A Study of Mental Map in Immersive Network Visualization

By

JOSEPH THOMAS KOTLAREK

THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

in

Computer Science

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Kwan-Liu Ma, Chair

Michael P. Neff

Joshua A. McCoy

Committee in Charge

2019

Contents

List of Figures

List of Tables

Abstract

A Study of Mental Map in Immersive Network Visualization

Mental maps represent how viewers store spatial and structural information in their mind. The quality and detail of a mental map is critical for viewers to get a better understanding of the data from a visualization presented to them. For graph visualization, where the layout of connected nodes constitutes the "structure" of the data, facilitating the creation of mental maps is thus important. To investigate the ability of users to create and maintain mental maps of network structures, we design and perform an experiment comparing the mental map creation and graph exploration with two different visualization methods: Traditional 2D and Immersive 3D. We find that mental map formation is easily disturbed by factors present in immersive visualization, though it may support an improved perceived user experience and strengths in other kinds of tasks.

Acknowledgments

This research is supported in part by the U.S. National Science Foundation through grant IIS-1528203 and the U.S. Department of Energy through grant DE-SC0012610.

Chapter 1

Introduction

Graphs are commonly used to represent relational data found in applications from social networks and biological networks to communication networks and power grids. Studying graph data is often done by visualizing graphs as node-link diagrams laid out two-dimensionally on a 2D display. 3D graph visualization is rarely used in practice due to its inherent occlusion problem. However, as Virtual Reality (VR) Head-Mounted Displays (HMDs) become increasingly high-resolution and affordable, it is worth reassessing the value of 3D graph visualization viewed with a VR HMD.

Stereoscopic displays allow users to see depth more naturally and effectively, while more advanced HMDs also have spatial tracking that allow them to partially or entirely negate the adverse effects of occlusion and foreshortening that can occur in 3D space. An effective graph visualization makes the graph's relational and structural information of interest more accessible and memorable. Would stereoscopic 3D visualization better facilitate perceiving such information? While occlusion is a concern, among others, a number of studies have shown that the additional visual channel (depth) can be used effectively for abstract data [21]. Significant research has been done surrounding graph layout techniques that leverage a third spatial dimension [18, 19, 21]. Adding a third spatial dimension with immersive technology not only permits real-world objects to be modeled more accurately, but also allows more freedom and depth (literally) in the representation of abstract objects.

Immersive visualization and analytics is an emerging field that aims to leverage the many benefits of mixed reality technology to information visualization, such as *spatial immersion*, *multi-sensory presentation*, and *increased engagement* [11]. The effectiveness of the immersion provided by VR has been explored in a wide but sometimes sparse breadth, in terms of what techniques are used to immerse users, and in what way it may elicit some positive gain. Our study aims to gain a better understanding of how graph visualization may benefit from immersive technology using HMDs. We begin by experimentally studying how users of 2D graph visualization versus stereoscopic 3D visualization may perceive information and perform tasks differently. In particular, we are interested in learning if users create and maintain mental maps of network structures, and how the mental maps could assist their tasks. Mental maps represent how viewers store spatial and structural information in their mind. The quality and detail of a mental map is critical for viewers to get a good understanding of the data from a visualization presented to them. A person who views a graph representation of a network forms a mental map over time that consists of information about topology, relative positions, and relative directions of elements in the graph [22]. We have designed an experiment to help characterize and understand the mental map created over immersive visualization. We find that mental map formation is easily disturbed by factors present in immersive visualization, though it may support an improved perceived user experience and strengths in other kinds of tasks.

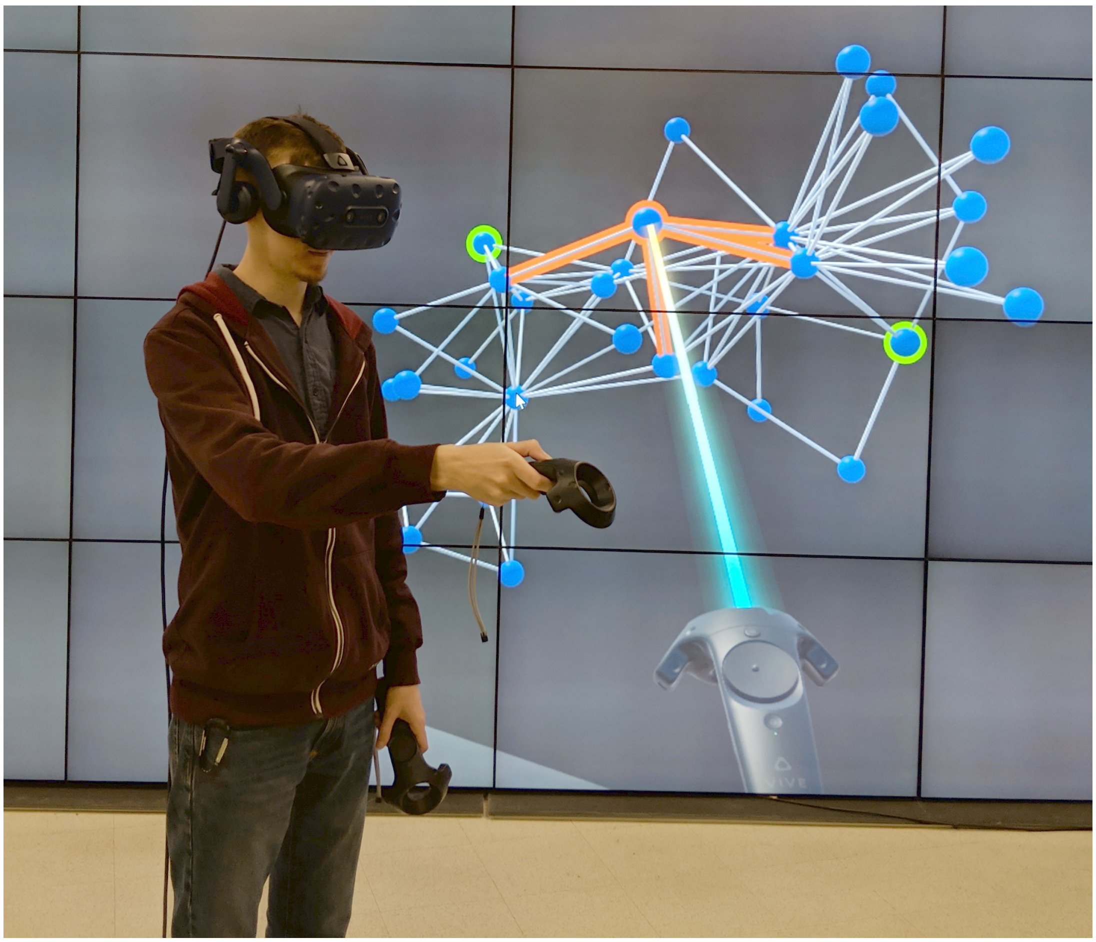


Figure 1.1. A user interacts with a 3D graph with the HTV Vive Pro HMD and controllers. The display mirrors the user's view. We compare this method of graph visualization with traditional desktop 2D graph visualization to determine how it effects users' mental maps.

Chapter 2

Background

Immersive environments, such as CAVEs and HMDs, have proven effective for information visualization and data analysis. Research by Ball and North [5] shows that high resolution tiled displays improve perception and navigation for visual tasks. Mania and Chalmers [20] (2001) studied memory in immersive and non-immersive spaces and found that immersion significantly improved recall for simple memory tasks. Krokos et al. [17] show more recently (2018) that recall can be improved with VR through techniques such as a memory palace metaphor, involving the virtual arrangement of objects to represent memories in a virtual environment. Krokos et al. attribute the success of the immersive memory palace in part to the sense of *presence* provided by immersion, further suggesting that recall should be improved in any suitably immersive experience, not just those involving mnemonic devices such as the memory palace metaphor [17]. Additionally, a study by Kwon et al. [19] shows that immersive graph visualization can clearly improve user performance on simple graph interpretation tasks. While immersive technology may have been eschewed in the past, studies such as these, in tandem with improvements in stereoscopic displays, mark the clear entrance of immersive visualization into the realm of practicality [16].

While keyboard and mouse are the ubiquitous standard for desktop interaction, recent studies show intuitive interaction in virtual reality environments can outperform traditional interaction systems, such as the work by Huang et al. [16] to create a VR gesture system for graph visualization. This is no surprise, as Büschel et al. [7] state, intuitive and low-effort interaction is key to leveraging the benefits of immersion, and keeping users invested in the environment. With different methods of interaction, however, different methods of visualization are also in order.

There has been a significant amount of research adapting virtual reality and other immersive technology to applications in a wide breadth of fields, including biomedical imaging [27], scientific visualization [28, 6, 9], education [31, 26] and collaboration [8]. These studies primarily focused on the application in their respective fields. While immersive scientific visualization was quick to establish itself, the impact of immersion on abstract data visualization remains largely unexplored. Indeed, very few studies have considered the mental map in immersive graph visualization. In the past year, however, more research into immersive visualizations with abstract data has been completed. Drogemuller et al. [10] evaluate navigation techniques for 3D graph visualizations in virtual reality. Greffard et al. [14] introduced an immersive visualization designed to preserve the mental map. The work in this paper differs from the works above in that we instead investigate the impact of immersion on the mental map.

Mental maps are typically used to measure the quality of a dynamic graph layout [1, 24, 23, 22], and the importance of mental map preservation in dynamic layouts has been investigated by several studies [24, 25]. Previous work by Archambault and Purchase [2] investigates mental map preservation in a traditional (non-immersive) environment to show it can help users orientation with tasks such as location and path finding. Herman et al. [15] emphasize the importance of considering *predictability* which is also referred to as *preserving the mental map* in dynamic graph layouts. As mentioned above, immersive environments also improve navigation and orientation [5, 8], but no study has yet combined these techniques.

Chapter 3

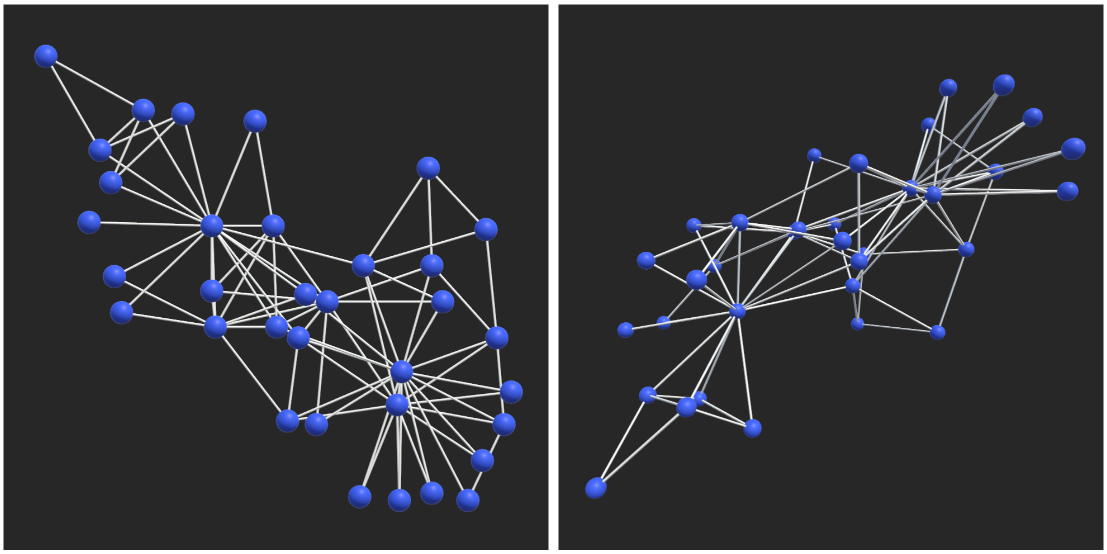
Considerations for Immersive Network Visualization

If our aim is to investigate the potential differences in mental map formation between these two visualization conditions (traditional and immersive), we need to take several aspects of design into careful consideration. That is, we define the mental map and how we can discover it. We determine how to lay out the network data being visualized, to maintain consistency between conditions. Finally, we design interaction to be simple and intuitive for both conditions.

3.1 Mental Map

To discover the mental map, past research has taken several different approaches. Tasks used in mental map discovery and preservation studies are classified into three different categories by Archambault and Purchase [3] as either *interpretation*, *memory*, or *change* tasks. These broad categories are outlined by Archambault and Purchase as follows. An interpretation task asks a question that requires the user to look at the graph and understand the structure in some way. An example of this would be to ask the user about node degree or paths through the graph. A memory task requires the user to recall information about the graph after viewing the visualization, for example redrawing the graph on paper or recalling if the graph is the same as another the user has seen. A change task asks how the graph changes over time, such as changes in degree, or changes in overall size.

In short, in order to discover the mental map, we want to test users' spatial memory and accuracy. The methods we elected to use to achieve this goal are described in the evaluation section.



2D 3D

Figure 3.1. Network layout comparison. On the left is a 2D layout, to be used with a traditional desktop display. On the right is a 3D layout of the same data, for use in an immersive environment. This is the dataset (D0) used for training participants.

3.2 Graph Layout

To earnestly discern the mental map of both immersive and traditional graph visualizations, we should implement the best case of both conditions. We limit ourselves to ball-stick representations, and use static Fruchterman-Reigngold [12] force-directed computed layouts for both, which is shown to be an effective method of graph representation [13, 29]. We do not use an egocentric layout for the immersive environment, such as the layout method presented by Kwon et al. [18], in order to keep as much in common between the two representations as possible. It is possible that an egocentric layout would significantly affect mental map formation, but it would also introduce another variable beyond the depth dimension. What does differ between the traditional and immersive layouts is the use of the third dimension in the immersive environments. Spatially-tracked immersive displays (such as the HTC Vive) allow us to take advantage of this dimension in a way that a traditional display cannot [30]. A side-by-side comparison of the layouts is shown in Fig. 3.1. Also visible in Fig. 3.1 is how the nodes and links use the same geometry and graphical shaders to minimize the number of differences between the two implementations. It is important to note that the 2D and 3D layouts are generated separately, i.e. the 2D layout is not simply a projection of the 3D layout.

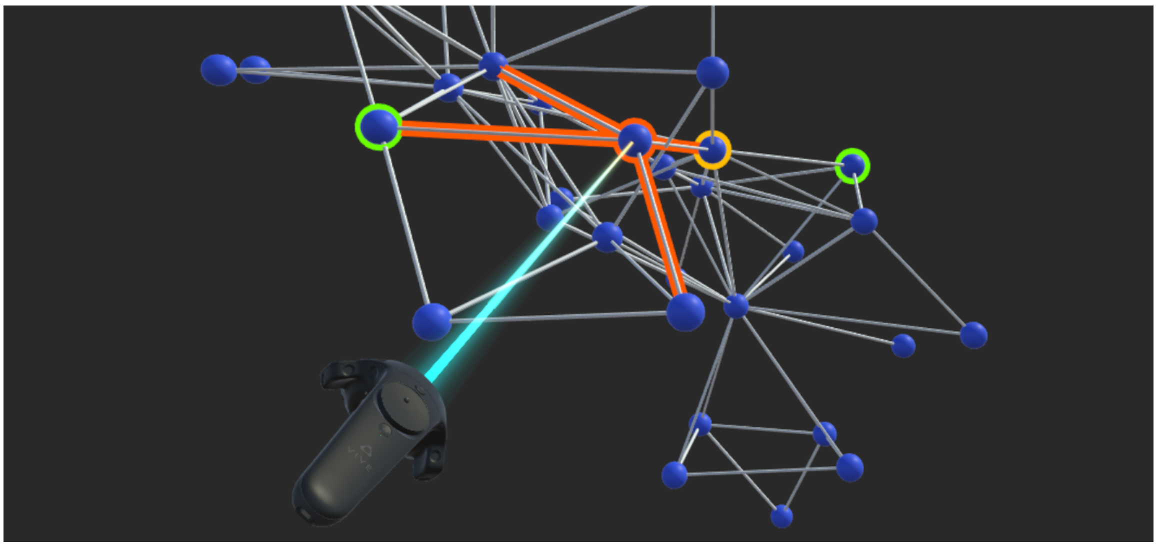


Figure 3.2. Network with selection process shown. The green nodes (far left, far right) are highlighted by the system to show task information to the user. The yellow node (middle-right) is a node selected by the user. The orange node (middle-left) is being hovered over by the user, ready to be selected. Edges connected to a node are also highlighted while the user hovers over it with the pointer.

3.3 Interactions

To complete the tasks in our user study, participants need to select nodes and we use that to infer the presence and quality of a mental map. To do this, we need to allow users to select nodes and provide some helpful information to them in the process. This is shown in Fig. 3.2.

While *selection* in a traditional 2D display with keyboard and mouse is a well established and straight-forward interaction, most commonly implemented as a left-click with the mouse, choosing a selection method for an immersive system is much more nuanced. Since motion tracking is a common feature of commercially available stereoscopic displays such as the HTC Vive and Oculus Rift, we leverage the ability to select objects by "pointing" with a tracked controller. With a virtual laser pointer, users of our system are able to select nodes by pressing the touch-sensitive pad. This virtual laser pointer technique is one of the most common ways to implement selection in virtual reality, and it was chosen for its accessibility and simplicity for most people with any amount of experience in VR [4]. Our laser pointer implementation does have two quality-of-life improvements to make the experience smoother. First, in order to reduce clutter in the visualization, the laser pointer is only visible while the user touches lightly on the touch-pad and a selection is counted when the user presses until an audible "click" is heard. Second, there is a built-in tolerance for selecting distant objects. This helps significantly, as it is possible with our system to have nodes that take up less than half a degree of the user's field of view, making them very difficult to select with precision. In our pilot study, we found that a cone of approximately two degrees at the apex made selection of distant objects much easier without causing the opposite problem of unintentionally selecting objects far from the pointer.

To match the natural ability to *scale* the graph with the mouse scroll wheel in the traditional display, we implement a pinch-zoom technique for the immersive version. To activate the pinch-zoom, the user must press a button on both tracked hand-held controllers, then stretch or contract the distance between the controllers to scale the graph by an equivalent factor. Again, this choice is motivated mostly by intuitiveness and popularity, as ideally users should have to learn as little about VR as they would have to about the traditional setup. Unfortunately, due to the relatively recent development of immersive technology, it is nearly impossible to approach the intuitiveness for users whose conception of the human-computer interface has been defined by the mouse and keyboard.

Since it would not make sense to introduce *rotation* to the 2D display, we also exclude this interaction from the VR version of the system. We felt it would potentially affect the results in an unfair way to add an additional control that users must learn for the VR version. However, the ability to look at the graph from different directions is necessary with the 3D layout to counteract possible occlusion issues. For this, we leverage the physical tracked space allowed by the immersive technology we use. By walking around the space, users can view the graph from any angle, including from within the graph itself, though they cannot rotate the graph inside the environment.

Chapter 4

Evaluation: Discovering the Mental Map

The main purpose of our study is to investigate the differences of *mental map (MM)* creation and preservation in *immersive virtual environments (IVE)* with graph visualization compared to standard desktop environments. To this end, we conduct a user study with 20 participants, each given three tasks to perform with each of the two visualization conditions.

4.1 Experiment Design

Our study is designed as a within-subjects experiment: *2 visualization conditions* $\times$ *3 graph sizes* $\times$ *3 tasks*. We evaluate three dependent variables in the study: task completion time, correctness rate, and number of interactions. *Task completion time* is counted from start to finish of each task, not including the participants time to read the description and learn the task. *Correctness rate* is the percentage of tasks correctly completed. *Number of interactions* is counted as the number of node highlights, selections, and manipulations made during a task.

4.2 Visualization Conditions

We consider two visualization conditions:

C1: Traditional (Desktop) Display with 2D graph layout. This condition uses a mouse for interaction and allows users to highlight nodes, select nodes, and navigate (pan and zoom) the graph.

C2: Immersive (VR) Display with 3D graph layout. This condition uses the HTC Vive Pro HMD with room-scale tracking, and controllers for interaction.

Both conditions use the same shaded blue sphere for all non-highlighted nodes.

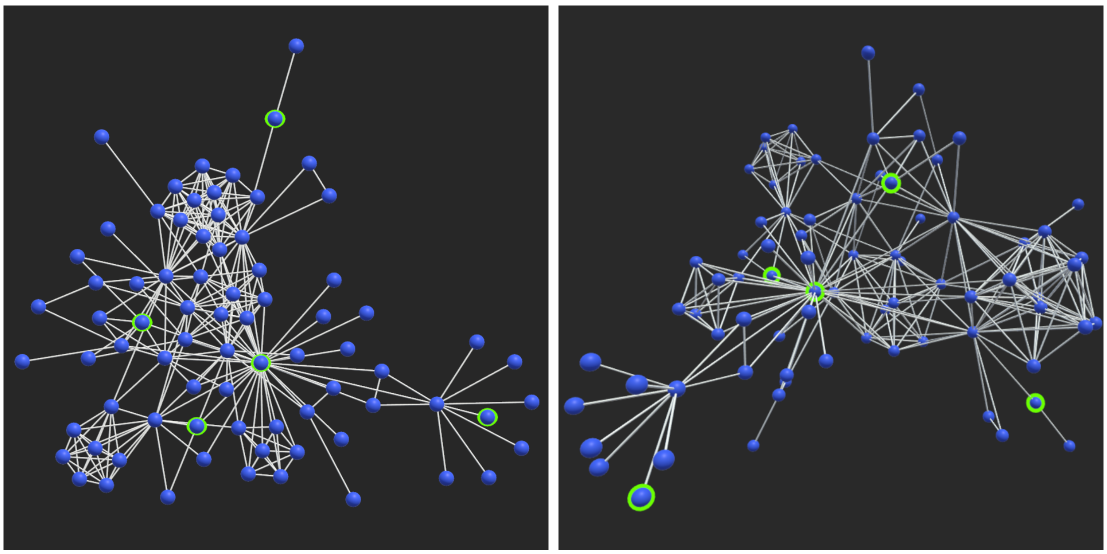


Figure 4.1. Example graph layouts with five highlighted nodes, for the recall nodes task (T2). Both images show graph D1, but the left uses a 2D layout, while the right uses 3D. For the recall nodes task (T2), the user sees this for 30 seconds, then they are asked which nodes were highlighted.

4.3 Tasks

We use three tasks in the study, which have been used to test participants formation of mental maps in graph visualization [1]. The participants are given a written description of the objective at the start of each task. The tasks are presented in the following order:

T1: Find a path. *Highlight nodes that form the shortest path from node A to Node B.* The participant is shown a graph with two pre-highlighted nodes. The goal of the task is to find a shortest path between the pre-highlighted nodes by selecting a set of nodes that forms the shortest path. If there are multiple shortest paths, the participant only needs to find one. Figure 3.2 also shows an example graph for this task.

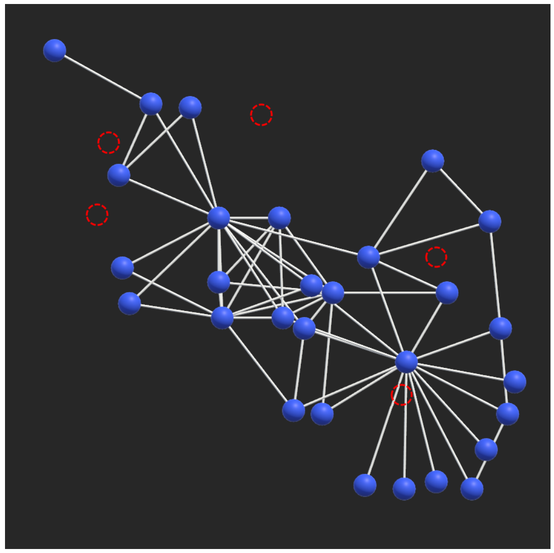


Figure 4.2. An example graph with five missing nodes (red circles), for the find differences task (T3). The red circles are not shown in the visualizations used in the study. The user sees this for 30 seconds, then they are asked which nodes were missing when shown the same graph without any missing nodes.

T2: Recall marked nodes. *Select the 5 nodes that were previously highlighted by the program.* Participants are shown a graph with 5 nodes pre-highlighted, which they can explore for 30 seconds. The graph is then removed, and the participants are shown a blank screen for 10 seconds. The graph is replaced in the same orientation, and participants must select the nodes that were previously marked by the program (Fig. 4.1).

T3: Find the differences. *Select nodes that were not a part of the original graph.* Participants are shown a graph that they may explore for 30 seconds, then they are shown a blank screen for 1 second. After this time, they are shown a graph with the same layout except for 5 additional nodes that they must identify as new. A wrong selection is weighted more heavily based on the distance from the nearest additional node that was not properly chosen (Fig. 4.2).

4.4 Network Data

We use three different graphs with different data sizes (i.e., number of nodes and edges) with one additional graph that was used for the training session. For C1, the graph is laid out to utilize the majority of a 27” display, with 16:9 aspect ratio. For C2, the graph is initially scaled to fill a one meter cube and placed in front of the user in the room-scale area. The data sets are:

- D0: Zacharys Karate Club, this graph consists of 34 nodes and 78 edges. This graph is used in the training session.
- D1: Les Miserables, the *small* graph, consists of 77 nodes and 254 edges.
- D2: Network Science, the *medium* graph, consists of 379 nodes and 914 edges. We use the largest component of the full network.
- D3: Power Grid, the *large* graph consists of 4,941 nodes and 6,594 edges.

4.5 Participants

We recruited 20 participants (9 males, 11 females) for our user study. The mean age of participants was 24.35, ranging from 18 to 34 years. Every participant was familiar with the concepts of virtual reality, and 16 participants had used a virtual reality device before this study. Of those 16, only two participants said they had used virtual reality devices extensively. Additionally, every participant was familiar with concept of a network, and 14 said they were familiar or experienced with network data structures. Four participants were in an undergraduate degree program, while the remaining 16 all possessed undergraduate degrees. 14 of the 16 were also pursuing postgraduate degrees.

Five participants had normal vision, while 15 had corrected vision. Of those 15, 13 wore glasses, and all but one were able to wear their glasses comfortably within the HTC Vive Pro HMD. The one participant that had to remove their glasses reported that they were still able to read text and see the nodes/edges clearly in the immersive environment. One participant had deuteranopic vision, but also reported that all colors used in the study were easily distinguishable.

4.6 Implementation

Both visualization conditions are implemented as Unity3D applications. Participants use the HTC Vive Pro HMD for the immersive portions of the study, and are instructed to stand in the middle of the room-scale environment at the start of each task, facing toward the area where the networks would be presented. The HTC Vive Pro HMD has a 2,880 $\times$ 1,600 px (1,440 $\times$ 1,600 px per eye) AMOLED display with a 90Hz refresh rate. The immersive environment was driven by a desktop computer with an Intel i7 6900K CPU and dual (SLI enabled) NVIDIA GeForce GTX 1080 GPUs. The environment was consistently rendered at 90 frames per second.

The tracked physical space for room-scale configuration measures 3.0 meters by 3.1 meters. The HTC Vive lighthouses are positioned 3.2 meters off the ground. The virtual environment is a simple space with a floor indicating the boundary of where users can walk. The immersive space and the traditional 2D display condition use the same dark gray background.

4.7 Procedure

After ensuring participants are aware of possible VR/HMD issues such as sickness or disorientation, we adjust the HMD fit and interpupillary distance (IPD) for each participant to provide optimal viewing condition. Participants were allowed as much time as necessary to ensure the HMD was fit comfortably without issue. None of the participants experienced any issue with the HMD.

Participants completed a pre-study questionnaire, followed by training, the full experiment, and a post-study questionnaire.

4.7.1 Questionnaire

All participants answered a two-part questionnaire. The first part covered participant demographics, including age, gender, education, colorblindness, perceived spatial reasoning skills, and VR/Visualization experience levels. The second part covered perceived task difficulty and the impact on graph size on task difficulty, as well as free response questions about participant preference and task completion strategies. These questions were based on the NASA-TLX and used seven point Likert scales.

4.7.2 Training

Participants were allowed up to 10 minutes to familiarize themselves with the visualization conditions and their respective interfaces, in a guided tour of the features of both applications. Each participant received training before performing each task for the first time, using a separate dataset. This training consisted of performing the same set of tasks of equal difficulty, with the correct answer available at the end. No data is used from the training phase.

4.7.3 Experiment

The order of the tasks is constant for each participant, always in order from T1 to T3. The order of the graphs increases from smallest (D0) to largest (D3), within each task. The order of the visualization conditions was counterbalanced such that half the participants started with C1 and the other half started with C2 to control for learning effects. For example, a participant that starts with C1 would perform tasks in order of C1-T1-D1, C1-T1-D2, ..., C1-T2-D1, C1-T2-D2, ..., then repeating everything with C2. The nodes chosen to pre-highlight for tasks T1 and T2 are the constant per graph, so the difficulty remains the same between participants.

Participants are encouraged to take short rests between tasks, for as long as they need. The visualization is fully reset between tasks. For C2 (Virtual Reality), the participants are instructed between tasks to re-orient themselves in the middle of the space and face the same direction so that movement in previous tasks does not affect the completion time or success rate of subsequent tasks.

4.8 Hypotheses

We expected the following results from our user study:

H1: For all the tasks, task completion time will be faster with traditional display (C1) than with immersive display (C2).

H2: For the shortest path task (T1), recall task (T2) and difference finding task (T3), C2 will outperform C1 in correctness rate.

H3: Participants will prefer immersive environments (C2) over traditional displays (C1) for exploring graph data.

4.9 Results

Overall, the results of the study confirm H1 and H3, while partially refuting H2. C1 is significantly faster than C2 for T1 and T2, though no significant difference was found for T3, this is likely due to T3 being too difficult overall, so H1 is able to be confirmed. For correctness rate, C1 is better for T2 and T3, while C2 is better for T1. The reasons for this are discussed in greater length in later sections, but we must at least partially refute H2 because it did not excel in every task. An overwhelming majority of participants (85%) favored C2 over C1 for general use, confirming H3.

Task-Dataset	C1 Time	C2 Time	C1 Corr.	C2 Corr.
T1-All	49.26	67.81	0.8056	0.9504
T1-D1	40.60	51.16	0.9917	1.0000
T1-D2	48.95	58.26	0.9250	0.9583
T1-D3	58.22	94.00	0.5000	0.8929
Task-Dataset	C1 Time	C2 Time	C1 Dist.	C2 Dist.
T2-All	31.92	50.70	0.0147	0.0686
T2-D1	19.41	32.25	0.0101	0.0576
T2-D2	27.38	54.59	0.0055	0.0732
T2-D3	48.96	65.26	0.0286	0.0749
Task-Dataset	C1 Time	C2 Time	C1 Dist.	C2 Dist.
T3-All	58.86	71.69	0.0986	0.2380
T3-D1	64.20	75.35	0.1014	0.2890
T3-D2	58.02	72.99	0.0714	0.2360
T3-D3	54.35	66.73	0.1232	0.1890

Table 4.1. Average completion times and correctness rates from the user study.

The average completion time and correctness rates are shown in Table 4.1. It is important to note that for tasks T2 and T3, lower distance indicates better performance.

4.9.1 Task Completion Time

Task completion time is measured with a timer built into the software created for this experiment. The timer starts when the user is able to start interacting with a network, and it ends when the user indicates they are finished by pressing the "Continue" button. None of the tasks directly indicate that a user has chosen correct answers, and so some users chose to be more meticulously than others when checking their work. As a result, there is high variance in the time required to complete all of the tasks. We compare each task separately because the completion time of each task varies.

On average for both conditions, participants completed T1 in 58.53s ($SD = 36.51$). They completed T1 faster with C1, taking only 49.25s ($SD = 24.77$), and slower with C2, taking 67.81s ($SD = 43.60$). A paired t-test shows that this difference is statistically significant ($p = 0.00041$). This result is shown in Fig. 4.3a.

T1 completion time for individual data sets (D1, D2, D3) is shown in Fig. 4.3b, and we compare pairs with t-tests using Bonferroni correction. For D1 and D2, participants did not complete T1 significantly faster with either C1 or C2 ($p_{D1} = 0.092, p_{D3} = 0.23$). For D3, participants completed T1 significantly faster with C1 than with C2, having average completion times of 58.22s ($SD = 23.79$) and 94.00s ($SD = 51.80$), respectively ($p = 0.0033$).

On average for both conditions, participants completed T2 in 41.30s ($SD = 28.25$). They completed T1 faster with C1, taking only 31.92s ($SD = 21.42$), and slower with C2, taking 50.70s ($SD = 31.18$), which is shown by t-test to be statistically significant ($p = 2.3e - 07$). This result is shown in Fig. 4.3c.

Differences between T2 completion time for individual data sets, is shown in Fig. 4.3d, and we compare pairs with t-tests using Bonferroni correction. For D1, participants completed T1 significantly faster with C1 than with C2, having average completion times of 19.41s ($SD = 11.88$) and 32.25s ($SD = 18.12$), respectively ($p = 0.00010$). For D2, participants completed T1 significantly faster with C1 than with C2, having average completion times of 27.38s ($SD = 13.54$) and 54.59s ($SD = 30.30$), respectively ($p = 0.00084$). For D3, participants completed T1 significantly faster with C1 than with C2, having average completion times of 48.96s (SD = 24.65) and 65.25s (SD = 34.26), respectively ($p = 0.014$).

The difference in completion time for T3 is statistically significant ($p = 0.020$) between C1 and C2 overall, with C1 having mean completion time of 58.86s ($SD = 34.54$) and C2 having 71.69s ($SD = 54.97$). However, the variance within data sets is such that there is no significant difference in completion time for individual pairs (Fig. 4.3e-f). Possible reasons for this are discussed in later sections.

4.9.2 Correctness Rate

Correctness rate is measured differently for each task. For T1 (shortest path task), the correctness is calculated as a composite metric. The participants were asked to find the shortest path between two highlighted nodes, and so the correct answer would be any complete path that is the same length as the shortest path between those two nodes. Thus, we use the shortest path length as a fraction of the participant's chosen path when the participant makes a complete path, and a correctness of zero is used for incomplete paths, shown in the following equation.

$$\text{correctness} = \frac{\text{length(shortest path)}}{\text{length(selected path)}} \tag{4.1}$$

With this method, the best correctness score a participant can receive is one, and any path that is complete but contains extra nodes will be between zero and one, approaching zero as more unnecessary nodes are added.

Overall for T1, participants succeeded in making a complete path in 89.17% of all trials, and that path was shortest in 80.83% of all trials. Participants performed significantly better in C2 with an average correctness rate of 0.95 out of 1.0 ($SD = 0.185$), compared to the average correctness rate with C1, at 0.81 ($SD = 0.387$). A t-test confirms that this result is significant ($p = 0.0065$). This is shown in Fig. 4.4a.

Significant differences also exist between T1 trials of data set D3 and the other data sets, shown in Fig. 4.4b, and we compare pairs with t-tests using Bonferroni correction. The average correctness rate for T1 with D3 and C1 is only 0.50 ($SD = 0.51$), which is significantly lower than participants using C1 with the other two data sets ($p = 0.0052$

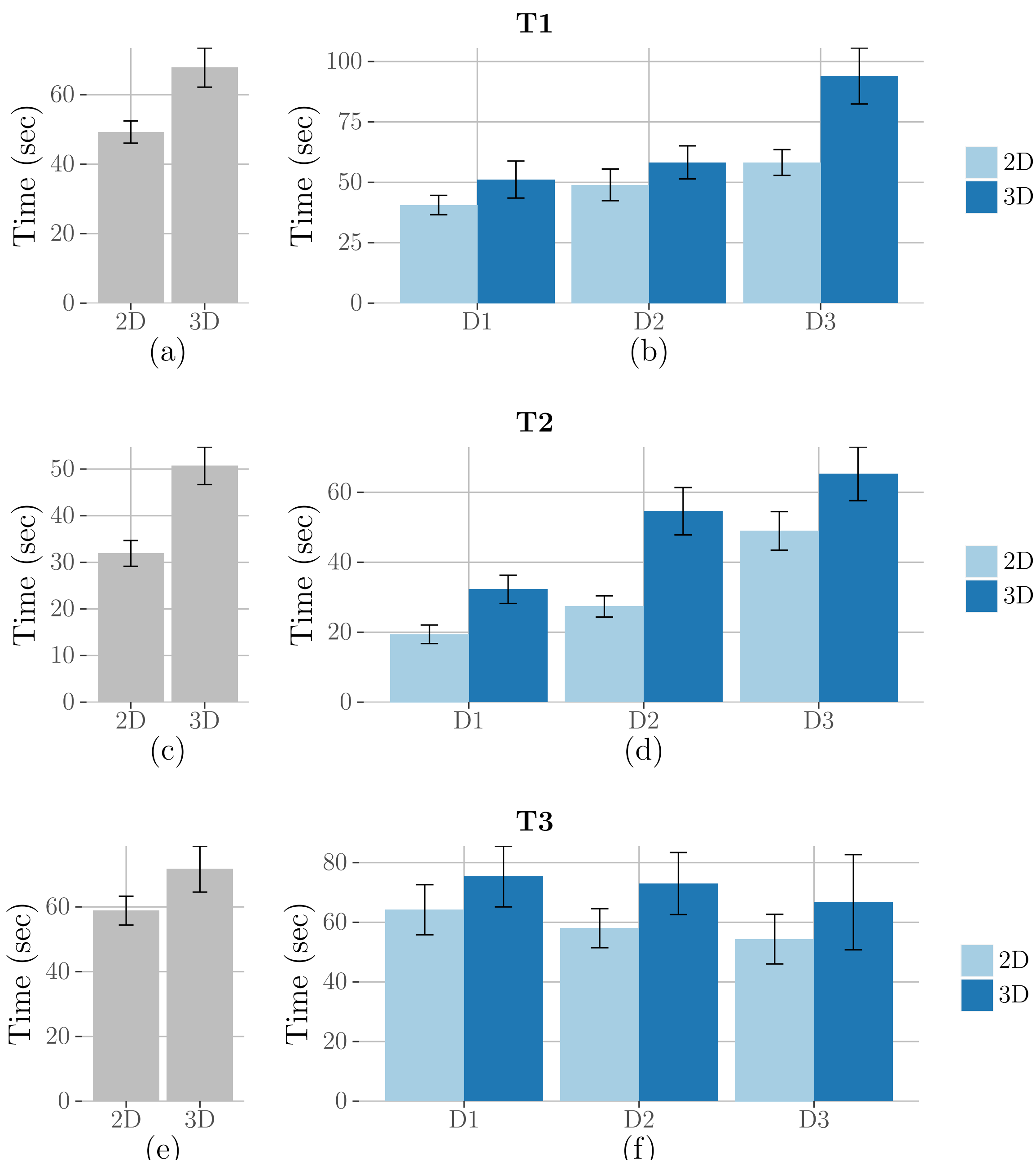


Figure 4.3. Summary of *completion time* for all tasks and data sets. (a) shows the average T1 completion time for each visualization condition, which is broken into individual data sets in (b). The same is shown below for T2 in (c) and (d), then again for T3 in (e) and (f).

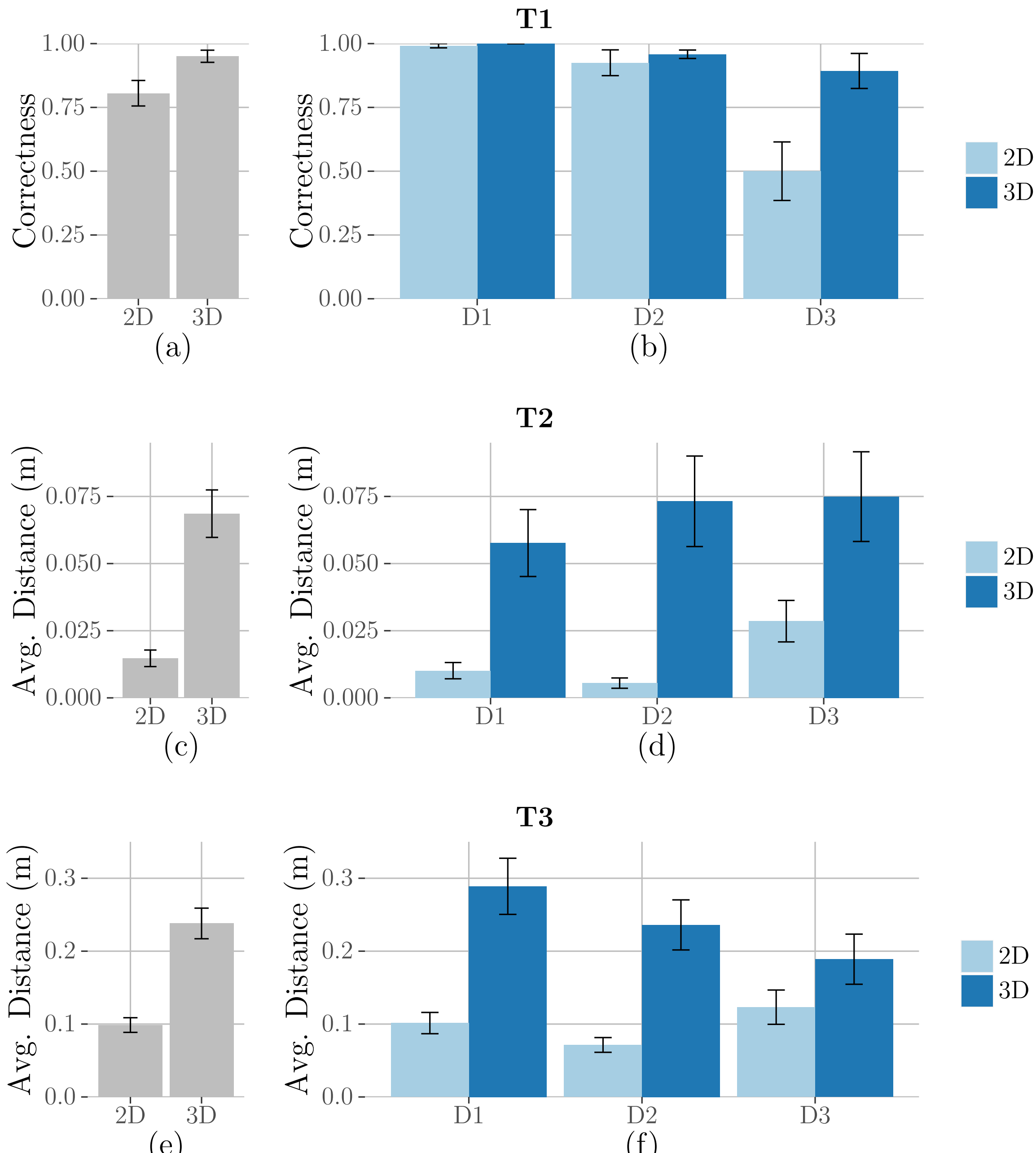


Figure 4.4. Summary of *correctness rate* for all tasks and data sets. (a) shows the average T1 correctness for each visualization condition, which is broken into individual data sets in (b). The same is shown below for T2 in (c) and (d), then again for T3 in (e) and (f). Note: Higher correctness is better for T1, while lower average distance is better for T2 and T3.

and $p = 0.0252$ for D1 and D2, respectively). The reason for this outlier in the data sets is observed to be occlusion due to edge crossing and is discussed in the next section. Participants also performed better with C2 than C1 for D3, with an average correctness rate of 0.89 ($SD = 0.31$) with C2 ($p = 0.0079$).

For T2 (recall nodes task) and T3 (find differences task), the correctness rate is calculated the same way because the goal of both tasks is for the participant to select some number of nodes, and there is only one correct answer for each task. For these tasks, we report the number of nodes that users chose correctly from memory.

Additionally, we provide a more interesting metric that represents on a continuous scale how close, in a spatial sense, participants were to the correct answer. To do this, we calculate the distance between each of the participants' answers and the nearest correct node, and take the average. Given as:

$$d(u,t) = \frac{\|u-t\|}{\overline{d_t}}, \quad \overline{d_t} = \frac{\Sigma_{v \in V}\|v-t\|}{|V|} \tag{4.2}$$

where V is the set of nodes in the network, u is a given node selected by the participant, and t is the target node, or nearest correct node. The distance is shown above as $d(u,t)$. Thus, a correct node will have a distance of zero, and incorrect nodes will have a positive decimal distance, with a larger distance revealing that a participant remembered less about the position of the node they were attempting to indicate. The distance is normalized such that the average distance of a node from the nearest correct answer is the same for the 2D and 3D layouts. The normalization factor is shown in the above equations as $\overline{d_t}$. This makes it possible to fairly compare distances from the two different spaces: immersive 3D and traditional 2D.

Participants correctly identified 62.67% of marked nodes in T2, and the average distance of guesses from the correct node was 0.0417 ($SD = 0.0578$) for all conditions and data sets. Participants performed T2 better with C1 than C2, correctly identifying 71.00% of nodes with C1, and only 54.33% of nodes with C2. Additionally, participants using C1 were able to guess closer to nodes even if their answer wasn't correct, with an average distance of 0.015 ($SD = 0.024$), which is depicted in Fig. 4.4c. Participants using C2 did not perform as well, with an average distance of 0.069 ($SD = 0.068$). This difference in correctness is statistically significant ($p = 1.6e - 07$).

Comparing T2 correctness between data sets shows that C1 is consistently better for the small and medium data sets, but the larger data set does not have a significant difference (Fig. 4.4d). We compare pairs with t-tests using Bonferroni correction. T2 average correctness with D1 was better with C1 than C2, with average distances of 0.010 ($SD = 0.014$) and 0.058 ($SD = 0.056$), respectively ($p = 0.00052$). T2 average correctness with D2 was also better with C1 than with C2, with average distances of 0.0055 ($SD = 0.0086$) and 0.073 ($SD = 0.075$), respectively ($p = 0.00050$). T2 average correctness with D3 also has a statistically significant difference, C1 had an average distance of 0.029 ($SD = 0.035$), while C2 had an average distance of 0.075 ($SD = 0.075$), showing an advantage for C1 ($p = 0.024$).

In T3, participants correctly identified only 8.33% of the additional nodes. This was higher for C1 alone, with which participants correctly identified 11.89% of additional nodes, and lower for C2, with which participants correctly identified only 7.67%. Due to the low number of correctly selected nodes, the average distance of the participants' guesses is much more accurate for determining how the viewing conditions affected performance. Participants achieved an average distance of 0.099 ($SD = 0.078$) for T3 with C1, while participants did not do as well for T3 with C2, with an overall average distance of 0.24 ($SD = 0.16$), which is shown also in Fig. 4.4e. This difference in correctness rate is statistically significant ($p = 3.1e - 09$).

Comparing T3 correctness between data sets shows the same result as that of T2: that C1 is consistently better for the small and medium data sets, but the larger data set does not have a significant difference (Fig. 4.4f), shown by pairwise t-tests. T3 average distance with D1 was better with C1 than with C2, at 0.10 ($SD = 0.065$) for C1 and 0.29 ($SD = 0.17$) for C2 ($p = 4.5e - 05$). T3 average distance with D2 was also better with C1 compared to C2, at 0.071 ($SD = 0.045$) for C1 and 0.24 ($SD = 0.15$) for C2 ($p = 4.8e - 05$). T3 average distance with D3 was not significantly lower with either C1 or C2 ($p = 0.053$).

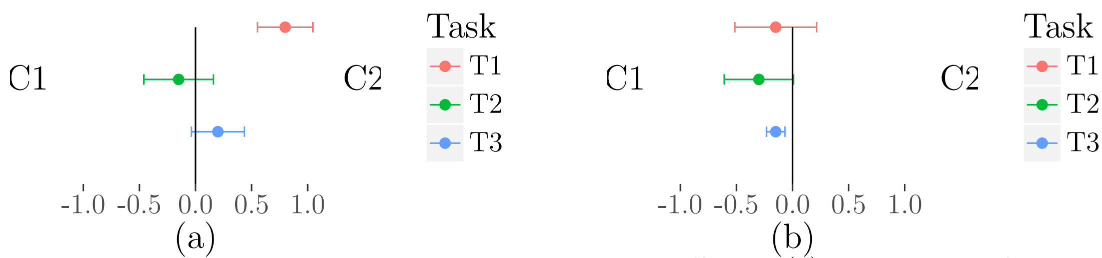


Figure 4.5. User responses to questions about task difficulty (a) and impact of graph size of difficulty (b). A positive result for (a) indicates that tasks were easier with C2 than C1. A positive result for (b) indicates that graph size had a larger impact on difficulty with C2 than C1. User responses to these questions were not found to be statistically significant.

4.9.3 User Feedback

Participants were asked to complete pre- and post-study questionnaires that provide more insight into what each person experienced between the two visualization conditions.

Seventeen out of the twenty participants (85%) that used the system answered that they preferred C2 (Immersive 3D Graph Visualization) over C1 (Traditional 2D Graph Visualization), showing a clear majority.

Users may have slightly favored C2 for T1 and T3, and preferred C1 for T2 in terms of difficulty, but the user reported difficulties were not found to be statistically significant (Fig. 4.5a). It is clear, however, that T1 was the easiest task, T3 was the hardest, while T2 was somewhere in between. Additionally, we were not able to discover any significant difference between C1 and C2 concerning user responses to the question of how much graph size impacted the difficulty of the tasks ($p > 0.05$, Fig. 4.5b).

To the question of what they liked about either C1 or C2, the following lists some of the participants' most common and insightful responses, starting with most the frequent sentiments:

- "[VR] makes it easier to grasp the structure."
- "3D graphs are easier [and/or] more interesting to explore"
- "Edge occlusion was much easier to resolve with in 3D than 2D space with VR."
- "Node selection is too much effort in VR"

- “2D layout is better for quick overview and memory”

Chapter 5

Discussion

The results suggest that one viewing condition did not dominate the other in terms of performance. Immersive environments performed better with path-finding, while traditional displays triumphed in the two short-term memory and change based tasks.

Finding the shortest path was significantly easier with the VR version, and this is the only task that VR excelled with. A significant portion of this can be attributed to the nature of 3D layouts not having any problem with *edge-crossings*. Since the 3D layouts used in the immersive space were completely devoid of any crossed edges, it was never the case that a user would be unable to tell whether the path they had highlighted was complete or not, as each edge was easy to follow through the environment from start to finish. In the case of the traditional 2D environment, and especially with the large data set (D3), it was common for edges to not only cross other edges, but also travel through nodes, making it unclear whether it was a single edge that went through the node, or two edges that happened to connect to the same node at opposite sides. If not for the edge-highlighting that was built into the system, it would be obvious that a 3D layout should do much better and any other factors would likely be insignificant; however, there is likely more at play here. The edge-highlighting allowed participants to see very precisely if two nodes were connected simply by hovering over it, and all participants were instructed in the use of this feature. In spite of this, participants with the 2D version still marked incomplete paths and complete but non-optimal paths more frequently than participants with the VR version. In short, the immersive experience helped participants complete the path finding task more accurately beyond the 3D nature of the layout. This could be evidence of the natural formation of a 3D mental map that helped participants understand the graph and connections between nodes better than could happen in the 2D traditional display environment.

Participants performed more slowly with immersive 3D, in most regards, compared to the traditional 2D setup. Part of this may have been due to participants' lack of familiarity with the controls and virtual environment. However, from the data collected about user movement and participant comments, it seems that the immersive environment also lends itself to fastidiousness. Participants in the immersive environment took more time not necessarily because they didn't find the answer as quickly, but because they often spent time meticulously inspecting their work before continuing to the next open-ended task, something they did not do with the traditional visualization conditions. This may suggest that VR and other immersive technologies are best suited for more in-depth or abstract tasks than the ones used in this experiment.

The mental map we discovered through this experiment seems to be limited to what is essentially a mental image. The simplicity of the mental map, and volatility, seems due to the short time frame of all the tasks. Since participants were given only 30 seconds to scrutinize each visualization, they only had time to construct a relatively crude mental map. This mental map was too fragile for immersive space, based on the overall far lower correctness in immersive 3D for both the recall nodes and find differences tasks. Users mentioned on several occasions that after having changed their perspective in the immersive environment, they lost their memory of the locations of marked nodes in the recall nodes task. We can conclude from the results that if it is possible to have a mental map that accurately represents a 3D environment, it requires more time and interaction to build than this experiment provided.

The difficulty of the *find differences* task seems to have contributed to some unexpected results. Users answered overwhelmingly that this task was the hardest. It is clear that some participants gave up entirely on completing the task for the largest data set, by either selecting randomly or not selecting any nodes. Despite a few participants exhibiting this behavior for both the VR and desktop versions, the results still indicate that overall the desktop version was easier, though this difficulty is likely the reason that there is no significant difference between completion time for either visualization condition. Additionally, for the recall nodes task, we did not find a significant difference in correctness rate. This may be due to either C1 losing its advantage when there are several thousand nodes, or the task being too difficult with networks of that size (>1000 nodes).

Chapter 6

Conclusion

Literature on the mental map in graph visualization is substantial but sparsely populated when it comes to the immersive design space, despite the rapid development of immersive visualization technologies. This work attempts to answer some questions about the mental maps users form under different circumstances, and how we can leverage this knowledge as creators of visualizations. While immersive 3D showed more success with path-finding, traditional 2D visualization has proven superior for memorization and dynamic graph tasks, at least for the depth of immersion we were able to reach.

The results of this experiment highlight the volatility and vulnerability of user's concentration and mental state while completing graph visualization tasks and, by extension, while interpreting real-world data. Additionally, it suggests that immersive technology may allow deeper and more natural understanding of abstract data structures, though it is more carefully considered and must be precisely designed.

To further investigate the use of immersive technology in understanding abstract data structures, such as graphs, the next step is to delve deeper into the individual processes for observing and understanding immersive environments with longer and more detailed experiments. These are processes such as how users mentally store relative positions and directions, and how they are able to turn that into an understanding of the data presented. Investigating the mental capabilities of humans in immersive environments will help us design better systems that more effectively leverage human perception.

References

[1] D. Archambault, H. Purchase, and B. Pinaud. Animation, small multiples, and the effect of mental map preservation in dynamic graphs. *IEEE Transactions on Visualization and Computer Graphics*, 17(4):539–552, 2011.

[2] D. Archambault and H. C. Purchase. Mental map preservation helps user orientation in dynamic graphs. In *International Symposium on Graph Drawing*, pages 475–486. Springer, 2012.

[3] D. Archambault and H. C. Purchase. The map in the mental map: Experimental results in dynamic graph drawing. *International Journal of Human-Computer Studies*, 71(11):1044 – 1055, 2013.

[4] F. Argelaguet and C. Andujar. A survey of 3d object selection techniques for virtual environments. *Computers & Graphics*, 37(3):121–136, 2013.

[5] R. Ball and C. North. Effects of tiled high-resolution display on basic visualization and navigation tasks. In *CHI'05 extended abstracts on Human factors in computing systems*, pages 1196–1199. ACM, 2005.

[6] S. Bryson. Virtual reality in scientific visualization. *Commun. ACM*, 39(5):62–71, 1996.

[7] W. Büschel, J. Chen, R. Dachselt, S. Drucker, T. Dwyer, C. Görg, T. Isenberg, A. Kerren, C. North, and W. Stuerzlinger. Interaction for immersive analytics. In *Immersive Analytics*, pages 95–138. Springer, 2018.

[8] M. Cordeil, T. Dwyer, K. Klein, B. Laha, K. Marriott, and B. H. Thomas. Immersive collaborative analysis of network connectivity: Cave-style or head-mounted display? *IEEE Transactions on Visualization and Computer Graphics*, 23(1):441–450, 2017.

[9] M. de Ridder, Y. Jung, R. Huang, J. Kim, and D. D. Feng. Exploration of virtual and augmented reality for visual analytics and 3d volume rendering of functional magnetic resonance imaging (fmri) data. In *Big Data Visual Analytics (BDVA 2015)*, pages 1–8. IEEE, 2015.

[10] A. Drogemuller, A. Cunningham, J. Walsh, M. Cordeil, W. Ross, and B. Thomas. Evaluating navigation techniques for 3d graph visualizations in virtual reality. In *2018 International Symposium on Big Data Visual and Immersive Analytics (BDVA)*, pages 1–10. IEEE, 2018.

[11] T. Dwyer, K. Marriott, T. Isenberg, K. Klein, N. Riche, F. Schreiber, W. Stuerzlinger, and B. H. Thomas. Immersive analytics: An introduction. In *Immersive Analytics*, pages 1–23. Springer, 2018.

[12] T. M. Fruchterman and E. M. Reingold. Graph drawing by force-directed placement. *Software: Practice and experience*, 21(11):1129–1164, 1991.

[13] H. Gibson, J. Faith, and P. Vickers. A Survey of Two-dimensional Graph Layout Techniques for Information Visualization. *Information Visualization*, 12(3–4):324–357, 2013.

[14] N. Greffard, F. Picarougne, and P. Kuntz. Immersive dynamic visualization of interactions in a social network. In *Challenges at the Interface of Data Analysis, Computer Science, and Optimization*, pages 255–262. Springer, 2012.

[15] I. Herman, G. Melançon, and M. S. Marshall. Graph visualization and navigation in information visualization: A survey. *IEEE Transactions on Visualization and Computer Graphics*, 6(1):24–43, 2000.

[16] Y.-J. Huang, T. Fujiwara, Y.-X. Lin, W.-C. Lin, and K.-L. Ma. A gesture system for graph visualization in virtual reality environments. In *2017 IEEE Pacific Visualization Symposium (PacificVis)*, pages 41–45. IEEE, 2017.

[17] E. Krokos, C. Plaisant, and A. Varshney. Virtual memory palaces: immersion aids recall. *Virtual Reality*, pages 1–15, 2018.

[18] O.-H. Kwon, C. Muelder, K. Lee, and K.-L. Ma. Spherical layout and rendering methods for immersive graph visualization. In *Visualization Symposium (PacificVis), 2015 IEEE Pacific*, pages 63–67. IEEE, 2015.

[19] O.-H. Kwon, C. Muelder, K. Lee, and K.-L. Ma. A study of layout, rendering, and interaction methods for immersive graph visualization. *IEEE transactions on visualization and computer graphics*, 22(7):1802–1815, 2016.

[20] K. Mania and A. Chalmers. The effects of levels of immersion on memory and presence in virtual environments: A reality centered approach. *CyberPsychology & Behavior*, 4(2):247–264, 2001.

[21] K. Marriott, J. Chen, M. Hlawatsch, T. Itoh, M. A. Nacenta, G. Reina, and W. Stuerzlinger.

[22] K. Misue, P. Eades, W. Lai, and K. Sugiyama. Layout adjustment and the mental map. *Journal of Visual Languages & Computing*, 6(2):183–210, 1995.

[23] S. C. North. Incremental layout in DynaDAG. In *Graph Drawing*, volume 1027 of *LNCS*, pages 409–418. Springer, Heidelberg, 1996.

[24] H. C. Purchase, E. Hoggan, and C. Görg. How important is the mental map? —an empirical investigation of a dynamic graph layout algorithm. In *Graph Drawing*, volume 4372 of *LNCS*, pages 184–195. Springer, Heidelberg, 2007.

[25] H. C. Purchase and A. Samra. Extremes are better: Investigating mental map preservation in dynamic graphs. In *Diagrammatic Representation and Inference*, volume 5223 of *LNCS*, pages 60–73. Springer, Heidelberg, 2008.

[26] I. Radu, E. Southgate, F. Ortega, and S. Smith. Summary: 2017 ieee virtual reality second workshop on k-12 embodied learning through virtual & augmented reality (kelvar). In *Virtual Reality Workshop on K-12 Embodied Learning through Virtual & Augmented Reality (KELVAR 2017)*, pages 1–2. IEEE, 2017.

[27] W. Usher, P. Klacansky, F. Federer, P.-T. Bremer, A. Knoll, J. Yarch, A. Angelucci, and V. Pascucci. A virtual reality visualization tool for neuron tracing. *IEEE transactions on visualization and computer graphics*, 24(1):994–1003, 2018.

[28] A. Van Dam, A. S. Forsberg, D. H. Laidlaw, J. J. LaViola, and R. M. Simpson. Immersive vr for scientific visualization: A progress report. *IEEE Computer Graphics and Applications*, 20(6):26–52, 2000.

[29] T. von Landesberger, A. Kuijper, T. Schreck, J. Kohlhammer, J. van Wijk, J.-D. Fekete, and D. Fellner. Visual analysis of large graphs: State-of-the-art and future research challenges. *Computer Graphics Forum*, 30(6):1719–1749, 2011.

[30] C. Ware and P. Mitchell. Visualizing graphs in three dimensions. *ACM Trans. Appl. Percept.*, 5(1):2:1–2:15, 2008.

[31] M. Zyda. From visual simulation to virtual reality to games. *Computer*, 38(9):25–32, 2005.