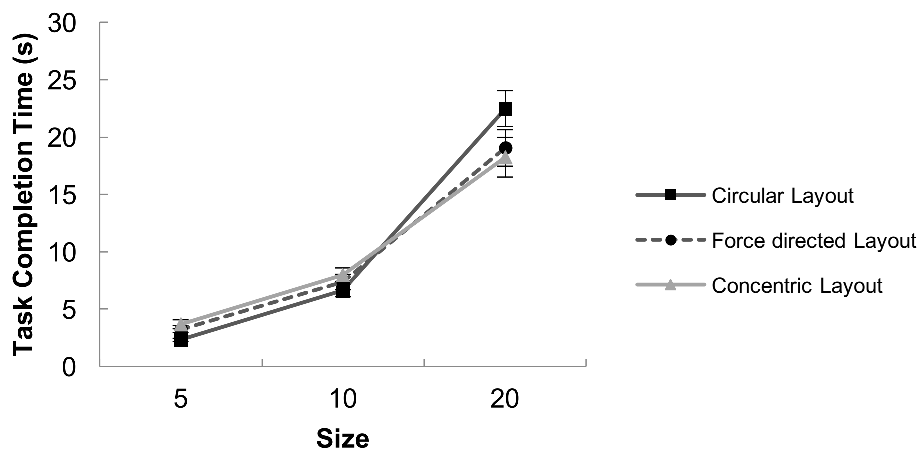


Figure 22. Layout, size, and their interaction effects on task completion time.

As illustrated in Figure 23and Figure 24, graph density's effect was significant at size 10 and 20, but not at size 5. Graphs with lower density had higher task accuracy and shorter task completion time.

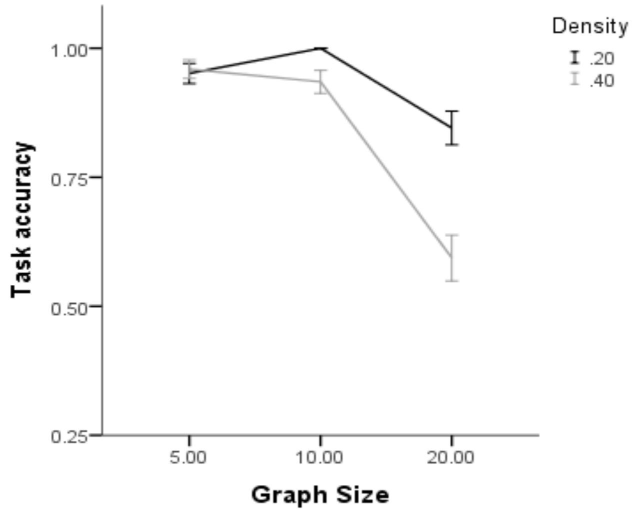


Figure 23. Graph size and density's effect on task accuracy.

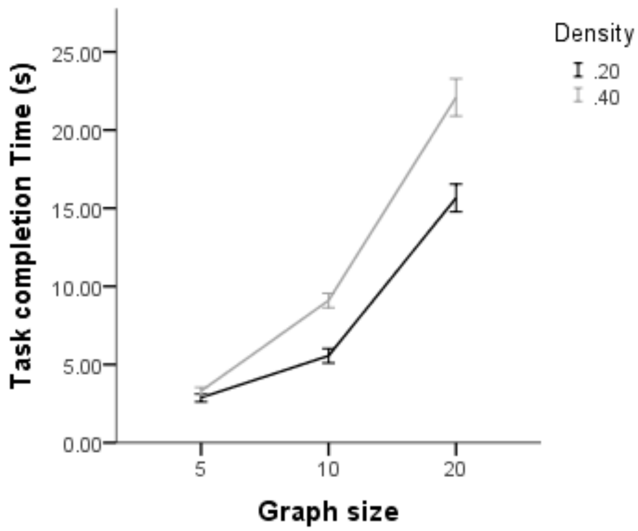


Figure 24. Graph size and density's effect on task completion time.

Eye tracking data analysis of node-locating tasks. Thirty participants' eye movement data were used for eye tracking analysis. Twelve participants' data were removed because of quality concerns (e.g. unreliable calibration, miss-record of data, etc.). Other trials containing missing data points (marked as '-1' in output spreadsheet of eye movement data) were also omitted before analysis.

AOI definition. Areas of Interest (AOIs) were defined for stimulus of node-locating tasks to facilitate analysis. As the process of participants searching for the most connected node was of concern, the target nodes of every task were defined as AOI 'T'. Padding area (r = 20 pixels) around the nodes was provided to tolerate inevitable error in gaze locating.

Metrics. Three eye tracking metrics were used to analyze the node locating data in order to examine the participants' graph reading behavior. The definition of these metrics was guided by the top-down cognitive model of graph comprehension as adopted by this research. The model divided the graph reading process into two sub-stages: target searching and target comprehension.

1. *Complete fixation time at AOI T*, indicating the time participants needed to 'read' the target node.
2. *Time till first fixation on AOI T*, indicating the time participants needed to 'search' the target.
3. *The duration between first fixation on AOI T* to task completion, indicating the 'comparison' and 'consideration' (disambiguation and comprehension process) before decision-making.

Complete fixation time on AOI T. Three-way ANOVA was conducted to investigate size, density, layout, and their interaction's effect on complete fixation time on AOI T (see Table 15 and as a summary of the results). Graph size and density as main effect were found to be significant, $F(2, 469) = 61.30$, $p < .05$ and $F(1, 469) = 8.57$, $p < .05$. The interaction between size, density, and layout also had significant effect on participants' fixation time on AOI T, $F(4, 469) = 3.24$, $p = .01$.

Table 14. Mean and standard deviation of three metrics.

Size			Fixation duration on T(ms)		Duration from locating T to task completion (ms)		Time till first fixation on T (ms)	
			Mean	SD	Mean	SD	Mean	SD
5	.20	Circular	409.50	471.66	688.67	1017.43	1061.33	889.87
		Concentric	604.78	401.94	2116.33	1617.98	1291.07	672.78
		Force	853.44	503.76	1188.78	1137.51	959.37	588.60
		Total	630.77	489.46	1355.97	1407.62	1105.56	725.62
	.40	Circular	769.52	441.42	1480.30	1296.01	1038.22	779.83
		Concentric	1292.38	1011.74	1709.03	1748.03	773.72	488.33
		Force	1082.70	647.94	1948.70	2873.36	1310.56	998.82
		Total	1054.08	769.05	1712.59	2056.67	1034.40	799.83
	Total	Circular	600.10	486.41	1107.76	1228.35	1049.10	825.00
		Concentric	960.86	847.16	1905.41	1683.83	1023.16	635.15
		Force	968.07	586.37	1568.74	2198.20	1134.96	831.13
		Total	849.00	680.77	1539.82	1775.76	1068.88	763.23
10	.20	Circular	1360.07	806.58	3892.75	3126.35	1142.96	978.97
		Concentric	1633.26	949.97	3318.41	2718.29	1866.78	1539.34
		Force	2082.00	1581.73	5302.40	5226.36	997.60	1194.72
		Total	1701.65	1201.77	4207.84	3946.41	1321.58	1295.35
	.40	Circular	2797.18	1285.03	5906.43	5199.69	3272.14	2379.66
		Concentric	2187.92	1565.93	8011.04	9268.05	5604.35	5970.37
		Force	1492.67	1323.93	5167.41	3952.19	3054.81	2389.78
		Total	2166.78	1478.47	6335.64	6517.78	3948.31	4028.30
	Total	Circular	2078.63	1286.75	4899.59	4370.69	2207.55	2098.65
		Concentric	1905.36	1307.09	5620.45	7113.38	3700.30	4677.63
		Force	1802.84	1482.41	5238.46	4626.21	1972.07	2112.98
		Total	1928.61	1359.94	5246.10	5446.45	2603.30	3233.63
20	.20	Circular	3911.48	3484.66	17107.69	18489.99	5306.10	5555.59
		Concentric	3939.29	2968.63	12798.18	10328.46	2844.68	3882.87
		Force	2769.48	2346.09	9617.88	7390.30	6222.12	5193.64
		Total	3572.80	3008.76	13352.67	13393.36	4744.89	5071.64
	.40	Circular	4727.96	6979.04	25908.83	33433.72	11757.83	10903.87
		Concentric	3875.08	3737.64	16653.27	22762.44	9269.81	13182.65
		Force	5967.64	6373.96	17453.93	17824.73	10296.07	11872.10
		Total	4888.68	5840.27	19788.55	25109.78	10403.76	11931.87
	Total	Circular	4281.21	5315.10	21093.11	26420.99	8227.64	8928.56
		Concentric	3908.37	3328.69	14654.33	17393.26	5938.26	10007.71
		Force	4459.08	5121.84	13757.68	14344.83	8374.40	9478.86
		Total	4214.29	4643.83	16490.16	20177.56	7503.59	9492.14
Total	.20	Circular	1991.89	2600.04	7674.67	13234.71	2609.28	3932.11
		Concentric	2082.04	2294.79	6159.60	7883.25	2011.13	2520.20
		Force	1887.07	1794.04	5263.61	6145.28	2577.85	3809.40
		Total	1986.98	2244.88	6360.62	9570.10	2398.57	3474.35
	.40	Circular	2690.75	4190.18	10470.38	21173.61	5086.58	7600.69
		Concentric	2408.85	2580.04	8528.81	15107.16	5051.43	8834.34
		Force	2885.72	4387.90	8302.99	12619.60	4953.11	8039.14
		Total	2662.46	3795.02	9086.11	16592.96	5029.59	8144.28
	Total	Circular	2336.95	3483.87	9055.05	17607.58	3832.45	6137.04
		Concentric	2244.44	2438.67	7336.94	12049.44	3521.96	6636.39

Size		Fixation duration on T(ms)		Duration from locating T to task completion (ms)		Time till first fixation on T (ms)	
		Mean	SD	Mean	SD	Mean	SD
	Force	2386.40	3379.06	6783.30	10011.42	3765.48	6383.26
	Total	2322.64	3128.30	7714.97	13578.39	3705.98	6378.57

As illustrated in Figure 25, size has significant effect across all three layouts, $F(2, 469) = 61.30$, $p < .05$. Complete fixation on AOI T increased significantly as graph size enlarged. Participants needed 2.29 seconds more on graph of size 20 than size 10, and 3.37 seconds more than size 5 to read the target nodes.

Table 15. Summarized ANOVA results.

Source	Dependent Variable								
	Fixation time on T (ms)			Time till first fixation on T (ms)			Duration between locating T to task completion (ms)		
	F	sig	η^2	F	sig	η^2	F	sig	η^2
Size	61.29	.000	0.21	61.53	.000	0.21	69.96	.000	0.23
Density	8.57	.004	0.02	30.12	.000	0.06	8.47	.004	0.02
Layout	0.08	0.93	0.00	0.14	0.87	0.00	1.73	0.18	0.01
Size * Density	1.32	0.27	0.01	10.77	.000	0.04	3.14	.044	0.01
Size * Layout	0.42	0.79	0.00	2.35	0.05	0.02	2.69	.031	0.02
Density * Layout	0.48	0.62	0.00	0.39	0.68	0.00	0.12	0.89	0.00
Size * Density * Layout	3.24	.012*	0.03	0.47	0.76	0.00	0.57	0.68	0.01

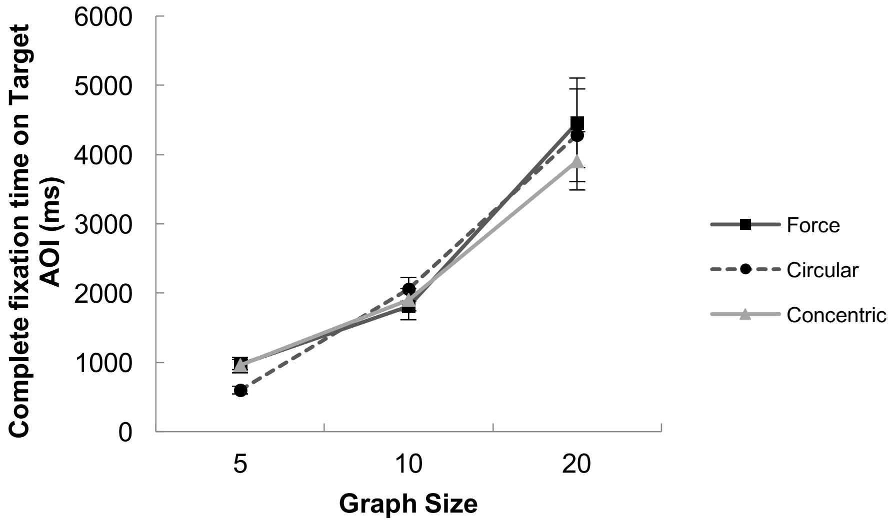


Figure 25. *The effect of graph size and layout on fixation time on AOI T.*

Time till fist fixation on AOI T. As illustrated in Figure 26, graph size has a significant effect $F(2, 469) = 61.53$, $p < .05$ on time till first fixation on AOI T. Participants needed more time to locate the target nodes as graph size increased.

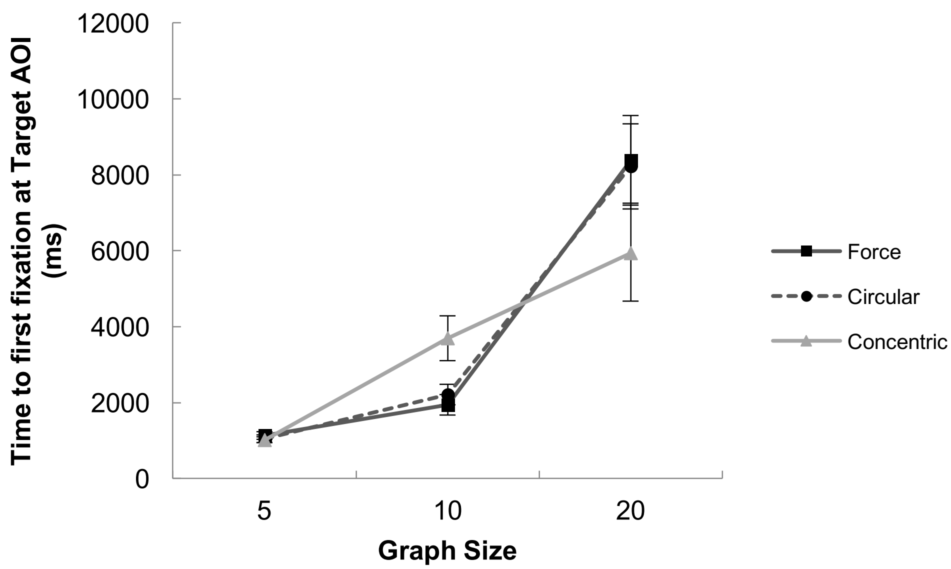


Figure 26. *The effect of graph size, layout and their interaction on time to first fixation at AOI T.*

The duration between first fixations on AOI T to task completion. As illustrated in Figure 27, graph size has significant effect on this duration time $F(2, 467) = 69.97$, p

< .05. Participants needed significantly more time to compare and confirm their decision as the graph size enlarged. At size 20, the circular layout needed significantly longer time to disambiguate than force-directed layout.

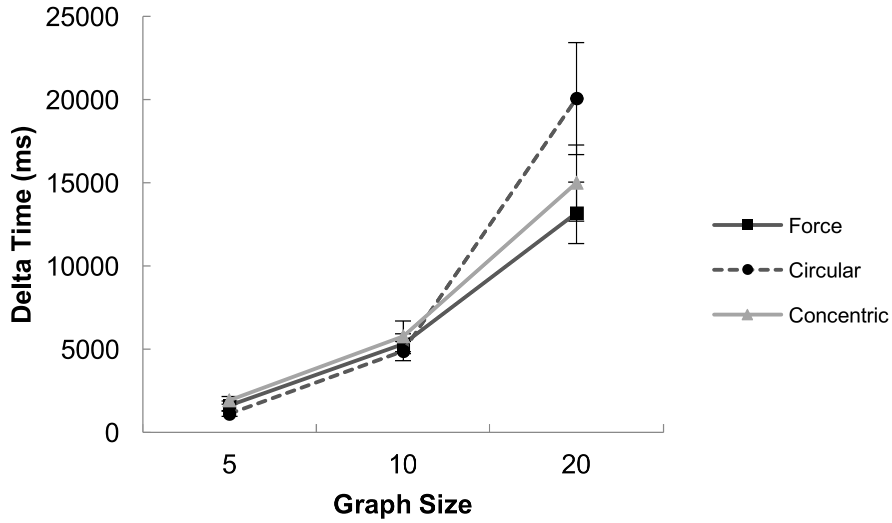
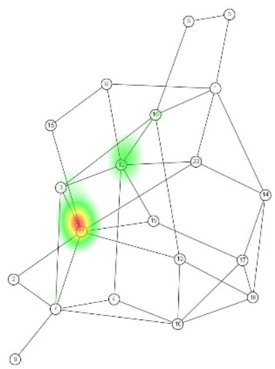


Figure 27. The effect of graph size, layout and their interaction on the duration between first fixations on target to task completion.

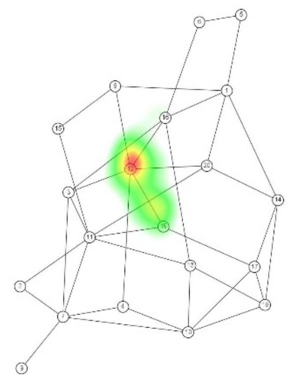
Error task analysis. The former analysis showed that larger graph size brought more difficulties for participants to complete the tasks correctly. Analysis on error trials would bring a deeper understanding on the cognitive process behind the wrong answers and hence the participants visual search strategies. The tasks with size 20 graphs had the lowest accuracy and longest completion time, and thus these 6 tasks (3 layouts × 2 densities) were the focus of following analysis. Scan-path, heat-map, and transition graph visualizations of eye movement data were used to facilitate the analysis.

N2 task, force-directed layout, size 20, density .2. The task was to find the most connected node in the graph (see Table 2). 80 % (24) of participants gave the correct answer of node No. 11, 6 participants gave the wrong answer of node No. 12.

Heat-map visualization (Figure 28. Heat-map visualizations of participant eye movement of N2 task. Note that the correct answer has been totally overlooked by participants who gave the wrong answer.) of participants' eye movement data showed participants who gave the wrong answers have totally ignored the node No. 11.

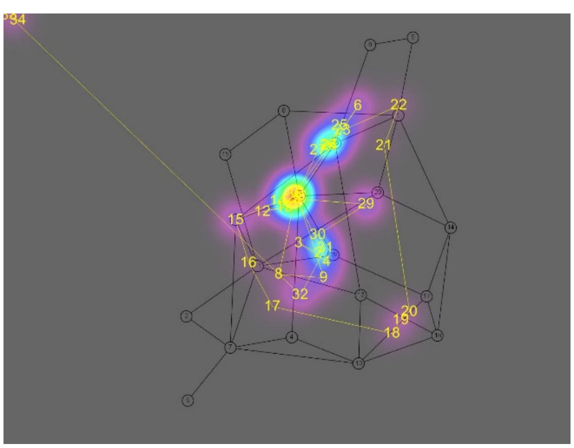


Heat-map visualization of participants' who answered No.11 node.

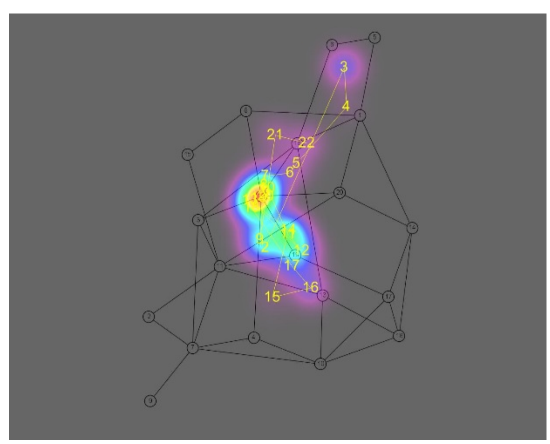


Heat-map visualization of participants' who answered No.12 node.

Figure 28. Heat-map visualizations of participant eye movement of N2 task. Note that the correct answer has been totally overlooked by participants who gave the wrong answer.



Scan-path visualization of participant S12 with N2 task.



Scan-path visualization of participant S23 with N2 task.

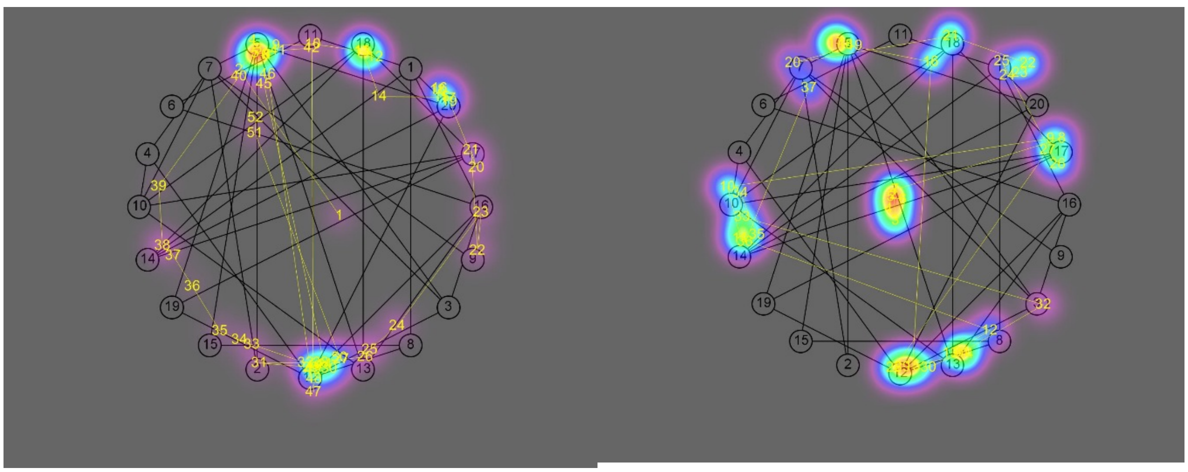
Figure 28. Scan-path visualization of participant S12 and S23.

N7 task, circular layout, size 20, density .2. This task (N7) was to find the most connected node (see table 2). 80% participants gave the correct answer, which was No. 5

node. Wrong answers given by other 6 participants included node 12 (4 answers), 13(1) and 7(1).

Visualization of eye tracking data revealed that S05 (Figure 29) had an interesting scan-path: participant S05 followed the circle in a clock-wise direction and compared node 5 with 12 several times before making his decision (wrong answer node 12).

Participant S12 had a different scan-path pattern compared to S05 (Figure 29): other than follow the visual lead of the circle, participant S12's eye movements jumped between dense areas to locate the answer.



Scan-path of S05 had circular pattern. He compared between 5 and 12 several times.

Scan-path of S12 showed jumps between dense areas.

Figure 29. Scan-path visualization of participant S05 and S12 on N7 task.

N10 task, concentric layout, size 20, density .2. Most participants (29 out of 30) gave the right answer (node 4). Only one participant gave the wrong answer of node 16. Analysis of scan-path visualization showed participants had a very efficient search process: they quickly went through several dense areas and then made their decision without too much back and forth comparison (see Figure 30).

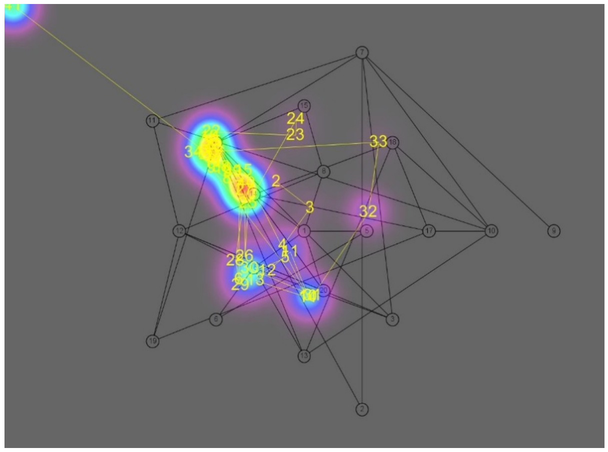
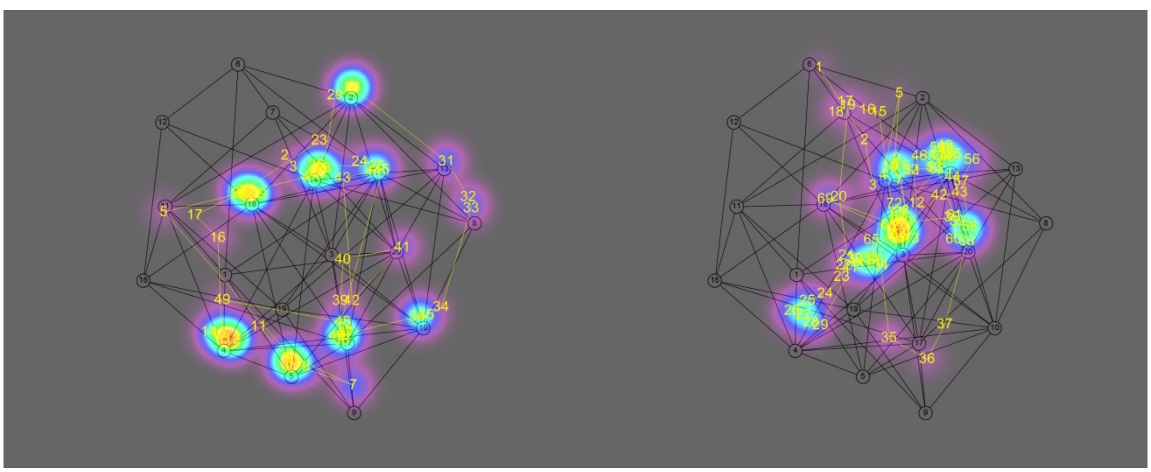


Figure 30. Scan-path visualization of S03 on N10 task.

N8 task, force-directed layout, size 20, density .4. As graph density increased, participants had more difficulties: compared to density .2, a graph with same size and layout at density .4 had more diverse wrong answers (8 different answers). From the eye tracking data, participants did look at the right answer, even compared it with others (see Figure 31). The pattern suggests that the density and overlapping required heavy cognitive loads making it perceptually difficult to discriminate.



Scan-path visualization of S17. Attention has been paid to the right answer.

Scan-path visualization of S25. Comparison has been conducted between several candidate nodes.

Figure 31. Scan-path visualization of S17 and S25 on N8 task.

N16 task, circular layout, size 20, density .4. Answers were distributed relatively evenly on node 7 (13 participants, correct answer) and node 20 (11 participants, wrong answers). These two nodes have nearly equivalent connection numbers (12 and 11 respectively). The highly dense edge crossings made it difficult for participants to visually count the connections. Further eye tracking data analysis revealed the common visual-search pattern for this task: participants always followed the circle 'lead' to search for the 'dense' nodes, the final answer was decided after a series of comparisons (see Figure 32). Moreover, the transitions maps using selected AOI (node 7, 20, 10, 17, 18, covered all answers given by participants) illustrated the different search patterns between participants who answered correctly and incorrectly (see Figure 33): The participants who gave the right answer seemed to have a more thorough comparison before making their decision (relatively equal attention had paid to both node 20 and 7, means they have compared these two nodes carefully.)

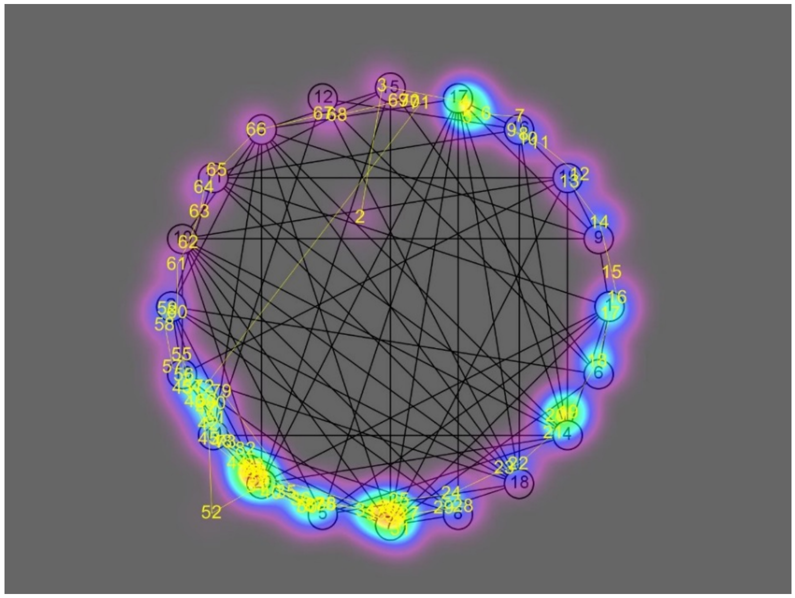
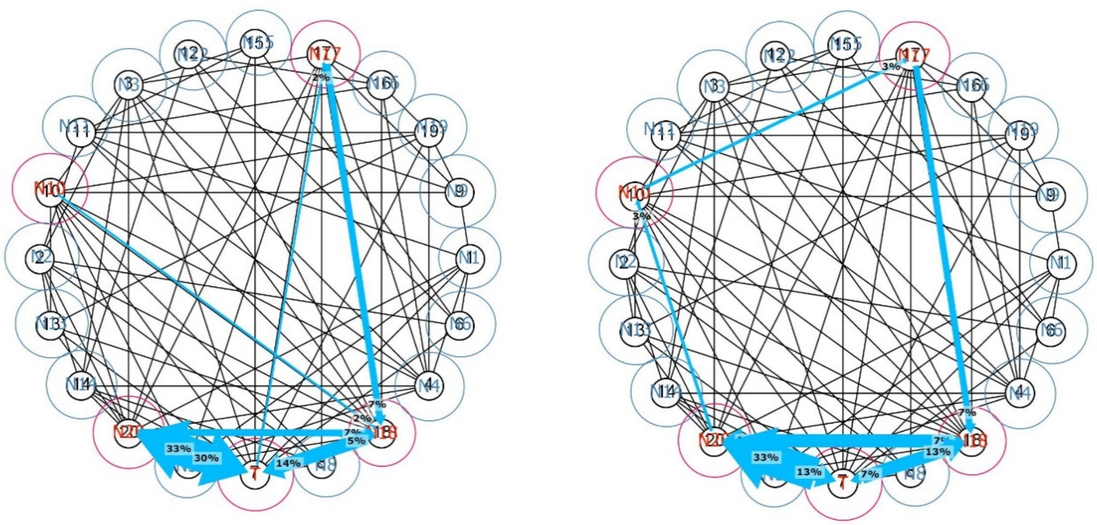


Figure 32. Scan-path visualization of participant S21. Note the formed circular pattern of the scan-path and the comparisons between several candidates (highlighted nodes).

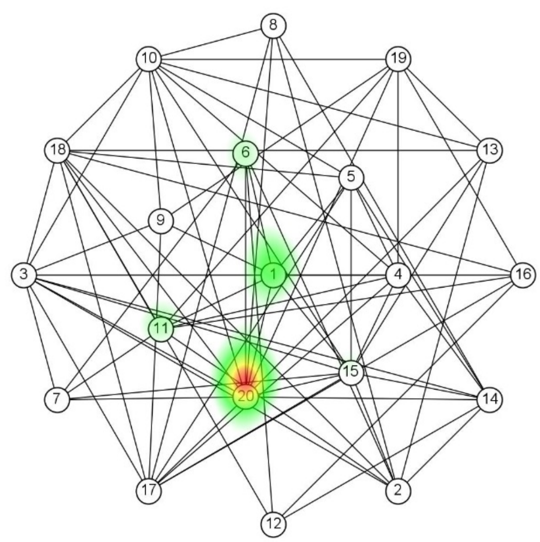


Transitions map of participants who answered node 7. Note there were comparisons between several competitors (transitions came and from node 7 to other nodes). The main transitions went toward node 20, implying the comparison and the final decision of 7.

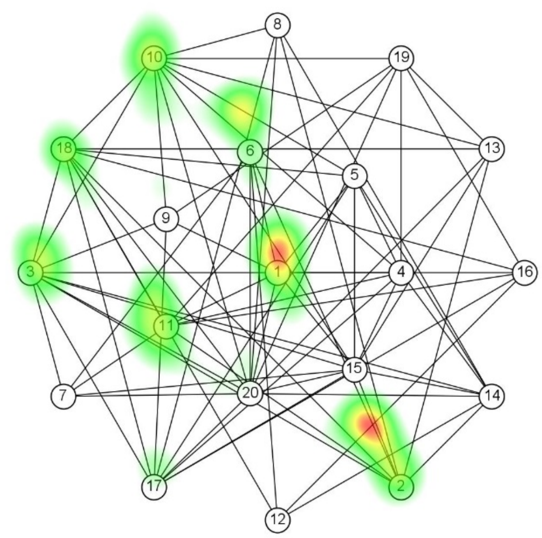
Transitions map of participants who answered 20. Note the main transitions went towards node 20 instead of node 7.

Figure 33. Transitions maps of participants who answered node 7(correct) and node 20 (incorrect).

N16 task, concentric layout, size 20, density .4. Heat-map comparisons between right and wrong answer groups showed the attention focus of these two groups: right answers focused on the comparison between limited candidates, whereas wrong answer had a more scattered attention pattern (see Figure 34).



Heat-map visualization of aggregated right answers.



Heat-map visualization of aggregated wrong answers. Note the more scattered focus points than the right answer group.

Figure 34. Heat-map visualization of the right answer group and wrong answer group.

Summary of node-locating tasks. Layout does not have a significant effect on accuracy and task completion time. The interaction between layout and graph size is significant on both accuracy and task completion time: at size 10, Circular and Force-directed layout has higher accuracy; at size 20, concentric layout has the highest accuracy and shortest time-on-task.

Density has a significant effect on task accuracy and task completion time. Less dense graphs had higher accuracy and shorter time-on-task. Larger and denser graphs resulted in more and diverse wrong answers. More quick answers were observed when participants conducting tasks with larger and denser graphs. The visualizations suggest that dense overlapping caused perceptual difficulty on counting connections. Participants made the tradeoff between accuracy and efficiency: for example, one might choose the 'looks most connected' node instead of really counting its connections and comparing it with other competitors.

Analysis of eye tracking data supported the findings of statistical analysis on task accuracy and completion time. The force-directed layout had the significant shorter disambiguation time than the other two layouts. Participants tended to make quicker decisions, even though the target had been totally ignored, implying force-directed layout gave participants confidence in their answers.

At size 20, the circular layout had a longer first-fixation-on-target time and longer disambiguation time. Eye tracking analysis showed that circular layout had a visual 'lead' effect: participants always followed the circle when they searched for the target. This made the target less 'ignorable' to participants, compared to force-directed layout. However, even though the target was noticed by participants and compared with other competitors, the right answer was not guaranteed. The dense overlapping made it hard to find the right answer even under careful consideration.

Concentric layout was found to be more 'insensitive' to size effect: at size 20, it had the shortest first fixation at target time and relatively shorter disambiguation time. Through eye tracking data analysis, it was found that concentric layout made the dense area stand out more. Participants always identified several dense areas quickly and made an efficient comparison before making a decision.

Task accuracy and completion time of path-finding task. Table 16 shows the mean and standard deviation of task accuracy and completion time of path-finding tasks under different conditions. A three-way ANOVA was conducted to investigate the effects of graph size, density, layout and their interaction on task accuracy and completion time. Participants' performance on spatial reasoning test was also included in the analysis as a covariate variable, and was found to have significant effect on task accuracy $F(2, 729) = 7.27$, $p < .05$, and task completion time $F(2, 729) = 10.50$, $p < .05$.The results of ANOVA analysis (see Table 17) revealed the significant effect of all investigated factors and their interactions on task accuracy and completion time except the three-way interaction between size, density, and layout.

For path-finding tasks, both layout and size had significant effect on task accuracy $F(2, 729) = 4.22$, $p < .05$ and $F(2, 729) = 71.29$, $p <$,05 respectively. Force-directed layout had higher task accuracy than circular and concentric layout at size 10 and 20 (see Figure 35). Both layout and size had significant effects on task completion time $F(2, 729) = 30.80$, $p < .05$ and $F(2, 729) = 379.27$, $p < .05$ respectively. The effect of layout by size interaction was also found to be significant for task accuracy $F(4, 729) = 12.87$, $p < .05$ and completion time $F(4, 729) = 5.81$, $p < .05$. At size 20, circular layout and concentric layout had the longest and shortest time-on-task respectively (see Figure 36).

Table 16. Mean and standard deviation of path-finding tasks' accuracy and completion time.

			Task No.	Accuracy		Time (s)	
Layout				Mean	SD	Mean	SD
Circular	5	.20	15	1.00	0.00	2.76	0.97
		.40	3	1.00	0.00	9.59	5.67
		Total		1.00	0.00	6.17	5.31
	10	.20	8	1.00	0.00	7.29	2.47
		.40	4	0.59	0.50	20.67	10.14
		Total		0.72	0.45	16.21	10.51
	20	.20	10	0.63	0.49	26.12	10.04
		.40	5	0.54	0.50	28.37	11.59

			Task No.	Accuracy		Time (s)	
Layout				Mean	SD	Mean	SD
		Total		0.59	0.50	27.25	10.83
	Total	.20		0.88	0.33	12.06	11.77
		.40		0.68	0.47	19.82	11.71
		Total		0.76	0.43	16.49	12.33
Concentric	5	.40	16	1.00	0.00	3.56	1.60
		Total		1.00	0.00	3.56	1.60
	10	.20	9	1.00	0.00	9.51	4.01
		Total		1.00	0.00	9.51	4.01
	20	.20	17	0.20	0.40	20.76	9.16
		.40	2	0.93	0.26	16.90	6.50
		Total		0.56	0.50	18.83	8.13
	Total	.20		0.60	0.49	15.13	9.02
		.40		0.98	0.15	8.01	7.44
		Total		0.82	0.38	10.86	8.81
Force	5	.20	11	1.00	0.00	2.71	1.45
		.40	1	1.00	0.00	5.22	3.96
		Total		1.00	0.00	3.96	3.22
	10	.20	18	1.00	0.00	5.15	1.96
		.40	13	0.76	0.43	12.12	6.98
		Total		0.88	0.33	8.63	6.19
	20	.20	7	0.73	0.45	20.41	11.09
		.40	6	0.71	0.46	23.38	9.77
		Total		0.72	0.45	21.90	10.49
	Total	.20		0.91	0.29	9.42	10.20
		.40		0.82	0.38	13.57	10.44
		Total		0.87	0.34	11.50	10.51
Total	5	.20		1.00	0.00	2.73	1.23
		.40		1.00	0.00	5.48	4.37
		Total		1.00	0.00	4.57	3.86
	10	.20		1.00	0.00	7.32	3.42
		.40		0.64	0.48	17.82	10.03
		Total		0.82	0.38	12.57	9.14
	20	.20		0.52	0.50	22.43	10.38
		.40		0.72	0.45	22.89	10.56
		Total		0.62	0.49	22.66	10.45
	Total	.20		0.82	0.38	11.84	10.75

Layout	Task No.	Accuracy		Time (s)	
		Mean	SD	Mean	SD
	.40	0.81	0.39	14.40	11.31
	Total	0.81	0.39	13.26	11.13

Table 17. Summary results of three-way ANOVA analysis of path-finding tasks.

Source	Dependent Variable					
	Time (s)			Accuracy		
	F	Sig.	η^2	F	Sig.	η^2
SRscore	10.49	.001	0.01	7.27	.007	0.01
Size	379.27	.000	0.51	71.29	.000	0.17
Density	44.16	.000	0.06	4.50	.034	0.01
Layout	30.80	.000	0.08	4.23	.015	0.01
Size * Density	13.53	.000	0.04	13.67	.000	0.04
Size * Layout	5.81	.000	0.03	12.87	.000	0.07
Density * Layout	9.23	.000	0.03	43.94	.000	0.11
Size * Density * Layout	2.86	0.06	0.01	0.81	0.45	0.00

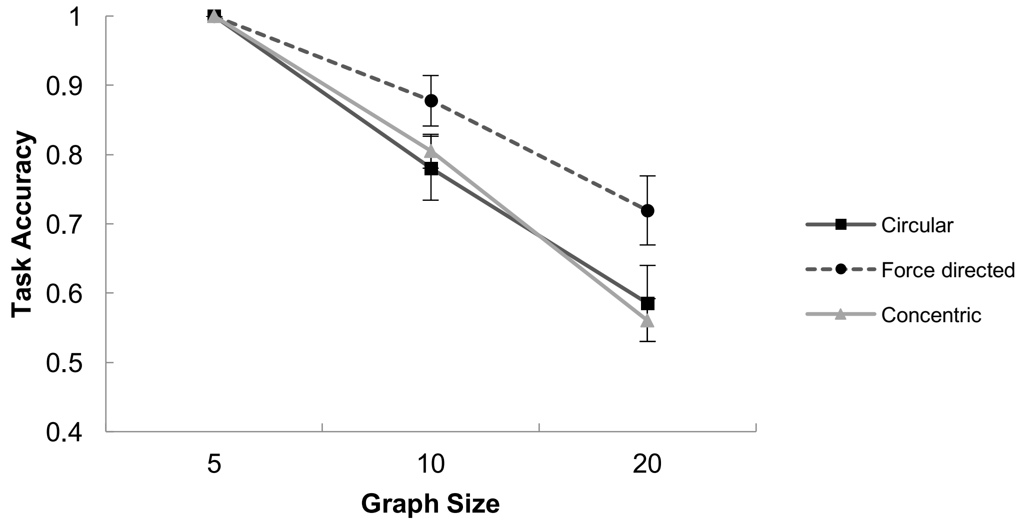


Figure 35. Layout and size effect on task accuracy.

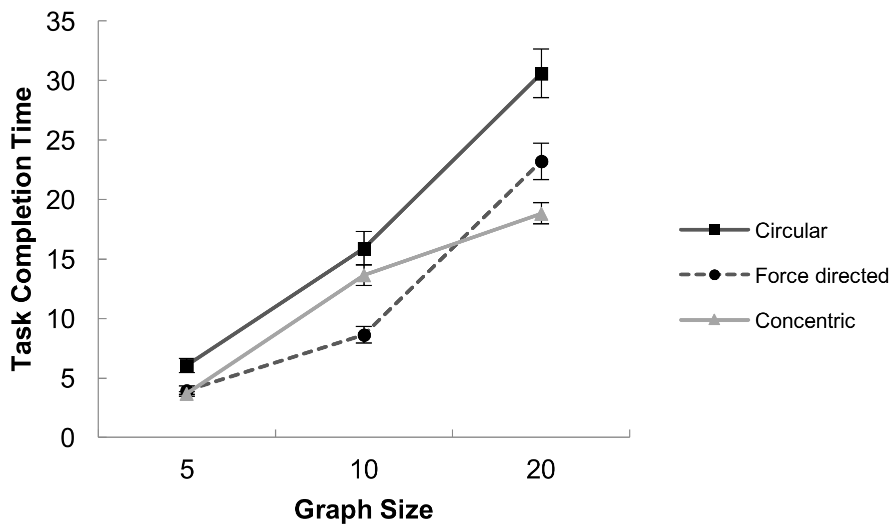


Figure 36. Layout and size effect on task completion time.

Interestingly, the three-way ANOVA analysis revealed that the concentric layout had higher task accuracy at density .4 than .2. Figure 38 shows that density had a significant effect on task completion time $F(2, 732) = 44.4.50$, $p < .05$. The significant effect of graph density is illustrated in Figure 38 ($F(2, 732) = 44.16$, $p < .05$).

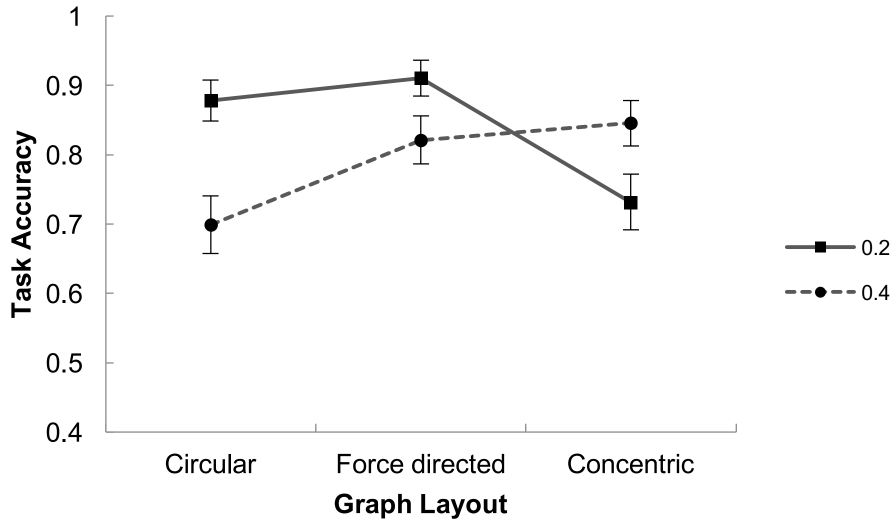


Figure 37. Density effect on task accuracy.

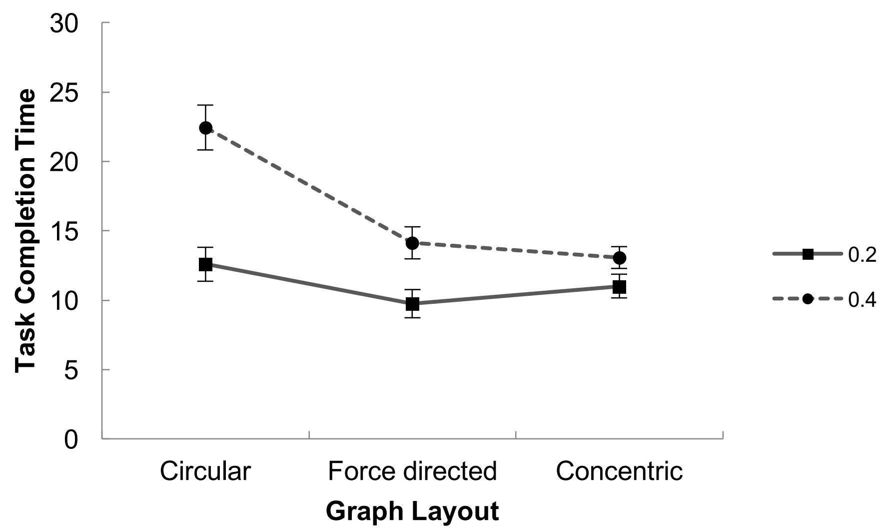


Figure 38. Density effect on task completion time.

Eye tracking data analysis of path-locating tasks. As in the node-locating task, 30 participants' eye movement data were used for eye tracking analysis. The '-1' value in the output spreadsheet of eye tracking data (means eye tracker either didn't record this measure or no fixation was recorded on this particular AOI.) were omitted from the analysis to make the result more interpretable.

AOI definition. The two designated end-nodes were defined as AOI T1 and T2 respectively.

Metrics. Note that the metric defined using T1 and T2 refer to their absolute values. For example, participants might look at T2 before T1, the value used for the time participants needed to find the second target after locating the first target was calculated as T1 minus T2 (instead of T2 minus T1). What is at issue in this study is the cognitive processing stage instead of the watching sequence.

The analysis of path-finding tasks' eye tracking data included the following metrics:

1. *Time till first fixation on the AOI T1 and T2*: Time to locate the first/second target (searching process)

2. *The duration between first fixation on AOI T1* to first fixation on AOI T2: the time participants needed to locate the second target after the first target was located.

3. *The duration between first fixation on T2 (or T1) to complete the task*: The time after two targets are located to task completion (comprehension and disambiguation process): This is the time participants needed to find the shortest path after they have located the two designated end-nodes. The value of this metric is not necessarily positive, as the negative value might imply that the participant completed the task even though the eye tracking data showed they did not look at the target nodes. There are two explanations for this negative value: first, participants actually 'looked' at these nodes and the eye tracker didn't catch this looking; second, participants found the path without looking at the end nodes. This is evidence of covert attention which needs further study.

Here, negative values were omitted from the analysis to make the result more interpretable.)

Time till first fixation on first target (used T1 if the value of T2 minus T1 is positive, use T2 if negative). One outlier (33135 ms, > 3 SD) was omitted from dataset before analysis. Table 18summarizes the mean and standard deviation of four eye tracking metrics under different conditions. Three-way ANOVA (see the summarized results in Table 19) showed that graph size had significant effect on the time participants needed to locate the first target $F(2, 397) = 16.87$, $p < .05$. At size 20, the force-directed layout was significantly faster (by .78 s) than concentric layout to locate the first target (see Figure 39).

Table 18. Mean (ms) and standard deviation of the eye movement metrics of path-finding tasks at different conditions.

			Time till first fixation on AOI T1 (ms)		Time till first fixation on AOI T2 (ms)		DeltaT1 (ms)		DeltaT2 (ms)	
Size			Mean	SD	Mean	SD	Mean	SD	Mean	SD
5	.20	Circular	582.82	314.95	1127.47	414.95	544.65	287.50	1651.00	1140.53
		Force	918.06	394.19	1715.38	456.50	797.31	495.87	819.25	825.96
		Total	745.36	389.08	1412.52	522.30	667.15	415.97	1247.73	1071.63
	.40	Circular	875.15	572.42	1787.22	1994.59	912.07	1494.81	7564.22	6456.27
		Concentric	740.52	586.29	1582.55	1008.92	842.02	783.95	2072.84	1454.05
		Force	1838.83	1199.00	3405.94	1890.58	1567.11	1164.91	1714.22	1667.85
		Total	1003.49	851.38	2013.42	1693.66	1009.92	1141.56	3666.24	4538.37
	Total	Circular	762.20	505.73	1532.32	1604.74	770.11	1189.36	5279.57	5845.58
		Concentric	740.52	586.29	1582.55	1008.92	842.02	783.95	2072.84	1454.05
		Force	1405.53	1014.32	2610.38	1633.90	1204.85	981.30	1293.06	1395.96
		Total	933.67	761.88	1850.88	1493.37	917.20	1008.41	3012.05	4055.48
10	.20	Circular	1201.26	713.88	2323.63	1246.96	1122.37	1102.83	5866.19	3746.09
		Concentric	1381.38	913.65	3094.33	1907.93	1712.96	1310.91	8118.46	5077.79
		Force	1080.74	802.67	2940.47	971.62	1859.74	867.79	2662.16	2203.11
		Total	1230.30	807.95	2755.30	1472.38	1525.00	1155.05	5768.73	4441.88
	.40	Circular	1579.70	934.34	4124.88	3313.98	2545.18	2758.71	17434.08	10612.45
		Force	1361.91	870.03	3004.17	1893.00	1642.26	1810.44	10547.74	7255.05
		Total	1511.08	914.21	3771.78	2973.87	2260.70	2521.75	15264.41	10154.10
	Total	Circular	1447.00	877.63	3493.27	2891.58	2046.27	2406.21	13377.81	10405.95
		Concentric	1381.38	913.65	3094.33	1907.93	1712.96	1310.91	8118.46	5077.79
		Force	1234.71	842.07	2975.36	1529.15	1740.64	1449.61	6980.45	6793.73
		Total	1373.64	872.25	3274.20	2407.84	1900.57	2002.22	10616.17	9195.49
20	.20	Circular	1908.44	2074.20	5909.04	5739.38	4000.59	5566.01	22865.63	14779.65
		Concent	2847.74	2107.54	6916.85	3806.02	4069.11	2833.15	14630.19	10337.36

Size			Time till first fixation on AOI T1 (ms)		Time till first fixation on AOI T2 (ms)		DeltaT1 (ms)		DeltaT2 (ms)	
			Mean	SD	Mean	SD	Mean	SD	Mean	SD
		ric								
		Force	2330.71	1241.00	5458.50	2607.61	3127.79	2118.48	19323.92	18625.80
		Total	2363.51	1888.84	6119.27	4291.83	3755.76	3832.78	18925.14	15014.31
	.40	Circular	3126.66	6177.86	6977.62	8415.57	3850.97	3563.89	33361.90	32810.48
		Concentric	2448.85	2206.07	5576.70	2998.86	3127.85	1986.71	17496.96	31913.01
		Force	1429.37	1268.58	5107.89	4683.61	3678.52	4251.50	26402.67	29812.56
		Total	2354.04	3947.31	5913.67	5885.70	3559.64	3378.72	25937.17	31864.55
	Total	Circular	2539.30	4673.44	6462.41	7205.34	3923.11	4595.34	28301.20	26063.76
		Concentric	2648.30	2146.38	6246.78	3460.55	3598.48	2469.74	16063.57	23539.92
		Force	1853.53	1323.53	5272.88	3816.53	3419.35	3397.16	23071.49	25189.03
		Total	2358.63	3114.86	6013.28	5160.32	3654.65	3595.82	22540.04	25322.15
Total	.20	Circular	1322.11	1442.54	3400.69	4120.82	2078.58	3785.06	11321.49	13123.93
		Concentric	2157.69	1800.03	5118.02	3594.47	2960.33	2524.90	11565.84	8843.28
		Force	1545.08	1133.73	3632.53	2365.11	2087.44	1732.34	8940.05	14662.02
		Total	1630.23	1497.36	3960.15	3540.29	2329.92	2909.66	10614.07	12623.26
	.40	Circular	1823.46	3375.60	4309.91	5349.24	2486.44	2940.26	19277.67	21015.23
		Concentric	1390.17	1648.15	3101.45	2789.08	1711.28	1758.56	7938.35	20891.23
		Force	1514.94	1130.04	3945.82	3394.56	2430.88	3078.33	14504.79	21662.54
		Total	1612.27	2463.69	3858.65	4238.72	2246.38	2706.90	14666.87	21599.34
	Total	Circular	1622.36	2772.43	3945.19	4901.50	2322.84	3300.89	16086.21	18635.31
		Concentric	1711.02	1747.81	3944.44	3291.91	2233.43	2192.28	9454.76	16971.48
		Force	1528.94	1127.36	3800.28	2954.12	2271.33	2539.66	11919.60	18874.61
		Total	1619.90	2105.81	3901.77	3952.97	2281.88	2791.79	12944.90	18421.59

Table 19. Summary of three-way ANOVA analysis results of path-finding tasks.

Dependent Variable	Source	F	Sig.	η^2
Time till first fixation on T1	Size	16.87	0.00	0.08
	Density	0.89	0.35	0.00
	Layout	0.06	0.94	0.00
	Size * Density	0.27	0.76	0.00
	Size * Layout	1.56	0.18	0.02
	Density * Layout	0.96	0.38	0.01
	Size * Density * Layout	2.97	0.05	0.01
Time till first fixation on T2	Size	46.17	0.00	0.18
	Density	0.92	0.34	0.00
	Layout	0.03	0.97	0.00
	Size * Density	0.33	0.72	0.00
	Size * Layout	1.25	0.29	0.01
	Density * Layout	1.40	0.25	0.01

Dependent Variable	Source	F	Sig.	η^2
	Size * Density * Layout	0.95	0.39	0.01
DeltaT1	Size	38.25	.000	0.16
	Density	0.34	0.56	0.00
	Layout	0.05	0.95	0.00
	Size * Density	0.19	0.83	0.00
	Size * Layout	0.46	0.77	0.00
	Density * Layout	0.94	0.39	0.01
	Size * Density * Layout	1.57	0.21	0.01
DeltaT2	Size	51.71	0.00	0.20
	Density	10.57	0.00	0.03
	Layout	4.66	0.01	0.02
	Size * Density	0.94	0.39	0.01
	Size * Layout	1.86	0.12	0.02
	Density * Layout	1.18	0.31	0.01
	Size * Density * Layout	0.02	0.99	0.00

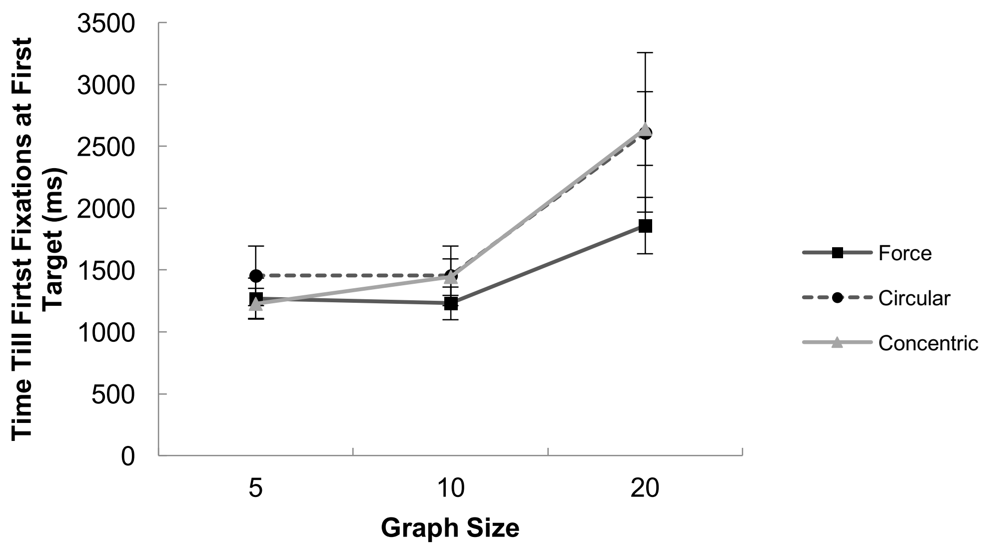


Figure 39. Time till first fixation on first target AOI.

Time till first fixation on second target (used T2 if the value of T2 minus T1 is positive, use T1 if negative). As illustrated in Figure 40, force-directed layout was significantly faster than both circular and concentric layout (by 2.06 s and .68 s) to locate the second target. At size 10, force-directed layout was faster by 1.41 s compared to circular layout. Participants almost always used the same time to locate the second target of graphs at size 20.

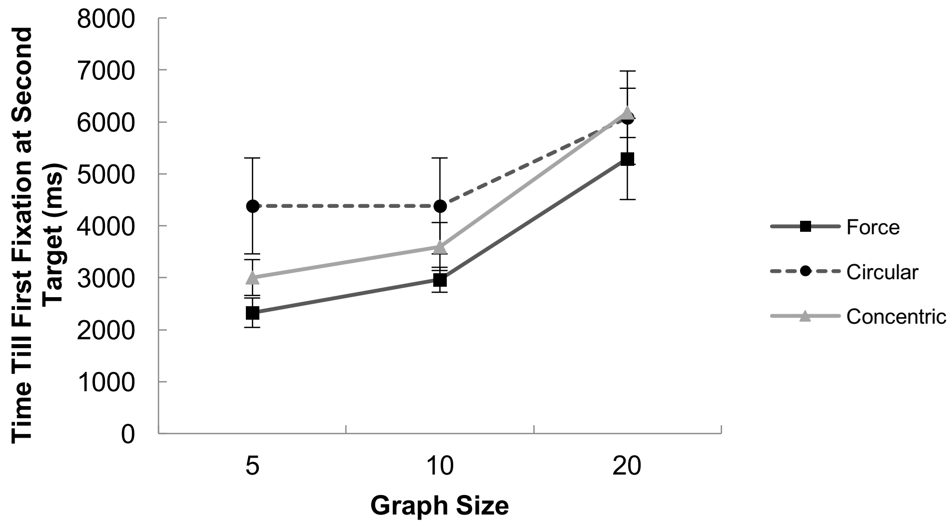


Figure 40. *Time till first fixation at second target.*

The duration between first fixations on AOI T1 to first fixation on AOI T2. Three-way ANOVA showed size had significant effect on the time participants needed to search for the second target after locating the first target, $F(2, 400) = 38.25$, $p < .05$. As illustrated in Figure 41, at size 5, force-directed layout was significantly faster than concentric and circular layout (by .72 s and 1.29 s respectively).

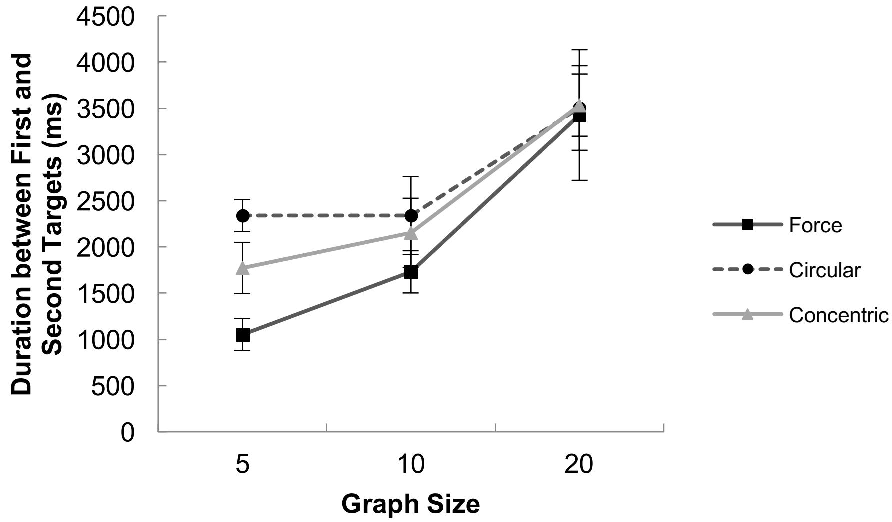


Figure 41. The duration between first fixations on AOI T1 to first fixation on AOI T2.

The duration between first fixations on T2 (or T1) to complete the task. Four outliers (> 3 SD, 174842 from circular layout at size 5, 10 and 20, 173658 from concentric layout at size 20) were moved from dataset before analysis. As illustrated in Figure 42, both size and layout had significant effects, $F(2, 396) = 51.71$, $p < .05$ and $F(2, 396) = 4.66$, $p < .05$. Circular layout needed significantly longer time than force-directed and concentric layout at all three size levels. Force-directed layout was most sensitive to size effect. Like the case in node-locating tasks, concentric layout seems to be more rersistent to the size effect: there was no significant size effect on concentric layout.

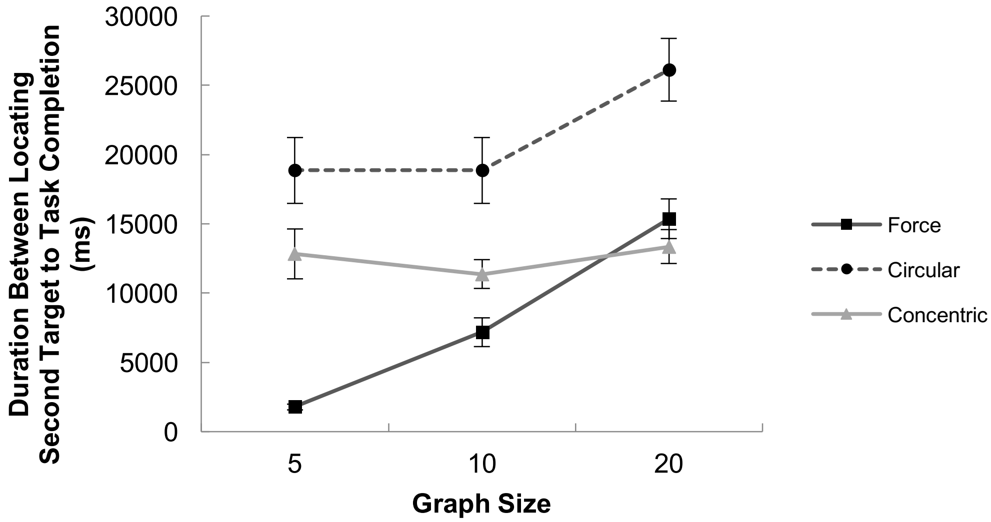


Figure 42. *The duration between first fixations on T2 (or T1) to task completion.*

Error tasks analysis. Like the case in node-located tasks, tasks with graphs size at 20 presented most difficulties to participants (lower accuracy and longer time-on-task). The following analysis was focused on the six path-finding tasks with graph at size 20.

P7 Task, force-directed layout, size 20, density .2. The P7 task asked the participants to find the shortest path between node 6 and 15. Figure 43 showed the 5 different paths considered by participants. The widths of the path were proportional to the number of participants who provided that answer. The closer the path was to the geometric path (straight line) of the two end-nodes, the more participants who chose that path. Further analysis of the eye tracking data revealed that participants' scan-path clearly had a three-stage processing pattern:

1. Exploring, fixations dispersed to almost all nodes.
2. Comparing, back-and-forth scan-path from one end-node to compare several possible paths.

3. Confirming, several back-and-forth scan-path on the selected path to confirm the verbally given answer.

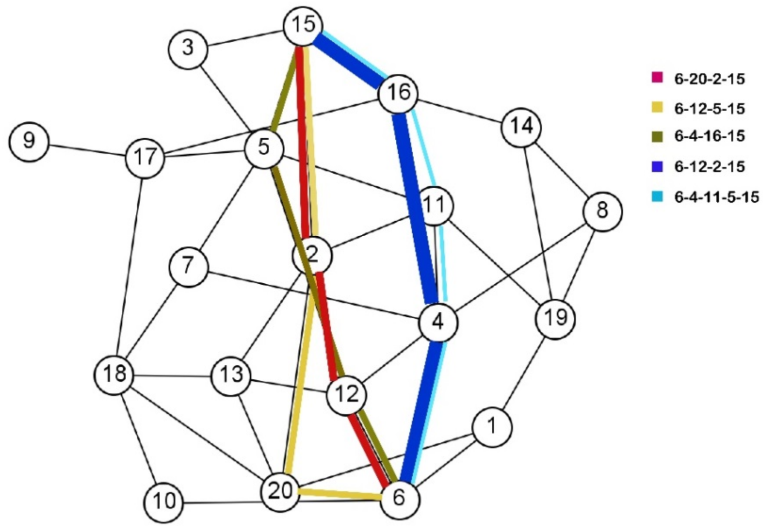
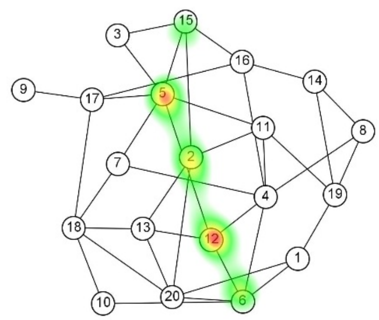
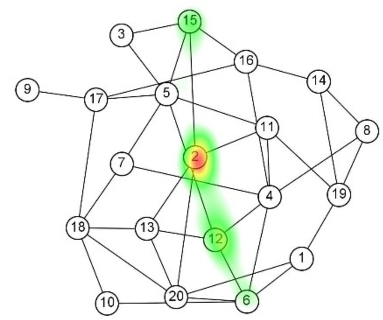


Figure 43. The visualization of participants' answers on P7 task.

Figure 44 shows two heat-maps of participants' eye movements. The heat-map on the right is evidence of 'careless' answer: participants did not notice that the path between node 2 and 12 was a false connection. Transitions map visualizations also supported this discovery (see Figure 45). The eye movement of participants who gave the right answers demonstrated a clear comparison stage before decision-making.

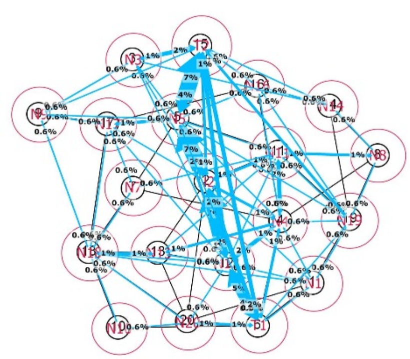


Heat-map visualization of participants’ eye movement who answered path 6-12-5-15 (right). Note that participants paid much attention to figure out whether node 5 and 12 were directly connected or connected through node 2.

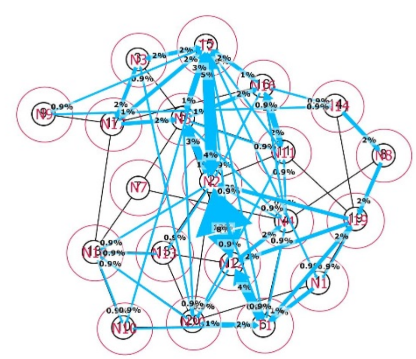


Heat-map visualization of participants’ eye movement who answered path 6-12-2-15 (wrong). Note that no comparison between node 2 and 5: participants didn’t notice that this is a false connection.

Figure 44. Heat-map visualization of participants’ eye movement of P7 task.



Transitions map of the eye movement of participants who answered path 6-12-5-15. Note that most transitions happened from node 12 to 2 (7%) and 2 to 5 (7%) to check the connectedness.



Transitions map of the eye movement of participants who answered path 6-12-2-15. Note that most transitions happened from node 12 to 2 (8%) but not very much from 2 to 5 (3%).

Figure 45. Transitions map visualizations of participants’ eye movement of P7 task.

P6 Task, force-directed layout, size 20, density .4. The P6 task was to find the shortest path between node 19 and 20. As illustrated in Figure *46*, as density increased, the answers became more diverse—evidence of density effect. Like task P7, the answers presented a geometric path tendency. Analysis of participants’ scan-paths showed that wrong answers had a relatively ‘intensive’ attention map; they tend to include the options they were confronted with first and do not go further to discover more possibilities (see Figure 47).

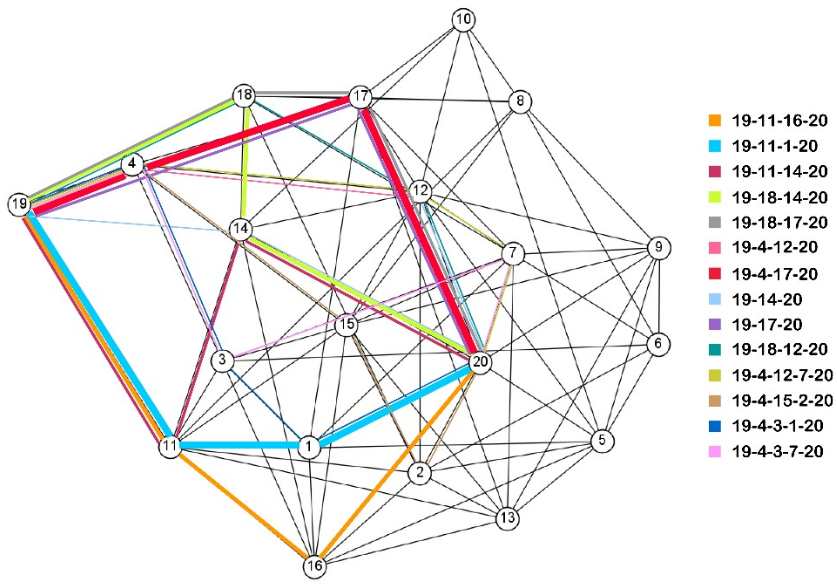
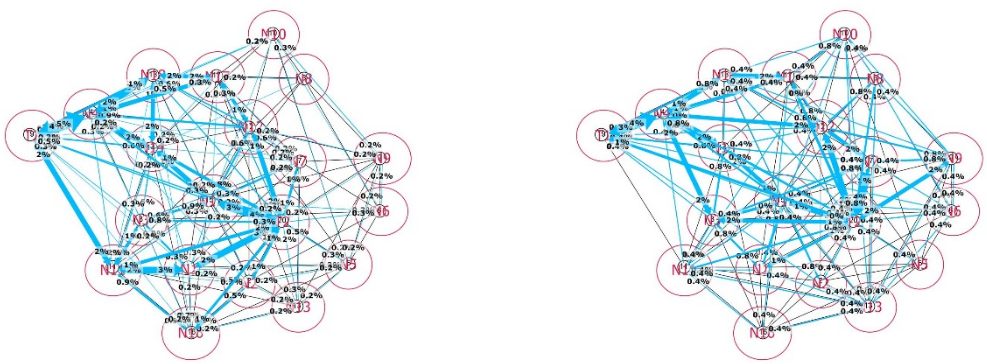


Figure 46*. The visualization of participants’ answers on P6 task.*



Transitions map visualization of participants’ eye movement who gave the right answer.

Transitions map visualization of participants’ eye movement who gave the wrong answer. Note that comparing the transitions map on the left, not too much attention transitions happened on the lower part the graph.

Figure 47. Transitions map visualizations of participants’ eye movement of P6 task.

P10 Task, circular layout, size 20, density .2. The task was to find the shortest path between node 10 and 11. As illustrated in Figure 48, the answers covered almost every possible path (went through 4 or 5 nodes) that connected node 10 and 11. Wrong answers were either counting that node 6 and 3 were connected or including one more node on the path than the shortest one (4 nodes). Two kinds of search strategies have been presented by the scan-path visualization of participants' eye movements: 1) first participants searched along the circle to get a general impression of the graph (exploration) then focused on the comparisons of several possible routes; 2) several interested areas have been compared (see Figure *49*).

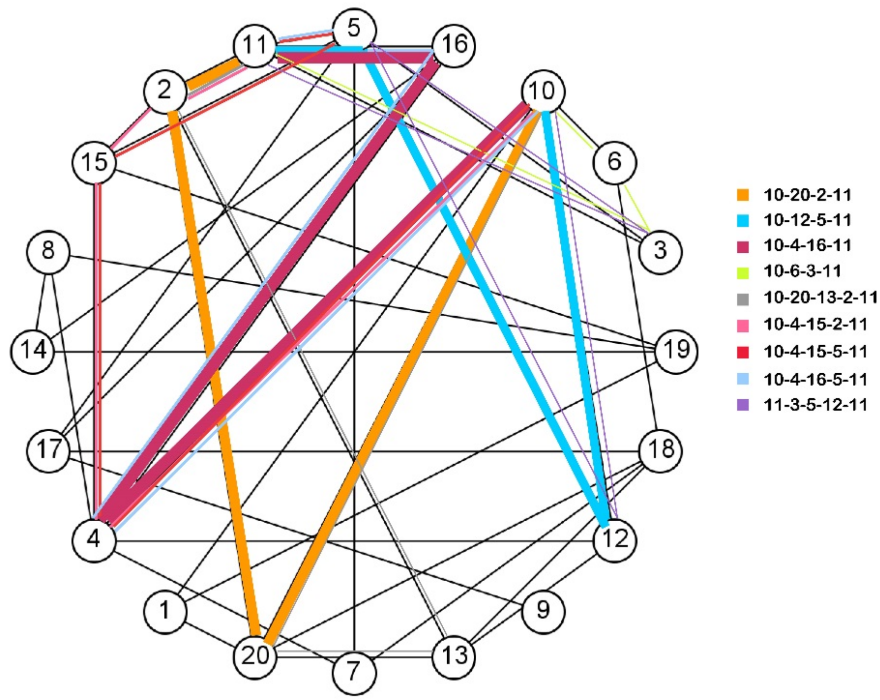
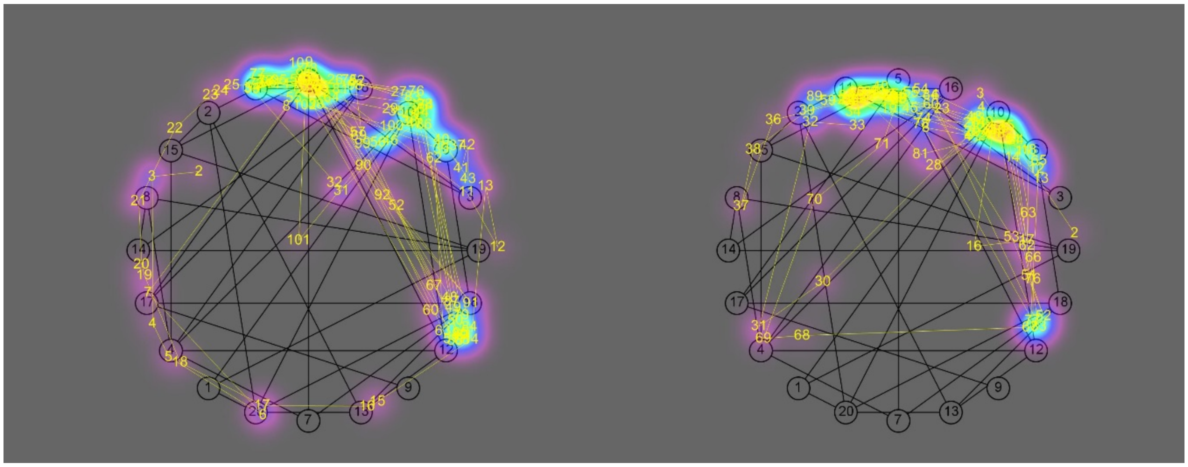


Figure 48. The visualization of participants' answers to P10 task.



The scan-path visualization of participant S30's eye movement on P10 task. First searched along the circle outliner to get a general impression.

The scan-path visualization of participant S03's eye movement on P10 task. Jumping between several interested areas to do the comparisons.

Figure 49. *Scan-path visualizations of participants' eye movement of P10 task.*

P5 Task, circular layout, size 20, density .4. P5 task asked participants to find the shortest path between node 9 and 20. Like the case for P10 task, answers covered almost all possible routes between node 9 and 20 (see Figure 50). Wrong answers were either caused by dense edge crossing and overlapping, or participant's got lost during the task (ended up on the wrong nodes).

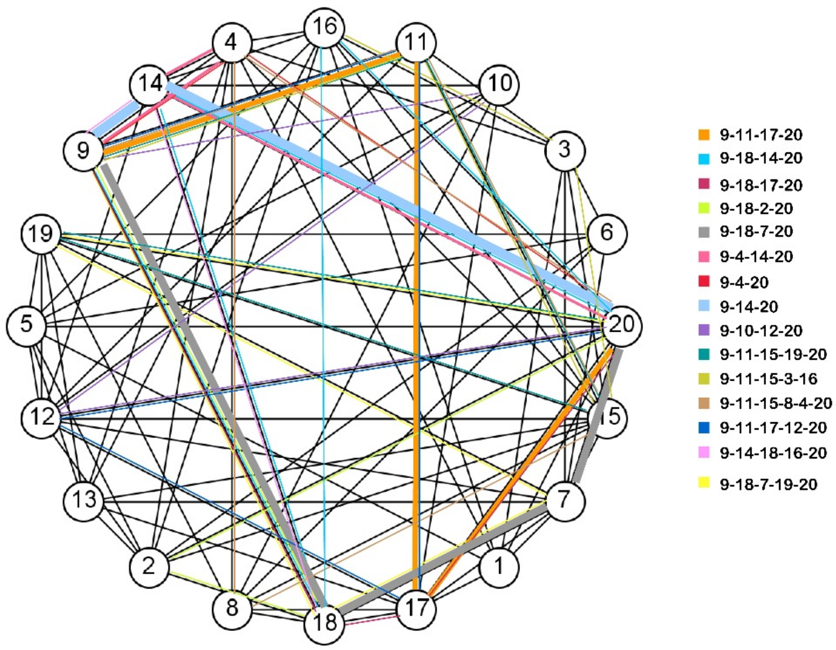


Figure 50. The visualization of participants' answers to P5 task.

P17 Task, concentric layout, size 20, density .2. P17 task was to find the shortest path between node 4 and 15. As illustrated in Figure 51, most participants (20) gave the wrong answer 4-16-15 because of the overlapping caused misreading. Participants who gave the wrong answer all had a sparse attention distribution pattern (see Figure 52).

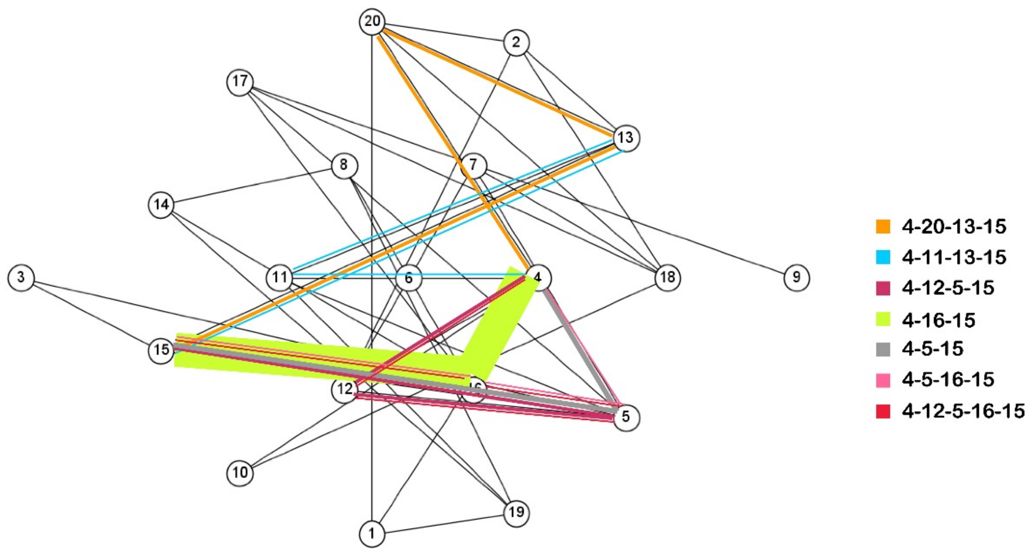
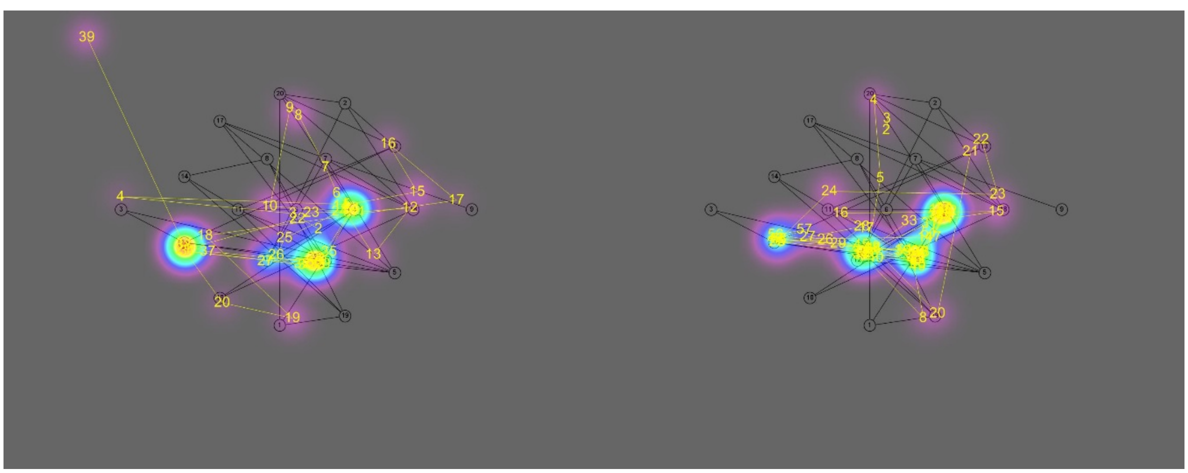


Figure 51. The visualization of the participants' answer to P17 task.



Scan-path visualization of the eye movement of participant S03.

Scan-path visualization of the eye movement of participant S05.

Figure 52. The scan-path visualization of participants' eye movements. Note that they have an economic scan-path pattern.

P2 Task, concentric layout, size 20, density .4. P2 task asked participants to find the shortest path between node 13 and 19. As illustrated in Figure 53, most participants' gave the correct answer 13-16-11-19 (26 out of 30)—another evidence of geometric scan-path tendency.

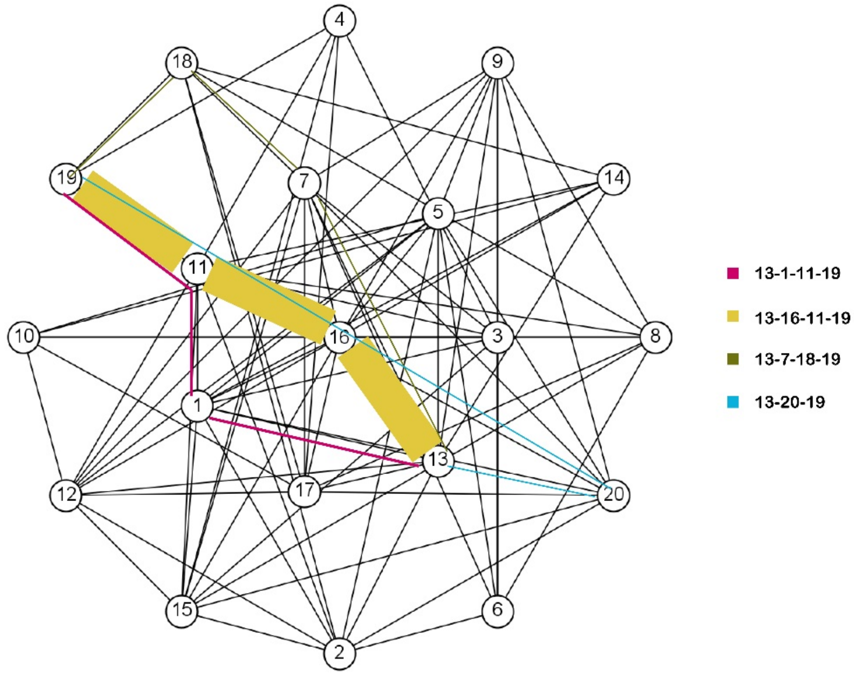


Figure 53. The visualization of participants' answers to P2 task.

Summary of path-finding tasks. For path-finding tasks, layout and size had significant effect on accuracy and task completion time. For all three sizes (5, 10, 20), force-directed layout has the highest accuracy, while the circular layout has the longest time-on-task and lowest task accuracy.

Density had a significant effect on task completion time, but not on task accuracy. The effect of density on task accuracy and completion time depends on graph layout. The accuracy on the concentric layout task did not drop with a denser graph (participants had significantly higher accuracy rates on density level of 0.4 than 0.2). At density level of 0.2, participants spent more time to complete tasks with a circular layout. At density level of 0.4, participants spent equal time on tasks of all three layouts.

At size 20, graphs with force-directed layouts, participants were significantly quicker than with the circular layout to locate the first target. Both size and layout had a significant effect on the comprehension process of path-locating tasks. Force-directed layout was most sensitive to size effect, whereas concentric was most insensitive to size effect. Circular layout needed significantly longer comprehension time than force-directed layout and concentric layout.

Edge crossing and node overlapping caused the most wrong answers of path-finding tasks. Scan-path analysis of participants' eye movements revealed clear geometric-path tendency when they searched for the paths between two nodes. Scan-path analysis also showed a three-stage-processing of path-locating tasks: exploring-comparing-confirming.

Summary

Spatial reasoning ability had a significant effect on participants' performance on node-locating and path-finding tasks: participants' who had higher score on the spatial reasoning test tended to have higher task accuracy and shorter task completion time.

For the node-locating task, participants had decreased task performance (longer completion time, lower accuracy) as graph size enlarged. Force-directed and circular layouts seemed to be more sensitive to the effect of size: the task accuracy of these two layouts dropped dramatically as graph size enlarged from 10 to 20. Concentric layout was insensitive to size effect: it had a relative constant accuracy across three size levels.

Eye tracking evidence showed participants needed more time to 'read' (here, counting connections) the target nodes as graph size enlarged. The larger and denser the graph, the more cognitive load was required from participants for information processing.

Concentric layout is 'insensitive' to size effect on task accuracy, which resulted in a higher accuracy than circular and force-directed layout at size 20. This was consistent with eye tracking data of concentric layout on time to locate targets: concentric layout had a more tempered response to size effect demonstrated by a shorter time to locate the target at size 20. Interestingly, the shorter target-locating time didn't necessarily bring a 'hasty' decision: concentric layout also has a higher accuracy rate compared to force-directed and concentric layout at size level of 20, implying that this is a confident and informative decision. Further analysis on participants' scan-paths showed that concentric layout made it easier for participants to discover the target and disambiguate between competitive answers.

Eye tracking analysis showed that participants had the shortest confirming/disambiguating time for force-directed graph at size 20. However, they didn't have the highest accuracy on task. Further error task analysis revealed that for N2 (forced-direct graph at size 20), the correct answer was totally ignored by participants who gave the wrong answer. Participants made quick/confident but incorrect answers.

The guiding effect of circular layout made the target more 'noticeable'. However, six participants (20%) gave the incorrect answer after comparison, inferring the problem is not coming from the target-noticeability of the circular layout, but visual difficulty resulting from dense overlapping.

Compared to node-locating tasks, path-finding tasks resulted in more/diverse wrong answers. This was partially because path-finding tasks had more than one answer. More importantly, path-finding was a different type of cognitive task which consumed more cognitive resources to find the answer.

Compared to node-locating tasks, the visual leading effect of circular layout was not the case for path finding tasks: Instead of going through every node one by one along the circle, participants were more likely to jump here and there leading by the connectedness. The search strategy was predetermined by participants according to the cognitive task type.

The scan-path analysis of path-locating tasks revealed three stages of processing: exploring-comparing-confirming (more discussion on this subject see Chapter 6).

CHAPTER 5

EXPERIMENT WITH DIA2

The previous chapter described the experiment to investigate the layout ,density, and size effects using abstract graphs and generic tasks. The experiment described here was designed to investigate the same layout and size effects using a real application and task-within-context.

The real application used here is Deep Insight Any Time Any Where (DIA2). DIA2 is a web-based visual analytics platform for searching, viewing, and analyzing the NSF research portfolio for 'casual experts' who have a high degree of training in their discipline, yet with little to no training in advanced visualization and analytics (Madhavan et al., 2015; Molnar et al., 2015). To help users get actionable insights, the visualization algorithms used by DIA2 were chosen not because of their novelty but the capability of producing familiar or self-explanatory representations to a casual expert. DIA2 currently archives data from January 1973 to March 2014 (only data from 1995 is exposed for searching purpose). The main visualizations used by DIA2 were force-directed (to visualize the collaboration network of PIs/coPIs within an institution or research topic, see Figure 54) or a concentric layout (to visualize the collaboration network of a specific PI/coPI, see Figure 55).

The graphs generated by DIA2 for this experiment were chosen to have two size levels: small graph, which includes less than 100 nodes; and large graph, which includes more than 100 nodes. Tasks that have been tested with abstract graphs were mapped and tested using the real application.

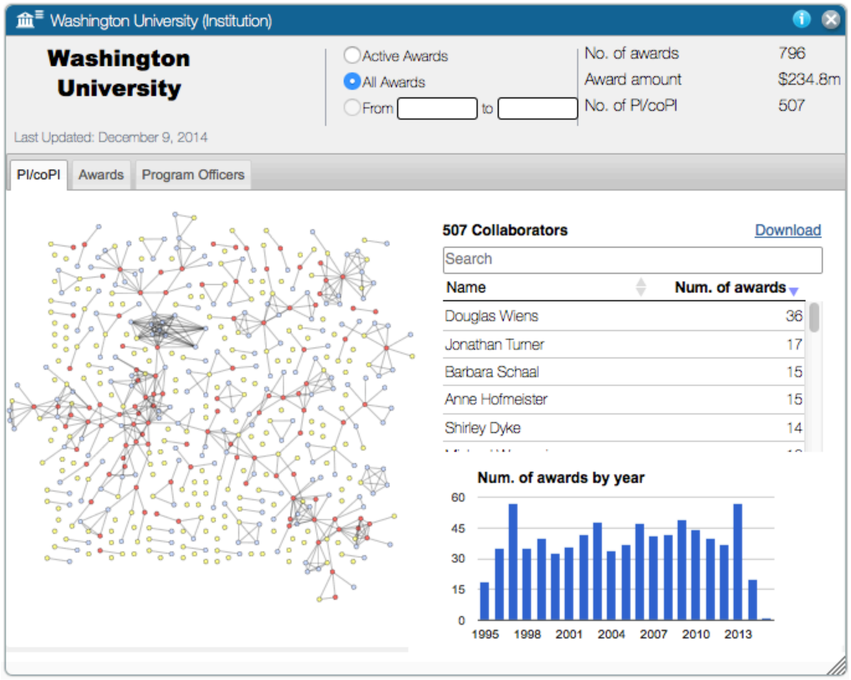


Figure 54. Collaboration network of PIs/coPIs of Washington University visualized using force-directed layout.

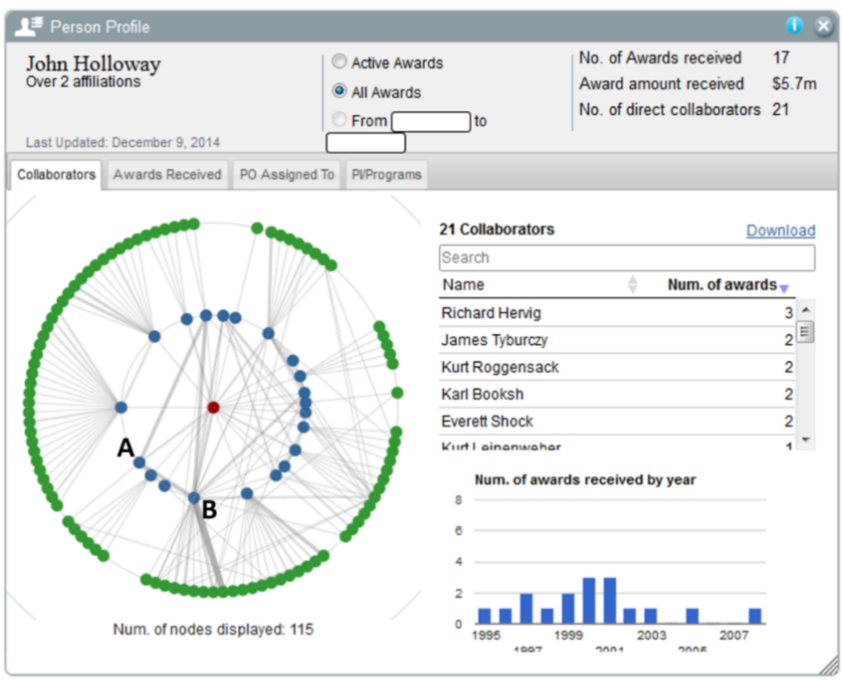


Figure 55. Collaboration network of a particular researcher using concentric layout. The red dot represents the targeted researcher. The blue dots are his/her first level collaboration, and green dots are the second level collaboration.

Methods

Participants. The same participants who participated in the experiment with abstract graphs automatically signed up for this second-stage experiment with DIA2.

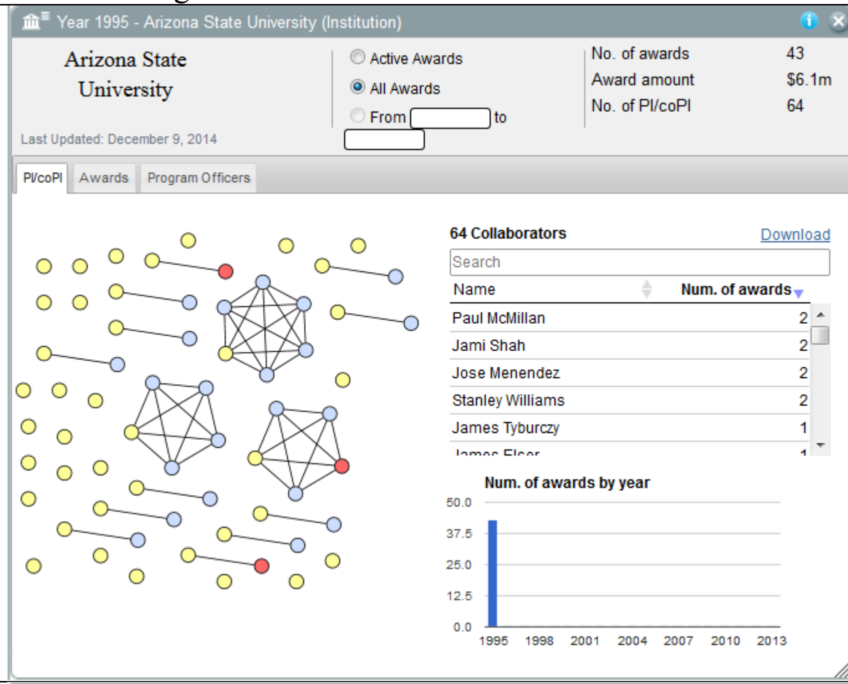
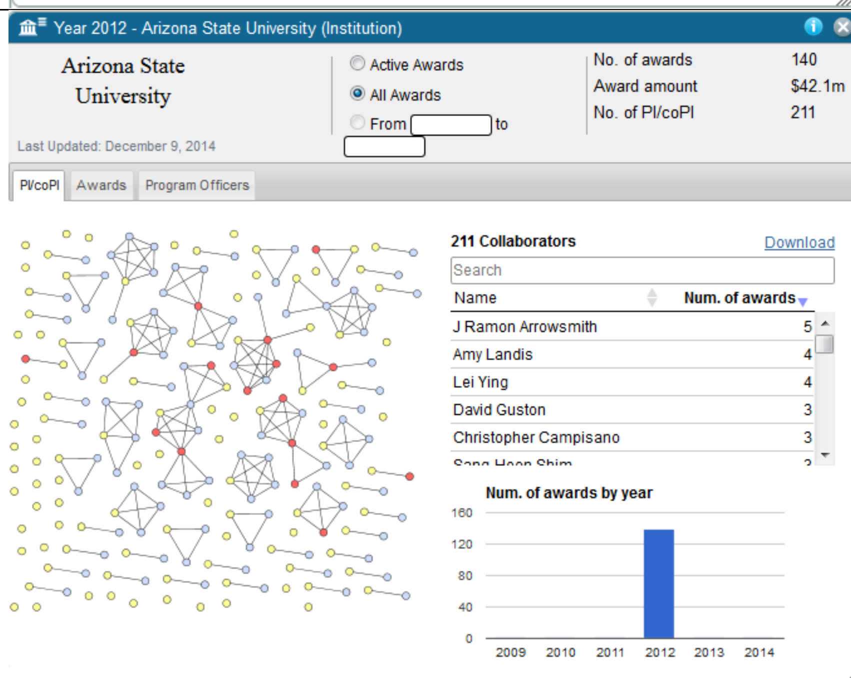
Apparatus. The same apparatus and set-up were used as in the last experiment with abstract graphs. An Eyetech VT2 model eye tracker, running QuickCapture software was used to capture participants' eye movement at a sampling rate of approximate 35 Hz. Stimuli were displayed using a Dell 20.8 inch screen with 1600 × 1200 resolution. Another screen, which was a duplication of the stimuli-display screen, was used for coordinator to control the experiment process. Eye tracking data were analyzed and visualized using the combination of open source software Ogama, Excel, and SPSS.

Graph Stimuli. All stimuli were screenshots from the DIA2 application with real data. In order to simulate the real context, the node-link diagram was presented with other information as a whole widget[6] (see Figure 54 and Figure 55).

Table 20 and Table 21 shows all eight stimuli used in the experiment.

[6] DIA2 uses widgets to present the concrete results for search inquires. A widget includes several tabs that each present a certain characteristic of the search results. For example, as Figure 54 shows, when a search is performed for NSF funded projects at Washington University, the information is presented across three tabs to include information on the collaboration network of funded Pi/coPIs, the awards they made, and the program officers who are managing these awards.

Table 20. Stimuli for DIA2 experiment node-locating tasks.

<table>
<tr><th>Layout</th><th>Size</th><th>Node-locating task</th></tr>
<tr><td rowspan="2">Force-directed</td><td>Small</td><td></td></tr>
<tr><td>Large</td><td></td></tr>
</table>

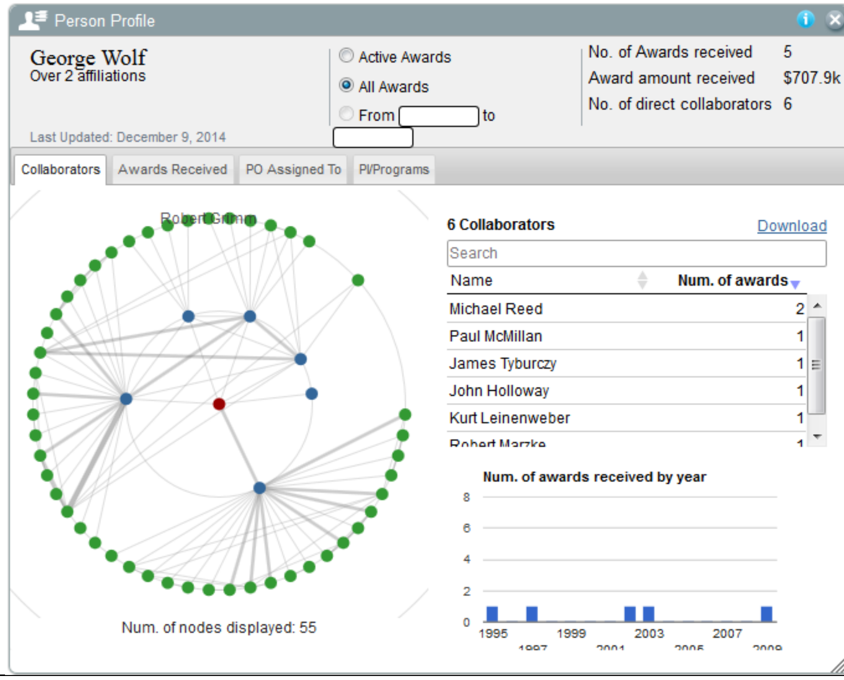
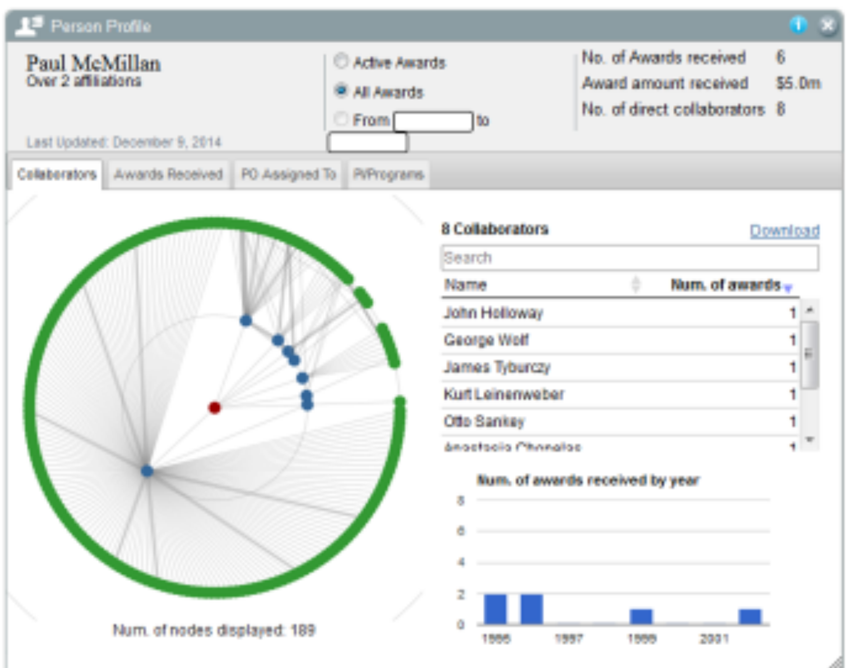
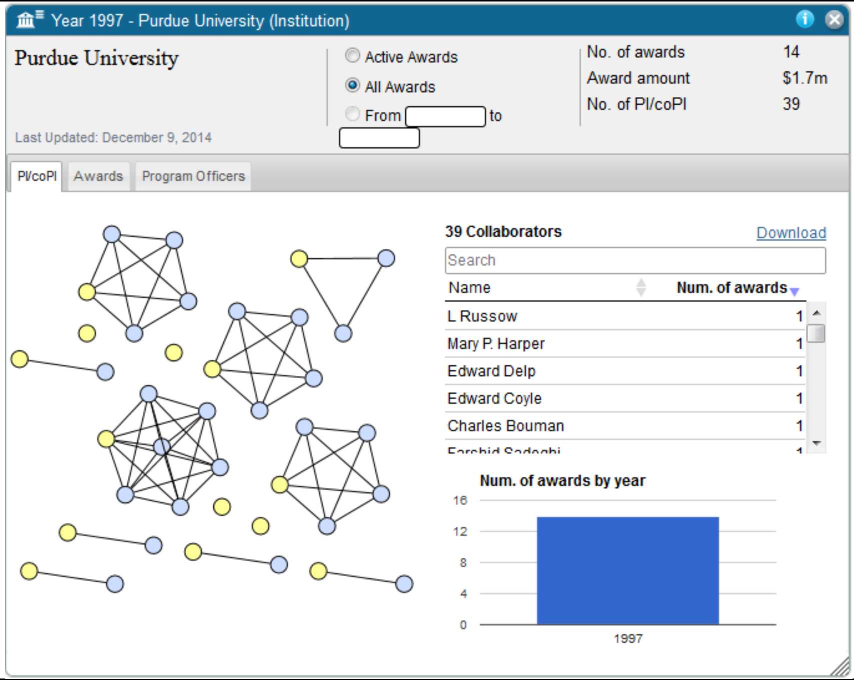
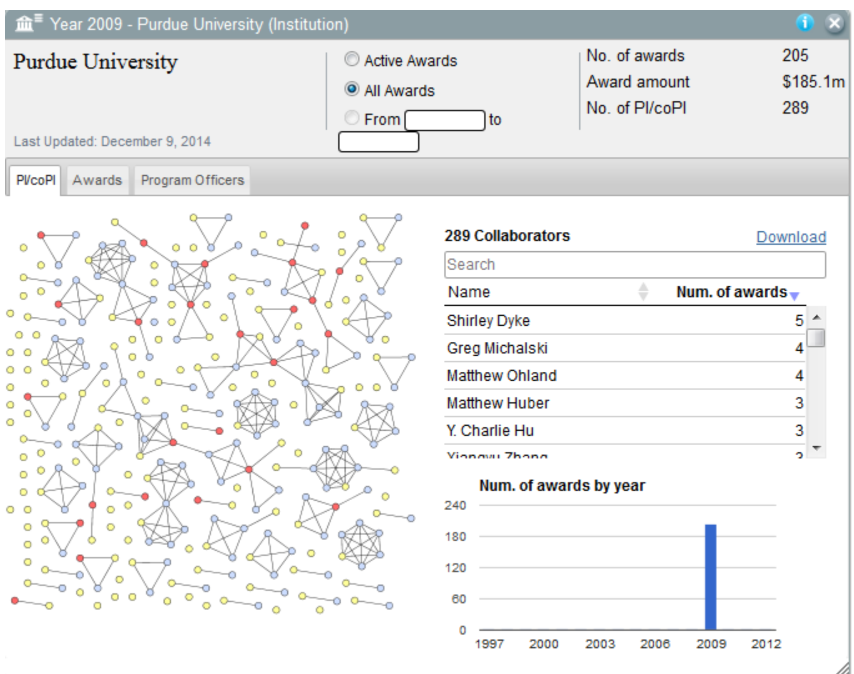
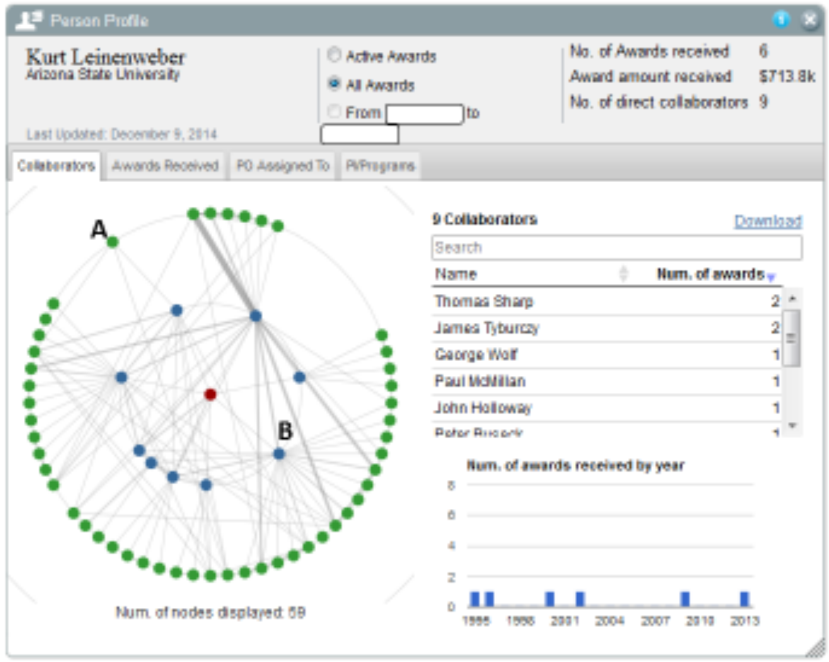
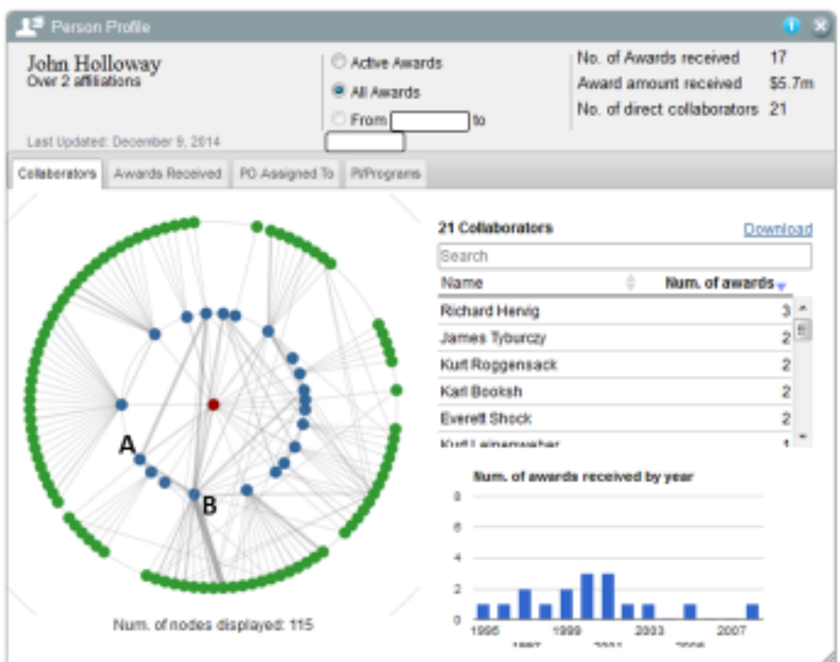
<table>
<tr><th>Layout</th><th>Size</th><th>Node-locating task</th></tr>
<tr><td>Concentric</td><td>Small</td><td>
</td></tr>
<tr><td></td><td>Large</td><td>
</td></tr>
</table>

Table 21. Stimuli for DIA2 experiment path-finding tasks.

<table>
<tr><th>Layout</th><th>Size</th><th>path-finding task</th></tr>
<tr><td>Force-directed</td><td>Small</td><td>
</td></tr>
<tr><td></td><td>Large</td><td>
</td></tr>
</table>

Layout	Size	Node-locating task
Concentric	Small	
	Large	

Tasks. The tasks used in this experiment were the mapping of the tasks in the last experiment with abstract graphs, except that in this experiment they are described using contextual information consistent with the real application. The generic nodes and paths used in the experiment with abstract graphs were replaced using meaningful entities (researchers and collaboration relationship). Comparing to read abstract graphs, participants had to understand the implications of tasks in order to complete the tasks. Moreover, in the purpose of comparing participants' performance on abstract graphs and graphs in real application, the same two kinds of tasks were includes for DIA2 experiment:

1. Node-locating task: find the most connected researchers.
2. Path-finding task: decide whether two researchers were connected, or try to find the 'introduction path' between researcher A and B (Though whom could researcher A be introduced to researcher B, if s/he want to build collaboration).

Table 22shows the eight DIA2 tasks used in the experiment.

Table 22. Tasks used for DIA2 experiment.

No.	Layout	Size	Task instructions
1	Force	Small	You will see the network of the Principal Investigators (PI) of Arizona State University in the year of 1996. Please identify the most connected researcher(s).
2	Force	Large	You will see the network of the Principal Investigators (PI) of Arizona State University in the year of 2012. Please identify the most connected researcher (s).
3	Force	Small	You will see the network of the Principal Investigators (PI) of Purdue University in the year of 1997. Could you tell me if researcher A collaborated with researcher B?
4	Force	Large	You will see the network of the Principal

No.	Layout	Size	Task instructions
			Investigators (PI) of Purdue University in the year of 2009. Could you tell me if researcher A collaborated with researcher B?
5	Concentric	Small	You will see the collaboration network *George Wolf*. Please identify the most connected researcher in this network.
6	Concentric	Large	You will see the collaboration network *Paul McMillan*. Please identify the most connected researcher in this network
7	Concentric	Small	You will see the collaboration network *Kurt Leinenweber*. If researcher A wants to collaborate with researcher B, whom should he/she ask for introduction?
8	Concentric	Large	You will see the collaboration network *John Holloway*. If researcher A wants to collaborate with researcher B, whom should he/she ask for introduction?

Procedure. This experiment commenced immediately after each participant completed the last experiment with abstract graphs. They were first given a general introduction to DIA2 and its visualizations. During this process the mapping between nodes-researchers and edge-collaboration relationship was established. Participants were then given several test questions to check whether they were ready to conduct the tasks. The eight tasks were presented in a random sequence. After completing all of the tasks, participants were invited to take a short interview to discuss the experience they had with DIA2.

Experimental design. In order to compare the layout and size effects between the abstract graph and real application, the experiment was designed as a repeated-measure experiment. The study presented 2 graph layouts (force-directed and concentric) × 2 graph sizes (large and small) × 2 task types (node-locating and path-finding) to each participant[7]. Each participant performed one trial of each layout, graph size, and task type, resulting in 8 trials in total.

Results

Participant S02's data was removed because of missing data. Forty-two participants' (35 male, 7 female) contributed responses for the statistical analysis on task accuracy and completion time.

Task accuracy and completion time. A two-way ANOVA (size by layout) was conducted. Like the case for the experiment with abstract graphs, participants' performance on spatial reasoning tasks was included in the analysis as a covariate variable. Results showed there was no significant effect of participants' performance on spatial reasoning tasks on their task accuracy and completion time (see Figure 56 and Figure 57). Also, no significant correlation was founded between participant's score of spatial reasoning test with their performance on graph reading tasks using DIA2.This results is in contrast with the findings for abstract graphs, in which participants' performance was significantly affected by their score on spatial reasoning test. One feasible explanation for this contradiction is that, for DIA2 tasks, other than spatial reasoning ability, more context factors contributed to the variations in the results. As which will be elaborated later in this chapter, the design attributions of the visualization (e.g. color coding, weighted path) significantly affected participants' understanding of the graphs. Therefore, the effect comes from participants' spatial ability has been 'diluted'.

[7] Constrained by the real data used by DIA2, this experiment only included two size levels and two graph layouts.

Table 23 summarizes the results of ANOVA analysis. Size and layout were found to have significant effect on task accuracy, $F(1, 334) = 4.51$, $p < .05$ and $F(1, 334) = 48.14$, $p < .05$.

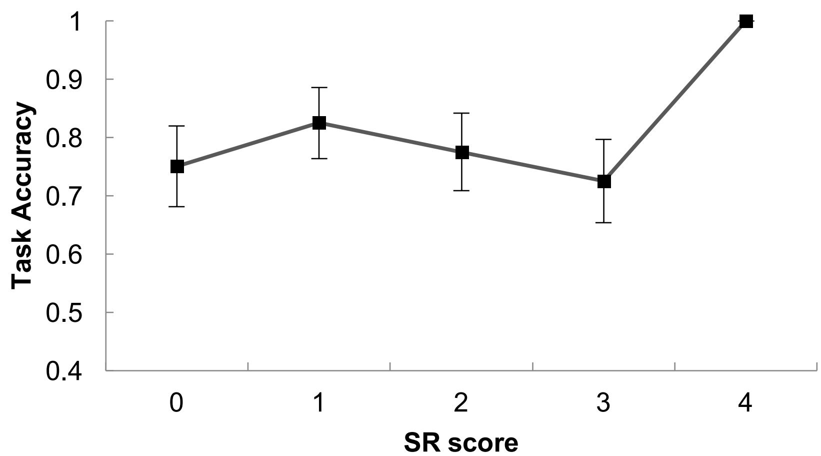


Figure 56.The effect of participants' spatial reasoning test performance on task accuracy.

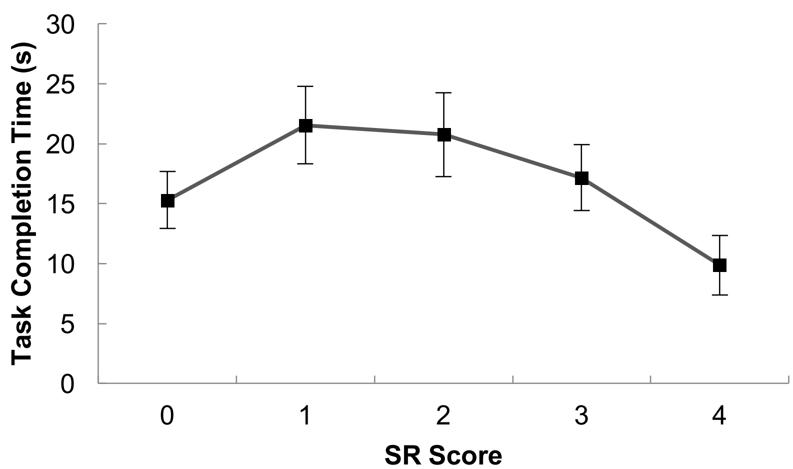


Figure 57. The effect of participants' spatial reasoning test performance on task completion time (s).

Figure 58 and Figure 59 show the task accuracy and completion time by task. Task 1, 2, 5 and 6 were node-locating tasks, while task 3, 4, 7 and 8 were path-finding tasks (see Table 22). Notice that Tasks 3 and 4 had 100% accuracy and shortest completion times.

Table 23. Summary results of ANOVA analysis of DIA2 tasks.

Source	Dependent Variable					
	Accuracy			Time (s)		
	F	Sig.	η^2	F	Sig.	η^2
SR Score	0.85	0.36	0.00	1.02	0.31	0.00
Size	4.51	0.03	0.01	13.77	0.00	0.04
Layout	48.15	0.00	0.13	0.03	0.87	0.00
Size * Layout	0.50	0.48	0.00	12.89	0.00	0.04

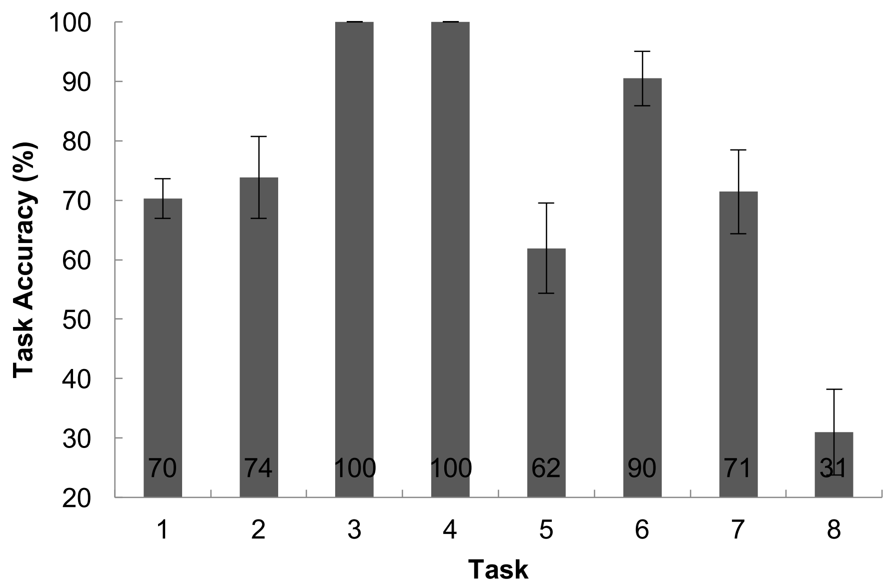


Figure 58. Task accuracy of each DIA2 task. Note that task 3 and 4 had 100% accuracy.

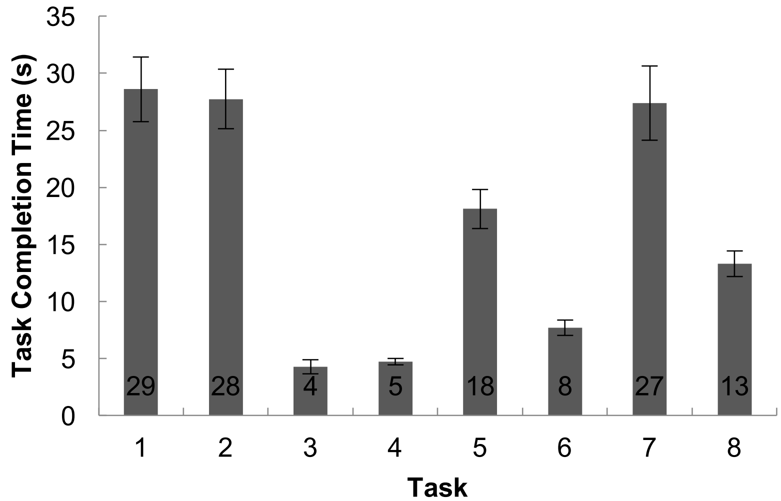


Figure 59. Task completion time of each DIA2 task. Note that task 3 and 4 had shortest completion time.

No significant difference on task accuracy was found between node-locating task and path-finding task. However, as illustrated in Figure 60, participants spent significantly more time (8.11 s on average) to complete node-locating task than path-finding task, $F(1, 334) = 23.25$, $p < .05$. These results are in contrast with the findings of the experiment with abstract graphs, in which path-finding tasks had longer time-on-task. By a further look, participants' performance on task 3 and 4 mainly contributed to this short task completion time. In task 3, 4, participants were asked to identify whether two designated researchers (A and B) were collaborated. Their answers were in the "Yes" or "No" format. The results shows participants completed these two tasks less than 5 seconds and all participants answered correctly.

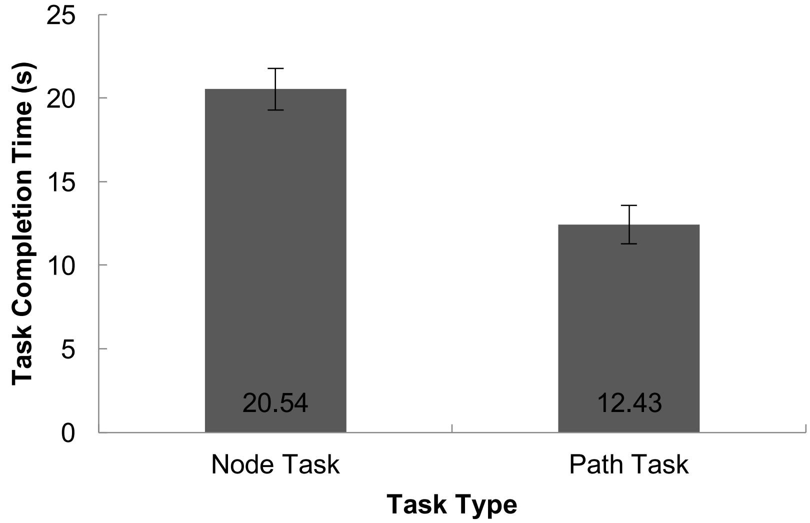


Figure 60. Average task completion time of DIA2 tasks by different task type.

For the node-locating task, a two-way ANOVA was conducted to investigate the size and layout effect on accuracy. No significant differences for both main effects (layout out and graph size) were found. However, the interaction was significant $F(1,164) = 18.28$, $p < .05$ (see Figure 61). As illustrated in Figure 62, the size and layout effects had significant effect on task completion time, $F(1,164) = 7.04$, $p < .05$ and $F(1,164) =51.49$, $p < .05$. The effect of layout on task accuracy and completion time was depend on graph size: for small graphs (include less than 100 nodes), force-directed layout had higher accuracy (95%) compared to concentric layout (62%), although this 33% higher accuracy was achieved assuming more mental effort (significantly longer task completion time). For larger graphs, however, concentric layout out-performed force-directed layout with a significantly shorter completion time and higher task accuracy. This result is consistent with the findings of experiment on abstract graphs: concentric layout made the target more 'noticeable', thus resulted in a higher task accuracy with less mental effort (shorter completion time).

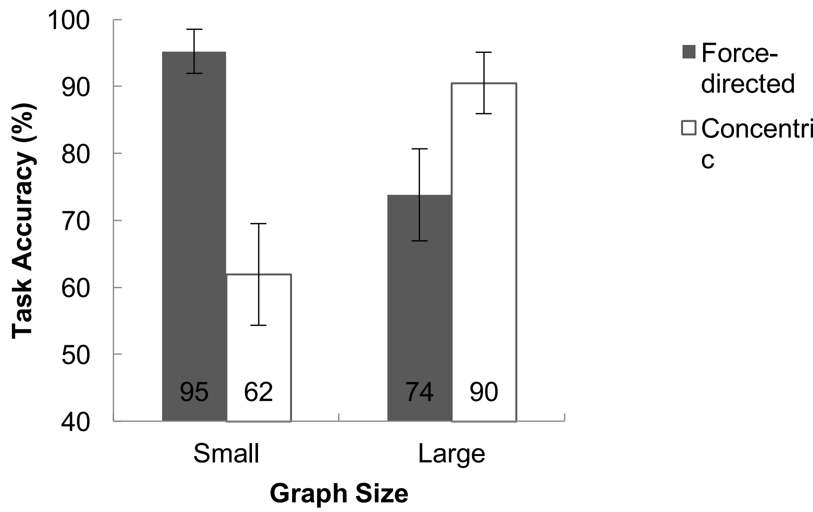


Figure 61. The effect of layout and size on accuracy of node-locating tasks of DIA2.

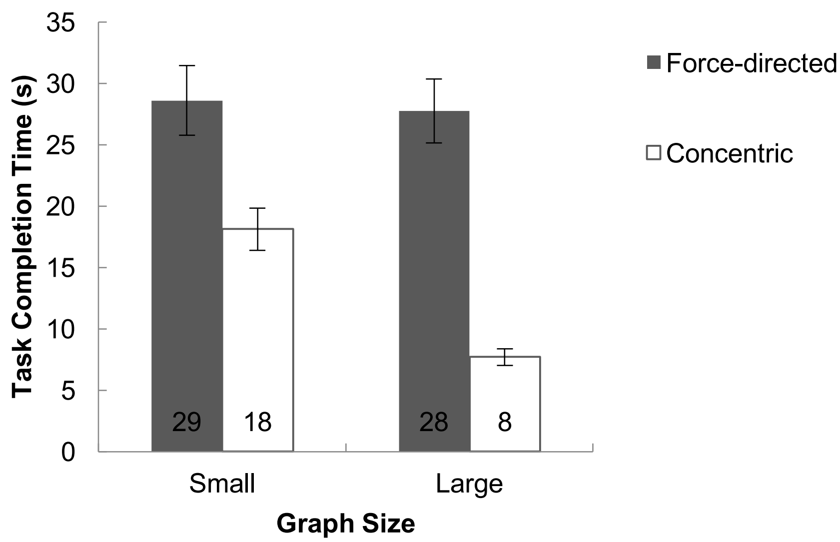


Figure 62. Size and layout effect on task completion time of node-locating tasks of DIA2.

For path-finding task, both main effects (size and layout) are significant on accuracy and task completion time (see Figure 63 and Figure 64). Forced-directed layout had significantly shorter completion time and higher task accuracy compared to concentric layout. This result is also consistent with the findings of experiment with abstract graphs.

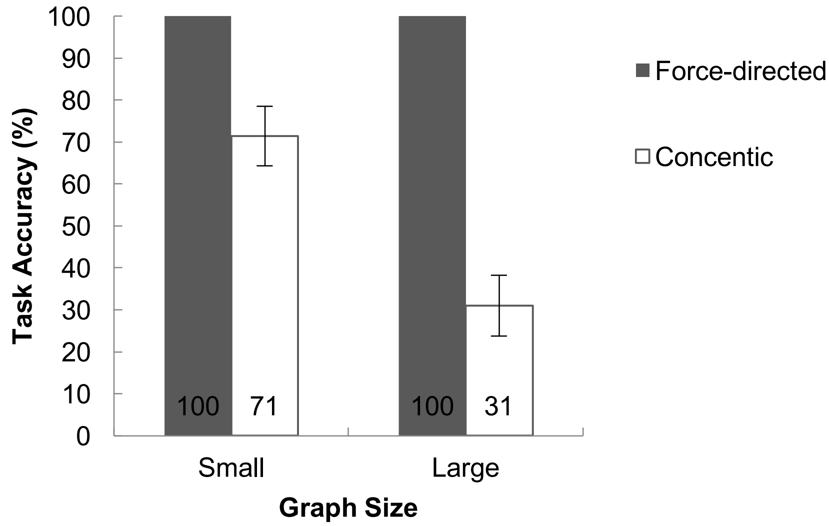


Figure 63. Layout and size effect on accuracy of path-finding task of DIA2.

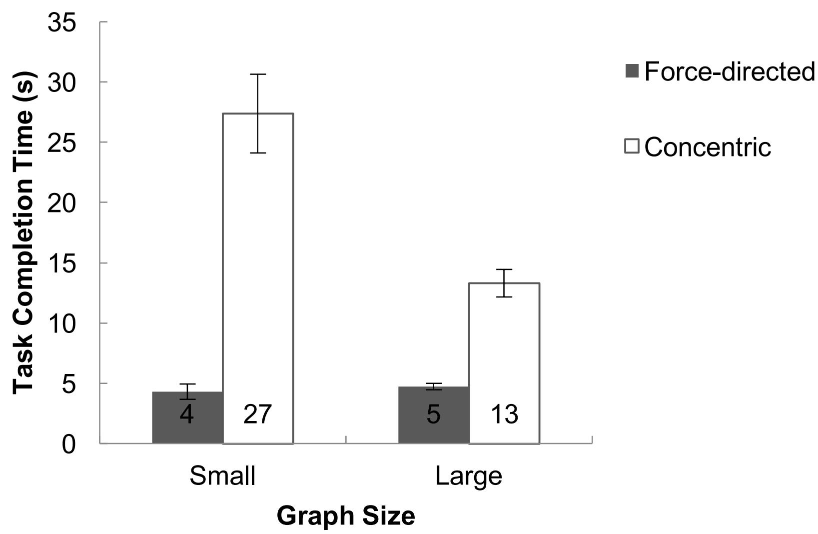


Figure 64. Layout and size effect on task completion time of path-finding task of DIA2.

Error analysis. Because the DIA2 tasks were affected by additional factors associated with the application context, additional analysis was necessary in order to interpret the results. AOIs were designated according to participants' answers and the purpose of analysis of each task. For example, for task 2 (find most connected researchers), participants answers were distributed into three different sub-groups. Three

AOIs were defined accordingly in order to analyze participants' attention transitions between these three subgroups (see Figure 69).

Considering the complexity of the tasks with application context and the objectives of the study, participants' answers were considered correct, as long as they mentioned one of the possible choices (if there were multiple correct answers). For example, for task 1, the right answers included 'the blue dots on that hexagon', 'the yellow dot', 'the group of dots on that hexagon'. Participants' verbal answers were coded for analysis purpose.

Task 1. Participants were asked to find out the most connected researcher(s) in this collaboration visualization. As illustrated in , 6 researchers formed a 'group'. They had the same connection number (5) which is higher than other researchers'. Thus, the correct answer of the task is the researchers within this group. Interestingly, participants tended to give their answer using the color-attributes of the dots. For example, one participant answered by 'the blue dots on that hexagon'. As mentioned before, participants' answers were considered correct as long as they belong to the possible answers. Figure 65 shows the distribution of participants' answers. Other than 'the dots on that hexagon' (coded '1' in Figure 66), correct answers also include 'the blue dots in that group' (coded '2') and 'the yellow dot' (coded '3'). Only one participant (S18) gave the wrong answer (coded '4'). During the post-task interview, he said the reason he chose this answer was that he thought 'color must mean something, and red made this stand out.' His eye tracking data (see Figure 67) showed that this is not a hasty answer: he actually compared it with the right answer (the group of dots coded '1'). Color-coding in the

DIA2 visualizations appears to influence users' decision-making process, but in this case the influence was in contrast with the intended goal.

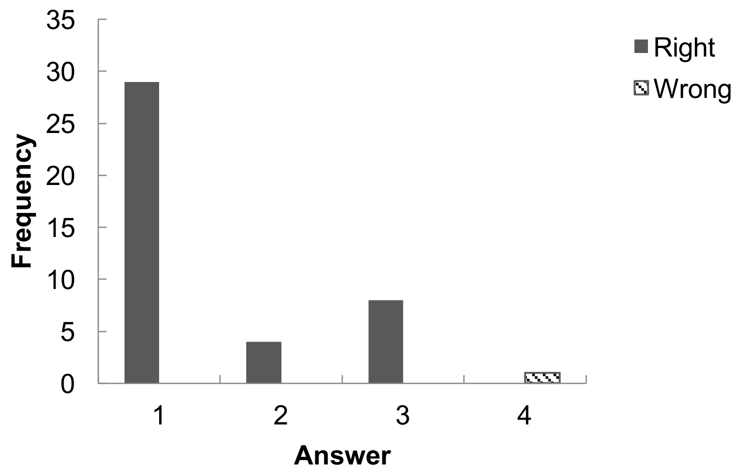


Figure 65. Distribution of participants' answers of DIA2 task 1.

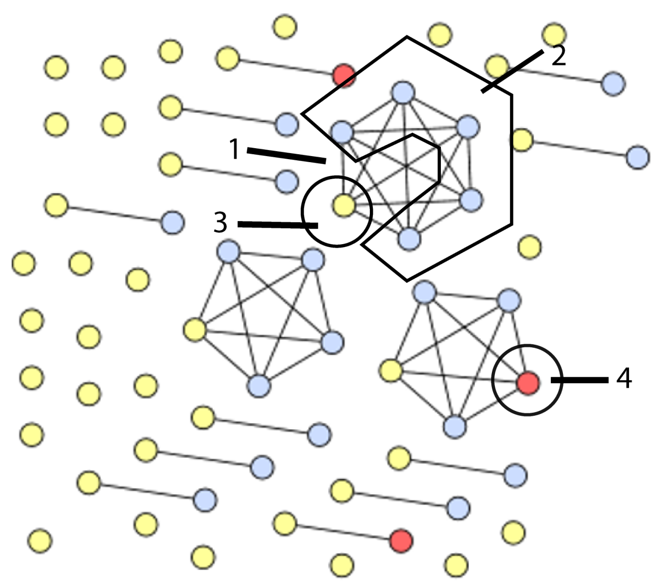


Figure 66. Participants' answers to DIA2 task 1. The answers were marked using different numbers (also the codes used for analysis).

Figure 67. *Scan-path visualization of participant S18's eye tracking data of DIA2 task 1.*

Task 2. demonstrated the five different answers for task 2. As participants' answers spread to the three dense sub-groups, in order to investigate the transitions of participants' attention, three AOIs (T1, T2 and T3 in Figure 69) were defined accordingly. A transitions map visualizing the participants' eye movements shows that attention transitioned between T1 and T3 and from T3 to T2 (but not vice visa). No transitions happened between T1 and T2 (see Figure 69).

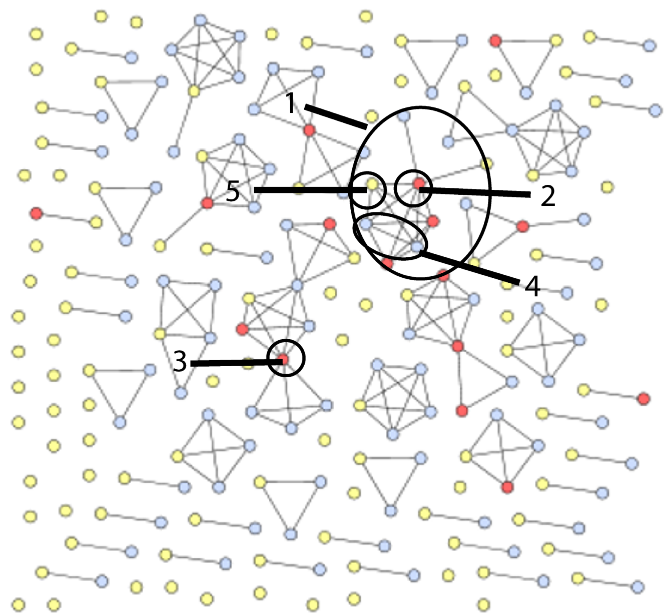


Figure 68. Participants' coded answers to DIA2 task 2.

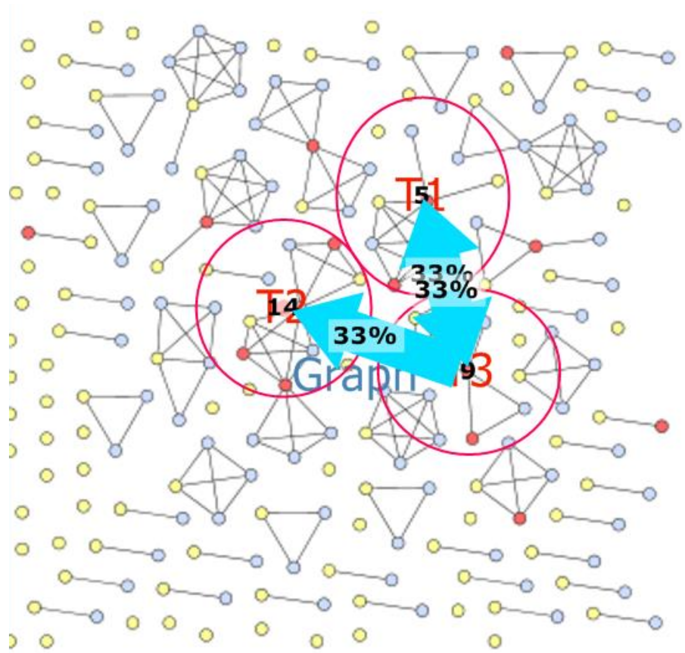


Figure 69. Transitions map of participants' eye movement of DIA2 task 2.

Task 3 & Task 4. These two tasks had 100% accuracy and shortest task completion time (see Figure 58 & Figure 59). Eye tracking analysis shows participants had a very efficienct scan-path on these two tasks.

Task 5. As in the DIA2 task 1, participants' answers were affected by color and their understanding of the context. Six participants answered 'the red dot' (coded '1', see Figure 70) because its color and central location. Heat-map evidence showed that they focused on the center area and just explored within the first level of connections. Their answers were mainly decided by their understanding of the color-coding (see Figure 71). Answer '2' has 20 connections (26 participants gave this answer), which is 3 more than the connections of answer'3' (10 participants gave this answer). The visual difference between '2' and '3' was subtle. Participants who answered '3' mainly focused on it and did not compare it with '1', and they did not count all the connections before they gave their answers (see Figure 72). Participants who gave the correct answer tended to have a more thorough comparison process before giving their answers.

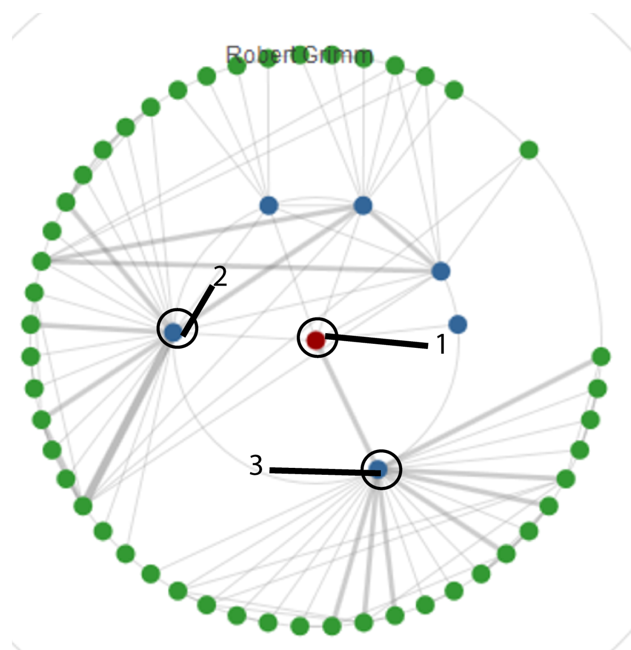


Figure 70. Participants coded answers of DIA2 task 5.

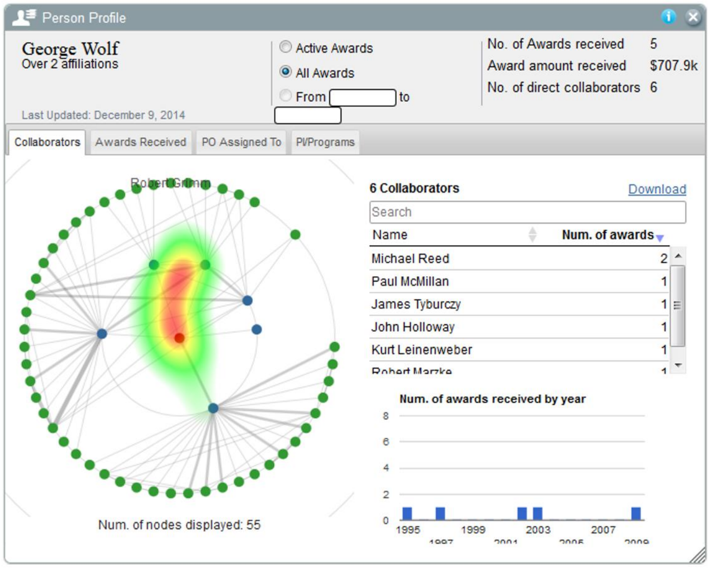
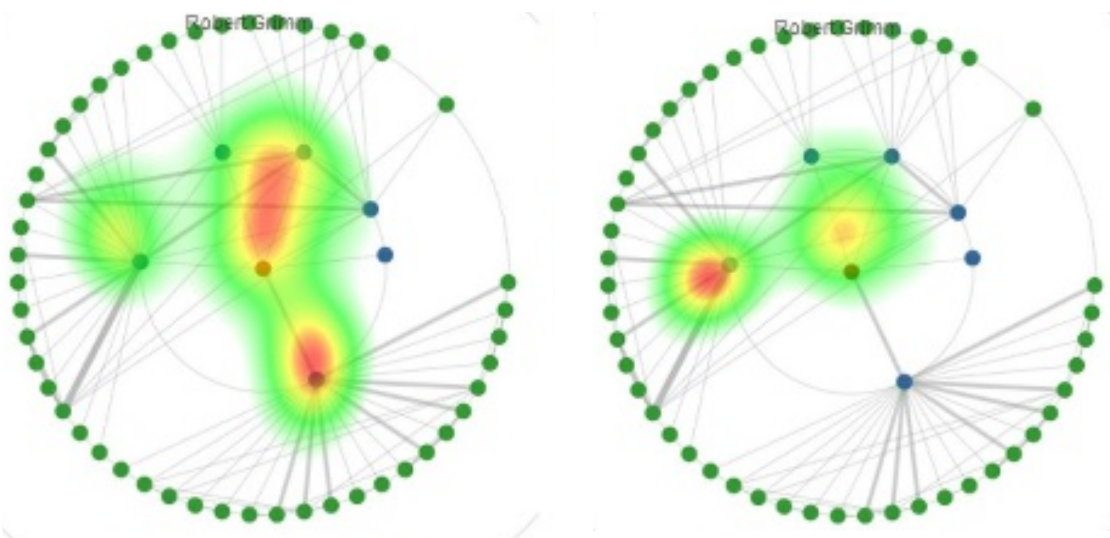


Figure 71. Heat-map visualization of participants' eye movement data that answered '1' to DIA2 task 5.



Heat-map visualization of participants' eye movement data who answered '2' to DIA2 task 5. Note that attention was distributed on both '2' and '3' areas.

Heat-map visualization of participants' eye movement data who answered '3' to DIA2 task 5. Note that attention has just focused on '3' areas.

Figure 72. Heat-map visualization of participants' eye movement data that answered '2' and '3'to DIA2 task 5. The correct answer had a more thorough comparison before decision-making.

Task 6. Most participants gave the correct answer to this task (38 out of 42) within a very short time (see Figure 74). The wrong answers were mainly affected by participants' understanding of the graph's color-coding (code '1' and '3', see Figure 73).

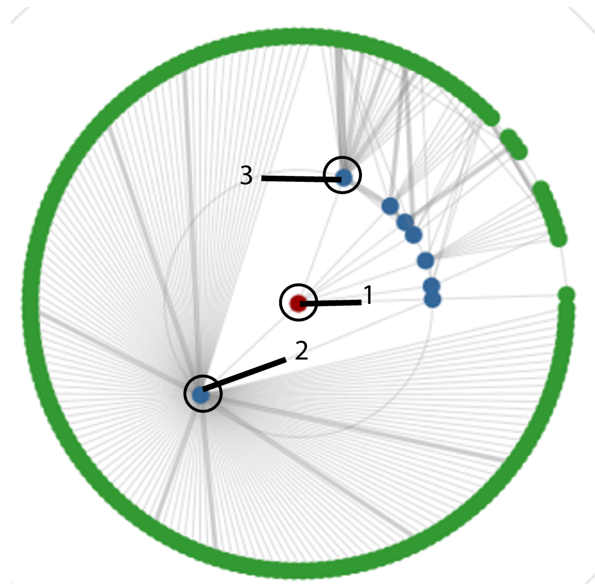


Figure 73. Participants coded answers to DIA2 task 6.

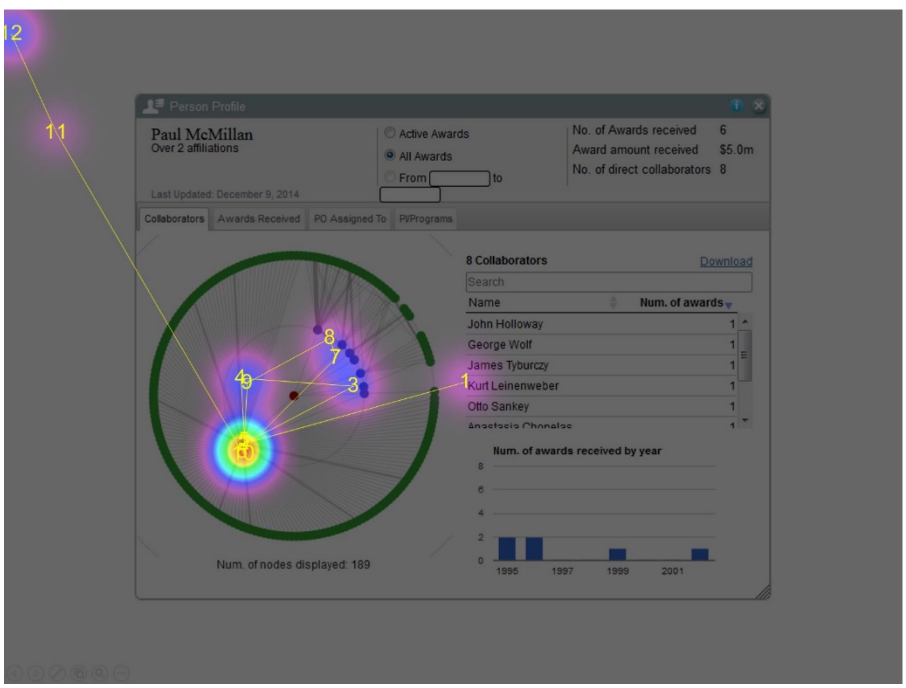


Figure 74. Scan-path visualization of participant S12's eye movement data on DIA2 task 6. Note the efficient pattern of the scan-path.

Task 7. Again, participants' answers were affected by their understanding of the visual attributes assigned to the graph (i.e. the color of the dots and the thickness of the lines): 8 participants answered ‘the path that go through the red dot’ (coded ‘1’, see Figure 75); 2 participants answered ‘the path with the thick line’ (coded ‘2 see Figure 75).

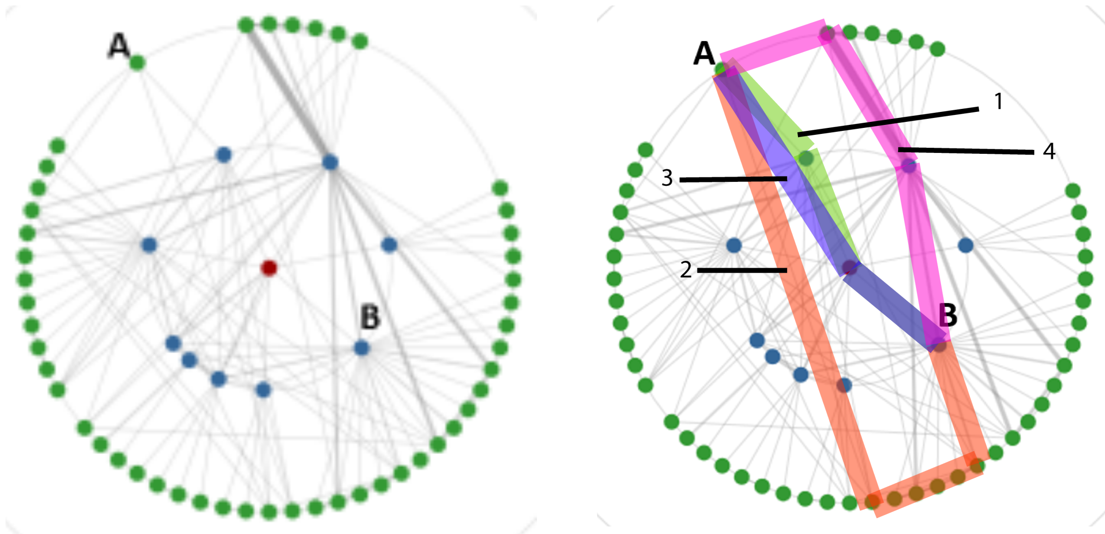


Figure 75. Participants’ coded answers to DIA2 task 7.

Task 8. This proved to be a tricky task because the two end nodes were already connected directly. This task was used to test participants' real understanding of the information. As predicted, most participants (23 out of 42) answered ‘the path goes through the red dot’ (code ‘1’, see Figure 76). Their choices seemed affected by the color-coding of the graph.

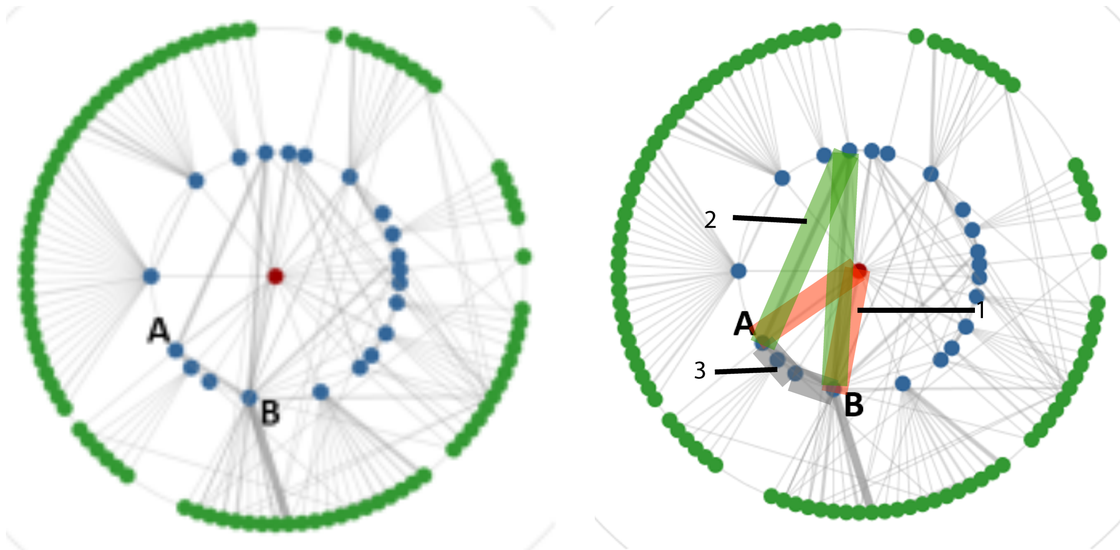


Figure 76. Participants' coded answers to DIA2 task 8.

Summary

The same size and layout effects were observed in DIA2 tasks. For the node-locating task, the performance of concentric and force-directed layouts were consistent with their performance with abstract node-link diagrams: accuracy dropped for force-directed layouts as graph size increased. The concentric was more insensitive to the size effect, even having higher task accuracy with larger graph size (the choice of stimuli for task 6 contributed to this high task accuracy). Eye tracking analysis provided more explanations of the findings:

1. There were more factors that affected participants' decisions in the DIA2 tasks than the controlled abstracted graph experiment. For example, the contextual information such as color-coding made a difference and more 'guessing' was observed in the tasks where contextual information was present.

2. Visual attributes of the graph (context) affected participants' performance on task. Participants usually elaborated their answers using color attributes. For example, one participant gave his answer to task D1 by saying 'the 'red dot' in the middle'. During the after-task interview, participants commented that they thought 'color must mean something', and even used their assumptions about colors to determine their answers. For example, one participant gave his answer to task D1 by saying: 'all the blue dots in that hexagon.' When he was asked why just blue dots, but not the yellow one during the interview, he answered: "I think yellow means bad, so I didn't choose it." The problem here was that the rationale for the color-coding used in the DIA2 visualizations was not evident and therefore, participants guessed, and most of the time, the color-coding led to misunderstandings.
3. In addition to the participants' eye movements on the node-link diagram, the general visual pattern on the whole widget was investigated in this study. The image used for stimuli was divided into four individual AOIs: Title, Graph, Table and Bar. Although participants had been instructed that the answer to the task can be found by just looking at the node-link diagram, participants' eye movements showed they cross-checked between other information areas (see Figure 77). Main transitions happened between the Graph and Table areas, implying that the user assumed a connection, or that the information was related. This finding can inform future development of DIA2 (and visualization designs for other applications). Specifically designers should: 1) provide task-related information and 2) provide the information in a way that

is consistent with users' mental model to help users efficiently and effectively comprehend the graph.

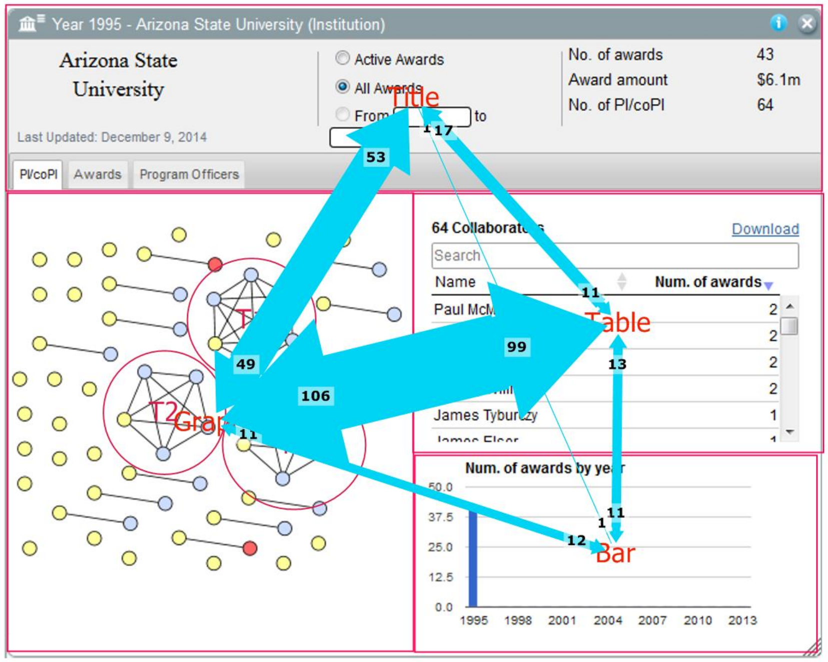


Figure 77. Transitions map visualization of participants' attention transitions between several AOIs.

CHAPTER 6

DISCUSSION & CONCLUSION

The effect of graph size, density, and layout on people's graph reading and comprehension has been investigated under different circumstances (using abstract graphs and a real application DIA2). Eye tracking methodology was used to understand people's cognitive processing which otherwise was difficult to observe using accuracy and completion time measures alone. Combined with pre-defined AOIs, analysis metrics and static visualizations of people's eye movements, eye tracking's utility on human graph reading behavior has been demonstrated.

Layout Effect on Graph Comprehension

Participants' abilities for spatial reasoning proved to be correlated with graph comprehension: participants who had higher scores on the spatial reasoning test tended to have higher task accuracy and shorter task completion times.

The data shows that larger and denser graphs resulted in longer completion times and lower accuracy rates. The heat-map and scan-path visualizations also indicated that participants had more "guessing" answer when reading larger and denser graphs. These findings suggest that denser graphs placed more cognitive demand on participants and made it difficult for them to extract information. Moreover, the findings suggest that specific layout contributed to better accuracy and completion times, even for large and dense graphs. The study found the specific features of the layouts including bend angles, crossing lines, and orientation of nodes influence graph reading behavior that can lead to successful task completion at different size and density levels.

Pre-defined metrics were used to identify different cognitive processes. In this study, the searching process was defined to be finished by locating the target nodes. The searching process was marked by the 'time to the first fixations in target AOI' eye tracking metric. After all elements needed to complete the task had been viewed by participants, the rest of the time needed to complete the task was designated as the comprehension process, during which participants disambiguated and confirmed their choices. This duration was calculated by subtracting the 'time to the first fixations in target AOI (in node-locating task) or second target AOI (in path-finding task)' from task completion time.

For node-locating graphs at size 20, force-directed layouts had a significantly shorter time on the 'comprehension' cognitive process. By examining the heat-map and scan-path visualizations of participants' eye tracking data, it was revealed that participants tended to make quick decisions with graphs in force-directed layout, even though the target might have been totally ignored. This implies that force-directed layouts made the target less 'noticeable' while providing a 'confidence' associated with participants' decision making. The circular layout had a significantly longer 'searching' and 'comprehension' process for node-locating task at graph size 20. Eye tracking analysis showed that the circular layout had a 'visual leading effect': participants tended to visually follow the circle while they were searching for the target. Compared to force-directed layouts, this traversal search behavior appeared to have the effect of making the target more noticeable to participants. However, even if the target was viewed by participants and compared with other nodes, a correct answer was not guaranteed. The dense overlapping made it difficult to find the right answer even with careful reading and repeated checking. Interestingly, the concentric layout was found to be more resistant to the size effect: it had a relatively constant accuracy rate over three graph sizes. At size 20, concentric layout even had the shortest 'target searching' time and highest task accuracy. Eye tracking analysis revealed that participants tended to have a more 'efficient' scan-path with the concentric layout, which seemed to have the effect of making the dense area 'stand out' more.

For path-finding tasks, participants needed to first locate the two designated end nodes and then find the shortest path between them. Eye tracking analysis revealed that the layout effect was significant at graph size 20 on locating the first target, whereas on locating the second target, the same effect was found at size 5. The assumption is that people can read information through parafoveal vision; thus the location of the second target could somehow be read while they were searching for the first target. However, this 'pre-read' advantage was 'diluted' as graph size increased. As a result, the layout effect on time to locate the second target was only significant at graph size five, but not 20. Force-directed layout was found to have the highest accuracy through all three graph-sizes. It also happened to be the most sensitive layout to the effect of graph size. The circular layout had the longest time-on-task and lowest task accuracy. Like the case for the node-locating task, the concentric layout was found insensitive to size. Through error task analysis, it was found that edge crossing and node overlapping caused the most task failures of path-locating tasks. Scan-path analysis of participants' eye movement data revealed clear geometric-path tendency (Huang, 2013): Participants tend to follow the paths that were near the geometric path, but not necessary the target path because edge crossing made the branches along the paths between two end nodes difficult to discern.

During the analysis, several 'out of expectation' results were observed. For example, the significant high accuracy of path-finding tasks using force-directed layout graphs in the pilot study, and the higher accuracy of concentric layout at density level .4 than .2 in the abstract graph experiment. By a closer look at participants' eye movement data, these exceptions were attributed to the randomly picked end nodes that simplified the task— either by centralizing the end nodes or designating the geo-matric path as target path. Presenting the same stimulus to all participants helped the study detect distinct patterns; however, it also limited the findings by adding noise to conclusions about different graph sizes, densities and layouts.

In summary, three layouts were compared by their advantages on accuracy, time size sensitivity, searching process and comprehension process. Red check marks and blue check marks represent the node-locating path-finding tasks, respectively (see Table 24). Force-directed layout out-performed others by its' shorter comprehension time on both N and P tasks. Also, concentric layout should be considered when visualize large datasets, since its relative insensitivity to the size effect.

Table 24.Advantages of different layouts on accuracy, time, size insensitivity, target searching and comprehension.

Node-locating

Path-finding

	Accuracy	Time	Size insensitivity	Searching	Comprehension
Force-directed	✓	✓		✓	✓ ✓
Circular		✓			
Concentric	✓	✓ ✓	✓ ✓	✓	✓

Node-link diagram in real application. DIA2 uses node-link diagrams to visualize collaboration networks. The DIA2 visualizations are consistent with the conventional use of node-link visualizations in current social media platforms (add reference). During a former usability study with DIA2, positive feedback about the use of node-link diagram was mentioned (Molnar et al, 2015). In Molnar et al.' s study, participants showed a good understanding of the information presented by the DIA2 visualizations. That is, most participants understood the graphs as collaboration networks of researchers without providing an explanation other than the general project background.

The same effect of size and layout was replicated with the experiment on DIA2. More factors were observed to have an effect on participants' decisions making in the real application context than the controlled experiment with abstract graphs. For example, more guessing answers based on participants' understanding of the visualization-attributes (e.g., color coding of the nodes and weight assignment of the link) were observed.

Transitions between AOIs other than graph areas implied that participants scan all the information provided even when they had been instructed to focus on a specific area. Furthermore, participants' eye movements demonstrated a frequent cross-checking behavior between the Graph area and the Table area. One explanation for this observation is that in users' mental model, these two kinds of information were highly related. Based on this finding, one suggestion could be made for a DIA2 developer: provide context relevant information in a way that can facilitate users' leveraging of existing mental models would facilitate their comprehension of complex visualizations.

Graph Reading Behavior

Although multiple stages of processing model of people's graph comprehension in general have been previously theorized, this study focused on node-link diagrams specifically. Eye tracking analysis provided detailed evidence regarding people's attention in addition to their behavior (data collected through observation) and attitude (data collected through interview). Static visualizations (heat-map, scan-path, and transitions map) of eye tracking data have been used to qualitatively investigate participants' visual search strategies.

Three-stage processing model. The study used a top-down model to define the eye tracking metrics; however, a bottom-up model of graph reading was developed based on observing participants' behaviors as part of this research study. Through the analysis of the video of participants' eye movement trajectories with path-finding tasks, a three-stage information processing model emerged. The three stages include:

1. Exploring, fixations dispersed to almost all nodes.

2. Comparing, back-and-forth scan-paths from one end-node to compare several possible paths.

3. Confirming, several back-and-forth scan-paths on the selected path to confirm the verbally give the answer.

To find the shortest path between two nodes, participants always started by an exploration of the whole graph to get a general impression. This exploring stage was evidenced by the dispersed fixations to all nodes in the graph. Also, analysis of participants' eye tracking data revealed that there was no significant size effect on the time needed to search for the second target after locating of the first target. One possible explanation for this is that during the exploration stage, the location of the second target had been somewhat noticed by the participants, hence they could conduct relatively faster searching for the second target. This exploring stage corresponded to the searching stage in the top-down model. The comprehension stage of the top-down model has been further refined into two different cognitive stages—comparing and confirming. These two sub-stages was differentiated by their distinct scan-path pattern: during the comparing stage, several branches from both ends nodes were explored as evidenced by the backtracking (regressive saccades) scan-path between the nodes that had multiple connections. Before verbally giving their answer, participants usually conducted several checks to confirm their answer. During this stage, the iterative scans of the candidate path were observed.

Straight line reading tendency. During the analysis of participants' eye movements, it was observed that the perception of angle was used differently for node-locating and path-finding tasks. For node-locating tasks, acute angles helped participants to identify the connections emanating from nodes, whereas for path-finding tasks for which participants tended to follow a straight line to find the path, acute angles made it difficult to notice the existence of a path. Therefore, the design of the angles in a graph should depend on the specific use context.

Seeing without looking. While looking at the comprehension process of path-finding tasks, it was observed that participants did not necessarily need to complete the task after 'looking at' all target nodes. Other than explaining this as a miss-catch of the eye tracker, another explanation was that participants mentally shifted their attention to the nodes without moving their eyes, in other words, they saw without looking. This may be eye tracking evidence of covert attention. How graph layout can facilitate or impede efficient covert attention is worthy of further research.

Eye Tracking Analysis for Visualization Evaluation

Eye tracking is a popular methodology in psychological research as the eye movement can be related to people's cognitive process. However, the application of eye tracking methods in visualization evaluation has been limited. It has been used only since 2003 primarily? to investigate online information retrieval. And in the limited eye tracking studies on visualization evaluation, the subjectively qualitative elaboration based on static visualization of eye tracking data was mainly conducted. The quantitative/objective analysis advantage of rich eye movement data has been under-utilized over the years. One contribution of this thesis is the comprehensive use of eye tracking analysis methods on people's graph reading behavior.

By using fine-grained metrics, complex graph reading processes were separated into interpretable and insightful cognitive stages and compared among different conditions (layout, size, etc.) to yield statistically significant results. Spatial and temporal visualization of eye tracking data were used for an analysis of errors to determine which features could be attributed to the task failure. For instance, this analysis informed whether it was because information had been layout in a confusing way so that people failed to notice it, or was it because the layout was so complicated that people saw it but did not understand the relationship with the task goal.

Lessons have been learned during the analysis of eye movement data. Heat-maps, scan-paths, and transitions maps were applicable depending on different analytic purposes. Heat-maps provide aggregated visualizations of attention for all participants.

They do not include the viewing –sequence dimension, and sometimes they are misleading in that all participants have the same overall visual pattern. Scan-paths present both spatial and temporal information, but these representations suffer from visual clutter when overlaying the data from several participants in a single representation. Transition maps provide a perspective of attention transitions between AOIs, but their interpretability depends on the definition of AOIs.

Limitations of eye tracking method. As recorded every 20-30 ms, the data produced from eye tracking generates huge datasets. Analyzing these data can be tedious and even misleading if not guided by clear assumptions and pre-defined metrics.

Also eye tracking analysis is subject to error and misinterpretation even under careful applications. First, the gaze samples collected at 35-120 Hz are filtered algorithmically to fixations around 3 Hz. The algorithm used to temporally and/or spatially disperse this gaze data is often unreported. Whether there are analytic differences resulting from the use of different dispersion algorithms may need further empirical research. Second, the definition of AOIs will affect the interpretation of participants' comprehension strategies. Several considerations on AOI specification include: the size of the AOI, the padding area to avoid the effect of gaze location error, etc. Third, the selection of metrics, though there are many metrics that can be used for analysis, should depend on the research questions. Fourth, the use of different eye movement data visualizations should also be determined by the research questions. For example, heat-map is more suitable to analyze aggregated behavior (multiple people's attention distribution), whereas scan-path is more suitable for analysis of individual's sequence strategy. Overlaying scan-paths of different participants' scan-path will make the result hard to interpret.

Suggestions for Designers

Based on the findings of this study, several suggestions could be made for algorithm and application designers.

Consider layouts relative to graph size. Force-directed and circular layouts seemed to be more sensitive to the effects related to graph size. The task accuracy of these two layouts dropped dramatically as graph size increased from 10 to 20. Concentric layout was insensitive to size effect. It has a relative constant accuracy across three size levels. Designers should consider this size-sensitivity as they design visualizations at different scales.

Choose reasonable graph size and density. There was a threshold at which participants' patience dropped dramatically and they tended to provide rushed answers with little regard for accuracy. Graph designers should choose a reasonable graph size to visualize their data. When the large size is inevitable, they should try to split the dimensions to be visualized in several graphs.

Use different crossing-angles according to task type. The layout effect was found significant for just path-finding tasks in the abstract graph experiment. This result was similar to Huang's findings (2007b) that the edge crossing affects human performance on path-finding tasks but not node-locating tasks. Acute angle inhibited human performance when path continuity was under concern. For node-locating tasks, acute angles help participants to distinguish the nodes. Designers should choose appropriate visualized angles according to the context of the specific task.

Provide informative but not overload vis-info. Analysis of the attentional transitions between several AOIs of DIA2 tasks revealed that participants' information search strategies were pre-determined by their mental models. Designers should provide information in the way that is consistent with users' mental models to help users efficiently comprehend the graph. For example, the place of related information should facilitate effortless cross-checking.

Future research

Working memory and graph reading. Working memory capacity has been empirically shown to have impact on people' ability to comprehend complex displays. This study confirmed this effect by the high correlation between participants' performance on spatial reasoning test and participants' graph reading performance. This study just used the aggregated score on the four kinds of tasks (perspectives, mirror images, spatial reasoning cubes, and organizing two dimensional shapes). It would be interesting to include more kinds of task types and examine how they impact people's graph reading behavior individually. In other words, examine how graph reading relates to specific abilities.

Experiments in context. The experiment with DIA2 has demonstrated that there are other effects related to details of the design of the graph. Future studies could be designed to look into the effect of syntactical information such as labels, color use, etc.

Visual analytic methods for eye tracking data. Eye tracking has been under-utilized in research evaluating visualizations. The analysis methods used for eye tracking data were mainly focused on qualitative elaboration of the real-time playback of the video of participants' eye movements. These methods are time-consuming and require subjective interpretation. The research advances from the area of data analytics for temporal and spatial datasets could and should be applied to future eye tracking studies on the evaluation of visualization.

REFERENCES

Aaltonen, A., Hyrskykari, A., & Räihä, K.-J. (1998). *101 spots, or how do users read menus? the SIGCHI conference* (pp. 132–139). New York, New York, USA: ACM Press/Addison-Wesley Publishing Co. doi:10.1145/274644.274664

Battista, G. D., Eades, P., Tamassia, R., & Tollis, I. G. (1998). *Graph Drawing: Algorithms for the Visualization of Graphs*. Prentice Hall PTR.

Bennett, C., Ryall, J., Spalteholz, L., & Gooch, A. (2007). The Aesthetics of Graph Visualization. *Computational Aesthetics*.

Brandes, U. (2001). Drawing on Physical Analogies. *Drawing Graphs*, *2025*(Chapter 4), 71–86. doi:10.1007/3-540-44969-8_4

Brandes, U., Kenis, P., & Wagner, D. (2003). Communicating centrality in policy network drawings. *Visualization and Computer Graphics, IEEE Transactions on*, *9*(2), 241–253. doi:10.1109/TVCG.2003.1196010

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. doi:10.1109/TVCG.2011.193

Buswell, G. T. (1935). How people look at pictures.

Byrne, M. D., Anderson, J. R., Douglass, S., & Matessa, M. (1999). *Eye tracking the visual search of click-down menus*. *the SIGCHI conference* (pp. 402–409). New York, New York, USA: ACM. doi:10.1145/302979.303118

COLEMAN, M. K., & PARKER, D. S. (1996). Aesthetics-based Graph Layout for Human Consumption. *Software: Practice and Experience*, *26*(12), 1415–1438. doi:10.1002/(SICI)1097-024X(199612)26:12<1415::AID-SPE69>3.0.CO;2-P

Davidson, R., & Harel, D. (1996). Drawing graphs nicely using simulated annealing. *ACM Transactions on Graphics (TOG)*, *15*(4), 301–331. doi:10.1145/234535.234538

Eye Tracking Methodology. (2007). Eye Tracking Methodology.

Fruchterman, T. M. J., & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience*, *21*(11), 1129–1164. doi:10.1002/spe.4380211102

Gansner, E. R., & North, S. C. (1998). Improved Force-Directed Layouts. *Graph Drawing*, *1547*(Chapter 28), 364–373. doi:10.1007/3-540-37623-2_28

Ghoniem, M., Fekete, J., & Castagliola, P. (1BC). A Comparison of the Readability of Graphs Using Node-Link and Matrix-Based Representations. *IEEE Symposium on Information Visualization*, 17–24. doi:10.1109/INFVIS.2004.1

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

Goldberg, J. H., Stimson, M. J., Lewenstein, M., Scott, N., & Wichansky, A. M. (2002). *Eye tracking in web search tasks: design implications*. *the symposium* (pp. 51–58). New York, New York, USA: ACM. doi:10.1145/507072.507082

Goldberg, J., & Helfman, J. (2011). Eye tracking for visualization evaluation: Reading values on linear versus radial graphs. *Information Visualization*, *10*(3), 182–195. doi:10.1177/1473871611406623

Gutwenger, C., & Mutzel, P. (1998). Planar Polyline Drawings with Good Angular Resolution. *Graph Drawing*, *1547*(Chapter 13), 167–182. doi:10.1007/3-540-37623-2_13

Huang, W. (2007). *Using eye tracking to investigate graph layout effects*. *Asia-Pacific Symposium on Visualisation 2007* (pp. 97–100). IEEE. doi:10.1109/APVIS.2007.329282

Huang, W. (2013). Establishing aesthetics based on human graph reading behavior: two eye tracking studies. *Personal and Ubiquitous Computing*, *17*(1), 93–105. doi:10.1007/s00779-011-0473-2

Huang, W. (2008, October 24). An Eye Tracking Study into the Effects of Graph Layout.

Huang, W., & Eades, P. (2005). How people read graphs (Vol. 45). Presented at the APVis '05: proceedings of the 2005 Asia-Pacific symposium on Information visualisation, Australian Computer Society, Inc.

Huang, W., & Huang, M. (2010). Exploring the relative importance of crossing number and crossing angle. *The 3rd International Symposium*, 10. doi:10.1145/1865841.1865854

Huang, W., Eades, P., & Hong, S. H. (2009). A graph reading behavior: Geodesic-path tendency. *Visualization Symposium*.

Huang, W., Eades, P., Hong, S.-H., & Lin, C.-C. (2010). Improving Force-Directed Graph Drawings by Making Compromises Between Aesthetics. *2010 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, 176–183. doi:10.1109/VLHCC.2010.32

Huang, W., Hong, S.-H., & Eades, P. (2006). Layout Effects on Sociogram Perception, *3843*(Chapter 24), 262–273. doi:10.1007/11618058_24

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*.

Körner, C. (2011). Eye movements reveal distinct search and reasoning processes in comprehension of complex graphs. *Applied Cognitive Psychology*.

Liu, Qing, Vorvoreanu, Mihaela, Madhavan, Krishna, and McKenna, Ann F. (2013). Designing Discovery Experience for Big Data Interaction: A Case of Web-Based Knowledge Mining and Interactive Visualization Platform, *In Design, User Experience, and Usability. Web, Mobile, and Product Design, Lecture Notes in Computer Science*, Springer-Verlag Berlin Heidelberg, v 8015, 543-552. Second International Conference, DUXU 2013, Held as Part of HCI International 2013, Las Vegas, NV, July 21-26, 2013.

Madhavan, Krishna P.C., Vorvoreanu, Michaela, Elmqvist, Niklas, Johri, Aditya, Ramakrishnan, Naren, Wang, Alan G., and McKenna, A. (2012). Portfolio Mining, *IEEE Computer*, October 2012, 45(10), 95-99

Mello-Thoms, C., Nodine, C. F., & Kundel, H. L. (2002). *What attracts the eye to the location of missed and reported breast cancers? the symposium* (pp. 111–117). New York, New York, USA: ACM. doi:10.1145/507072.507095

Molnar, Andreea, McKenna, Ann F., Liu, Qing, Vorvoreanu, Mihaela and Madhavan, Krishna (2015). Using Visualization to Derive Insights from Research Funding Portfolios, *IEEE Computer Graphics & Applications*, May/June 2015, 6-12

Papakostas, A., & Tollis, I. G. (2000). Efficient Orthogonal Drawings of High Degree Graphs. *Algorithmica*, *26*(1), 100–125. doi:10.1007/s004539910006

Pohl, M., Schmitt, M., & Diehl, S. (2009). Comparing the readability of graph layouts using eyetracking and task-oriented analysis, 49–56. doi:10.2312/COMPAESTH/COMPAESTH09/049-056

Purchase, H. (1997). Which aesthetic has the greatest effect on human understanding? *Graph Drawing*, *1353*(Chapter 23), 248–261. doi:10.1007/3-540-63938-1_67

Purchase, H. C. (1998). Performance of layout algorithms: Comprehension, not computation. *Journal of Visual Languages & Computing*.

Purchase, H. C. (2002). Metrics for graph drawing aesthetics. *Journal of Visual Languages & Computing*.

Purchase, H. C., Carrington, D., & Allder, J.-A. (2002). Empirical Evaluation of Aesthetics-based Graph Layout. *Empirical Software Engineering*, *7*(3), 233–255. doi:10.1023/A:1016344215610

Purchase, H. C., Cohen, R. F., & James, M. (1996). Validating graph drawing aesthetics. *Graph Drawing*, *1027*(Chapter 44), 435–446. doi:10.1007/BFb0021827

Purchase, H. C., McGill, M., & Colpoys, L. (2001). Graph drawing aesthetics and the comprehension of UML class diagrams: an empirical study. Presented at the Proceedings of the 2001 ….

Purchase, H. C., Pilcher, C., & Plimmer, B. (2012). Graph Drawing Aesthetics—Created by Users, Not Algorithms. *Visualization and Computer Graphics, IEEE Transactions on*, *18*(1), 81–92. doi:10.1109/TVCG.2010.269

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. doi:10.1037/0033-2909.124.3.372

Rayner, K., Pollatsek, A., Ashby, J., & Clifton, C., Jr. (2012). *Psychology of Reading*. Psychology Press.

Reingold, E. M., & Tilford, J. S. (1981). Tidier Drawings of Trees. *Software Engineering, IEEE Transactions on*, *SE-7*(2), 223–228. doi:10.1109/TSE.1981.234519

Sibert, J. L., Gokturk, M., & Lavine, R. A. (2000). *The reading assistant: eye gaze triggered auditory prompting for reading remediation*. *the 13th annual ACM symposium* (pp. 101–107). New York, New York, USA: ACM. doi:10.1145/354401.354418

Sugiyama, K., Tagawa, S., & Toda, M. (1981). Methods for Visual Understanding of Hierarchical System Structures. *Systems, Man and Cybernetics, IEEE Transactions on*, *11*(2), 109–125. doi:10.1109/TSMC.1981.4308636

Tamassia, R. (1987). On Embedding a Graph in the Grid with the Minimum Number of Bends. *Siam Journal on Computing*, *16*(3), 421–444.

Tamassia, R., Di Battista, G., & Batini, C. (1988). Automatic graph drawing and readability of diagrams. *Systems, Man and Cybernetics, IEEE Transactions on*, *18*(1), 61–79. doi:10.1109/21.87055

Taylor, M., & Rodgers, P. (2005). Applying graphical design techniques to graph visualisation. *... Visualisation*.

Ware, C., Purchase, H., Colpoys, L., & McGill, M. (2002). Cognitive Measurements of Graph Aesthetics. *Information Visualization*, *1*(2), 103–110. doi:10.1057/palgrave.ivs.9500013

Yarbus, A. L., Haigh, B., & Rigss, L. A. (1967). *Eye movements and vision*.

APPENDIX A
[STUDIES ON CRITERIA VALIDATION]

Study	Subject	Independent Variable	Dependent Variable	Stimuli	Task	Conclusion
(Purchase et al., 1996)	49 for dense graph 35 for sparse graph	3 level of bends/crossings/symmetry	accuracy	9 dense graph (16 nodes/28 edges, 3 level of edges bends, 3 level of crossings, and 3 level of symmetry) 9 sparse graph (16 nodes/18 edges)	1. How long is the shortest path between two given nodes? 2. What is the minimum number of nodes that must be removed to disconnect two given nodes? 3. What is the minimum number of edges that must be removed to disconnect two given nodes?	The results indicated that minimizing edge bends and minimizing edge crossings are both important aids to human understanding.
(Purchase, 1997)		Aesthetics (bends, crosses, angles, orthogonality, and symmetry)	1. response time 2. accuracy	10 graphs	1. How long is the shortest path between two given nodes? 2. What is the minimum number of nodes that must be removed to disconnect two given nodes? 3. What is the minimum number of edges that must be removed to disconnect two given nodes?	Edge crossing has most impact on graph understanding.
(Purchase, 1998)	55 computer science students	8 graph drawing algorithms implementing different aesthetics	1. response time 2. accuracy	8 graphs (17 nodes/29 edges) drawn using different algorithms	1. How long is the shortest path between two given nodes? 2. What is the minimum number of nodes that must be removed to disconnect two given nodes? 3. What is the minimum number of edges that must be removed to disconnect two given nodes?	The study concluded that despite different aesthetic bases, algorithms aiming to satisfy one or two aesthetic criteria produce drawings with comparable effectiveness.

Study	Subject	Independent Variable	Dependent Variable	Stimuli	Task	Conclusion
(Ware et al., 2002)	43 students majored in Computer Science and Information Systems	3 level of shortest path, continuity, number of crossing edges on the shortest path, the total number of crossed edges, …	1. response time 2. accuracy	20 graphs	Find the shortest path between two pre-specified nodes	The results suggested that after the length of the path the two most important factors are continuity and edge crossings.
(W. Huang, Hong, & Eades, 2006)	27 postgraduates majored in computer science	2 networks (Krackhardt's advice network: directed graph with 14 nodes and 23 edges vs.fictionalized collaboration network from the former one: undirected graph with 14 nodes and 23 edges) 5 different layouts (circular, hierarchical, radial, group ,free) 2 level of crossing number (minimum vs. many)	1. subjective rating on usability and preference 2. response time 3. accuracy	12 drawings: advice network: 5 layouts * 2 level of crossing number collaboration network: random layout at 2 level	Domain specific tasks 1. Importance: find 3 most important actors and rate them according to their importance levels 2. Presence of social groups: determine how many groups are in the network	No significant effect of layout and edge crossing was found in the study

Study	Subject	Independent Variable	Dependent Variable	Stimuli	Task	Conclusion
(W. Huang, 2008)	22 postgraduate students	7 level of crossing angle (10, 15, 20, 30, 50, 70, 90 degrees)	response time	128 graphs 16*(7+1)	Follow the path from one end to the other and determine the length of the path	The performance of path tracing tasks improves with increases in the crossing angle; the rate of improvement slows down when the angle size is close to 90 degrees
(W. Huang, Eades, Hong, & Lin, 2010)	43 undergraduate students	2 different force-directed algorithms (Classical spring vs. BIGANGLE)	1. response time 2. accuracy 3. mental effort	100 pairs randomly generated using BIGANGLE and Classical	To find the shortest path between two nodes.	The study concluded that effectiveness can be improved when algorithms are designed by making compromises between aesthetics, rather than trying to satisfy one or two of them to the fullest.
(W. Huang & Huang, 2010)	32 students	crossing number and crossing angle	1. response time 2. accuracy 3. mental effort (subjective rating using 9-point scale) 4. visualization efficiency	200 graphs with size between 10 and 50, using spring algorithm.	Find the shortest path between two pre-specified nodes	2 IV together explained 33% of variance, with about 38% of the explained variance being attributed to the crossing angle.

Study	Subject	Independent Variable	Dependent Variable	Stimuli	Task	Conclusion
(Ghoniem, Fekete, & Castagliola, 1BC)	36 subjects consisted of post-graduate students and confirmed researchers in the fields of computer science	2 different layouts (node-link vs. matrix-based) 3 different sizes (20/50/100 vertices) 3 different densities (0.2/0.4/0.6)	1. response time 2. accuracy		1. Approximate estimation of the number of nodes in the graph, referred to as "nodeCount". 2. Approximate estimation of the number of links in the graph, referred to as "edgeCount". 3. Finding the most connected node, referred to as "mostConnected". 4. Finding a node given its label, referred to as "find- Node". 5. Finding a link between two specified nodes, referred to as "findLink". 6. Finding a common neighbor between two specified nodes, referred to as "findNeighbor". 7. Finding a path between two nodes, referred to as "findPath".	The results favored the node-link graph on path-finding task. When graphs were bigger than twenty vertices, the matrix-based visualization performed better than node-link diagrams on most tasks.
(W. Huang & Eades, 2005)*	13 postgraduates	2 different layouts (radial layout vs. circle layout) 2 different aesthetic focus (with cross on the shortest path vs. without cross on the shortest path) 3 different graph (with context info: social network data: 11 nodes/15 edges; 9 nodes/13 edges; 10 nodes/14 edges)	1. response time 2. accuracy 3. eye movement data (video data only)	12 graphs with context info	Tasks with context info 1. "What is the separation level between the two highlighted families?" (shortest path) 2. "Do the two highlighted managers have Friend's friend relationship?	Edge crossing introduce extra eye movement and confusion

Study	Subject	Independent Variable	Dependent Variable	Stimuli	Task	Conclusion
(W. Huang, 2007)*	16 participants	to test the effect of edge crossing: 3 crossing condition (no crossings; nearly-90-degree crossings; small-angle crossings) to replicate geometric-path tendency 3 condition (distance from the geometric path)	1. response time 2. accuracy 3. eye movement data (video data only)	6 graphs in total 3 graphs for E1 (32 nodes/ 43 edges) 3 graphs for E2 (20 nodes/32 edges)	1. Path task: find the shortest path between two highlighted nodes; 2. node+path task: find the shortest path between two nodes which are not highlighted; 3. node task: find the most connected node	Small angles can slow down and trigger extra eye movements, causing delays for path search tasks, whereas crossings have little impact on node locating tasks.
(W. Huang, Eades, & Hong, 2009)*	13 students/33 students		1. response time 2. accuracy 3. eye movement data (video data only)	24 drawings (2 nodes are highlighted, 2 paths between these two nodes, no branches on each path to avoid confounding effects of path continuity)	1. findpath: find a shortest path between two nodes	Experiment 1 validated geodesic-path tendency; Experiment 2 found when graphs are drawn with branch links on the path leading away from the target node, graph-reading performance can be significantly improved.
(Pohl, Schmitt, & Diehl, 2009)*	5 for pre-study, 36 for main study	3 different layouts (force-directed, hierarchical, and orthogonal) 3 graph sizes (10, 15 and 20 nodes with an average degree of 3)	1. response time 2. accuracy 3. eye movement data (heatmap visualization)	9 graphs (3 layouts * 3 sizes)	1. warm up: identify the existence of the node with a given label; 2. find path between two nodes; 3. find isomorphic sub graphs 4. find 4-clique nodes 5. find most connected nodes	Force-directed layout outperformed the other layouts for most tasks, while for the rest tasks all three layouts performed equally well.

Study	Subject	Independent Variable	Dependent Variable	Stimuli	Task	Conclusion
(Körner, 2011)*	12 students	2 planarity (non-crossed vs. crossed) 2 slopes (upright vs. slanted) 2 levels (horizontal vs. non-horizontal)	1. response time 2. eye movement data	8 hierarchical graphs (with context info: computer file system) 9 nodes/10 edges		Distinct search and reasoning processes were identified; Search was mostly unaffected by graph properties; Reasoning was affected by edge crossing; The layout and linguistic factors affect comprehension.
(Burch, Konevtsova, Heinrich, Hoeferlin, & Weiskopf, 2011)*	38 students	3 tree layouts (traditional, orthogonal, and radial) 4 orientations (just for traditional and orthogonal layouts) 3 level of marked nodes (3, 6, and 9 nodes)	1. response time 2. accuracy 3. eye movement data (heatmaps, scanpath, and AOI)	9 graphs (radial layout, and 4 orientations of orthogonal and traditional layouts)	Find the least common ancestor	Traditional and orthogonal layouts significantly outperform radial layouts; By analyzing the scan-path of eye movement data, the authors found that participants cross-checked more often in the radial than in the non-radial layouts.