CHORDination: Evaluating Visual Design Choices in Chord Diagrams for Network Data

Kai Wang*
Xi'an Jiaotong-Liverpool University
Suzhou, China

Shuqi He*
Xi'an Jiaotong-Liverpool University
Suzhou, China

Wenlu Wang
Xi'an Jiaotong-Liverpool University
Suzhou, China

Jinbei Yu
Xi'an Jiaotong-Liverpool University
Suzhou, China

Yu Liu
Xi'an Jiaotong-Liverpool University
Suzhou, China

Lingyun Yu†
Xi'an Jiaotong-Liverpool University
Suzhou, China

ABSTRACT

Chord diagrams are widely used for visualizing data connectivity and flow between nodes in a network. They are effective for representing complex structures through an intuitive and visually appealing circular layout. While previous work has focused on improving aesthetics and interactivity, the influence of fundamental design elements on user perception and information retrieval remains under-explored. In this study, we explored the three primary components of chord diagram anatomy, namely the nodes, circular outline, and arc connections, in three sequential experiment phases. In phase one, we conducted a controlled experiment (N=90) to find the perceptually and information optimized node widths (narrow, medium, wide) and quantities (low, medium, high). This optimal set of node width and quantity sets the foundation for subsequent evaluations and were kept fixed for consistency. In phase two of the study, we conducted an expert design review for identifying the optimal radial tick marks and color gradients. Then in phase three, we evaluated the perceptual and information retrieval performance of the design choices in a controlled experiment (N=24) by comparing four chord diagram designs (baseline, radial tick marks, arc color gradients, both tick marks and color gradients). Results indicated that node width and quantity significantly affected users' information retrieval performance and subjective ratings, whereas the presence of tick marks predominantly influenced subjective experiences. Based on these findings, we discuss the design implications of these visual elements and offer guidance and recommendations for optimizing chord diagram designs in network visualization tasks.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in visualization**; **Visualization design and evaluation methods**.

---

*The first two authors contributed equally to this research.
†Corresponding author

KEYWORDS

Chord diagram, Network data, User perception

1 INTRODUCTION

Network data is ubiquitous across many research domains, including social networks [1, 33, 39], biological systems [14], and transportation systems [12, 45]. These complex network scenarios require effective visualization techniques to represent the interconnected relationships and flows. Among various techniques, chord diagrams stand out for their intuitive circular layout and ability to display bidirectional relationships compactly [2].

The anatomy of a chord diagram consists of several key components (Figure 1). The *circular outline* forms the backbone of a chord diagram, providing a structural foundation. Data entities are represented as segments called *nodes* along this outline. Chords, or *arc connections*, are the links that connect between nodes. These elements, combined with color and text labels, create visually appealing and space-efficient designs for network visualization.

Despite their popularity, chord diagrams have some known drawbacks such as visual clutter and difficulty in accurately perceiving connection weights due to overlapping connections. Gutwin et al.'s work [15] suggested that chord diagrams are less effective than its alternative sankey diagrams. However, how the design variations of chord diagrams influence user perception and information seeking performance remains under-explored. Most prior research has focused on improving the aesthetic and interactive features of these diagrams rather than systematically evaluating the impact of basic design elements [33]. For example, Haghnazar et al. [26], Kakaraparty [23] and Kriebel [28] employed ribbon design and color design for aesthetic improvements. Finnegan et al. [14] and Kriebel [28] implemented hover highlighting for interactivity.

To address this research gap, we chose to examine one primary design consideration per component: the **node width and quantity** of the node segments, the use of **color gradients** on the arc connections, and the presence of **radial tick marks** on the circular outline. Specifically, we formulated the following research questions:

- **RQ1**: How do variations in node width and quantity affect the readability of chord diagrams?
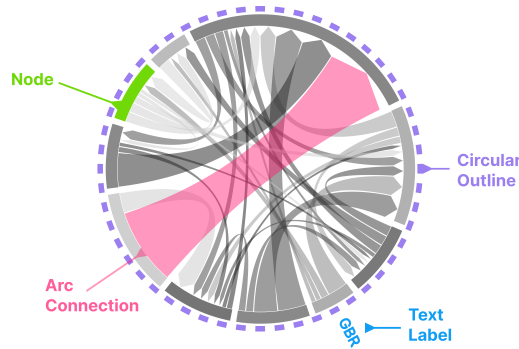
Figure 1: The anatomy of a chord diagram, with key elements highlighted and labeled.

- **RQ2**: What are the key design considerations for color gradients and tick marks when added in chord diagrams?
- **RQ3**: How do color gradients and radial tick marks influence the perception of chord diagrams?

We designed three experimental phases to investigate these factors and derive optimal design choices. The first phase focused on determining the optimal node parameters for presentation and foundational readability (N=90). The second phase aimed to narrow down the tick marks and color gradients design with expert review. The third phase examined the effects of the pre-selected radial tick marks and color gradients designs (N=24). Ultimately, CHORDination, aims to derive practical design guidelines by systematically investigating how design components influence user perception and information acquisition in chord diagrams.

2 RELATED WORK

2.1 Network Data Visualization

As an important branch of information visualization, network data visualization shows data relationships without lengthy explanations. Techniques include matrix charts, node-link diagrams, word clouds, and alluvial diagrams [20]. Network data is typically represented by nodes and edges, where nodes represent entities and edges represent relationships between these entities [36].

Node-link diagrams use geometric shapes to represent nodes and lines to represent edges [13]. For instance, edges are typically drawn as straight lines in social network visualizations, whereas more complex geometries and curved edges are used in dense networks to reduce occlusion [10]. However, node-link diagrams are criticized for producing visual clutter with complex data [24].

Adjacency matrix representations are particularly suitable for identifying clusters and communities within networks [13, 35]. However, as matrices do not directly visualize paths or connections, they may be less straightforward than node-link diagrams [5].

Node-link diagrams and matrices can be combined together, allowing for the display of both the global structure and local details of networks [13]. This method is advantageous in handling complexity in the inter-community structures and local densities [4]. However, implementing it requires more sophisticated algorithms and computational resources [3].

2.2 Chord Diagrams for Network Data

Chord diagrams originated as a variant of Cartesian graphs, known as radial diagrams [6]. Its circular appearance offers better scalability [21] to accommodate more data within the same space. Additionally, due to the centralized nature of circular diagrams, users' visual attention tends to focus on the center of the circle [7, 17]. In a chord diagram, links can be either bidirectional or unidirectional. For unidirectional links, the direction of the chord represents the flow of data, while bidirectional links are more complex. This experiment adopts unidirectional links for simplicity.

2.3 Visual Design Choices

Node quantity is a critical design consideration for network visualizations. In a comprehensive evaluation by Komarek et al. [25], it was tested that chord diagrams can display up to 100 sets of data while maintaining aesthetics and readability.

Moreover, different aspects of color choices can enhance chord diagram readability. Specific color choices can be reserved for encoding anomalies or specific features [22]. Additionally, color settings such as transparency [31] and brightness [29] can be adjusted to reveal overlapping elements and highlight differences.

More recently, tick marks have been studied in visualizations for providing a visual reference and enhancing accuracy estimation. For example, Kosslyn et al. [27] suggested that adding dense tick marks can calibrate axis. Likewise, it has been demonstrated in donut charts that tick marks can improve accuracy of estimating proportions [7]. Teng-Yun Ch et al. [9] used outer ring tick marks for facilitating value comparisons in chord diagrams.

2.4 Relevant Tasks & Performance Metrics

Chord diagrams are evaluated through tasks related to visual search, numerical relationship comparison, and path following. Studies suggested that strong color contrasts and consistent layouts avoiding acute-angle crossings help users locate information quickly and improves readability [38, 44]. Visual comparison is crucial in data analysis as users need to draw meaningful conclusions from comparing quantities [40]. Clear and easily comparable graphic designs such as aligned bar charts or proportionate pie charts can help users perform quantity comparisons more accurately [11]. Path following in flow diagrams involves tracing links between entities. Using straight lines instead of curves and reducing line crossings can significantly improve path-following accuracy [19, 43].

User performance metrics are essential for assessing the effectiveness of data visualization design. Studies commonly use completion time [8, 18] and error rate [34, 37, 42] to measure user's efficiency and accuracy in performing tasks. Subjective evaluations are also important for assessing user experience. The NASA Task Load Index (NASA TLX) is widely used for assessing workload [16]. Additionally, user satisfaction questionnaires offer direct feedback and emotional responses regarding the visualization tools [30].

3 STUDY DESIGN AND METHODOLOGY

3.1 Sequence of Experimental Phases

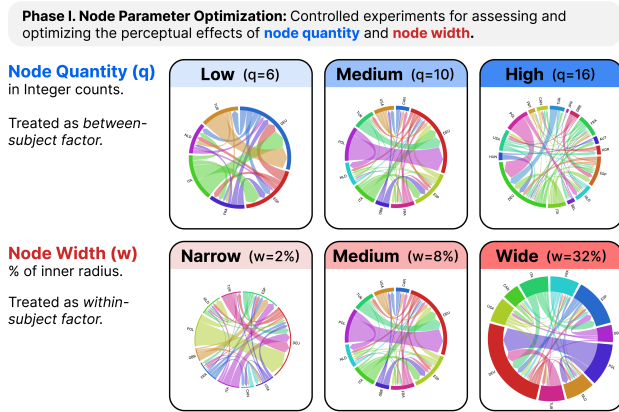The study consisted of three distinct phases of evaluations.

Figure 2: Experimental conditions of Phase I.

The first phase focused on **node parameter optimization**, which involved a controlled experiment aimed at determining the optimal combination of node width and node quantity. Three node width conditions (narrow, medium, wide) and three node quantity conditions (low, medium, high) were evaluated. Participants performed readability and information retrieval tasks using chord diagrams with varying node width and quantity configurations. The results from this experiment informed the selection of the optimal node width and quantity settings for subsequent phases.

The second phase, an **expert design review**, aimed for eliciting feedback on the design of tick marks and color gradients for chord diagrams. A diverse set of design variations were created and presented to five visualization experts. Their feedback and preference were recorded for guiding the selection of a single optimal tick mark and color gradient scheme for further evaluation in phase three.

The final phase, **design choices evaluation**, investigated the effects of the selected radial tick marks and color gradients design using a controlled user study. Participants performed tasks using chord diagrams with the optimal node width and quantity settings from the first phase, combined with the tick mark and color gradient designs chosen based on the expert review session.

3.2 Dataset and Tasks

For the study, we utilized a migration dataset [41], a typical example of network data visualized with chord diagrams. This dataset was selected for its social relevance and practical real-world application. This dataset provides information on the number of people migrating between countries over specified periods. In the created sample chord diagrams, countries of origin and destination were represented as nodes with three-letter acronym text labels. Migration events were depicted as chord connections between nodes, and the direction of migration was shown with arrows on the arc connections. These visual elements were consistently applied throughout the study.

We assessed the chord diagram designs using five tasks that represent typical analysis goals for network data (Figure 3), adapted from Gutwin et al.'s study [15]. Tasks ranged from basic retrievals to complex comparisons, presented in a counter-balanced order to mitigate order effects.
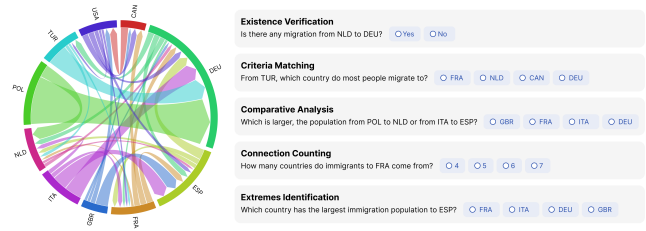

Figure 3: Summary of tasks. Left: A sample chord diagram stimuli. Right: Five types of task questions.

- **Existence verification**: Identifying if a specific node or connection existed. For example, determining whether a connection existed between the United States (USA) and Canada (CAN).
- **Criteria matching**: Identifying a node or connection that matched a specific criterion, such as finding the country with the most incoming migration.
- **Comparative analysis**: Comparing two elements, such as determining which connection represented a larger migration flow between two countries.
- **Connection counting**: Determining the number of incoming or outgoing connections associated with a particular node.
- **Extremes identification**: Identifying the nodes or connections with the maximal or minimal quantities, such as the country with the highest or the lowest net immigration.

3.3 Chord Diagram Generation and Variation

Baseline Chord Diagram Template. We used a consistent baseline template in D3.js for all diagrams, following standard conventions with color encoding for countries and directional arrows for migration flows [32]. To mitigate learning effects, the radial arrangement of the nodes was randomized across tasks. The baseline diagram was modified to create different variants for experimental conditioning.

Varying Node Parameters. The number of nodes can impact a chord diagram's complexity. To optimize the quantity of nodes, we defined three experimental conditions: *low (6 nodes)*, *medium (10 nodes)*, and *high (16 nodes)* (Figure 2 top row). The number of countries in the dataset was varied based on the baseline template while keeping other parameters identical.

Similarly, we categorized three levels of node widths: *narrow*, *medium* and *wide* (Figure 2 bottom row). These widths were quantified in increments relative to the inner radius of the chord diagram's circular outline. The *narrow* condition was set as 2% of the inner radius, the *medium* condition was set as 8%, and the *wide* condition was set to 32%. These increments, set at four-fold increases, were designed to create clear visual distinctions among the levels. In our experiments, node quantity was a between-subject factor, while node width was a within-subject factor.

Varying Color Gradients. To assess how color gradients impact the interpretability of chord diagrams, we created four gradient variations (Figure 4 top row) for expert review.

The *transparency gradient* linearly adjusts the opacity of arc connections to indicate direction. As the data flows from the origin to the destination, the opacity faded from 75% to 50%.
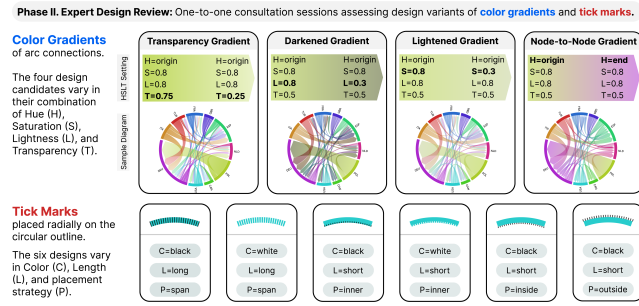
Figure 4: Experimental Condition of Phase II.

In *darkened gradient*, the color of the arc connections transitions to a darker shade in lightness as the flow progresses from the origin to the destination.

The *lightened gradient* involves gradually lightening the color saturation of the arc connections as the flow moves from the origin to the destination.

The *node-to-node gradient* creates a linear color transition from the origin to the destination, with each arc starting with the origin node's color and gradually blending into the destination node's color.

Adding Radial Tick Marks. To evaluate data comparison and value reading, we integrated various tick marks into the circular outline of our baseline chord diagram template. We developed six tick mark designs (Figure 4 bottom row), differing in color, length, and placement. The tick marks were designed in black and white, for providing contrast against the node segments. Some designs spanned the full node width, while others were shorter for visual subtlety. The tick marks were also placed in several different ways: inside the circular outline, along the inner edge of the circular outline, superimposed to span the node segments, along the outer edge of the circular outline, or at the outside of the circular outline.

3.4 Evaluation Methods

We adopted a hybrid evaluation methodology combining objective and subjective data. The study was approved by the university ethics committee prior to its conduct.

Controlled Experiments in Phase I & III. In Phases I (section 4) and III (section 6), quantitative questionnaires were designed to evaluate participants' viewing experiences with different chord diagram designs. The questionnaires consisted of three parts:

- **Practice Set:** Served as a training set and introduced the format of the questionnaire with a simplified sample chord diagram (10 nodes and 11 chords). Each page displayed a sample chord diagram at the top, followed by a multiple-choice question. Participants were required to answer correctly to proceed.
- **Objective Performance:** The second section extended the practice set to measure objective performance with various chord diagram stimuli. Five questions per task per experimental condition were presented. Performance metrics included the time taken to complete each question (from when it appeared on the screen to when the participant confirmed their answer correctly) and the counts of error occurrences.

- **Subjective Experience:** Participants completed the NASA Task Load Index (TLX) to and provided rankings and feedback on the diagram design based on ease of information retrieval, accuracy, and overall preference.

Qualitative Design Review in Phase II. Phase II involved qualitative consultation sessions with visualization experts (E1 to E5). One-on-one interviews, conducted in the experts' native languages, lasted 10-15 minutes each. The open-ended interviews began with a brief study overview, followed by a presentation of six design candidates. Subsequently, experts discussed the potential impact of these designs on the five information tasks and concluded by selecting one most-preferred color gradients and tick mark designs, respectively.

4 I. NODE PARAMETER OPTIMIZATION

In phase one, we recruited 112 participants from the university. After discarding 22 responses due to outlying completion times (over 15 minutes or less than one second per question), we had 90 valid responses (age: 25 ± 6; gender distribution: 52 females, 35 males, 2 others, 1 undisclosed). Among these participants, 6 were very familiar with data visualizations, 29 were familiar, 39 had some knowledge, and 16 were completely unfamiliar.

Completion time and error occurrences for each question were collected, aggregated and averaged by tasks, resulting in 1,350 objective measures (30 participants $\times$ 3 node quantities $\times$ 3 node widths $\times$ 5 tasks). Additionally, we gathered 270 sets of subjective measures from the NASA TLX and preference rankings (30 participants $\times$ 3 node quantities $\times$ 3 node widths). These data were analyzed with mixed ANOVA and Pearson Chi-Squared tests, respectively, with detailed results available in Appendix B.

4.1 Node Quantity Impacted Performance Metrics

Fewer nodes led to shorter completion time in existence verification and comparative analysis. As shown in Figure 5 (left), for existence verification tasks, low node quantity resulted in the shortest completion times compared to medium ($p < 0.001$) and high node quantities ($p = 0.024$). Similarly, for comparative analysis, low node quantity resulted in significantly shorter completion times than medium node quantity ($p = 0.042$). No significant effects in completion time were found for criteria matching, extremes identification, or connection counting across different node quantities.

Medium node quantity resulted in fewer errors. As shown in Figure 5 (middle), node quantity significantly affected error occurrences ($F_{2,87} = 8.246$, $p < 0.001$). Both low and medium node quantities had significantly lower error occurrences compared to high node quantity ($p = 0.007$ and $p < 0.001$, respectively). Medium node quantity had the lowest mean error occurrences, though the difference between low and medium was not statistically significant.

4.2 Interaction Effects Between Node Width and Quantity in Subjective Experience

Statistically significant interaction effects between node quantities and node widths were identified for mental demand ($F_{4,174} = 2.564$, $p = 0.040$), physical demand ($F_{4,174} = 2.670$, $p = 0.034$), and perceived effort ($F_{4,174} = 2.540$, $p = 0.042$) (Figure 5 (right)).
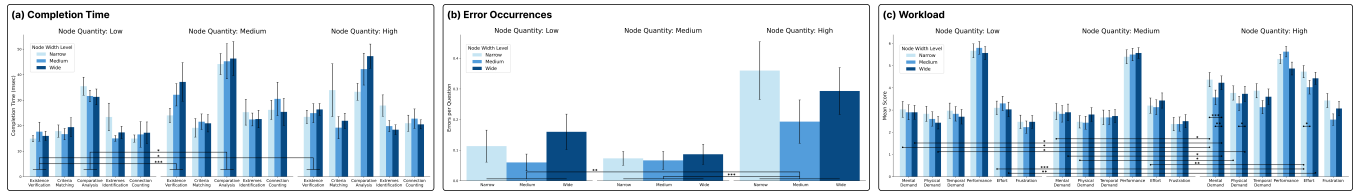
Figure 5: Phase I Results: Average completion time, error occurrences and subjective ratings of workload of different node quantities and widths across tasks. Significance levels: $p < .05$(*), $p < .01$(**), and $p < .001$(***).

Fewer Nodes Led to Lower Mental and Physical Demands. For narrow nodes, significantly lower mental demands were observed between low and high quantities ($p = 0.025$), as well as between medium and high quantities ($p = 0.014$). Besides, medium quantity showed a significantly lower physical demand to high quantity ($p = 0.013$). For wide nodes, low quantity also led to significantly lower mental demand ($p = 0.017$) and physical demand ($p = 0.012$) than high quantity. Similarly, medium quantity resulted in lower mental demand compared with high quantity ($p = 0.017$).

Lower Node Quantity Reduced Perceived Effort. Low node quantity with narrow node width led to significantly lower perceived effort compared to high node quantity ($p < 0.001$) and medium node quantity ($p = 0.002$). For wide node width, low node quantity resulted in significantly lower perceived effort compared to high node quantity ($p = 0.007$).

Medium Node Widths Resulted in Less Frustration. Node width made a significant difference in frustration ($p = 0.019$). Post hoc analysis revealed medium width led to significantly lower frustration compared to narrow nodes ($p = 0.016$).

Medium Node Widths Reduced Mental, Physical Demands and Effort with High Node Quantity. For high node quantity, medium node width consistently resulted in lower workload. Medium width led to significantly lower mental demand compared to narrow ($p < 0.001$) and wide ($p = 0.005$). Similarly, medium width reduced physical demand for high node quantity compared to wide nodes ($p = 0.043$). Additionally, medium width resulted in significantly lower perceived effort when compared to narrow nodes ($p = 0.014$).

4.3 Preferences and Qualitative Feedback

Over 56% of participants preferred the medium width for better speed (56.7%, N = 51), better accuracy (60%, N = 54), and overall perception (62.2%, N = 56). No significant association was found between the preference rankings of node quantities and widths. The participants mentioned that the node width was particularly important in comparative tasks, especially when the flows were of similar size. One participant commented, "It is tough to compare the smaller ones (flows); mostly it's a guess which one is larger". Participants with more nodes felt that the many links in the chord diagrams increased their visual burden. One said, "If a country has too many migration lines, it visually becomes a bit chaotic."

In summary, fewer nodes led to faster completion times for tasks like existence verification and comparative analysis, showing that simplifying visual complexity improves information retrieval speeds without losing accuracy. Medium node quantities reduced errors,

suggesting a good balance between detail for accurate analysis and reduced cognitive load. Interactions between node quantities and widths also affected perceived workload, highlighting how these elements together influence user experience. Overall, medium node width was favored for its speed, accuracy, and overall perception. Therefore, we proceeded with medium node width and medium node quantity in the subsequent phases.

5 II. EXPERT DESIGN REVIEW

We invited five visualization experts to provide design insights on color gradients and tick marks (Appendix Figure 1).

Evaluation of Color Gradients. Three experts favored transparency gradient for its effectiveness across various tasks. They highlighted that changes in transparency maintained a high level of stylistic consistency and visual effects. E1 noted, "The transparency gradient just looks overall brighter".

The darkened gradient was less favored due to its dimmer appearance when blending multiple colors. The lightened hue gradient received mixed reviews. While some experts appreciated its visual appeal, others were concerned about its reduced saturation in dense diagrams. The node-to-node gradient was less favored because of the potential visual complexity when multiple colors were involved. Overall, the consensus leaned towards transparency gradients for their clarity and readability.

Evaluation of Tick Mark Designs. All experts agreed that the tick mark colors should have high contrast against the background to ensure readability. White tick marks   were thought to provide better legibility against blue and purple backgrounds.

On the other hand, opinions diverged on the placement strategies of the tick marks. Two experts preferred tick marks spanning the entire node width for consistent visual cues . However, one expert criticized this approach because this placement "made the nodes appear disjointed" (E3). One expert applauded the placement approach inside the circular outline , arguing that placing the tick marks in proximity to the arc connections "reduced the effort required during comparison," (E2), yet two experts criticized it for "occupying additional space and overlapping with some arc connections." Two experts preferred white tick marks along the inner edge , which combines the advantages of proximity to the chords while not invading the space of the chords themselves. None of the experts favored placing the tick marks outside the circular outline due to concerns that it "occupies extra space and is too distant from the chords upon comparison" (E1, E2, E4). After synthesizing their opinions, we finalized the design choices as white tick marks

Figure 6: Experimental conditions of phase III.

along the inner edge of the circular outline. This combination was selected for clarity, contrast and readability.

6 III. DESIGN CHOICES EVALUATION

Building on findings from Phase I and II, we assessed chord diagram perception under four design configurations (Figure 6):

- **Baseline:** Chord diagram template with 10 nodes and medium node width, serving as reference before any design alterations.
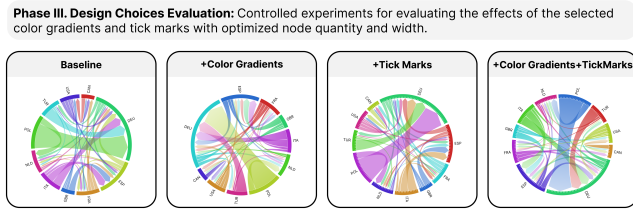- **Baseline + Color Gradient:** Arc connections were altered with a transparency gradient.
- **Baseline + Tick Marks:** White tick marks were added onto the nodes, extending outward from the inner edge of the nodes to $\frac{1}{3}$ of node length.
- **Baseline + Color Gradients + Tick Marks:** Combined color gradients and tick mark designs with the baseline.

We recruited 24 participants from the university (age: 24 ± 2; gender distribution: 13 females, 11 males), all of whom had normal visual acuity and varying experience with visualization (1 very familiar, 10 familiar, 13 somewhat knowledgeable). In total, we collected 480 sets of task completion time and error occurrences data (24 participants × 4 conditions × 5 tasks), and 96 sets of subjective measures on workload and preference (24 participants × 4 conditions). Additionally, we compiled interview records totaling 63 minutes. While we acknowledge the relatively small sample size, this number was chosen to balance statistical power with the ability to conduct in-depth interviews for rich qualitative insights into user experience. For data analysis, we conducted the Friedman test due to non-normality of the dependent variables across the four conditions.

Despite rigorous experimental design, the Phase III results (Figure 7) did not reveal statistically significant differences in completion time, error occurrences, or workload among the four conditions ($p > 0.1$) (Appendix C). This outcome may suggest several key points about the experimental design and metric selection.

Factors Influencing Experimental Sensitivity. The lack of statistically significant results could be due to several factors. Firstly, the sensitivity and specificity of the chosen metrics, particularly completion time and error rates, may not have effectively captured the subtle effects of design changes on user performance. This metric granularity may have limited the chance to accurately reflect the cognitive processes involved in interpreting chord diagrams. Additionally, the visual tasks used may not have been distinct enough to show noticeable performance differences among the design conditions. Future studies might benefit from either more pronounced design changes or a variety of tasks designed to highlight specific design features. Lastly, the homogeneity of the participant pool, marked by

similar backgrounds as students and age group, potentially diluted the observable impact of the design modifications. A more diverse group of participants could likely yield more significant differences to enhance the generalizability and sensitivity of the results.

Apart from the broader challenges in experimental sensitivity, several interesting trends emerged in the observation of key metrics.

Task-Specific Completion Times. The completion times varied across different tasks in a consistent manner. Notably, the *comparative analysis* tasks consistently required the longest completion times, potentially indicating that the inherent complexities of specific tasks dictate the time required for completion. This observation is consistent with the time recorded in Phase I, suggesting that complex tasks demand more time irrespective of design changes.

Error Occurrences Across Designs. The error occurrences peaked in the *color gradient* condition. However, after the addition of tick marks in the *combined* condition, the error occurrences returned to a lower level (Figure 7). This reduction may suggest that tick marks potentially provided visual cues that aided in the interpretation of the gradients, which almost rescued the initial increase in errors.

Preferences and Qualitative Feedback. Feedback on adding tick marks was generally positive. About 37.5% (N=9) of participants believed that chord diagrams with tick marks helped them find answers more quickly and accurately. Many participants (N=14) noted that tick marks allowed for more accurate comparisons of chord widths.

Some participants (N=6) noted areas for improvement. One mentioned that reading values from tick marks was not straightforward, suggesting, "It would be better to directly label the numeric values instead of making me count them." (P14). There were also issues with precision: "Some chords are narrower than the smallest tick mark, making it hard to use the marks to speed up comparisons" (P20). Additionally, tick marks did not always align perfectly with the chords, "Not all tick marks start at one end of the chord, which might introduce errors in comparisons" (P4).

A few participants (N=4) criticized tick marks for increasing cognitive load and distraction. One mentioned that "When not comparing tasks, tick marks distract my attention; having the option to display tick marks would be better"(P10).

Regarding the transparency gradients, some participants (N=4) did not notice their presence and questioned the difference between graphs with and without these gradients. Only four users explicitly expressed the benefits of transparency gradients for distinguishing flow direction, describing it as "a psychological hint" of directionality (P1). However, the majority of users (N=13) preferred a simpler design. As one participant (P11) pointed out, "Similar colors become indistinguishable after transparency gradients."

7 LIMITATIONS OF THE STUDY

Challenges in Color Gradients and Tick Marks Design. In the experiment, many users either did not notice the gradient colors or found them unhelpful. This may be due to the gradient scheme not being visually prominent enough. In future design iterations, more diverging and easily distinguishable gradient schemes should be considered. As for the tick marks, the added detail from tick marks might lead to unnecessary cognitive burdens in simpler tasks. Tick marks should be used selectively according to task complexity.
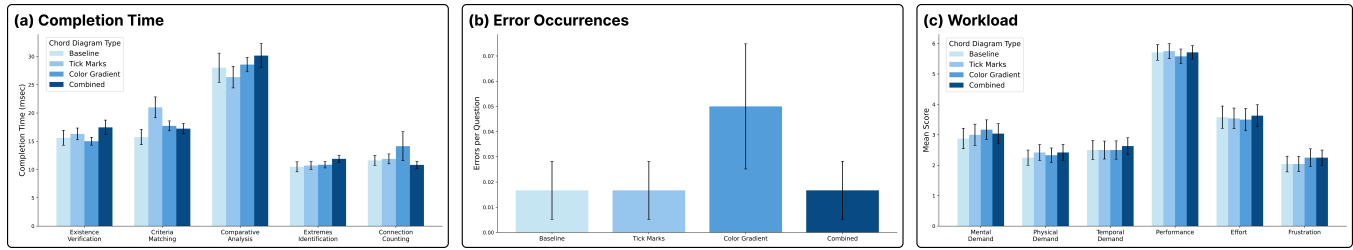
Figure 7: Phase III Results: Average completion time, error occurrences and subjective ratings of workload of different design choices across tasks. Significance levels: $p < .05$(*), $p < .01$(**), and $p < .001$(***).

Difficulty in Tasks and Questions. This study utilized five tasks to evaluate chord diagram designs. However, one particular task, comparative analysis, consistently took longer across different conditions. This indicates that our current tasks might not fully encompass the spectrum of difficulty or adequately represent all potential user interactions. To address this, future research should expand the range of tasks to more effectively assess design variants across varied difficulty levels, particularly those involving comparative numerical analysis. These questions are likely to place greater demands on the precision of tick marks and the visual clarity of the chord diagrams.

Dataset and Generalizability. In this study, we utilized a specific migration dataset for its real-life application as a type of network data. It is important to note that the findings may have limited generalizability due to the dataset's specific characteristics. Future studies could enhance the applicability of these results by incorporating a diverse range of datasets, including those with varying sizes, and complexities, thereby strengthening the validity of the recommendations for broader use in network data visualization.

8 PRACTICAL GUIDELINES

Finding the Sweet Spot: Optimizing Node Width and Quantity. Our findings suggest that using a medium number of nodes ($n = 10$) and medium widths generally offers the best readability and user satisfaction under the current experimental conditions. While the exact number may vary, the rule of thumb is to strike a balance between providing sufficient detail and avoiding visual clutter. When working with larger datasets, designers can consider implementing selective filtering to maintain this optimal balance. For instance, filtering functionalities enable users to focus on a selected subset of nodes at a time. An overview-detail approach can also be effective, presenting a full chord diagram for context alongside a more detailed view of selected nodes, where users can interpret data up close.

Navigating Directionality: Customizing Color Gradients and Beyond. When approaching the directional representation of flows, designers can consider using color gradients or alternative methods with flexibility and customization. For example, the study tested different styles of color gradients but not their mapping polarity. Experimenting with whether more vibrant colors are placed at the source or destination node could enhance intuitive understanding of data flow direction. Offering users the ability to customize the direction-color gradient mapping can also be beneficial. This allows users to adjust the visualization to fit their individual preferences.

Making Comparison Easier: Tick Marks Where It Counts. Tick marks enrich the user's subjective experience by offering clear visual references. These markers are especially useful in scenarios that demand data comparison. For tasks that require a broad overview or coarse comparisons, incorporating tick marks can help viewers quickly gauge differences between values. However, when precise data values are crucial and exact quantities are statistically relevant, designers can display numerical values directly on the diagram. Designers can explore different techniques for placing the tick marks or numerical values within a chord diagram: interactive elements such as hover tooltips, toggle switches or zooming features can be implemented to selectively reveal specific data when a user focuses on a segment or connection.

One Size Does Not Fit All: CHORDinating Design Elements. The interaction effects observed between node quantity and width, as well as the combined influence of tick marks and color gradients, demonstrate that different design elements can interact in complex ways that influence overall effectiveness of visualizations. Instead of simply layering design strategies, testing how different design elements work together holistically can reveal insights into how they influence user perception and task performance. For instance, while medium node widths generally improve readability, their effectiveness can be contingent on the node quantity present. Similarly, the benefits of color gradients in indicating direction can be amplified or obscured by how tick marks are implemented. Therefore, designers should conduct regular user testings and use iterative design processes to fine-tune how elements like node size, tick marks, and color gradients combine.

9 CONCLUSION

This study explored the key elements of chord diagram design and their impact on user perception and information acquisition. Through three experimental phases, we assessed the effects of node width and quantity, as well as radial tick marks and color gradients. Node width did not significantly affect performance metrics, but impacted subjective experiences. Medium node width was preferred by the majority of users. Increasing the number of nodes extended task completion time and increased error occurrences, especially in tasks involving comparison and existence verification. Tick marks improved the perceived accuracy of data interpretation, while color gradients, despite aiming to enhance understanding of data flow, had limited practical effects. Future research should optimize these to improve the usability and effectiveness of chord diagrams.

ACKNOWLEDGMENTS

REFERENCES

[1] Guy J. Abel and Nikola Sander. 2014. Quantifying Global International Migration Flows. *Science* 343 (2014), 1520 – 1522. https://doi.org/10.1126/science.1248676

[2] Inc. An Idera. 2024. FusionCharts: Chord Diagrams.

[3] Lorenzo Angori, Walter Didimo, Fabrizio Montecchiani, Daniele Pagliuca, and Alessandra Tappini. 2020. Hybrid Graph Visualizations With ChordLink: Algorithms, Experiments, and Applications. *IEEE Transactions on Visualization and Computer Graphics* 28 (2020), 1288–1300. https://doi.org/10.1109/TVCG.2020.3016055

[4] Vladimir Batagelj, Franz-Josef Brandenburg, Walter Didimo, Giuseppe Liotta, Pietro Palladino, and Maurizio Patrignani. 2010. Visual Analysis of Large Graphs Using (X,Y)-Clustering and Hybrid Visualizations. *IEEE Transactions on Visualization and Computer Graphics* 17 (2010), 1587–1598. https://doi.org/10.1109/TVCG.2010.265

[5] Michael Behrisch, Benjamin Bach, Nathalie Henry Riche, Tobias Schreck, and Jean-Daniel Fekete. 2016. Matrix Reordering Methods for Table and Network Visualization. *Computer Graphics Forum* 35 (2016), 693–716. https://doi.org/10.1111/cgf.12935

[6] Michael Burch and Daniel Weiskopf. 2014. *On the Benefits and Drawbacks of Radial Diagrams*. Springer New York, New York, NY, 429–451. https://doi.org/10.1007/978-1-4614-7485-2_17

[7] Xiao Cai, Konstantinos Efstathiou, Xiao Xuan Xie, Yingcai Wu, Y. Shi, Lingyun Yu, and Lingyun Yu. 2018. A Study of the Effect of Doughnut Chart Parameters on Proportion Estimation Accuracy. *Computer Graphics Forum* 37 (2018), 300–312. https://doi.org/10.1111/cgf.13325

[8] Stuart K. Card, Allen Newell, and Thomas P. Moran. 1983. *The Psychology of Human-Computer Interaction*. L. Erlbaum Associates Inc., USA.

[9] Teng-Yun Cheng, Sam Yu-Chieh Ho, Tsair-Wei Chien, and Willy Chou. 2023. Global research trends in artificial intelligence for critical care with a focus on chord network charts: Bibliometric analysis. *Medicine* 102, 38 (2023), e35082. https://doi.org/10.1097/MD.0000000000035082

[10] Se-Hang Cheong and Yain-Whar Si. 2019. Force-directed algorithms for schematic drawings and placement: A survey. *Information Visualization* 19 (2019), 65 – 91. https://doi.org/10.1177/1473871618821740

[11] William S. Cleveland and Robert McGill. 1985. Graphical Perception and Graphical Methods for Analyzing Scientific Data. *Science* 229 (1985), 828 – 833. https://doi.org/10.1126/science.229.4716.828

[12] Tori Clifford. 2018. Chord Diagrams: 5 Inspirational Chord Diagrams: Data Visualization.

[13] Walter Didimo, Giuseppe Liotta, and Fabrizio Montecchiani. 2024. *Chapter 1: Network data visualization*. Edward Elgar Publishing, Cheltenham, UK, 2 – 11. https://doi.org/10.4337/9781803921259.00007

[14] Amy Finnegan, Saumya S. Sao, and Megan J Huchko. 2019. Using a Chord Diagram to Visualize Dynamics in Contraceptive Use: Bringing Data Into Practice. *Global Health: Science and Practice* 7 (2019), 598 – 605.

[15] Carl Gutwin, Aristides Mairena, and Venkat Bandi. 2023. Showing Flow: Comparing Usability of Chord and Sankey Diagrams. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 825, 10 pages. https://doi.org/10.1145/3544548.3581119

[16] S. G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in psychology* 52 (1988), 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

[17] Shuqi He, Yuqing Chen, Yuxin Xia, Yichun Li, Hai-Ning Liang, and Lingyun Yu. 2024. Visual harmony: text-visual interplay in circular infographics. *Journal of Visualization* 27 (2024), 1–17. https://doi.org/10.1007/s12650-024-00957-3

[18] Jeffrey Heer and Michael Bostock. 2010. Crowdsourcing graphical perception: using mechanical turk to assess visualization design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. Association for Computing Machinery, New York, NY, USA, 203–212. https://doi.org/10.1145/1753326.1753357

[19] Danny Holten and Jarke J. van Wijk. 2010. Evaluation of cluster identification performance for different PCP variants. In *Proceedings of the 12th Eurographics / IEEE - VGTC Conference on Visualization* (Bordeaux, France) *(EuroVis'10)*. The Eurographs Association & John Wiley & Sons, Ltd., Chichester, GBR, 793–802. https://doi.org/10.1111/j.1467-8659.2009.01666.x

[20] Mohaiminul Islam and Shang Jin. 2019. An Overview of Data Visualization. *2019 International Conference on Information Science and Communications Technologies (ICISCT)* 7 (2019), 1–7.

[21] Shumaila Islam. 2021. Circle: a symbol of objective and subjective realms. *Journal of Immersive Media and Creative Arts* 1, 1 (2021), 79–114. https:

[22] Mikel Iturbe., Iñaki Garitano., Urko Zurutuza., and Roberto Uribeetxeberria. 2016. Visualizing Network Flows and Related Anomalies in Industrial Networks using Chord Diagrams and Whitelisting. In *Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2016) - IVAPP*. INSTICC, SciTePress, Rome, ITA, 99–106. https://doi.org/10.5220/0005670000990106

[23] Sashank Kakaraparty. 2020. Create beautiful and interactive Chord Diagrams using Python.

[24] Stephen G. Kobourov. 2013. Force-Directed Drawing Algorithms. In *Handbook of Graph Drawing and Visualization*. Chapman & Hall/CRC, USA, 382–408.

[25] Ales Komarek, Jakub Pavlik, and Vladimir Sobeslav. 2015. Network Visualization Survey. In *Computational Collective Intelligence*, Manuel Núñez, Ngoc Thanh Nguyen, David Camacho, and Bogdan Trawiński (Eds.). Springer International Publishing, Cham, 275–284. https://doi.org/10.1007/978-3-319-24306-1_27

[26] Roozbeh Haghnazar Koochaksaraei, Ivan Reinaldo Meneghini, Vitor Nazário Coelho, and Frederico Gadelha Guimarães. 2017. A new visualization method in many-objective optimization with chord diagram and angular mapping. *Knowledge-Based Systems* 138 (2017), 134–154. https://doi.org/10.1016/j.knosys.2017.09.035

[27] Stephen M. Kosslyn. 2006. *Graph Design for the Eye and Mind*. Oxford University Press, UK. https://doi.org/10.1093/acprof:oso/9780195311846.001.0001

[28] Andy Kriebel. 2024. How to create interactive chord diagrams.

[29] Yei-Soon Lee, Julie Chi Chow, Tsair-Wei Chien, and Willy Chou. 2023. Using chord diagrams to explore article themes in 100 top-cited articles citing Hirsch's h-index since 2005: a bibliometric analysis. *Medicine* 102, 8 (2023), e33057. https://doi.org/10.1097/MD.0000000000033057

[30] James R. Lewis. 2014. Usability: Lessons Learned . . . and Yet to Be Learned. *International Journal of Human–Computer Interaction* 30, 9 (2014), 663–684. https://doi.org/10.1080/10447318.2014.930311

[31] Johan Mazel, Romain Fontugne, and Kensuke Fukuda. 2014. Visual comparison of network anomaly detectors with chord diagrams. In *Proceedings of the 29th Annual ACM Symposium on Applied Computing* (Gyeongju, Republic of Korea) *(SAC '14)*. Association for Computing Machinery, New York, NY, USA, 473–480. https://doi.org/10.1145/2554850.2554886

[32] Vinh Xuan Nguyen. 2012. CircoSonic: A SONIFICATION OF CIRCOS, A CIRCULAR GRAPH OF TABLE DATA. In *Proceedings of the 18th International Conference on Auditory Display*. International Community for Auditory Display, Atlanta,GA,USA, 105–112.

[33] Michael Nicholas, Daniel Archambault, and Robert S Laramee. 2014. Interactive Visualisation of Automotive Warranty Data Using Novel Extensions of Chord Diagrams. In *EuroVis - Short Papers*, N. Elmqvist, M. Hlawitschka, and J. Kennedy (Eds.). The Eurographics Association, Strasbourg, France, 115–119. https://doi.org/10.2312/eurovisshort.20141167

[34] Donald A. Norman. 2013. *The Design of Everyday Things: Revised and Expanded Edition*. Basic Books, New York, NY, USA, Chapter 5, 137–158. To Err is Human.

[35] Mershack Okoe, Radu Jianu, and Stephen Kobourov. 2019. Node-Link or Adjacency Matrices: Old Question, New Insights. *IEEE Transactions on Visualization and Computer Graphics* 25, 10 (2019), 2940–2952. https://doi.org/10.1109/TVCG.2018.2865940

[36] Joshua O'Madadhain, Danyel Fisher, Padhraic Smyth, Scott White, and Yan-Biao Boey. 2005. Analysis and visualization of network data using JUNG. In *Journal of Statistical Software*, Vol. 10. Citeseer, USA, 1–35.

[37] Catherine Plaisant. 2004. The challenge of information visualization evaluation. In *Proceedings of the Working Conference on Advanced Visual Interfaces* (Gallipoli, Italy) *(AVI '04)*. Association for Computing Machinery, New York, NY, USA, 109–116. https://doi.org/10.1145/989863.989880

[38] Ruth Rosenholtz, Yuanzhen Li, Zhenlan Jin, and Jonathan Mansfield. 2010. Feature congestion: A measure of visual clutter. *Journal of Vision* 6 (2010), 827–827. https://doi.org/10.1167/6.6.827

[39] Nuwan I. Senaratna. 2020. World Bilateral Migration.

[40] David Simkin and Reid Hastie. 1987. An Information-Processing Analysis of Graph Perception. *J. Amer. Statist. Assoc.* 82, 398 (1987), 454–465. https://doi.org/10.1080/01621459.1987.10478448

[41] Stat technology | © OECD. 2022. International Migration Database.

[42] Edward R. Tufte. 2018. *The Visual Display of Quantitative Information*. Graphics Press, USA, Chapter 2, 51–77. Graphical Integrity.

[43] Colin Ware and Robert J. Bobrow. 2005. Supporting Visual Queries on Medium-Sized Node–Link Diagrams. *Information Visualization* 4 (2005), 49 – 58. https://doi.org/10.1057/palgrave.ivs.9500090

[44] Colin Ware, Helen C. Purchase, Linda Colpoys, and Matthew J. McGill. 2002. Cognitive Measurements of Graph Aesthetics. *Information Visualization* 1 (2002), 103 – 110. https://doi.org/10.1057/palgrave.ivs.9500013

[45] Wei Zeng, Chi-Wing Fu, Stefan Müller Arisona, Alexander Erath, and Huamin Qu. 2014. Visualizing Mobility of Public Transportation System. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 1833–1842. https://doi.org/10.1109/TVCG.2014.2346893

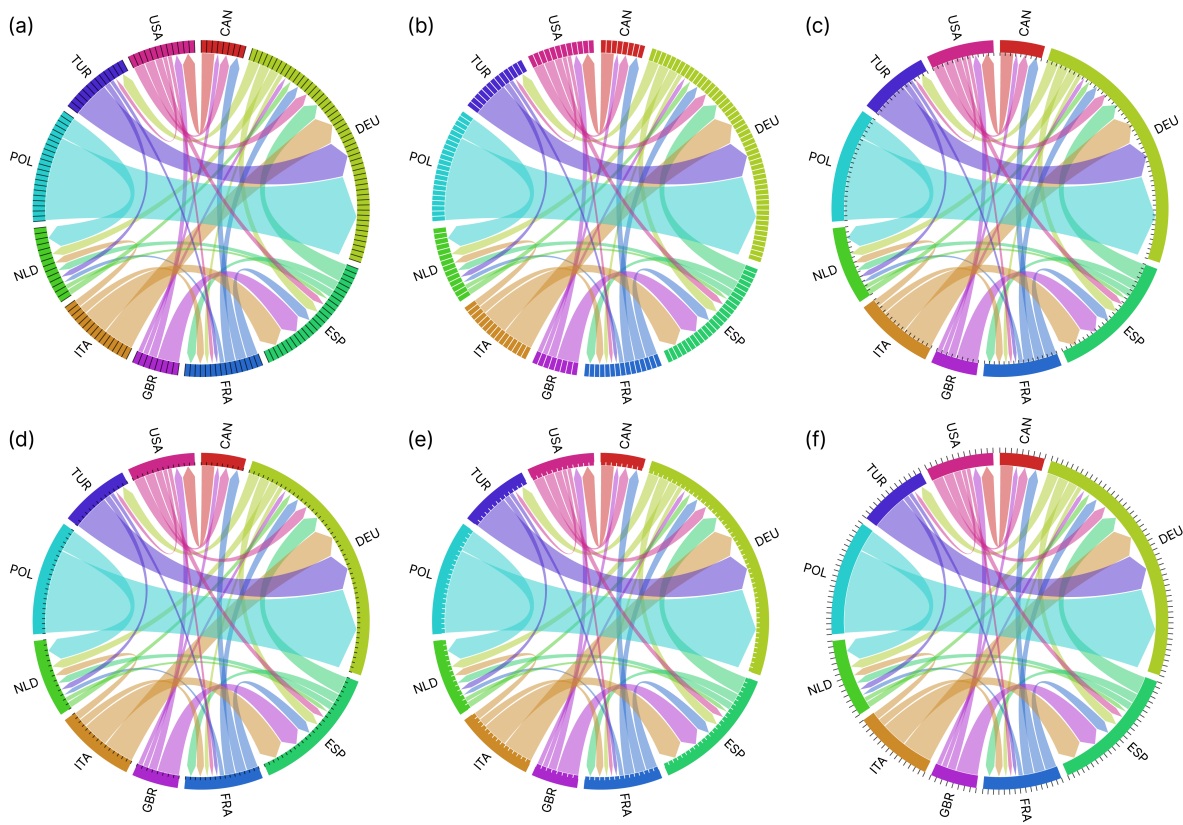# A   TICK MARKS DESIGN CANDIDATES IN PHASE II

Figure 1: Six design candidates of tick marks that were presented to the experts for review.

B SUMMARY OF STATISTICAL RESULTS FROM PHASE I

B.1 Completion Time For Different Node Quantities and Node Widths

| | | Low Node Quantity | | | Medium Node Quantity | | | High Node Quantity | | | ANOVA | | |
		Narrow	Medium	Wide	Narrow	Medium	Wide	Narrow	Medium	Wide	Interaction	Node Quantity	Node Width
Existence	Mean	14.915	17.638	15.981	24.011	32.109	37.141	23.444	24.912	26.353	$F_{(3.55,154.439)} = 0.792$	$F_{(2,87)} = 10.908$	$F_{(1.775,154.439)} = 1.871$
Verification	SE	1.244	3.727	1.819	2.805	4.391	7.578	2.609	3.722	2.340	$p = 0.519$	$\boldsymbol{p < 0.001}$	$p = 0.162$
Criteria	Mean	17.865	16.685	19.417	19.147	21.543	20.924	33.930	19.219	21.905	$F_{(3.092,134.489)} = 1.083$	$F_{(2,87)} = 2.010$	$F_{(1.546,134.489)} = 0.687$
Matching	SE	2.414	2.232	3.665	3.628	3.052	3.394	10.342	4.350	2.967	$p = 0.359$	$p = 0.140$	$p = 0.469$
Comparative	Mean	35.466	31.621	31.279	44.147	45.239	46.301	33.273	42.165	47.250	$F_{(4,174)} = 1.377$	$F_{(2,87)} = 3.232$	$F_{(2,174)} = 0.732$
Analysis	SE	3.542	2.281	3.091	4.128	6.871	6.666	3.318	6.248	4.735	$p = 0.244$	$\boldsymbol{p = 0.044}$	$p = 0.482$
Extremes	Mean	23.404	14.999	17.351	25.285	22.422	22.592	27.883	19.822	18.415	$F_{(3.234,140.673)} = 0.371$	$F_{(2,87)} = 1.411$	$F_{(1.617,140.673)} = 3.541$
Identification	SE	5.413	1.131	2.413	5.066	2.526	3.451	4.174	2.190	1.924	$p = 0.789$	$p = 0.249$	$\boldsymbol{p = 0.041}$
Connection	Mean	14.941	16.567	17.229	26.193	30.498	25.398	20.951	22.751	20.507	$F_{(2.568,111.728)} = 0.165$	$F_{(2,87)} = 2.685$	$F_{(1.284,111.728)} = 0.327$
Counting	SE	1.416	5015	4.224	3.767	6.545	5.298	3.170	3.810	1.779	$p = 0.895$	$p = 0.074$	$p = 0.624$

| | | Post hoc Analysis | | | | | |
		Low-Medium	Low-High	Medium-High	Narrow-Medium	Narrow-Wide	Medium-Wide
Existence	Mean Difference	-14.909	-8.725	6.184			
Verification	SE	3.207	3.207	3.207			
	p	$\boldsymbol{< 0.001}$	$\boldsymbol{0.024}$	0.171			
Comparative	Mean Difference	-12.440	-8.107	4.333			
Analysis	SE	4.968	4.968	4.968			
	p	$\boldsymbol{0.042}$	0.319	1.000			
Extremes	Mean Difference				6.443	4.072	-0.372
Identification	SE				3.008	3.054	1.949
	p				0.105	0.150	1.000

B.2 Error Occurrences For Different Node Quantities and Node Widths

| | | Low Node Quantity | | | Medium Node Quantity | | | High Node Quantity | | | ANOVA | | |
		Narrow	Medium	Wide	Narrow	Medium	Wide	Narrow	Medium	Wide	Interaction	Node Quantity	Node Width
Average	Mean	0.113	0.060	0.160	0.073	0.067	0.087	0.360	0.193	0.293	$F_{(4,174)} = 0.785$	$F_{(2,87)} = 8.246$	$F_{(2,174)} = 2.123$
Error	SE	0.052	0.027	0.057	0.022	0.029	0.033	0.094	0.071	0.077	$p = 0.537$	$\boldsymbol{p < 0.001}$	$p = 0.123$

| | | Post hoc Analysis Results | | |
		Low-Medium	Low-High	Medium-High
Average	Mean Difference	0.036	-0.171	-0.207
Error	SE	0.054	0.054	0.054
	p	1	$\boldsymbol{0.007}$	$\boldsymbol{< 0.001}$

B.3 Subjective Ratings on Workload

		Low Node Quantity			Medium Node Quantity			High Node Quantity			ANOVA		
		Narrow	Medium	Wide	Narrow	Medium	Wide	Narrow	Medium	Wide	Interaction	Node Quantity	Node Width
Mental	Mean	3.033	2.900	2.900	2.933	2.833	2.900	4.367	3.567	4.233	$F_{(4,174)} = 2.564$		
Demand	SE	0.357	0.337	0.312	0.365	0.372	0.369	0.323	0.331	0.309	$\boldsymbol{p = 0.040}$		
Physical	Mean	2.833	2.600	2.433	2.467	2.433	2.800	3.767	3.300	3.733	$F_{(4,174)} = 2.670$		
Demand	SE	0.339	0.324	0.282	0.278	0.290	0.312	0.320	0.319	0.339	$\boldsymbol{p = 0.034}$		
Temporal	Mean	2.967	2.833	2.700	2.667	2.667	2.733	3.867	3.133	3.600	$F_{(4,174)} = 1.727$	$F_{(2,87)} = 2.369$	$F_{(2,174)} = 2.296$
Demand	SE	0.344	0.326	0.329	0.330	0.316	0.287	0.321	0.298	0.351	$p = 0.146$	$p = 0.100$	$p = 0.104$
Performance	Mean	5.667	5.800	5.567	5.400	5.500	5.567	5.300	5.633	4.867	$F_{(4,174)} = 1.567$	$F_{(2,87)} = 1.411$	$F_{(2,174)} = 2.437$
	SE	0.312	0.285	0.290	0.313	0.283	0.243	0.210	0.237	0.291	$p = 0.185$	$p = 0.472$	$p = 0.090$
Effort	Mean	3.100	3.300	3.033	3.200	3.133	3.433	4.733	4.033	4.433	$F_{(4,174)} = 2.540$		
	SE	0.305	0.322	0.327	0.344	0.335	0.341	0.275	0.309	0.274	$\boldsymbol{p = 0.042}$		
Frustration	Mean	2.467	2.233	2.467	2.367	2.367	2.500	3.433	2.567	3.067	$F_{(4,174)} = 1.847$	$F_{(2,87)} = 1.829$	$F_{(2,174)} = 4.082$
	SE	0.302	0.278	0.287	0.301	0.320	0.299	0.317	0.257	0.318	$p = 0.122$	$p = 0.169$	$\boldsymbol{p = 0.019}$

Post hoc Analysis Results				
		Narrow-Medium	Narrow-Wide	Medium-Wide
Frustration	Mean Difference	0.367	-0.171	-0.207
	SE	0.128	0.138	0.139
	p	**0.016**	1.000	0.124

Pairwise comparisons for each of the three Node Quantity

		Narrow Node Width			Medium Node Width			Wide Node Width		
		Low-Medium	Low-High	Medium-High	Low-Medium	Low-High	Medium-High	Low-Medium	Low-High	Medium-High
Mental	Mean Difference	0.100	-1.333	-1.433	0.067	-0.667	-0.733	0	-1.333	-1.333
Demand	SE	0.493	0.493	0.493	0.491	0.491	0.491	0.469	0.469	0.469
	p	1.000	**0.025**	**0.014**	1.000	0.533	0.416	1.000	**0.017**	**0.017**
Physical	Mean Difference	0.367	-0.933	-1.300	0.167	-0.700	-0.867	-0.367	-1.300	-0.933
Demand	SE	0.444	0.444	0.444	0.440	0.440	0.440	0.441	0.441	0.441
	p	1.000	0.115	**0.013**	1.000	0.346	0.156	1.000	**0.012**	0.112
Effort	Mean Difference	-0.100	-1.633	-1.533	0.167	-0.733	-0.900	-0.400	-1.400	-1.000
	SE	0.437	0.437	0.437	0.455	0.455	0.455	0.446	0.446	0.446
	p	1.000	**< 0.001**	**0.002**	1.000	0.333	0.154	1.000	**0.007**	0.083

Pairwise comparisons for each of the three Node Width

		Low Node Quantity			Medium Node Quantity			High Node Quantity		
		Narrow-Medium	Narrow-Wide	Medium-Wide	Narrow-Medium	Narrow-Wide	Medium-Wide	Narrow-Medium	Narrow-Wide	Medium-Wide
Mental	Mean Difference	0.133	0.133	0	0.100	0.033	-0.067	0.800	0.133	-0.667
Demand	SE	0.199	0.184	0.204	0.199	0.184	0.204	0.199	0.184	0.204
	p	1.000	1.000	1.000	1.000	1.000	1.000	**< 0.001**	1.000	**0.005**
Physical	Mean Difference	0.233	0.400	0.167	0.033	-0.333	-0.367	0.467	0.033	-0.433
Demand	SE	0.207	0.189	0.173	0.207	0.189	0.173	0.207	0.189	0.173
	p	0.788	0.110	1.000	1.000	0.242	0.112	0.080	1.000	**0.043**
Effort	Mean Difference	-0.200	0.067	0.267	0.067	-0.233	-0.300	0.700	0.300	-0.400
	SE	0.241	0.230	0.229	0.241	0.230	0.229	0.241	0.230	0.229
	p	1.000	1.000	0.745	1.000	0.937	0.583	**0.014**	0.584	0.254

C SUMMARY OF STATISTICAL RESULTS FROM PHASE III

C.1 Completion Time for Different Design Choices

		Baseline	Tick Marks	Color Gradient	Combined	Friedman Test
Existence Verification	Mean	15.626	16.305	15.016	17.461	$\chi^2(3) = 2.000$
	SE	1.319	1.044	0.696	1.294	$p = 0.572$
	Mean Rank	2.250	2.580	2.420	2.750	
Criteria Matching	Mean	15.757	21.004	17.727	17.239	$\chi^2(3) = 6.100$
	SE	1.367	1.844	0.876	0.923	$p = 0.107$
	Mean Rank	2.040	2.960	2.540	2.460	
Comparative Analysis	Mean	28.022	26.355	28.572	30.156	$\chi^2(3) = 4.350$
	SE	2.578	1.916	1.283	2.134	$p = 0.226$
	Mean Rank	2.290	2.170	2.790	2.750	
Extremes Identification	Mean	10.487	10.716	10.851	11.898	$\chi^2(3) = 4.650$
	SE	0.886	0.705	0.562	0.647	$p = 0.199$
	Mean Rank	2.040	2.670	2.500	2.790	
Connection Counting	Mean	11.633	11.890	14.137	10.826	$\chi^2(3) = 0.750$
	SE	0.886	0.881	2.578	0.646	$p = 0.861$
	Mean Rank	2.330	2.580	2.630	2.460	

C.2 Error Occurrences for Different Design Choices

		Baseline	Tick Marks	Color Gradient	Combined	Friedman Test
Average Error	Mean	0.017	0.017	0.050	0.017	$\chi^2(3) = 3.375$
	SE	0.012	0.012	0.025	0.012	$p = 0.337$
	Mean Rank	2.440	2.440	2.690	2.440	

C.3 Subjective Ratings on Workload

		Baseline	Tick Marks	Color Gradient	Combined	Friedman Test
Mental Demand	Mean	2.875	3.000	3.167	3.042	$\chi^2(3) = 3.669$
	SE	0.337	0.351	0.322	0.332	$p = 0.300$
	Mean Rank	2.270	2.380	2.790	2.560	
Physical Demand	Mean	2.250	2.417	2.333	2.417	$\chi^2(3) = 2.012$
	SE	0.250	0.262	0.238	0.269	$p = 0.570$
	Mean Rank	2.330	2.500	2.520	2.650	
Temporal Demand	Mean	2.500	2.500	2.500	2.625	$\chi^2(3) = 0.833$
	SE	0.313	0.295	0.301	0.275	$p = 0.841$
	Mean Rank	2.480	2.460	2.420	2.650	
Performance	Mean	5.708	5.750	5.583	5.708	$\chi^2(3) = 1.903$
	SE	0.259	0.250	0.240	0.221	$p = 0.593$
	Mean Rank	2.520	2.670	2.350	2.460	
Effort	Mean	3.583	3.542	3.500	3.625	$\chi^2(3) = 0.195$
	SE	0.366	0.340	0.366	0.360	$p = 0.978$
	Mean Rank	2.480	2.440	2.520	2.560	
Frustration	Mean	2.042	2.042	2.250	2.250	$\chi^2(3) = 3.988$
	SE	0.259	0.252	0.290	0.250	$p = 0.263$
	Mean Rank	2.330	2.350	2.650	2.670	