*Erschienen in: IEEE Transactions on Visualization and Computer Graphics ; 28 (2022), 11. - S. 3651-3661*
*https://dx.doi.org/10.1109/TVCG.2022.3203001*



# Visual Comparison of Networks in VR

Lucas Joos, Sabrina Jaeger-Honz, Falk Schreiber, Daniel A. Keim, and Karsten Klein

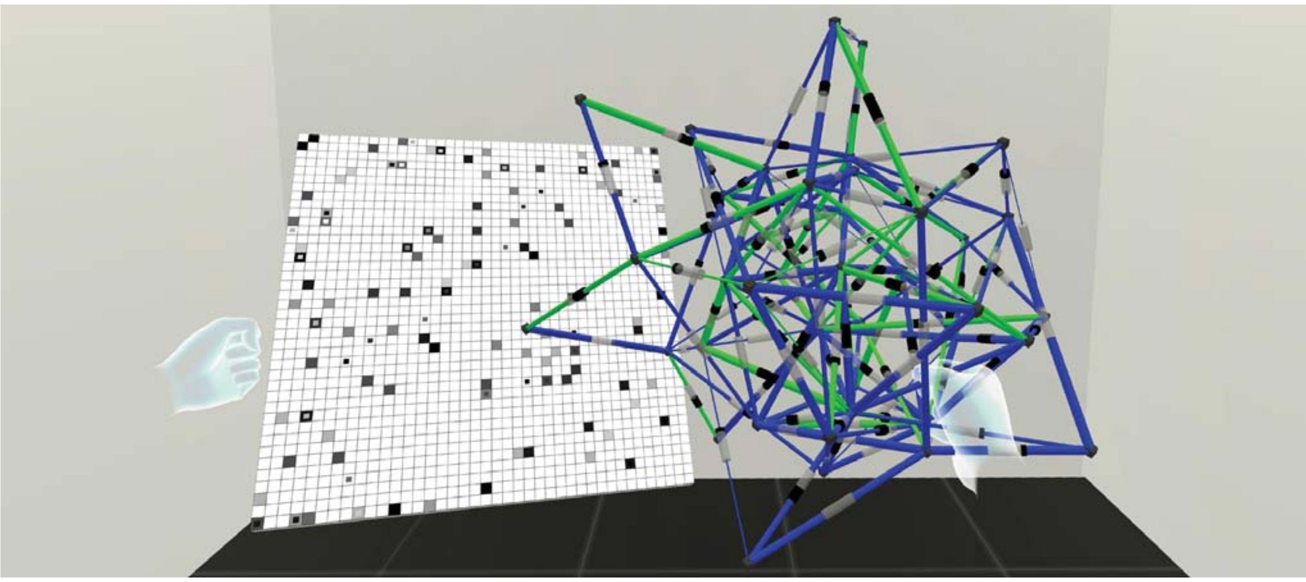


Fig. 1: A matrix representation (left) and a node-link diagram representation (right), both supporting the simultaneous visual comparison of two weighted networks in an immersive environment.

**Abstract**—Networks are an important means for the representation and analysis of data in a variety of research and application areas. While there are many efficient methods to create layouts for networks to support their visual analysis, approaches for the comparison of networks are still underexplored. Especially when it comes to the comparison of weighted networks, which is an important task in several areas, such as biology and biomedicine, there is a lack of efficient visualization approaches. With the availability of affordable high-quality virtual reality (VR) devices, such as head-mounted displays (HMDs), the research field of immersive analytics emerged and showed great potential for using the new technology for visual data exploration. However, the use of immersive technology for the comparison of networks is still underexplored. With this work, we explore how weighted networks can be visually compared in an immersive VR environment and investigate how visual representations can benefit from the extended 3D design space. For this purpose, we develop different encodings for 3D node-link diagrams supporting the visualization of two networks within a single representation and evaluate them in a pilot user study. We incorporate the results into a more extensive user study comparing node-link representations with matrix representations encoding two networks simultaneously. The data and tasks designed for our experiments are similar to those occurring in real-world scenarios. Our evaluation shows significantly better results for the node-link representations, which is contrary to comparable 2D experiments and indicates a high potential for using VR for the visual comparison of networks.

**Index Terms**—Network comparison, virtual reality, weighted graphs, immersive analytics

## 1 Introduction

Networks are used in a large variety of application areas, such as biology, software engineering, and social science, to model, visualize, and communicate information. Consequently, a large body of research has been devoted to visual network analysis, see e.g. [11, 48, 54, 61]. An important task in many use cases is the comparison of networks, e.g. to compare brain activity networks of cohorts of healthy and diseased individuals, protein interaction networks under different conditions, financial networks, or friendship networks on social media platforms. Networks can be compared in multiple ways, for instance by applying different network measures or reduction techniques, or by visual comparison. Several approaches have been proposed so far to support visual network comparison, but it still remains a grand challenge for network visualization research that requires further experimental evidence. Graph statistics and metrics can facilitate comparison [37, 44], but cannot cover all aspects and use cases, and might fail to foster insight into the network structures [15, 35, 37]. Combinations with visual network representations might have the potential to exploit the complementary strengths of both approaches, but require further design efforts and might increase cognitive load as well as the potential for interpretation, coordination, and interaction errors with multiple views or representations. Node-link diagrams and matrix representations are the two main idioms for the visual depiction of networks. While node-link diagrams are dominating both in practical use and as a research topic, several studies have shown that matrices, despite a trade-off in space usage, can have advantages for certain tasks and settings, including comparison [1, 25]. A particular strength of the matrix representation is the use case of dense networks [22], also using combined representations [9, 31], as well as the depiction of connectivity patterns [12].

Immersive environments (IE) are gaining popularity for data analy-

- *Lucas Joos, Sabrina Jaeger-Honz, Daniel A. Keim, and Karsten Klein are with University of Konstanz. E-mail: {lucas.joos | sabrina.jaeger | keim | karsten.klein}@uni-konstanz.de.*
- *Falk Schreiber is with University of Konstanz and Monash University. E-mail: falk.schreiber@uni-konstanz.de.*

sis [39,40], with increasing quality and availability of hardware devices and software tools, but the fundamentals of visual network analysis in IE are not yet investigated systematically. Mixed reality technologies such as virtual reality (VR) and augmented reality (AR) headsets have been recently investigated for visual network analysis, and have shown promising potential [14, 17, 38, 42, 60]. However, there is still very restricted experimental evidence on effective and efficient network comparison in such environments. In addition, while from a conceptual perspective 3D node-link representations are a straightforward extension of their 2D counterpart, the extension of static adjacency matrix representations to 3D is less obvious. In particular, the issue of occlusion, due to the more compact representation, needs to be handled, e.g. by introducing additional interaction operations [6].

In this work, we present an experimental investigation into network representations for comparison tasks in VR, which was conceived in the context of brain activity network analysis [20, 33]. Our aim is to assess the usefulness of different comparative network representations in VR to support visual analysis of networks similar to such real-world networks. To this end, our contributions are as follows: We present multiple encodings for the visual comparison of two weighted networks in an immersive environment and evaluate them experimentally with data and tasks comparable to real-world applications. Based on the evaluation, we retrieve a clear favorite and thus support researchers considering immersive environments for the comparison of their network data with their design decisions. Moreover, the difference between our results and the results of a comparable 2D evaluation demonstrates how the use of immersive technology in combination with visual encodings making use of the extended 3D design space can mitigate issues with traditional 2D visualizations. Therefore, our work also contributes to the field of immersive analytics.

## 2 Related Work

Our research touches several relevant aspects, including network visualization in general, visual network comparison, as well as data visualization in immersive environments, in particular in VR.

**Network comparison and visualization metaphors.** Research on visual network comparison has mainly focused on the development of practical approaches, e.g. for specific types of networks [33, 53, 55, 57, 66], and for structural overviews by abstraction or aggregation, see e.g. [10, 35, 67]. In an attempt to advance the methodology, Gleicher et al. [27] proposed a taxonomy of visual designs for comparison, which groups designs into three categories: juxtaposition, superposition, and explicit encodings. Gleicher [26] also provided a framework facilitating the design of visual comparison solutions, based on four considerations that help characterizing tasks, challenges, and potential solutions for a given scenario. Javed et al. [34] proposed a model of composite visualization views and presented corresponding strategies.

The visualization of dynamic networks can be considered a special case due to the specific interrelation and temporal order of the network states under comparison. Initial results indicate that difference maps might outperform the presentation of the network evolution as time slices for certain settings [4], and that mental map preservation can improve task performance [2, 3]. Graphdiaries [7] introduced animated transitions that help the user to focus on changes between consecutive time steps. Cui et al. [18] proposed a static flow visualization approach for the analysis of dynamic graph changes.

While there is a variety of proposed concepts for the comparison of networks in 2D or 2.5D such as small multiples [2] or 2.5D stacking [13], there is no commonly accepted method that fits a larger range of use cases. Most approaches are limited to a small number of networks and make assumptions on the structural features of the networks or the differences, such as sparsity. Notable results have in particular been achieved for the comparison of different metaphors: Alper et al. [1] investigated the use of adjacency matrices for the comparison of two weighted networks in the context of brain activity. They found that certain representations using adjacency matrices can outperform node-link representations for tasks related to brain connectivity analysis, however, limited to pairs of networks and a small range of network characteristics, in particular regarding scale. Okoe et al. [49] compared node-link and adjacency matrix representations in 2D in a crowd-sourced study, and report advantages of both representations depending on the task. Ren et al. [51] conducted a crowd-sourced study to compare differences in human understanding of node-link and matrix representations, finding better accuracy and task time for node-link diagrams and also differences in learning during the study.

**Networks in VR and AR.** The use of VR and AR technologies for network analysis has received increased attention recently in the context of immersive analytics [24, 45, 58]. Immersive analytics (IA) is concerned with the design and evaluation of immersive environments for data analysis and aims at supporting smooth workflows where analysts are immersed in their data throughout the analysis process. Recent work includes both application-oriented research, e.g. in the context of connectome and brain activity analysis [50, 66], and more fundamental investigations, e.g. on navigation [23, 60], the influence of encodings [14], or the difference between immersive environments [17]. Perceptual aspects and also the influence of interaction operations might play a bigger role in IE for network comparison, in particular in 3D, but this has not been in the focus of research yet. For example, one issue in the transfer of existing results for dynamic data is the use of the third dimension in approaches like the space-time cube [6, 29]. When comparing for example node-link and adjacency matrix representations across 2D and 3D, the node-link representations can naturally be extended to make use of 3D (where the effect of this extension still needs to be investigated), while adjacency matrices cannot be extended in the same way. Thus, the advantage of the latter that was measured in 2D might vanish when switching from 2D to 3D. There is a limited amount of work on perception of networks, in particular for IE. Several works investigated the impact of motion and depth cues and found benefits of stereoscopic 3D (S3D) visualizations [28, 63, 64]. Büschel et al. investigated the influence of the edge encoding in AR on task performance, concluding that in general different styles can be used [14]. Vogogias et al. [62] investigated designs for encoding multiple types of edges in matrices and found task-dependent performance differences. Soni et al. and Kypridemou et al. investigated the influence of different layout methods on the perception of graph properties in 2D [43, 59]. Soni et al. [59] investigated the smallest noticeable difference in density and local clustering coefficients. They concluded that density perception did not differ significantly across algorithms. This might differ strongly in S3D, due to the influence of the viewing perspective and depth distribution. There is substantial work for general perception in IE, including a review reporting benefits and shortcomings [46], a classification of issues in AR [41], and an investigation of graphical perception for immersive analytics of point clouds [65]. Several of the identified issues are also of significance for network analysis, such as the difficulty in estimation of depth and distances [5, 52], and interindividual differences [21]. The former is important when distance is related to structural properties, e.g. for distance-based layouts, and the latter might be emphasized in IE, e.g. deficiencies in 3D perception [32].

## 3 Comparative Network Representations in VR

As the discussion of existing approaches for the visualization of networks in the previous sections shows, the two most common network visualization metaphors, i.e. matrices and node-link diagrams, both come with advantages and disadvantages depending on the use case. These representations are also used to compare networks, but existing approaches mainly rely on juxtapositional comparison as it is easy to implement. However, juxtapositional comparison methods come with disadvantages, including the mental matching effort and that more visualizations are required. Thus, more space is required and users are forced to constantly move their focus between the representations to compare. Both of these issues do not apply to superpositional comparison methods and explicit encodings. However, in contrast to side-by-side comparison strategies, superpositional representations require techniques to ensure that they are perceivable and understandable. For the comparison of networks, especially if they are weighted, the use of adequate comparison techniques beyond juxtaposition is in our opinion still underexplored. The few approaches, such as [1, 36, 47], rely on

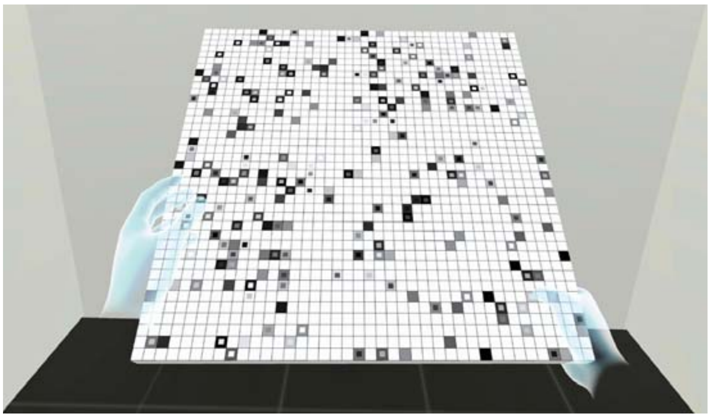


Fig. 2: Matrix representation of two networks in the 3D environment. A cell corresponds to an edge between two nodes and their weights mapped on the inner and outer part of the scale using a grayscale.

two-dimensional visualizations only making use of a less rich design space compared to visualizations in IE and can suffer from overplotting for complex data. A design space that comes with enough opportunities to map relevant attributes in an adequate way and the readability of these representations are both crucial for the successful comparison of networks. As immersive visualizations have the potential to overcome some of the perceptual issues and design space limitations, but might have further effects that counteract these benefits, we explore how immersive visualizations could be used to facilitate the visual comparison of weighted networks. Our aim is to investigate how comparative 3D network representations have to be designed to support common network comparison tasks and to determine for which tasks a certain representation is beneficial. For this purpose, we came up with different approaches of 3D node-link diagrams supporting the comparison of two networks at the same time using explicit encodings. We evaluated these representations and their applicability for visual 3D network comparison in two consecutive user studies.

### 3.1 Matrix Model

Classical 2D matrix visualizations of networks most often consist of a two-dimensional grid, where each cell represents an edge between two nodes [11]. To present more information than the mere existence of edges, color scales or individual visual encodings like glyphs are commonly used. Transforming 2D matrices into the 3D space is not trivial and there has not been much research concerning this issue, with few exceptions, such as the MatrixCube approach by Bach et al. [8], for instance. The authors replace fields by cubes and stack them on top of each other. The cube size or color could be used to map additional attributes like edge weights. The approach has been developed to find the main tendencies of many different network states. However, for the comparison of only two networks, the approach introduces occlusion while barely making use of the opportunities of the third dimension, which makes it hard to retrieve details. Therefore, we did not follow the MatrixCube method but tried to develop less complex 3D representations suiting the use case of simultaneously comparing two weighted networks. We created different 3D matrix models where 2D fields were replaced by stacked cubes combined with different designs for weight encodings. However, we could not observe any advantages of these models compared to classical 2D matrices. Moreover, using the third dimension introduced occlusion and hampered the readability of the representation, as the observer must be more careful to find the desired cube corresponding to a matrix field within the 3D space without confusing it with other cubes. As we could not find any indications that 2D matrices profit from a transformation into the three-dimensional space, we decided not to focus on 3D encodings for matrices. Furthermore, not using actual 3D matrices preserves comparability to the study of Alper et al. [1], which found the matrix representation to outperform node-link diagrams. A study favoring 3D node-link diagrams over 3D matrices would have issues to explain that the outcome is not the result of a less readable and less understandable encoding compared to standard 2D matrices. For these reasons, we decided to use a baseline 2D matrix representation placed on a three-dimensional board, so that the visualization could be moved and rotated in the three-dimensional space (see Fig. 2). While not optimized for 3D, this representation allows to investigate if the advantages observed in the literature for the matrix compared to the node-link representation still hold when the latter one is lifted to 3D. The matrix representation should encode the edge weights of two networks and allow their comparison as well as identifying details from the original networks. Alper et al. [1] compared different 2D matrix representations encoding two weighted networks at the same time and allowing comparison. The representation which turned out to be the best, made use of the structure of a regular 2D matrix but each field was split into an inner part and an outer part. The weights of the corresponding edges in the two networks to compare were mapped on these two parts of each matrix cell using grayscale values. We could not observe any disadvantages of using these matrices in our 3D environment. The actual environment made use of a different wall color compared to the one shown in the images, where it is adjusted for printing. Thus, the matrix representation could be perceived without interference with the background.

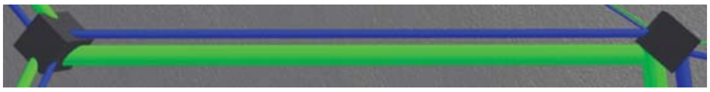
### 3.2 Node-Link Diagram Models

Networks represented as node-link diagrams in the three-dimensional space are most commonly visualized by three-dimensional node objects like spheres or cubes that are linked by lines or tubes as edges. As the focus of our investigation is the impact of encodings for differences in the edge weights, we simply use black cubes for nodes. The node placement in the three-dimensional space is done using a stress minimization graph layouting algorithm implemented in the OGDF graph drawing library [16]. For the edge representations, we generate straight tube objects. We only consider straight edge representations, as suggested by the experimental results of Bueschel et al. [14]. Similar to the matrix representation, the design objective for the node-link diagram representation is to visualize two weighted networks within one node-link diagram. Therefore, each edge should encode two weights while preserving the ability to retrieve absolute values, differences, and the corresponding networks. We present seven different edge representations matching those requirements:

**Parallel Edge Model** The approach by Alper et al. [1] for non-juxtapositional node-link diagrams relied on two parallel lines for every edge and a color scale encoding the weights of two networks. We adopted this idea for our three-dimensional node-link diagram and created the *Parallel Edge Model* (see Fig. 3a and Fig. 4a) consisting of two 3D tubes drawn next to each other. We use the color hue (green and blue) to visualize the network affiliations and map the absolute edge weight on the tube radius.

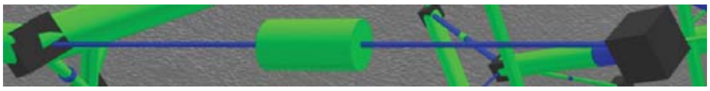
**Split Edge Model** Instead of drawing parallel tubes, one could also split the edge in half and use one half per network to visualize the edge weights. We called this approach *Split Edge Model* (see Fig. 3b and Fig. 4b). As before, the color hue determines the network that is represented by the edge half. Furthermore, the absolute edge weights are mapped to tube radius and color saturation, as this combination appeared to be most promising in internal tests. For small edge weights, a small green or blue ring is placed at the beginning of an edge half to ensure identification of network affiliation.

**Chunk Edge Model** Simply drawing two tubes per edge to map two network weights would hide one of the tubes within the other one. To solve this issue, our *Chunk Edge Model* (see Fig. 3c) consists of one long tube mapping the lower edge and one short tube mapping the higher value. For this model, we use the color hue for the network identification and the tube radius for the weight value mapping. The combination of a long and a short tube avoids the issue of occlusion while giving users the opportunity to estimate relative differences and absolute values.

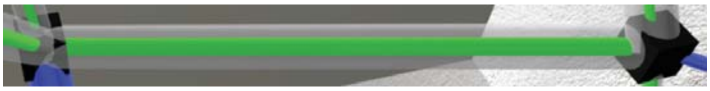


(a) Parallel Edge Model

(b) Split Edge Model



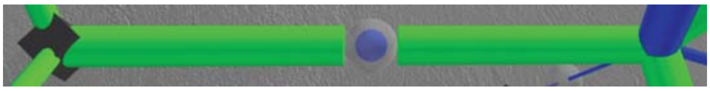
(c) Chunk Edge Model



(d) Inner Outer Edge Model

(e) Cylinder Glyph Model

(f) Cube Glyph Model



(g) Sphere Glyph Model

Fig. 3: Our node-link diagram edge models designed to encode two edge weights at the same time.

**Inner Outer Edge Model** Similar to the *Chunk Edge Model*, the objective of our *Inner Outer Edge Model* (see Fig. 3d) is to solve the issue of occlusion when there are two tubes per edge. Instead of modifying the tube length, this approach is based on opacity modification. The lower weight value is mapped to an opaque tube which is surrounded by a translucent tube representing the higher value. The weight is mapped on the tube radius and the color hue of the inner tube is used to identify the network affiliation.

**Cylinder Glyph Model** Instead of visualizing only absolute values, we also pursued an approach called *Cylinder Glyph Model* (see Fig. 3e and Fig. 4c), which visualizes one absolute value and the difference to the other edge weight. For that purpose, the edge with the higher weight is drawn as a radius-mapped tube connecting two vertices, but with a gap in the middle containing a translucent cylinder. The difference between both edge weights is indicated by the filling of the cylinder. A difference of 0 leads to an empty cylinder, a difference of 1 to an entirely filled cylinder. The color hue of the tube and cylinder visualize the network affiliation.

**Cube Glyph Model** For the *Cube Glyph Model* (see Fig. 3f), we rely on the same concept as for the *Cylinder Glyph Model* with the adaption that a cube glyph is used instead of a cylinder. The weight difference is now depicted by a colored cube within a translucent cube, where the volume of the inner cube is used to map the difference.

**Sphere Glyph Model** The *Sphere Glyph Model* (see Fig. 3g) is based on the same approach as the *Cube Glyph Model*. But instead of a cube glyph, this method makes use of a sphere glyph consisting of a translucent sphere and a colored sphere in the center of the outer sphere with a mapping of weight difference to volume.

After conceiving and implementing the above representation models, we performed a first informal test to get some impression of similarities and potential issues of the models. We used this initial judgment to select suitable candidates for a pilot study, as comparing seven different models with different tasks and different complexities would overcharge the participants. Our preselection proceeded as follows: Firstly, we sorted out the *Inner Outer Edge Model*, since the translucent tube and its radius could not be perceived well in the VR environment, as well as the *Chunk Edge Model*, which was similar to the *Split Edge Model*, but less understandable and biased with regard to the perception of the different network weights. Our glyph models were very similar, so we decided for the one that could be perceived most accurately in VR, which was the *Cylinder Glyph Model*. This preselection led to three remaining edge models to compare, namely the *Parallel Edge Model*, the *Split Edge Model*, and the *Cylinder Glyph Model*.

## 4 Pilot User Study

We designed and conducted a pilot user study to compare and evaluate the different node-link diagram encodings that have been previously described. Examples of these three models applied to an entire node-link diagram are shown in Fig. 4. The first objective of this study was to determine how suitable the developed models were to solve tasks related to the comparison of two weighted networks. This includes several questions that we wanted to investigate:

- Were the participants able to solve a certain task *correctly* with a given model?
- How *fast* were the participants when solving a certain task with a given model?
- What is the *preference* of participants, when they could decide for a certain model to solve a given task?

To investigate these questions, we designed five different tasks related to the comparison of networks derived from functional magnetic resonance imaging (fMRI) but abstracted to weighted graphs. That way, participants without a medical background could solve the tasks. We furthermore generated data with different complexities that are similar to real-world data derived from fMRI scans, as described in Section 4.2. For all tasks and users, we measured the correctness of answers and the answer time. Furthermore, a questionnaire has been provided to gather qualitative feedback on individual preferences and general comments.

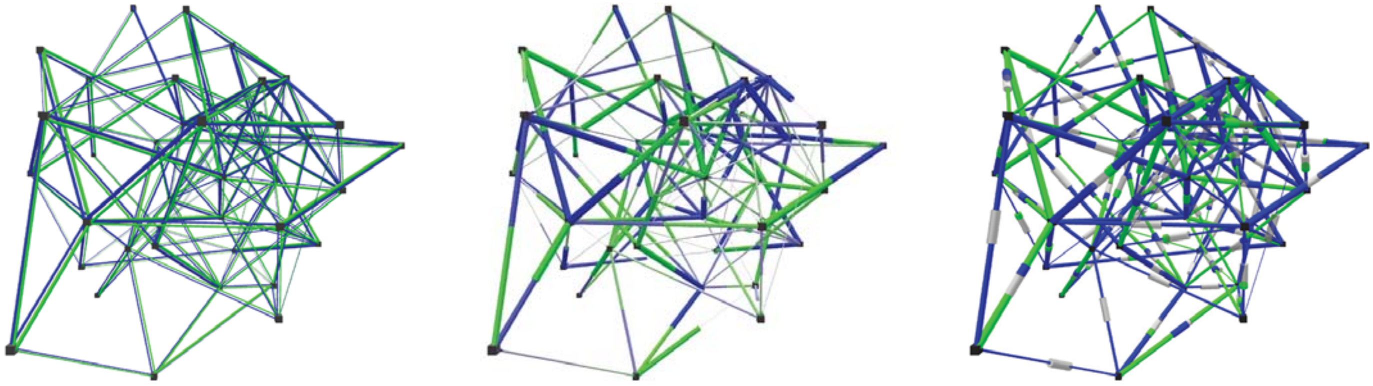
The second objective of the pilot user study was to evaluate our study setup and to find out, where the tasks, data, comparison models, and study procedure required adaptions for the main user study.

### 4.1 Tasks

Research related to fMRI data comes with certain challenges and demands that can be mainly abstracted to different comparison tasks regarding weighted graphs [1, 18, 33, 66]. Alper et al. made use of a literature survey and individual expert interviews with seven experts to determine the main challenges that neuroscientists have to solve when dealing with fMRI comparison [1]. Based on these main challenges, they derived three abstract tasks that can be summarized as follows:

1. Evaluating the dominance of a network at a given node in terms of higher weight accumulated over all incident edges.
2. Assessing topology differences regarding the common neighbors of two given nodes.
3. Finding the region containing the edges with the highest accumulated weight differences between both networks.

We incorporated the first task introduced by Alper et al. in a similar way into our user study as **T4**. We simplified their third task and incorporated it as **T2**, asking participants to find the edge with the highest weight difference between both networks. Instead of covering connectivity differences by comparing topological differences, as Alper et al. do with their second task, we cover connectivity differences by examining weight differences, since higher connectivity is represented by higher edge weights in a weighted graph and vice versa. We included a task (**T3**), in which participants had to evaluate the weight difference



(a) (b) (c)

Fig. 4: Examples for the node-link diagram representations as used for the pilot user study with the *Parallel Edge Model* (a), the *Split Edge model* (b), and the *Cylinder Glyph Model* (c).

of the edges connecting two given nodes, which inherently corresponds to comparing the connectivity between two nodes. Furthermore, we included a task (**T1**) for estimating weight differences (or connectivity differences) on a whole-network level, i.e. estimating which of the two networks is the one with more edges with weights higher than the weights of the other network. The last task added to the pilot user study (**T0**) examines how suitable the different models are to get an impression of how similar or dissimilar the different networks are. In this context, similarity is defined as a threshold on the sum of differences in the weight as outlined below in the description of Task **T0**. The tasks are summarized in Table 1 and explained in more detail in the following:

**T0: Global Similarity Task** For this task, participants examined a node-link diagram with one of the edge models representing weights of two networks and had to estimate whether the networks were similar or dissimilar. For the similar network pairs, a weight change of $|E| \cdot 0.05$ (where $|E|$ is the number of edges in the network) has been arbitrarily applied to the second network while for dissimilar networks the weight change was $|E| \cdot 0.15$ to create a sufficient gap between the differences in the similar and dissimilar pairs. The answer options were *Similar*, *Dissimilar* and *Don't know*. We added the last option to prevent participants from choosing any option if they could not figure out the answer. The goal of this task was to measure how well similarity of two networks could be identified with the given representations, which is relevant, for instance, when analyzing two consecutive scans or differences in cohorts. Demo examples prior to the actual study made sure that participants have seen examples for networks that are similar or dissimilar according to our definition.

**T1: Overall Dominance Task** Before describing this task, we start by defining the term *dominance*. In this context, we use the term *dominance* for edges to describe that one network $N_A$ has a higher weight compared to the other network $N_B$ at a certain edge $e_k$, or in other words: $N_A$ dominates $N_B$ at edge $e_k$. On the level of networks, dominance means that a network $N_A$ contains more dominant edges than the other network $N_B$ or in other words: $N_A$ dominates $N_B$. For this task, participants examined a node-link diagram with one of the edge models representing weights of two networks and had to estimate which network was the dominant one. We ensured that one randomly chosen network in each pair of networks to compare contained 20% more dominant edges than the other one. To answer this task the participants had to choose the color of the dominant network, namely *Blue Network* or *Green Network*. As for **T0** we also included a *Don't know* option to prevent guessing. With regard to the evaluation of differences between cohorts, for instance, it is beneficial for users to get an impression of whether edge weights are in general higher for a certain network and which network it is. Therefore, we included this task in the user study.

**T2: Highest Difference Task** For this task, participants examined a node-link diagram with one of the edge models representing weights of two networks and had to find the edge, where the weights of the two networks differed the most. The networks used in this task have been designed in such a way that there was always exactly one edge with the maximum weight difference leading to only one correct answer. To answer the task the participants had to enter the 3-digit number of the edge having the highest weight difference. The edge numbers could be displayed by clicking a controller button. Getting an impression of regions, where two networks highly differ can be beneficial in neuroscience, for instance, as a starting point for further explorations investigating reasons for major connectivity differences. With this task, we aim at testing the ability to find such a starting point for further exploration using the different edge models.

**T3: Single Edge Difference Task** For this task, participants examined a node-link diagram with one of the edge models representing weights of two networks, where one edge was highlighted. To solve the task, participants had to estimate the weight difference between the two networks at the highlighted edge. The task could be answered by choosing an estimated weight difference between 0 and 1 with a step size of 0.1, which leads to 11 answer options. This task measured how precisely differences could be perceived with the different representations.

**T4: Local Dominance Task** For this task, participants examined a node-link diagram with one of the edge models representing weights of two networks, where one node was highlighted. The participants had to assess whether the blue or green network dominated at the given node based on the edge weights of the incident edges. Thus, all the edges connected to the given node and their weights had to be examined and summed up. The network with the higher accumulated weight had to be selected. It was ensured that there was always exactly one correct answer. To answer this task, the participants had to choose the color of the dominant network at the given node, namely *Blue Network*, *Green Network* or *Don't know*. This task measured how well local differences and trends could be assessed with the examined representations.

### 4.2 Data

For the user study, we generated synthetic weighted networks, which were inspired by networks derived from actual fMRI data. This allowed us to conduct the study under controlled and equal conditions without decoupling the study from the actual use case. The generated networks contained 40 nodes with edge densities of 8 and 16 percent. Edge weights varied between 0 and 1 with a step size of 0.1. In order to focus on weight differences and not to introduce potentially confounding factors, we generated connected networks only, i.e. with only a single connected component. This generation concept corresponds to a threshold filtering approach as used for brain activity analysis, where some of the edges (in our case pairs of edges) are either of weight zero or below the threshold and are thus not part of the analysis [19]. For each of the generated networks, we created a further network to compare to.

Table 1: Study tasks and their characteristics.

| Task | Name | Description | Answer | Network Properties |
|---|---|---|---|---|
| T0 | Global Similarity Task | Estimate similarity between two networks | Similar / Dissimilar / Don't know | Same networks with random total weight change of $\lvert E\rvert \cdot \{0.05 \mid 0.15\}$ |
| T1 | Overall Dominance Task | Assess which network is the dominant one (number based) | Network identifier / Don't know | 20% more dominant edges in one network |
| T2 | Highest Difference Task | Find the edge with the highest difference between the networks | 3-digit edge number | Existence of only one unambiguous edge with the highest difference |
| T3 | Single Edge Difference Task | Estimate the edge weight difference at a given edge | Decimal number from 0 to 1, 0.1 steps | No network adaption required |
| T4 | Local Dominance Task | Assess which network is the dominant one (weight based) at the highlighted node | Network identifier / Don't know | Unambiguous weight-based dominance |

Similar to Alper et al. [1], the comparison network has been created by copying the original one and perturbing 70% of the edge weights. Thus, we ensured that there were enough edges of the same weight in both networks, as it occurs in fMRI networks, too. The networks were generated in advance and the same for all subjects.

### 4.3 Apparatus

Our VR environment consisted of a Unity3D application running on a standard VR-capable setup. Users perceived the application using an Oculus Rift CV1 HMD allowing stereoscopic 3D vision. For the interaction, the Oculus Touch controllers where used along with the headset position and rotation. The user study was conducted in a laboratory at the University of Konstanz.

### 4.4 Interaction

The virtual environment allowed interaction by using two controllers, one per hand. The controller positions, rotations, and the recognized gestures were visualized in the virtual environment using semi-translucent virtual hands (see Fig. 2). By collapsing all fingers of a hand (i.e. forming the hand into a fist), the participants could perform the *grab* gesture. It could be used to rotate and move grabbed objects, namely node-link diagrams for the pilot user study. Moreover, the study setup allowed to point at virtual objects with a virtual laser ray directed by one of the controllers. The virtual ray could be activated by performing the *pointing* gesture (index finger is stretched, all other fingers are collapsed) and triggered actions when a further controller button was clicked. Hence, participants could start tasks, press keys on virtual keypads, and confirm their answers. Besides the controller interaction, participants could walk around and rotate their heads to change their perspectives.

### 4.5 Procedure

For the pilot user study, the participants were invited for individual sessions. All participants signed a consent form before the actual study procedure began. Since the experiment was conducted prior to the COVID-19 pandemic the study did not require special safety measures. The study procedure started with a detailed explanation of the visualizations, the tasks, and the study setup. Questions could be posed at any time. After the explanation, the participants mounted the HMD and solved training examples for each task. Then, the actual study started. By using the controllers the participants were able to move and rotate the node-link representations in the VR environment. Before a sub-task could be solved, a virtual start button had to be clicked by using the virtual ray and pressing a controller button, which made the visualization visible and started a timer. After figuring out the solution, the answers could be given directly in the application using task-dependent answer boards and the pointing gesture to activate an answer option. For each of the five tasks, all three edge representations were tested with networks of 8% and 16% density. We tested every condition three times leading to $5 \cdot 3 \cdot 2 \cdot 3 = 90$ sub-tasks in total that had to be solved. While the order of the tasks was fixed, we randomized the order of the conditions within each task. For this purpose, the order of the network representation models (i.e. the three different edge encodings) and the order of the networks used to test the conditions were randomized. However, the networks with low density always appeared before the higher density networks to ensure that the difficulty increases within each task. Together with the training session in the beginning, this approach minimized training effects. After solving all tasks the participants were asked to fill out a questionnaire regarding personal questions like their age as well as questions concerning their experience with the application and the representations. The study took approximately one and a half hours and breaks could be done anytime.

### 4.6 Participants

We invited five unpaid participants associated with the Computer Science department of our institution. Three of them were male and two female. Their age varied between 16 and 27, the mean age was 23.6, the median 25, and the standard deviation 4.4. Participants younger than the country's legal age were allowed to take part in the user study and all required measures were taken. Four persons indicated the right hand as their dominant hand, one person the left hand. The participants had normal or corrected-to-normal vision and did not declare any perceptual issues with the application. A majority of four participants stated that they had prior knowledge concerning network analysis. Three of the participants indicated that they had prior experience with VR hardware and applications.

### 4.7 Results

With a sample size of five participants, statistical tests are not meaningful for the pilot user study. Therefore, we use the mean accuracies, completion times, and the user feedback to get an impression of the different models and to evaluate the overall study setup. The pilot study results are shown in Table 2. The combined accuracies for all tasks as well as the single task accuracies show that the *Cylinder Glyph Edge Model* led to the best results consistently for both densities. Only for T0, the other representations led to better results. The completion time results do not show a clear favorite across the tasks and the time differences do not exceed a few seconds.

The analysis of the questionnaires filled out by the participants did not indicate problems with the application or study setup, except for a too small edge label size. All participants preferred the *Glyph Edge* representation for the given tasks. A majority of three participants considered the *Parallel Edge* representation at least for one task to be not suitable and criticized the high level of occlusion. With regard to physical discomfort, a few users reported headache ($n = 1$), motion sickness ($n = 1$), or tiredness due to standing for a long time ($n = 1$). For this reason, more breaks were incorporated in the second study and the study duration has been decreased. Although all participants preferred the *Glyph Edge Model*, there were comments criticizing the coloring of cylinders leading to visual interference. We incorporated this feedback into the main user study.

Table 2: The mean accuracy and mean answer time results of the pilot user study for tasks 0 to 4 and their aggregation (=). The conditions were *Parallel Edge Model* (P), *Split Edge Model* (S), and *Cylinder Glyph Model* (G) with network edge densities of 8% and 16%. For each condition, the best result is marked.

| | Accuracy | | | | | | Answer Time (s) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 8% | | | 16% | | | 8% | | | 16% | | |
| T | P | S | G | P | S | G | P | S | G | P | S | G |
| 0 | 0.73 | **0.80** | 0.73 | **0.73** | 0.53 | 0.67 | **17.9** | 22.3 | 19.9 | **16.3** | 18.7 | 16.9 |
| 1 | 0.87 | 0.67 | **0.93** | 0.67 | 0.53 | **0.93** | 18.2 | **14.1** | 18.4 | 16.9 | **14.4** | 18.8 |
| 2 | 0.13 | 0.07 | **0.67** | 0.07 | 0.13 | **0.33** | 30.7 | 32.1 | **28.7** | 35.9 | **30.2** | 34.4 |
| 3 | 0.53 | 0.67 | **0.73** | 0.00 | 0.33 | **0.67** | 14.6 | **14.0** | 14.6 | 14.8 | 11.1 | **10.4** |
| 4 | 0.73 | 0.80 | **0.93** | 0.53 | 0.60 | **0.93** | 13.0 | 12.9 | **12.3** | 16.8 | **11.0** | 11.8 |
| = | 0.60 | 0.60 | **0.80** | 0.40 | 0.43 | **0.71** | 18.9 | 19.1 | **18.8** | 20.1 | **17.1** | 18.4 |

### 4.8 Discussion

In our pilot user study the *Glyph Edge* representation led to the best results regarding the accuracy and user preference. The answer time analysis did not show a clear favorite. Therefore, we stick to the *Glyph Edge* representation for the main user study.

Besides these findings, the initial study also revealed some weaknesses of the setup. Firstly, the performance of the participants as well as their comments during and after the study suggest that T0 did not work as intended. Users had difficulties getting a feeling for our similarity definition. A more intuitive similarity metric and a more exhaustive training session could have led to better results. However, the task remains much more abstract and difficult to communicate compared to the other tasks leading to more reasonable results. Besides task-related findings, the pilot user study revealed issues with the length of the study. The participant feedback on the study length and physical comfort indicated that the study was too long and contained too many sub-tasks. Especially participants without experience in VR reported discomfort after wearing the headset for the entire study duration. Although participants could always ask for a break, obligatory breaks and a shorter study duration should be considered to avoid physical discomfort and to increase the comparability of the individual results.

## 5 Main User Study

The objective of the main user study was to compare 3D node-link diagrams with matrix representations, which both used a certain encoding to support the visual comparison of two networks within the same representation. Analogously to the pilot study, the main questions this study investigated were whether participants were able to solve a certain task with a given model *correctly*, how much *time* users need for that and which model they *preferred* considering a certain task.

For the matrix representation, the model favored and used by Alper et al. has been chosen (see Section 3.1). For the 3D node-link diagrams we made use of our *Glyph Edge* representations, since this representation was preferred by the participants and led to the best accuracy results. However, there were participant comments on the visual interference of the cylinder filling with other tubes, since the same colors were used. We reconsidered our *Glyph Edge* design and agreed with the criticism. Thus, we changed the cylinder filling color to black for the main user study (see Fig. 1). An internal test suggested that the color change reduced the color interference without introducing further issues.

### 5.1 Tasks

For the main user study, we included the same tasks as before except for T0. By omitting T0, we aimed for a study time reduction and reacted to the unclear results for this task. Tasks requiring a network identifier as answer had the same answer options for the node-link diagrams as in the pilot study. For matrix representations, the answer options were *Outer Network*, *Inner Network*, and *Don't know*. These answer options correspond to our matrix drawing strategy, which maps all edge weights of one network to the outer part of each matrix cell and the edge weights of the other network to the inner part of the matrix cells.

### 5.2 Data & Apparatus

For the pilot user study, we generated and used artificial network data inspired by real-world fMRI data. As we could not observe any issues with the data itself or the different data complexities, we made use of the same complexities and network generation techniques to create data for the main user study. The pilot user study did not reveal any issues with the study apparatus. Hence, we relied on the same apparatus for the main study as for the pilot study.

### 5.3 Interaction

The study environment used for the main user study supported the same interaction opportunities as the pilot study. Thus, the participants were able to trigger actions using the pointing gesture as well as moving and rotating the data representations, namely the node-link diagram and the 3D board visualizing the 2D matrix.

### 5.4 Procedure

The main user study had a similar structure compared to the pilot user study. The study also started with an explanation phase followed by a training session with demo tasks. Then, the actual study started, where participants solved four tasks. Each task had to be done using the matrix model and the node-link representation. Moreover, we tested two different network complexities and repeated each of the resulting conditions four times leading to $4 \cdot 2 \cdot 2 \cdot 4 = 64$ sub-tasks in total. While the order of the tasks was fixed, the order of conditions and networks was arranged analogously to the pilot study in order to counterbalance training effects. Whenever the task or representation changed, the participants were asked to unmount the headset and to fill out a NASA TLX test [30] for the previous condition. The NASA TLX was explained in the explanation phase preliminary to the study. In total, each participant was asked to fill out eight NASA TLX tests leading to eight planned breaks during the study. That way, we gathered additional quantitative data and solved the issue of physical discomfort due to long phases of headset-wearing. After the actual study phase, the participants filled out a questionnaire similar to the one of the pilot user study. The average study duration was one hour.

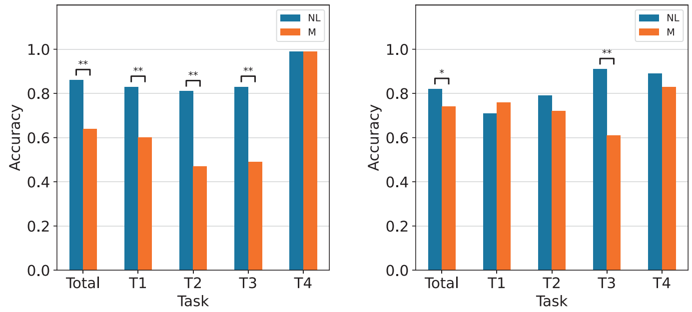
### 5.5 Participants

For the main study, we invited students and employees of the Computer Science and Mathematics department of our institution, of which 18 persons volunteered to take part. 15 participants were male and three participants were female. The gender imbalance occurred by chance and is similar to the male-female proportion of the Computer Science and Mathematics students. The participants were compensated with 10€ for the user study. Their age varied between 18 and 26 with a mean of 22.2, a median of 22, and a standard deviation of 2.6. All of them were right-handed and had normal or corrected-to-normal vision. One person reported a weakness in perceiving red color (protanomaly), but the training did not indicate issues with perceiving the application elements and the performance of the user was very similar compared to the other participants. Five participants already had experience with VR and all participants had at least some fundamental knowledge about (weighted) networks, node-link diagrams, adjacency matrices, and network analysis. All participants received the same explanation and had the ability to pose questions at any time. We made sure that the participants had enough knowledge about networks, our custom representations, and the tasks to complete the study.

### 5.6 Results

In the following, we present the study results regarding the accuracy, mean answer time, task load, and qualitative feedback. Significant results are divided into three categories: $***$ for $p < 0.001$, $**$ for $p < 0.01$ and $*$ for $p < 0.05$.

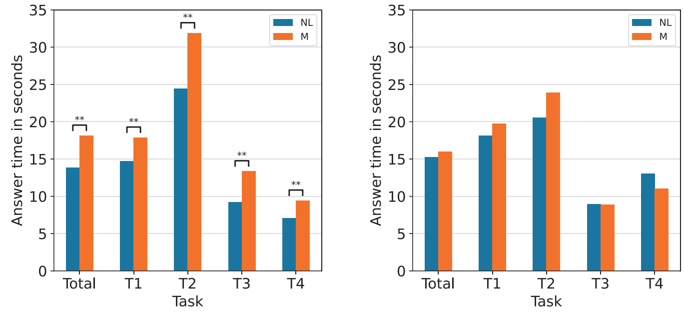
Accuracy  The accuracy results for both tested data complexities are shown in Table 3 and visualized in Fig. 5. We made use of Fisher's exact test to evaluate the significance of differences between the conditions. For the networks of 8% density, the node-link diagram representation led to significantly better results for each task except for T4, where




(a) Edge density: 8%

(b) Edge density: 16%

Fig. 5: The mean accuracy results of the main user study.




(a) Edge Density: 8%

(b) Edge Density: 16%

Fig. 6: The mean answer time results of the main user study.

almost no wrong answers occurred regardless of the representation. Considering all answers independently of the task, the node-link representation outperformed the matrix representation significantly. For the higher edge density, the node-link representation also outperformed the matrix representation significantly regarding all answers. For the single tasks, there are only significant differences for T3, where the node-link diagram representation outperformed the matrix representation.

**Answer Time** The results for the mean answer time are shown in Table 3 and visualized in Fig. 6. We applied a Shapiro-Wilk test [56], which indicated that the answer time results differ significantly from a normal distribution. Since the t-test may only be applied to normally distributed data, we evaluated the significance of answer time differences with the Wilcoxon signed-rank test. For networks with an edge density of 8%, the significance test showed that the node-link diagrams significantly outperformed the matrix representation for every task with regard to answer time. For the networks with 16% edge density, we could not find significant answer time differences for any of the tasks.

**NASA TLX Test** The task load results for all tasks combined are shown in Table 4 and visualized in Fig. 7. We applied a Shapiro-Wilk test [56] on the task load results, which indicated that the result distributions significantly differ from normal distributions. Therefore, Wilcoxon signed-rank tests were applied to evaluate the significance of differences between the results. Regarding the results aggregated for all tasks, the node-link diagram led to significantly better results for effort, frustration, mental demand, and performance. The physical demand ratings were significantly higher for the node-link diagrams compared to the matrix results, while there was no significant difference for the temporal demand. We do not discuss the task-individual results in detail, since they are very similar to the aggregated results across all tasks without relevant differences.

**Qualitative Results** The user study ended with a questionnaire similar to the one used for the pilot user study. Besides person-related questions, the participants were asked to comment on the different representations, the tasks, and their experiences with the study procedure. The large majority of the participants did not encounter any issues with the application or study setup. There were only a few comments by individuals concerning discomfort wearing glasses under the HMD ($n = 2$), the frequency of NASA TLX tests requiring to unmount and mount the VR headset ($n = 2$), or dizziness after the user study ($n = 1$). However, we do not consider these complaints to be general problems, since they did not apply to most of the participants. Furthermore, all participants were able to perceive the application as desired and could solve the tasks without issues. Except for the two participants criticizing the frequency of NASA TLX tests, there were no comments against the planned breaks (8 in total) and no participant indicated that the study duration was too long. Regarding the network representations, all participants preferred the node-link diagram visualization over the matrix representation. User comments on the matrix representation suggest that looking at the visualization for a long time can be exhausting for the eyes and that it produces too much input ($n = 2$). Further comments indicated that users had issues with the difference estimation using a grayscale as implemented in the matrix representation. One of the participants had the impression that the outer part of matrix cells was perceived stronger by the eyes than the inner part. Some participants ($n = 3$) expressed that the matrix representation was helpful to get an overview, but for the investigation of detailed information, the node-link diagram representation was advantageous.

### 5.7 Discussion

The objective of the main user study was to evaluate how suitable the matrix representation and the node-link diagram are for the visual comparison of two weighted networks. With regard to the accuracy, the node-link encoding clearly outperformed the matrix representation for the lower edge density. For the higher density, the node-link diagram still outperformed the matrix visualization, but the accuracy differences are remarkably smaller. One would assume that the accuracy results decrease with the higher density. However, this is not the case for our

Table 3: The mean accuracy and mean answer time results of the main user study for tasks 1 to 4 and their aggregation (=). The conditions were *Matrix* (M) and *Node-Link Diagram* (NL) with network edge densities of 8% and 16%. Significant pairwise differences between the conditions are indicated with asterisks. The bold value indicates a significantly better result for a certain condition.

| | Accuracy | | | | Answer Time (s) | | | |
|---|---|---|---|---|---|---|---|---|
| | 8% | | 16% | | 8% | | 16% | |
| T | M | NL | M | NL | M | NL | M | NL |
| 1 | 0.60 | **0.83** ** | 0.76 | 0.71 | 17.9 | **14.7** ** | 19.8 | 18.1 |
| 2 | 0.47 | **0.81** ** | 0.72 | 0.79 | 31.9 | **24.4** ** | 23.9 | 20.6 |
| 3 | 0.49 | **0.83** ** | 0.61 | **0.91** ** | 13.4 | **9.3** ** | 8.9 | 9.0 |
| 4 | 0.99 | 0.99 | 0.83 | 0.89 | 9.4 | **7.1** ** | 11.0 | 13.0 |
| = | 0.64 | **0.86** ** | 0.74 | **0.82** * | 18.2 | **13.9** ** | 16.0 | 15.3 |

Table 4: Mean results of the NASA TLX test accumulated over all tasks of the main study. The lower the value on the 0-100 scale the better the result. Significant pairwise differences between the conditions are indicated with asterisks. The bold value indicates a significantly better result for a certain condition.

| Nasa TLX Domain | Matrix mean (sd) | Node-Link mean (sd) |
|---|---|---|
| Effort (EF) | 46.2 (21.8) | **35.1** (20.1) ** |
| Frustration (FR) | 35.0 (22.5) | **21.1** (16.1) ** |
| Mental Demand (MD) | 46.5 (21.2) | **30.6** (19.4) ** |
| Performance (PF) | 47.4 (20.5) | **36.4** (23.1) ** |
| Physical Demand (PD) | **21.1** (16.2) ** | 28.3 (20.6) |
| Temporal Demand (TD) | 38.2 (21.6) | 34.3 (22.6) |

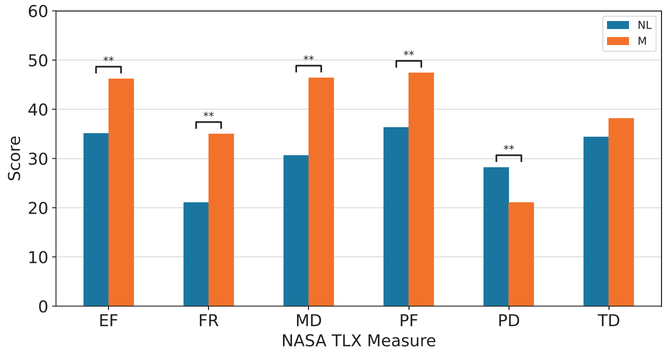



Fig. 7: The aggregated NASA TLX results of the main user study.

study, where the node-link accuracy results stay relatively similar across the different densities and the matrix accuracy results even increase with a higher edge density. A possible explanation for the latter finding is that training effects led to a higher accuracy for the matrices even though the edge density increased. It is also surprising that the accuracy result for the node-link diagrams did not decrease with an increasing edge density. Other user studies dealing with 2D node-link diagrams experienced significantly worse results with an increasing network complexity [1, 25]. Therefore, a possible explanation–besides training effects–for our complexity-stable results is that the additional dimension and interaction opportunities provided by our setup counteracted the issues coming with a higher network complexity.

With regard to the answer time, the node-link representation outperformed the matrix visualization across all tasks for the networks with 8% edge density while there were no significant differences between the models for the networks with 16% edge density. Analogously to the accuracy results, the answer time results for the more complex networks surprise since the matrix model answer times mainly decrease with an increasing edge density and the node-link diagram results remain similar or increase slightly. These results are in accordance with our explanations for the accuracy results.

The NASA TLX results show that the participants experienced a significantly lower task load for the node-link representation considering effort, frustration, mental demand, and performance. These findings match with the accuracy and answer time results, which indicate at least for the lower-complexity networks that users were more successful and faster when using the node-link variant. However, the participants did not experience significant differences with regard to the temporal demand. One explanation could be that the node-link diagram representation required rotation and movement to see the entire network without occlusion but presented relevant information in a direct way, while the matrix visualization presented information in a more abstract way but could be examined without movement or rotation to counter occlusion. This might have led to the impression that both effects counterbalanced each other leading to a similar temporal demand for the representations when used for solving tasks. For the physical demand, the matrix led to a significantly better result. This is not surprising, as the node-link representation benefits from rotation and translation leading to different perspectives, which is not the case for the flat matrix model.

In addition to the quantitative results, the qualitative user feedback shows a clear preference for the node-link diagram representation. The setup of our user study is comparable to the study of Alper et al. [1], since we compared similar network representations, tasks, and data sets. However, the results of our 3D study are very different compared to their results with a 2D setup. In their study, the matrix representation outperformed the node-link diagram for all tasks considering accuracy, answer time, and user preference. Furthermore, the node-link diagram performance dropped heavily with an increasing graph complexity while the matrix representation results were less affected by the network complexity. The authors explained this behavior with increasing occlusion for the node-link diagrams. Since our results concerning the comparison of matrix representations and node-link diagrams are very different, one could argue that our approach for a 3D node-link diagram comparing two networks at the same time was successful and could benefit from the opportunities offered by the immersive environment. Especially the issue with occlusion and perceptual problems could be mitigated by allowing interaction and stereoscopic vision. These advantages may also explain why our node-link representation did not lead to a performance drop for more complex networks.

## 6 Conclusion & Future Work

Approaches for visual comparison of weighted graphs are still underexplored, although this is a relevant topic in many application areas, such as biology, sociology, and medicine. The most common visual representations for such networks are matrices and node-link diagrams. Although node-link diagrams give a good impression of network topologies and have the ability to preserve spatial properties like the placement of brain regions, they come with issues such as occlusion, especially with an increasing network complexity and in two-dimensional setups. Therefore, existing research on visual comparison of weighted networks considered matrix representations to be advantageous compared to node-link representations. In this work, we investigated how weighted networks could be visually compared in an immersive 3D VR environment and how existing representation techniques could benefit from the additional capabilities. For this purpose, we developed seven different approaches for node-link diagrams encoding two networks at the same time using the extended design space of an immersive setup. A pilot user study compared the performance of three preselected node-link diagram models using tasks and data related to use cases from fMRI research. The node-link diagram representation that performed best in the pilot study was used in our main study to compare it against a comparative matrix representation. We found significant differences for these models regarding accuracy, answer time, task load, and user preference, which were in favor of the node-link diagram representation. These results are contrary to the results in a similar two-dimensional setup and promising considering the use of immersive technology joint together with 3D-adapted visual representations. Therefore, we show how the visual comparison of networks can benefit from the capabilities provided by an immersive setup. Our results should encourage researchers working with network data to consider immersive environments for exploring their data and support their design decisions when creating an immersive setting by suggesting the use of certain node-link encodings to support common comparison tasks. Furthermore, our work generally encourages to make use of immersive technology in combination with visual representations exploiting the extended design space of S3D to facilitate the exploration of data. We hope that our approaches and evaluation results shed some light on the underexplored research area considering the comparison of weighted networks, especially in the context of IA and with regard to real-world use cases.

Overall there are still some limitations in our work. In the main user study, there is a gender imbalance between participants. This imbalance occurs since most participants are students with a background in Computer Science or Mathematics where there is a similar proportion of gender as in the main user study which should be improved in future studies. Since this work only considered undirected, weighted graphs with a common topology, we plan to expand our research by investigating how the exploration of other kinds of networks relevant for further use cases could also profit from immersive settings and the broadened design space. In addition, the synthetic data which was used in this study to increase comparability and reproducibility of the results should be supplemented by real-world data and additional network densities to draw direct conclusions for the application areas. Especially for the matrix representation, which can be helpful for getting an overview and does not introduce occlusion with an increasing network complexity, there is still a lack of adequate 3D representations exploiting the S3D design space without introducing new perceptual drawbacks. Thus, representations like the matrix visualization should be also included in future work dealing with network comparison in a 3D environment.

## Acknowledgments

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 251654672 – TRR 161.

## References


[1] B. Alper, B. Bach, N. Henry Riche, T. Isenberg, and J.-D. Fekete. Weighted Graph Comparison Techniques for Brain Connectivity Analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pp. 483–492. ACM, New York, NY, USA, 2013. doi: 10.1145/2470654.2470724

[2] 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. doi: 10.1109/TVCG.2010.78

[3] D. Archambault and H. C. Purchase. Mental Map Preservation Helps User Orientation in Dynamic Graphs. In W. Didimo and M. Patrignani, eds., *International Symposium on Graph Drawing*, pp. 475–486. Springer, 2012. doi: 10.1007/978-3-642-36763-2_42

[4] D. Archambault, H. C. Purchase, and B. Pinaud. Difference Map Readability for Dynamic Graphs. In U. Brandes and S. Cornelsen, eds., *Graph Drawing*, pp. 50–61. Springer, 2011. doi: 10.1007/978-3-642-18469-7_5

[5] C. Armbrüster, M. Wolter, T. Kuhlen, W. Spijkers, and B. Fimm. Depth Perception in Virtual Reality: Distance Estimations in Peri- and Extrapersonal Space. *CyberPsychology & Behavior*, 11:9–15, 03 2008. doi: 10.1089/cpb.2007.9935

[6] B. Bach, P. Dragicevic, D. Archambault, C. Hurter, and S. Carpendale. A Review of Temporal Data Visualizations Based on Space-Time Cube Operations. In R. Borgo, R. Maciejewski, and I. Viola, eds., *EuroVis - STARs*. The Eurographics Association, 2014. doi: 10.2312/eurovisstar.20141171

[7] B. Bach, E. Pietriga, and J. Fekete. GraphDiaries: Animated Transitions and Temporal Navigation for Dynamic Networks. *IEEE Transactions on Visualization and Computer Graphics*, 20(5):740–754, 2014. doi: 10.1109/TVCG.2013.254

[8] B. Bach, E. Pietriga, and J.-D. Fekete. Visualizing Dynamic Networks with Matrix Cubes. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pp. 877–886, 2014. doi: 10.1145/2556288.2557010

[9] V. Batagelj, F. J. Brandenburg, W. Didimo, G. Liotta, P. Palladino, and M. Patrignani. Visual Analysis of Large Graphs Using (X,Y)-Clustering and Hybrid Visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 17(11):1587–1598, 2011. doi: 10.1109/TVCG.2010.265

[10] G. D. Battista, F. Frati, M. Patrignani, and M. Tais. Schematic Representation of Large Biconnected Graphs. In *Proceedings Graph Drawing and Network Visualization 2020*. Springer, 2020. doi: 10.7155/jgaa.00560

[11] F. Beck, M. Burch, S. Diehl, and D. Weiskopf. A Taxonomy and Survey of Dynamic Graph Visualization. *Computer Graphics Forum*, 36(1):133–159, 2017. doi: 10.1111/cgf.12791

[12] M. Behrisch, B. Bach, N. Henry Riche, T. Schreck, and J.-D. Fekete. Matrix Reordering Methods for Table and Network Visualization. *Computer Graphics Forum*, 35(3):693–716, 2016. doi: 10.1111/cgf.12935

[13] U. Brandes, T. Dwyer, and F. Schreiber. Visual Understanding of Metabolic Pathways Across Organisms Using Layout in Two and a Half Dimensions. *Journal of Integrative Bioinformatics*, 1(1):11–26, 2004. doi: 10.1515/jib-2004-2

[14] W. Büschel, S. Vogt, and R. Dachselt. Augmented Reality Graph Visualizations. *IEEE Computer Graphics and Applications*, 39(3):29–40, 2019. doi: 10.1109/MCG.2019.2897927

[15] H. Chen, U. Soni, Y. Lu, V. Huroyan, R. Maciejewski, and S. G. Kobourov. Same Stats, Different Graphs: Exploring the Space of Graphs in Terms of Graph Properties. *IEEE Trans. Vis. Comput. Graph.*, 27(3):2056–2072, 2021. doi: 10.1109/TVCG.2019.2946558

[16] M. Chimani, C. Gutwenger, M. Jünger, G. W. Klau, K. Klein, and P. Mutzel. The Open Graph Drawing Framework (OGDF). *Handbook of graph drawing and visualization*, 2011:543–569, 2013.

[17] 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. doi: 10.1109/TVCG.2016.2599107

[18] W. Cui, X. Wang, and N. H. Riche. Let It Flow: A Static Method for Exploring Dynamic Graphs. *IEEE Pacific Visualization Symposium*, pp. 121–128, 2014. doi: 10.1109/PacificVis.2014.48

[19] F. De Vico Fallani, V. Latora, and M. Chavez. A Topological Criterion for Filtering Information in Complex Brain Networks. *PLOS Computational Biology*, 13(1):1–18, 01 2017. doi: 10.1371/journal.pcbi.1005305

[20] F. De Vico Fallani, J. Richiardi, M. Chavez, and S. Achard. Graph analysis of functional brain networks: practical issues in translational neuroscience. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 369(1653), 2014. doi: 10.1098/rstb.2013.0521

[21] N. Diersch and T. Wolbers. The potential of virtual reality for spatial navigation research across the adult lifespan. *Journal of Experimental Biology*, 222(Suppl 1), 2019. doi: 10.1242/jeb.187252

[22] K. Dinkla, M. A. Westenberg, and J. J. van Wijk. Compressed Adjacency Matrices: Untangling Gene Regulatory Networks. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2457–2466, 2012. doi: 10.1109/TVCG.2012.208

[23] A. Drogemuller, A. Cunningham, J. Walsh, B. H. Thomas, M. Cordeil, and W. Ross. Examining virtual reality navigation techniques for 3D network visualisations. *Journal of Visual Languages & Computing*, 56:100937, 2020. doi: 10.1016/j.cola.2019.100937

[24] A. Fonnet and Y. Prié. Survey of Immersive Analytics. *IEEE Transactions on Visualization and Computer Graphics*, 27(3):2101–2122, 2021. doi: 10.1109/TVCG.2019.2929033

[25] M. Ghoniem, J. D. Fekete, and P. Castagliola. A Comparison of the Readability of Graphs Using Node-Link and Matrix-Based Representations. In *IEEE Symp. on Information Visualization*, pp. 17–24, 2004. doi: 10.1109/INFVIS.2004.1

[26] M. Gleicher. Considerations for Visualizing Comparison. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):413–423, 2018. doi: 10.1109/TVCG.2017.2744199

[27] M. Gleicher, D. Albers, R. Walker, I. Jusufi, C. D. Hansen, and J. C. Roberts. Visual comparison for information visualization. *Information Visualization*, 10(4):289–309, 2011. doi: 10.1177/1473871611416549

[28] N. Greffard, F. Picarougne, and P. Kuntz. Visual Community Detection: An Evaluation of 2D, 3D Perspective and 3D Stereoscopic Displays. In M. van Kreveld and B. Speckmann, eds., *Graph Drawing*, pp. 215–225. Springer, 2012. doi: 10.1007/978-3-642-25878-7_21

[29] T. Hagerstrand. Diorama, path and project. *Tijdschrift voor economische en sociale geografie*, 73(6):323–339, 1982. doi: 10.1111/j.1467-9663.1982.tb01647.x

[30] S. G. Hart and L. E. Staveland. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. A. Hancock and N. Meshkati, eds., *Human Mental Workload*, vol. 52 of *Advances in Psychology*, pp. 139–183. North-Holland, 1988. doi: 10.1016/S0166-4115(08)62386-9

[31] N. Henry, J. Fekete, and M. J. McGuffin. NodeTrix: a Hybrid Visualization of Social Networks. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1302–1309, 2007. doi: 10.1109/TVCG.2007.70582

[32] R. F. Hess, L. To, J. Zhou, G. Wang, and J. R. Cooperstock. Stereo Vision: The Haves and Have-Nots. *i-perception*, 6(3), 2016. doi: 10.1177/2041669515593028

[33] S. Jaeger, K. Klein, L. Joos, J. Zagermann, M. de Ridder, J. Kim, J. Yang, U. Pfeil, H. Reiterer, and F. Schreiber. Challenges for Brain Data Analysis in VR Environments. In *2019 IEEE Pacific Visualization Symposium (PacificVis)*, pp. 42–46, 2019. doi: 10.1109/PacificVis.2019.00013

[34] W. Javed and N. Elmqvist. Exploring the Design Space of Composite Visualization. In *2012 IEEE Pacific Visualization Symposium*, pp. 1–8, 2012. doi: 10.1109/PacificVis.2012.6183556

[35] Z. Jin, N. Chen, Y. Shi, W. Qian, M. Xu, and N. Cao. Trammelgraph: visual graph abstraction for comparison. *Journal of Visualization*, 2021. doi: 10.1007/s12650-020-00706-2

[36] M. John, M. Baumann, D. Schuetz, S. Koch, and T. Ertl. A Visual Approach for the Comparative Analysis of Character Networks in Narrative Texts. In *2019 IEEE Pacific Visualization Symposium (PacificVis)*, pp. 247–256, 2019. doi: 10.1109/PacificVis.2019.00037

[37] S. Kairam, D. MacLean, M. Savva, and J. Heer. GraphPrism: Compact Visualization of Network Structure. In *Advanced Visual Interfaces*, 2012. doi: 10.1145/2254556.2254651

[38] J. Kotlarek, O.-H. Kwon, K.-L. Ma, P. Eades, A. Kerren, K. Klein, and F. Schreiber. A Study of Mental Maps in Immersive Network Visualization. In *2020 IEEE Pacific Visualization Symposium (PacificVis)*, pp. 1–10. IEEE, 2020. doi: 10.1109/PacificVis48177.2020.4722

[39] M. Kraus, J. Fuchs, B. Sommer, K. Klein, U. Engelke, D. Keim, and F. Schreiber. Immersive Analytics with Abstract 3D Visualizations: A Survey. In *Computer Graphics Forum*. Wiley Online Library, 2021. doi: 10.1111/cgf.14430

[40] M. Kraus, K. Klein, J. Fuchs, D. A. Keim, F. Schreiber, and M. Sedlmair. The Value of Immersive Visualization. *IEEE computer graphics and*

*applications*, 41(4):125–132, 2021. doi: 10.1109/MCG.2021.3075258

[41] E. Kruijff, J. E. Swan, and S. Feiner. Perceptual issues in augmented reality revisited. In *2010 IEEE International Symposium on Mixed and Augmented Reality*, pp. 3–12, 2010. doi: 10.1109/ISMAR.2010.5643530

[42] O. Kwon, C. Muelder, K. Lee, and K. 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. doi: 10.1109/TVCG.2016.2520921

[43] E. Kypridemou, M. Zito, and M. Bertamini. The Effect of Graph Layout on the Perception of Graph Properties. In *EuroVis 2020 - Short Papers*. The Eurographics Association, 2020. doi: 10.2312/evs.20201039

[44] O. Macindoe and W. Richards. Graph Comparison Using Fine Structure Analysis. In *Proceedings of the 2010 IEEE Second International Conference on Social Computing*, SOCIALCOM '10, pp. 193–200. IEEE Computer Society, USA, 2010. doi: 10.1109/SocialCom.2010.35

[45] K. Marriott, F. Schreiber, T. Dwyer, K. Klein, N. Henry Riche, T. Itoh, W. Stuerzlinger, and B. H. Thomas, eds. *Immersive Analytics*, vol. 11190 of *LNCS*. Springer, 2018. doi: 10.1007/978-3-030-01388-2

[46] J. P. McIntire, P. R. Havig, and E. E. Geiselman. Stereoscopic 3D displays and human performance: A comprehensive review. *Displays*, 35(1):18–26, 2014. doi: 10.1016/j.displa.2013.10.004

[47] S. Murugesan, K. Bouchard, J. Brown, M. Kiran, D. Lurie, B. Hamann, and G. H. Weber. State-based network similarity visualization. *Information Visualization*, 19(2):96–113, 2020. doi: 10.1177/1473871619882019

[48] C. Nobre, M. Streit, M. Meyer, and A. Lex. The State of the Art in Visualizing Multivariate Networks. *Computer Graphics Forum (EuroVis)*, 38:807–832, 2019. doi: 10.1111/cgf.13728

[49] M. Okoe, R. Jianu, and S. G. Kobourov. Node-link or Adjacency Matrices: Old Question, New Insights. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–1, 2018. doi: 10.1109/TVCG.2018.2865940

[50] B. Pester, O. Winke, C. Ligges, R. Dachselt, and S. Gumhold. Immersive 3D Visualization of Multi-Modal Brain Connectivity. In K. Vrotsou and J. Bernard, eds., *EuroVis Workshop on Visual Analytics (EuroVA)*. The Eurographics Association, 2021. doi: 10.2312/eurova.20211091

[51] D. Ren, L. R. Marusich, J. O'Donovan, J. Z. Bakdash, J. A. Schaffer, D. N. Cassenti, S. E. Kase, H. E. Roy, W.-y. S. Lin, T. Höllerer, and et al. Understanding node-link and matrix visualizations of networks: A large-scale online experiment. *Network Science*, 7(2):242–264, 2019. doi: 10.1017/nws.2019.6

[52] R. S. Renner, B. M. Velichkovsky, and J. R. Helmert. The Perception of Egocentric Distances in Virtual Environments - A Review. *ACM Computing Surveys (CSUR)*, 46(2):1–40, Dec. 2013. doi: 10.1145/2543581.2543590

[53] L. Sancho-Chavarría, C. Gómez-Soza, F. Beck, and E. Mata-Montero. Diaforá: A Visualization Tool for the Comparison of Biological Taxonomies. In J. L. Crespo-Mariño and E. Meneses-Rojas, eds., *High Performance Computing*, pp. 423–437. Springer International Publishing, Cham, 2020. doi: 10.1007/978-3-030-41005-6_29

[54] S. Schöttler, Y. Yang, H. Pfister, and B. Bach. Visualizing and Interacting with Geospatial Networks: A Survey and Design Space. *Computer Graphics Forum*, 40(6):5–33, 2021. doi: 10.1111/cgf.14198

[55] F. Schreiber. Comparison of Metabolic Pathways Using Constraint Graph Drawing. In *Proceedings of the First Asia-Pacific Bioinformatics Conference on Bioinformatics 2003 - Volume 19*, APBC '03, pp. 105–110. Australian Computer Society, Inc., 2003.

[56] S. S. Shapiro and M. B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611, 1965. doi: 10.1093/biomet/52.3-4.591

[57] R. Sharan, S. Suthram, R. M. Kelley, T. Kuhn, S. McCuine, P. Uetz, T. Sittler, R. M. Karp, and T. Ideker. Conserved patterns of protein interaction in multiple species. *Proceedings of the National Academy of Sciences*, 102(6):1974–1979, 2005. doi: 10.1073/pnas.0409522102

[58] R. Skarbez, N. F. Polys, J. T. Ogle, C. North, and D. A. Bowman. Immersive Analytics: Theory and Research Agenda. *Frontiers in Robotics and AI*, 6:82, 2019. doi: 10.3389/frobt.2019.00082

[59] U. Soni, Y. Lu, B. Hansen, H. C. Purchase, S. Kobourov, and R. Maciejewski. The Perception of Graph Properties in Graph Layouts. *Computer Graphics Forum*, 37(3):169–181, 2018. doi: 10.1111/cgf.13410

[60] J. Sorger, M. Waldner, W. Knecht, and A. Arleo. Immersive Analytics of Large Dynamic Networks via Overview and Detail Navigation. In *2nd International Conference on Artificial Intelligence & Virtual Reality*, pp. 144–151. IEEE, 2019. doi: 10.1109/AIVR46125.2019.00030

[61] R. Tamassia. *Handbook of Graph Drawing and Visualization*. Chapman & Hall/CRC, 1st ed., 2016.

[62] A. Vogogias, D. Archambault, B. Bach, and J. Kennedy. Visual Encodings for Networks with Multiple Edge Types. In *Proceedings of the International Conference on Advances Interfaces (AVI)*, 2020. doi: 10.1145/3399715.3399827

[63] C. Ware and G. Franck. Evaluating Stereo and Motion Cues for Visualizing Information Nets in Three Dimensions. *ACM Transactions on Graphics*, 15(2):121–140, 1996. doi: 10.1145/234972.234975

[64] C. Ware and P. Mitchell. Visualizing Graphs in Three Dimensions. *ACM Transactions on Applied Perception*, 5:1–15, 2008. doi: 10.1145/1279640.1279642

[65] M. Whitlock, S. Smart, and D. A. Szafir. Graphical Perception for Immersive Analytics. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 616–625. IEEE, 2020. doi: 10.1109/VR46266.2020.00084

[66] R. Xu, M. M. Thomas, A. Leow, O. A. Ajilore, and A. G. Forbes. TempoCave: Visualizing Dynamic Connectome Datasets to Support Cognitive Behavioral Therapy. In *2019 IEEE Visualization Conference (VIS)*, pp. 186–190. IEEE, 2019. doi: 10.1109/VISUAL.2019.8933544

[67] V. Yoghourdjian, T. Dwyer, K. Klein, K. Marriott, and M. Wybrow. Graph Thumbnails: Identifying and Comparing Multiple Graphs at a Glance. *IEEE Transactions on Visualization and Computer Graphics*, 24(12):3081–3095, 2018. doi: 10.1109/TVCG.2018.2790961